Math 183
Statistical Methods

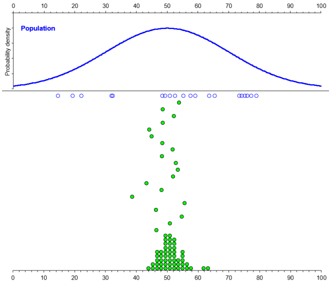
Eddie Aamari
S.E.W. Assistant Professor

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

Today: Chapter 4 (continued)

- Refresher on confidence intervals
- Hypothesis testing
- Null hypothesis framework and how it mimics real-life thinking
- Using confidence intervals to answer hypothesis testing questions
- Types of errors in hypothesis testing (Type I and Type II errors)

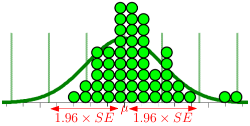
Refresher

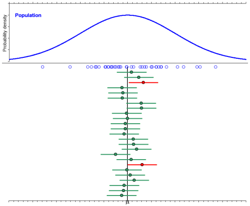


1. Draw a sample, find $\bar{x}$ and s_x
2. Recognize that sampling variations make $\bar{x}$ an imperfect measure for μ.
3. Study the sampling distribution, which shows the variability in $\bar{x}$
4. Discover that the sampling distribution is $N\left(\mu, \sigma/\sqrt{n}\right)$
5. Use this information to build a confidence interval of μ from $\bar{x}$, that shows your uncertainty.

Another visualization of this process on Brown University's website

Refresher



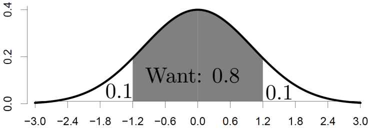


6. If we stand at μ and reach a certain number of SE's each way, we will grab a certain fraction of the sample means.
7. Conversely, if we reach out this same distance from each sample mean, we will grab μ a certain fraction of the time
8. Introduce language like "95% confident". This requires reaching out $1.96 \times SE$'s, because on any normal curve, 95% of the points are within $1.96SE$'s of the mean
9. To be "95% confident in an interval" means to have chosen a width (when building the interval), that if this width were used repeatedly for many intervals, 95% of those intervals would capture the true mean μ

A Complete Example

You are curious how many Facebook friends your friends have. You draw a random sample of 30 and go to their Facebook pages and look at their friend counts. The average is 325 with a standard deviation 25. Build an 80% confidence interval for the average Facebook-friend-account of all our friends.

1) Find the critical value z^*




```
> qnorm(0.9,mean=0,sd=1,
    lower.tail=T)
[1] 1.281552
```

2) Build the interval

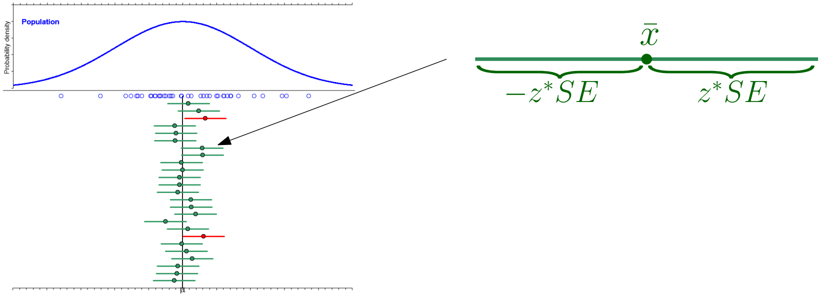
$$\bar{x} \pm z^* SE = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$
$$\simeq \bar{x} \pm z^* \frac{s_x}{\sqrt{n}}$$

We get $325 \pm 1.28155 \frac{25}{\sqrt{30}}$.

3) Interpret in English

We are 80% confident that the true mean is between 319.15 and 330.85 FB friends.

Anatomy of a Confidence Interval



Always remember that your interval is one of an infinite sequence of possible (virtual) intervals that could have arisen. The confidence level suggests what percentage of intervals in this family will capture the true population parameter.

Confidence Intervals in Hypothesis Testing

Hypothesis: A claim that or may not be true.

Examples:

- $\mu_{\text{cancer rate SD citizens}} \neq \mu_{\text{cancer rate in the US}}$
- $\mu_{\text{weight of you friends}} \neq \mu_{\text{weight of the US population}}$
- $\mu_{\text{cholesterol using new drug}} < \mu_{\text{cholesterol using placebo}}$
- $\mu_{\text{happiness increase from new drug}} > \mu_{\text{happiness increase from placebo}}$

Suppose we know that $\mu_{\text{happiness increase from placebo}} = 1.9$ and in a small sample of patients given the new drug, we find $\bar{x}_{\text{new drug}} = 2.3$ (people rate their happiness change on a -5 to 5 scale).

