DSC 152:
Applied Statistical Data Analysis and Inference

Lecture #4
The t-Test and Statistical Power

Thursday, April 9
Spring Quarter 2026
Peter Chi

Statistical Power

Last couple of lectures

- The Type I Error rate of a statistical test is the probability of rejecting H_0 when H_0 is actually true.

- We can estimate the Type I Error rate via simulation

Now, in general terms, what is statistical power?

https://pollev.com/chi

Statistical Power

Statistical power is a function of these four things:

- Sample size
- Statistical significance level
- Standard deviation
- Effect size

The first 3 of these should be relatively apparent as to what they are (in principle). But what is effect size?

Statistical Power

Example: UCSD students' sleep again

Suppose that prior to any data collection, we want to know what our statistical power is for detecting a difference from an average of 6 hours of sleep.

- Sample size: 5
- Statistical significance level: $\alpha = 0.05$
- Standard deviation: ?
- Effect size: ???

Statistical Power

Example: UCSD students' sleep again

The challenging thing about power calculations in practice is that they need to be done prior to any data collection, which makes it difficult to know what the standard deviation and the effect size should be.

Three overall approaches

- Literature review
- Pilot study
- Posthoc

Statistical Power

Example: UCSD students' sleep again

Three overall approaches

- Literature review
  - Look through previous studies to find anything similar
- Pilot study
  - Collect a small preliminary set of data, use the estimates from that
- Posthoc
  - Do power calculations after your data collection and statistical analyses, using estimates from your actual data as inputs

Posthoc is obviously the least desirable, and I am philosophically against this practice in general (but just be aware that it is out there in real research studies – I often get asked by collaborators to do exactly this).

Statistical Power

Example: UCSD students' sleep again

Let us suppose that from a literature review, we believe that the variance of the hours of sleep a UCSD student gets is 1 hour, and that the average UCSD student might only sleep for 4 hours per night. That is, if X_i is the amount of sleep that the i^{th} student gets, then we will proceed by assuming that:

$$X_i \sim N(\mu = 4, \sigma^2 = 1)$$

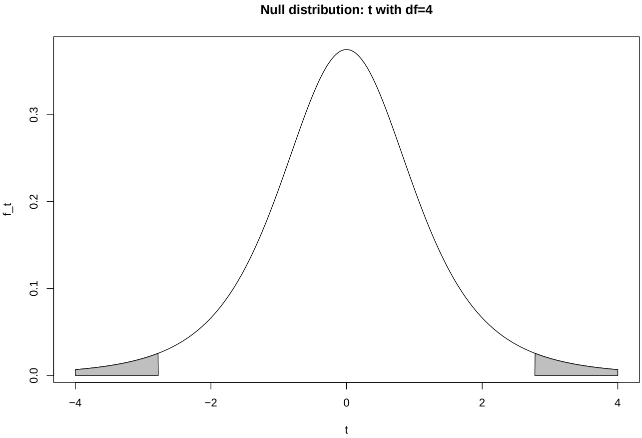
A power calculation proceeds by determining the probability that we would reject H_0 if the above values are actually true.

What is the distribution of t_s under this reality?

$$t_s = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \dots$$

Statistical Power

Example: UCSD students' sleep again



The statistical power of the test here is the probability of observing a value of t_s that is in the rejection region, if $X_i \sim N(\mu = 4, \sigma^2 = 1)$.

Statistical Power

Example: UCSD students' sleep again

What is the distribution of t_s under this reality?

$$t_s = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{\overline{x} - 6}{s/\sqrt{5}}$$

but we are operating under the assumption that the mean of $\overline{X}$ is 4, so this does not actually have a t distribution. But...

$$\begin{aligned} &= \frac{\overline{x} - 4 + 4 - 6}{s/\sqrt{5}} \\ &= \frac{\overline{x} - 4}{s/\sqrt{5}} + \frac{4 - 6}{s/\sqrt{5}} \\ &= \frac{\overline{x} - 4}{s/\sqrt{n}} - \frac{2}{s/\sqrt{5}} \end{aligned}$$

The first quantity after the = sign is a t random variable with df=4. The second quantity is a *random shift* to the left...

