Math 183
Statistical Methods

Eddie Aamari
S.E.W. Assistant Professor

eaamari@ucsd.edu
math.ucsd.edu/~eaamari/
AP&M 5880A

Today: Chapter 5 (continued)

- Find the power of a test given the description of the setting
- Find the sample size needed to achieve a certain power level and confidence level
- Power calculation examples

How Big Should My Sample(s) Be?

There is no easy answer to this question because you must balance many different (and competing) factors:

- Cost of larger sample (money, time)
- Risk to participants (new drug, dangerous procedure)
- Need to get sampling distribution look a certain way
- Downside of missing a real effect

Notice, as n gets large, the sampling distribution approaches its theoretical shape, but costs and dangers might skyrocket.

The first two are context-dependent.

We focus on the last using power analysis.

Power

Recall that power is

$$1 - \beta = P(\text{rejecting } H_0 | H_0 \text{ is false}).$$

Since hypothesis testing revolves around assuming H_0 is true, the power is not immediately apparent in what we've done before.

Importantly, what we often care about is the power of a test:
If there really is something interesting going on, how likely am I to find it (to end up with H_A instead of H_0)?

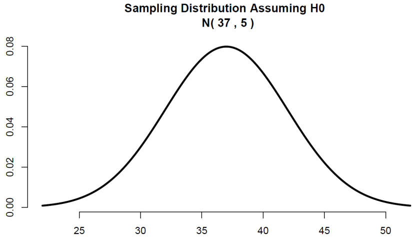
Sample size is tied to power, and power is tied to the ability to uncover a true effect. Missing a true effect could be annoying...!

Your friend claims that UCSD students send, on average, 37 text messages a day. You think it is lower and decide to run a hypothesis test.

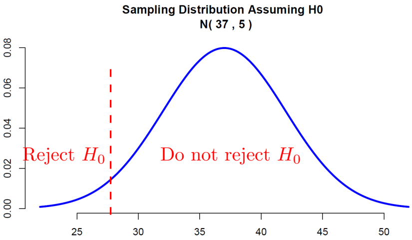
Let μ be the average text messages sent, per day, by UCSD students.

$$H_0\colon\ \mu = 37$$
$$H_A\colon\ \mu < 37$$

Assuming H_0, we get a sampling distribution centered at 37, with some SE (here, just assume 5).

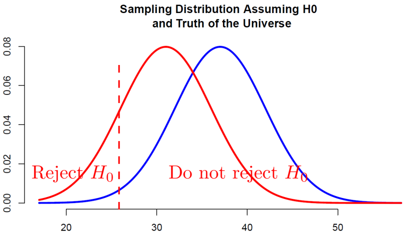


Remark: The distribution is a T-distribution which we assume is normal for ease of future calculations.
In this course, always use a Normal Curve when doing power calculations.



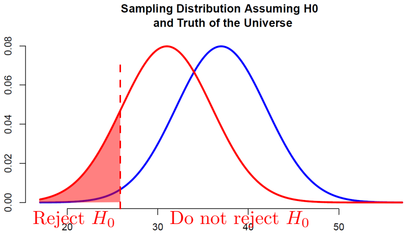
When we draw a sample, it lands somewhere on this distribution. If it is in the bottom 5% of means, the p-value will be less than 0.05, and we reject H_0 (if $\alpha = 0.05$).

Assume for a moment that μ is truly 31. When we draw a sample, it really comes from the red curve, and we explore it relative to the blue curve.



What area in the picture is the power $1-\beta = P(\text{reject } H_0 | \text{ true effect})$?

Visualizing Power



The power is the area under the truth curve that falls into the rejection zone(s) based on the H_0 curve. Our sample comes from the truth of the universe, but we decide how to interpret it based on assuming H_0.

You choose your sample size based on the power you want from your test. You decide on your power based on how horrible it would be to overlook a true effect.
Common power levels $1 - \beta$: between 0.8 and 0.9.

Remark: Here, we assumed we knew the true value, which is never the case!