Since variation occurs when we take any sample, how do we decide whether a difference we see is natural variation or a genuine difference?
Said differently: Is the difference we see significant?

Sample Variation or True Effect?

Hypothesis testing is the rigorous way statisticians have devised to sort out how confident we can be that sample variation is not the cause.

Step 1: Write down a **Null hypothesis**. Usually, this is a statement that says nothing is happening.
It is always a claim about some population parameter!
Examples:

- H_0: $\mu_{\text{cancer rate SD citizens}} = \mu_{\text{cancer rate in the US}}$
- H_0: $\mu_{\text{weight of you friends}} = \mu_{\text{weight of the US population}}$
- H_0: $\mu_{\text{cholesterol using new drug}} = \mu_{\text{cholesterol using placebo}}$
- H_0: $\mu_{\text{happiness increase from new drug}} = \mu_{\text{happiness increase from placebo}}$

Alternative Hypothesis

Step 2: Write down an **Alternative hypothesis**. This is what you suspect might be true and is what you hope to show.
Examples:

- H_A: $\mu_{\text{cancer rate SD citizens}} \neq \mu_{\text{cancer rate in the US}}$
- H_A: $\mu_{\text{weight of you friends}} \neq \mu_{\text{weight of the US population}}$
- H_A: $\mu_{\text{cholesterol using new drug}} < \mu_{\text{cholesterol using placebo}}$
- H_A: $\mu_{\text{happiness increase from new drug}} > \mu_{\text{happiness increase from placebo}}$

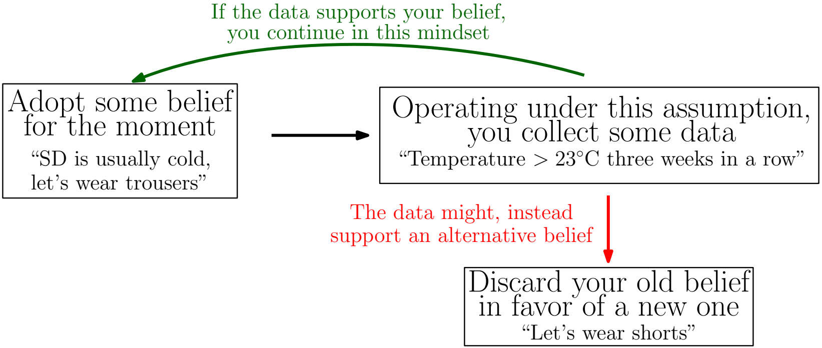
A **one-sided alternative hypothesis** will use a $<$ or $>$ sign.
It's when you're hoping your average is on a certain side of the comparison average.

A **two-sided alternative hypothesis** will use a $\neq$ sign.
It's when you're just wondering if the mean is different than the comparison mean.

The kind of alternative hypothesis you use depends on what you are guessing/hoping might be true (before you collect any data).

Mimicking real-life

How do we decide between H_0 and H_A?
Answer: How we often decide between beliefs in real life:



Notice that you are comparing the data from your life against some belief that you hold temporarily (here, wearing trousers). Perhaps the data support it, perhaps they support movement to an alternative.

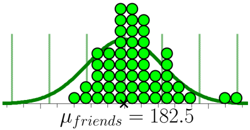
Collecting Data

<u>Step 3</u>: For the moment, assume the **null hypothesis** is the law of the land. Collect some data under this assumption and see if it supports your assumption (H_0) or the alternative (H_A).

You read that the average Ameican weights is 182.5 pounds. You weigh 20 random friends and get a mean of 175 pounds, with a standard deviation of 14 pounds.
Is the average weight of **all** your friends different from the national average?

H_0: $\mu_{\text{weight of friends}} = 182.5$
H_A: $\mu_{\text{weight of friends}} \neq 182.5$

Assuming H_0 is true, we got a sample from a sampling distribution centered at $\mu = 182.5$.



Confidence Intervals Help Us Decide between H_0/H_A

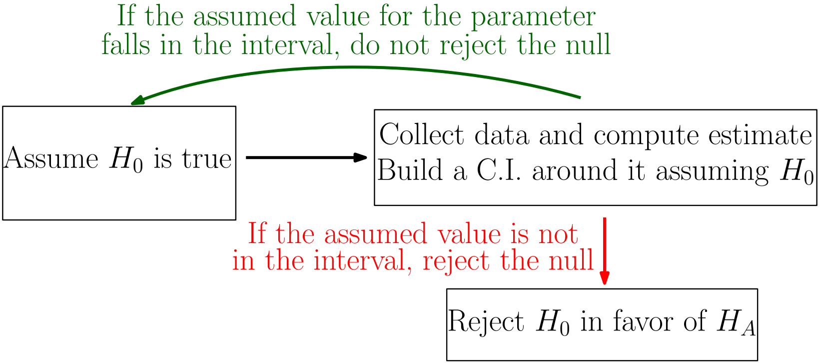
If we build a 95% confidence interval around the sample mean, we can give a good sense of what the data believe μ_{friends} should be.