Statistical Power

Example: UCSD students' sleep again

What is the distribution of t_s under this reality?

$$t_s = \frac{\overline{x} - 4}{s/\sqrt{n}} - \frac{2}{s/\sqrt{5}}$$

The left quantity is a t random variable with df=4. The right quantity is a *random shift* towards smaller values. . .

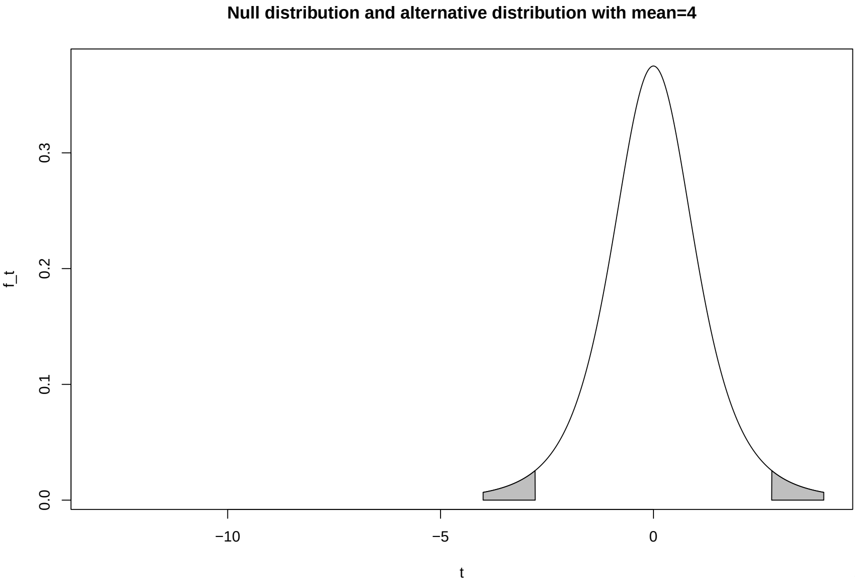
This means that, under this particular effect size and with $\sigma = 1$, the distribution of t_s follows a non-central t-Distribution with a noncentrality parameter of:

$$\delta = -\frac{2}{1/\sqrt{5}} = -2\sqrt{5}.$$

. . . what?