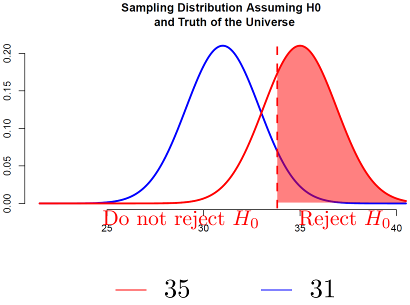
Are You Happy?

You are running a hypothesis test to see if Math 183 students are happier than average Americans. You learn the average Happiness Index of Americans is 31 with a standard deviation of 12.
You draw a sample of size $n = 40$. If the true Happiness Index of Math 183 students is 35, how powerful is your test?

1) Set up the parameter and the hypotheses
Let μ be the average Happiness Index of Math 183 students.
Set

$$H_0\ \mu = 31 \text{ and } H_A;\ \mu > 31.$$

Draw the two sampling distributions (based on H_0 and Truth of Universe).
Find the SE and the zone(s) where H_0 is rejected and not rejected (relative to the H_0 sampling distribution) by finding the “rejection fence”.



With $\sigma = 12$ and $n = 12$,
$SE = \frac{12}{\sqrt{40}} \simeq 1.897$.

With $\alpha = 0.05$ and a one-sided alternative, we get p-values less than α when the sample mean is in the top 5% strangest means.

```
> qnorm(0.95, mean = 31, sd =
    1.897)
[1] 34.12029
```

3) Shade the area under the Truth of Universe sampling distribution that lies in the rejection zone(s). This area is the power.

```
> pnorm(34.12, mean = 35, sd  = 1.897, lower.tail = F)
[1] 0.6786368
```

With a sample size of 40, we would uncover the true effect only 68% of the time. 32% of the time, the universe will give us a sample mean that makes us keep believe in H_0!

Question: How do we do if no-one tells us the Truth of Universe?

Two-Sided Alternatives, Hidden Truth

As a statistician in a pharmaceutical company, you must decide if a new drug works or not. You divide 200 subjects into equal treatment and control groups, and measure the weight changes in each group.

1) Define parameters and hypotheses

Let μ_T be the average weight change among all people who could ever take the drug.
Let μ_C be the weight change among people who don't take the drug.

$$\text{We care about } \mu_T - \mu_C.$$

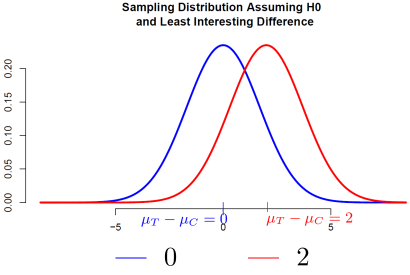
Set

$$H_0\colon\ \mu_T - \mu_C = 0$$
$$H_A\colon\ \mu_T - \mu_C \neq 0$$

2) If you are not told the truth of the universe, decide the smallest value of $\mu_T - \mu_C$ you care about and assume this is the truth.

Perhaps a difference of 2 pounds is enough to make us excited about the drug.

Assuming a weight change standard deviation of 12 pounds (a common number in weight change studies), find the power of your test with 200 participants.

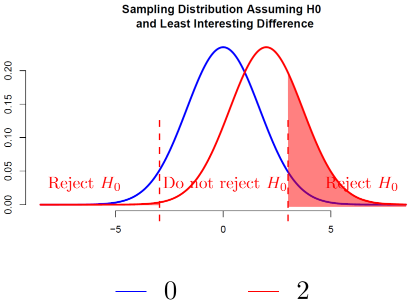


Recall that

$$SE_{\bar{x}_T - \bar{x}_C} = \sqrt{\frac{s_T^2}{n_T} + \frac{s_C^2}{n_C}}$$

$$\simeq \sqrt{\frac{12^2}{100} + \frac{12^2}{100}} \simeq 1.697.$$

Remember: When you explore a different idea, then the sampling distribution has a different SE.