A 95% C.I. has the form $\bar{x} \pm 1.96SE$. For us, this is $175 \pm 1.96\frac{14}{\sqrt{20}}$.

From our data, we are 95% confident that μ_{friends} is in $[168.86, 181.14]$.

Because 182.5 is not in this interval, we <u>reject H_0</u> in favor of H_A: it appears the friend's average weight is different than the American average.

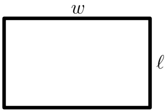
Null Hypothesis Testing Framework



Note that our data <u>do not prove</u> the null is true, nor that the alternative is true.
The data simply suggests which we should adopt moving forward.

Example

The Greeks were famous for the "golden rectangle": $w/\ell \simeq 0.618$.



Did they discover an "aesthetic standard", or did future societies just copy Greek ideas?

$$H_0\text{: } 0.618 \text{ is some standard}$$
$$H_A\text{: Greeks influenced others}$$

Researchers visited the Shoshoni Indians and measured the w/ℓ ratios of 20 hand-sewn beaded rectangles.
They found $\bar{x} = 0.661$, $s_x = 0.093$. Construct a 95% C.I. for the true w/ℓ ratio of Shoshoni beaded rectangles.

We know the C.I. is $\bar{x} \pm z^* SE \simeq 0.661 \pm 1.96 \times \frac{0.093}{\sqrt{20}} = [0.620, 0.702]$.

Since 0.618 is not in the C.I., this suggests the "aesthetic standard" may not be so standard...

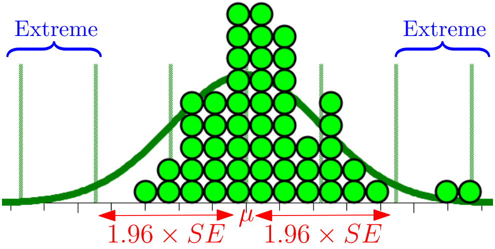
What Can Go Wrong?

There are Four possible scenarios we must think about related to hypothesis testing:

1. We reject H_0 when H_0 is actually true — Type I Error
2. We do not reject H_0 when H_0 is actually true — Awesome!
3. We reject H_0 when H_0 is actually false — Awesome!
4. We do not reject H_0 when H_0 is actually false — Type II Error

Truth \ Test Conclusion	Do not reject H_0	Reject H_0
H_0 true	OK	Type I Error
H_A true	Type II Error	OK

How Often Do We Make These Errors?



Suppose for a moment that H_0 really is true. Any time we draw a sample we get a green dot, and if we create a 95% C.I. around the dot, we know the interval captures the true mean μ 95% of the time.

Only those extreme samples with sample means more than $1.96SE$'s away will yield dots with C.I.'s that don't capture the true mean μ.

How Often Do We Make These Errors?

We know that 5% of the sample means will fall outside the interval $\mu \pm 1.96SE$, so we will make a Type I error 5% of the time (if we use 90% C.I.'s)

Hence, if we use 95% confidence intervals,

$$P\left(\text{Making a Type I Error}\right) = P\left(\text{Reject } H_0 | H_0 \text{ is true}\right) = 0.05.$$

The value 0.05 is given the letter α and called the **significance level**.

We also care about

$$P\left(\text{Making a Type II Error}\right) = P\left(\text{Do not reject } H_0 | H_0 \text{ is false}\right).$$

No nice formula for this probability: it is given the letter β.

The **power** of a test is

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

Type I and II Errors in Real Life

In the US legal system, you are either innocent (H_0) or guilty (H_A).

Type I Error: You reject H_0 when, in fact, H_0 is true.
This means you convict an innocent person: "Wrongful conviction"

How you might reduce Type I Error: Require more proof to convict.
Notice that this increases Type II Error.

Type II Error: You don't reject H_0 when, in fact, H_0 is false.
This means you set free a guilty person: "Guilty person set free"

How you might reduce Type II Error: Require less proof to convict.
Notice that this increases Type I Error.

Moral: For most actions,

$$\text{Type I Error} \searrow \text{ when Type II Error} \nearrow \text{ (and vice versa).}$$

Thinking About Errors in a Tree Diagram