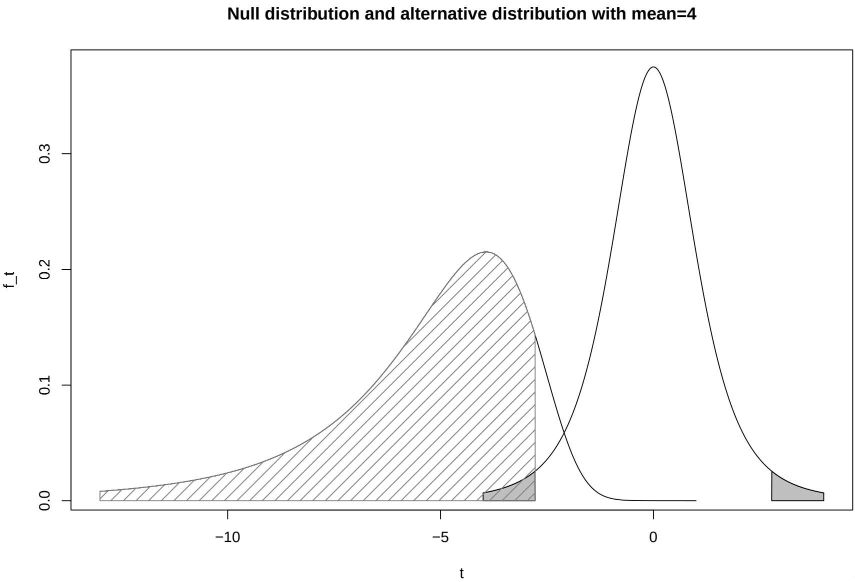
Statistical Power

Example: UCSD students' sleep again



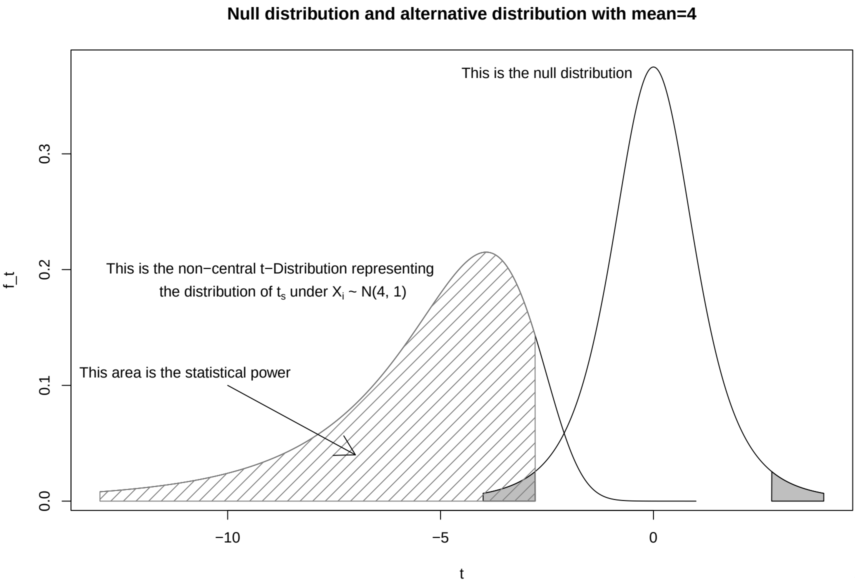
Statistical Power

Example: UCSD students' sleep again



Statistical Power

Example: UCSD students' sleep again



Statistical Power

Example: UCSD students' sleep again

So then using the non-central t-Distribution function in R, we can obtain it. First, the rejection region starts at:

```r
qt(0.025, df=4)
```

```
## [1] -2.776445
```

We can then put this value into the `pt` function with the appropriate non-centrality parameter:

```r
pt(-2.776445, df=4, ncp=-(2*sqrt(5)))
```

```
## [1] 0.9088849
```

Note: technically, there's also a right tail! But under this H_A, the probability of observing a t_s over there is practically 0.

Statistical Power

Example: UCSD students' sleep again

There is also a built-in function in R:

```
power.t.test(n=5, delta=2, sd=1, sig.level=0.05,
             type="one.sample", alternative="two.sided")
```

```
##
##      One-sample t test power calculation
##
##               n = 5
##           delta = 2
##              sd = 1
##       sig.level = 0.05
##           power = 0.9088849
##     alternative = two.sided
```

Statistical Power

Example: UCSD students' sleep again

Conclusion:

If UCSD students' sleep truly follows a $N(\mu = 4, \sigma^2 = 1)$ distribution, then our statistical power to detect $H_A\colon \mu \neq 6$ with a sample size of n=5 is approximately 0.9089.

Lingering questions:

- What if H_A is true, but not with exactly $X_i \sim N(\mu = 4, \sigma^2 = 1)$?
- Is there a better way that we can accomplish what was done on the previous slides?

Let's answer the 2nd one first, then swing back to the 1st one.

Statistical Power

Everything we just did is fine, but:

- it can be a bit conceptually difficult to understand
- it involves a lot of memorization or looking stuff up in order to do it
- it doesn't readily transfer to other situations; if you are doing a different statistical test (such as something in a regression setting), you have to learn/look up basically an entirely new procedure and/or find the right function for that situation (and hope that you use it correctly)
- it relied on the distribution of sleep following a normal distribution!

So what might we do instead?

Statistical Power

Simulation to estimate statistical power

Another valid way to estimate statistical power is to do it via simulation. The advantages are:

- We are already good at simulation in general, so it is conceptually easier to understand any simulation procedure than it is to understand e.g. what a non-central t-Distribution is.
- Once you learn how to simulate an estimate of statistical power in one situation, it is easier to transfer that knowledge to any other situation than it is to learn the appropriate theory-based procedure for that situation.
- If we simulate, we are not limited to H_A that assume the data follow the normal distribution; we can consider practically any distribution whatsoever.

Statistical Power

Simulation to estimate statistical power

So then how do we do it?