With a two-sided test and $\alpha = 0.05$, we reject H_0 with results in the upper and lower 2.5% strangest results.

```
> qnorm(0.975, mean = 0, sd =
      1.697)
[1] 3.326059
```

Shading under the red curve, we get two areas: one in the far left tail, and a larger area on the right (in this case).

```
> pnorm(3.32, mean = 2, sd = 1.697,lower.tail = F)
[1] 0.2183307
> pnorm(-3.32, mean = 2, sd = 1.697,lower.tail = T)
[1] 0.0008594345
```

The power of our test, for a *Least Interesting Event* of 2 pounds and $\alpha = 0.05$ is

$$1 - \beta = 0.21833 + 0.00086 = 0.21919 \simeq 21.9\%.$$

This is very "underpowered". There is a good chance you will not end up at the alternative hypothesis even when there is a true effect!

Deciding on a Sample Size for a Given Power

The factors

- α value (often 0.05)
- Desired power of test (often 0.8 – 0.9)
- Smallest change that is interesting (varies)
- Nature of test

altogether, help you decide the sample size.

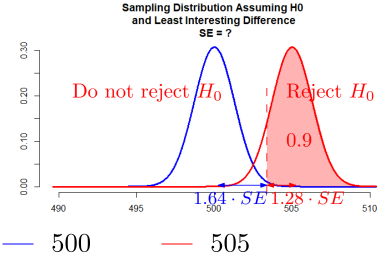
Setup: You are curious if your new MCAT course raises scores (score range 472 and 528, with 500 average).
You want to be 95% sure in your hypothesis test conclusion and desire a setup that has 90% power.

You really only care if the course raises grades 5 points (on average).

Question: How many students must you put through your course to meet these demands?

Let μ be the average score of people who take your course.

Set H_0: $\mu = 500$ and H_A: $\mu > 500$.



Since we don't know n, we can't find the SE!
You must do the whole problem in SE's.

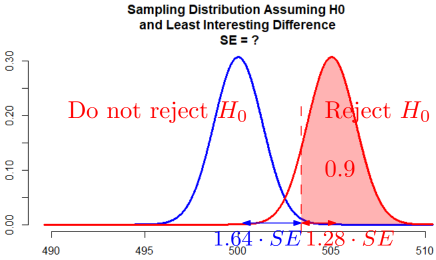
How many SE's above 500 is the rejection fence?

```
> qnorm(0.95, mean=0, sd =1)
[1] 1.644854
```

How many SE's below 505 must we got to get the shaded area to be 0.9?

```
> qnorm(0.1, mean=0, sd =1)
[1] -1.281552
```

So $1,64 \cdot SE + 1.28 \cdot SE$ span a distance of 5.



Thus, $2.92 \cdot SE = 5$.

These sampling distributions have $SE = \dfrac{\sigma}{\sqrt{n}}$.

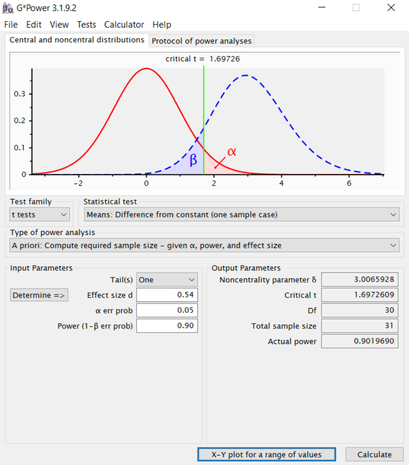
Online, you read MCAT scores have $\sigma = 9.5$.

So $2.92 \cdot \dfrac{9.5}{\sqrt{n}} = 5$, which yields $n = 30.78$.

We should put 31 people through our course!

Do people Actually Do That?

Yes... Helped with tools!



Few people think carefully about sample size. They choose n based on cost/dangers rather than giving their test a certain degree of power...

Those who care about power calculations often use software to help decide based on their test, α, power level, and minimum interesting difference ("effect size d")