- Simulate samples of size 5 under the $N(\mu = 4, \sigma^2 = 1)$ distribution
- Do a t-test with $H_0\colon \mu = 6$ for those simulated data, get the p-value
- Determine whether the p-value for t_s is less than 0.05
- Repeat many times, count the proportion of times that t_s does give a p-value that is less than 0.05

Your Turn #1

Let's do it together. Open up an R Markdown file for today's Daily Check and start putting in this skeleton code:

```
count <- 0
reps <- 10000

for(i in 1:reps){
  sleep <- ...

  pval <- ...
  if(pval < 0.05){
    count <- count + 1
  }
}
```

Your Turn #1

If done correctly, we should get something close to:

```r
count / reps
```

```
## [1] 0.9074
```

Statistical Power

Example: UCSD students' sleep again

Conclusion:

If UCSD students' sleep truly follows a $N(\mu = 4, \sigma^2 = 1)$ distribution, then our statistical power to detect $H_A\colon \mu \neq 6$ with a sample size of n=5 is approximately 0.9089 from the theory-based approach, and 0.9074 from the simulation approach.

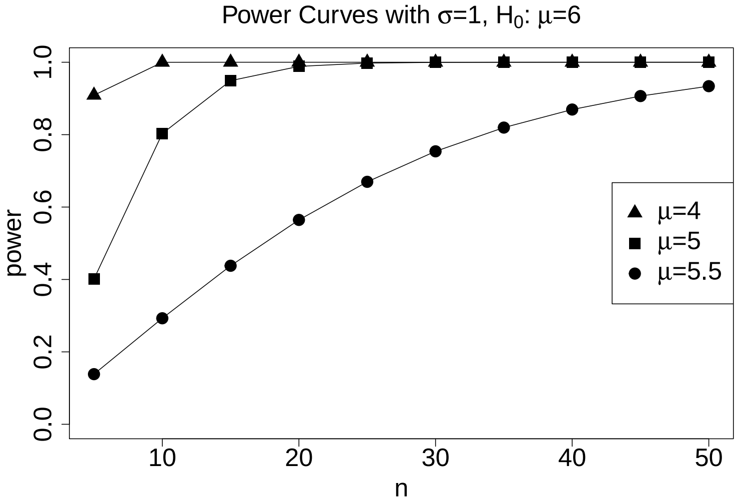
Lingering questions:

- What if H_A is true, but not with exactly $X_i \sim N(\mu = 4, \sigma^2 = 1)$?
- Is there a better way that we can accomplish what was done on the previous slides?

Now, back to the first question here.

Power Curves

Since we don't ever have full confidence in *one single* effect size / variance combination, practitioners typically report power calculations over a wide range of possibilities, presented as power curves.



Power Curves

What do we observe in the power curves shown on the previous slide?

https://pollev.com/chi

Power Curves

What do we observe in the power curves shown on the previous slide?

https://pollev.com/chi

Now, how do we make them?

You will do / are doing more of it in Lab 2, but for now, we will give you the foundations of how to do it.

First, recall the code that we just wrote in the Your Turn on Slide 20.

- That was hardcoded for one specific sample size, effect size, and variance.
- Now, we will want to write a function that can take any combination of inputs for these
- It should also allow the user to specify the number of `reps`
- For now, we will not change the fact that it simulates from a normal distribution, nor the $\alpha = 0.05$ significance level

Your Turn #2

Fill in the required code below:

```r
sim_power_t <- function(n, delta, sd, reps=10000){
  count <- 0

  ...



  return(count / reps)
}
```

Your Turn #2

If written correctly, running the function should give something close to these outputs:

```
sim_power_t(n=5, delta=2, sd=1)
```

```
## [1] 0.9033
```

```
sim_power_t(n=10, delta=1, sd=1)
```

```
## [1] 0.8018
```

Summary of today's Daily Check

- Your Turn #1 from Slide 20
- Your Turn #2 from Slide 25

Upload your pdf output file to Gradescope by midnight tonight.

Recap and Looking Ahead

Recap

- Statistical power is the probability of correctly rejecting H_0. We want it to be high! (80% is a standard rule of thumb).
- We learned how to do simulation to estimate statistical power
- It depends on a variety of inputs that are not always obvious as to what they should be (see Slides 3 to 6).

Looking Ahead

- In Lab 2, you will use the function you just wrote to reconstruct the power curves shown in Slide 23
- Question: if power always increases with sample size, then will there always be a sample size at which we would detect even a negligible difference – that is, a difference that is so small that we wouldn't actually care about it even if it is real?