EDP308: STATISTICAL LITERACY

The University of Texas at Austin, Fall 2020

RAZ: Rebecca A. Zárate, MA

Overview

- Why does this matter?
- How Unlikely is Unlikely?
  - z and t Cut Offs
- The Logic of Hypothesis Testing
    - Confirmation Bias
  - Null Hypothesis
  - Alternative Hypothesis
- One vs. Two Tailed Tests
- Practice Formulating Hypotheses

Why does this matter?

- Distributions help up determine the probability of observing a sample mean $\bar{x}$ of a certain value, and they help us quantify the uncertainty around those sample means using confidence intervals.
- Why does this matter?
  - It is not uncommon to observe one student with poor academic performance or one building with high levels of lead in its water. This may not be ground breaking news or require attention. But what if we saw an entire city (or state) of students with poor academic performance or high lead content in the water? This is much more unlikely and may be worth of further investigation…

But HOW unlikely does something have to be for us to conclude there IS something going on here?

How Unlikely is “Unlikely”?

- A fairly common convention in statistics is to test how likely something is. For example, assuming the water is fine, the probability of seeing this high of lead content in water, is less than 5%.

Do you think this is a fluke or do you think something is wrong with the water?

What if the probability was less than 1%?

How Unlikely is “Unlikely”?

- If the probability of seeing this high of lead content in water, is less than 5%, we would say “the results were statistically significant at the .05 level.”
  - In statistics we call this “statistically significant.”
- There are three main levels of significance:
  - .10, which is equivalent to a less than 10% chance
    - p value < .10
  - .05, which is equivalent to a less than 5% chance
    - p value < .05

    Most common.
  - .01, which is equivalent to a less than 1% chance
    - p value < .01

.05 is most common, but it depends on the context. In high stakes situations, like pharmaceutical drug testing, .01 may be used.

Unlikeliness with Known σ

- When we are working with the z-distribution, the z-values ("cut offs) associated with those chances are known and we can look them up on a z-table:
- 90% confidence has z = 1.65
  - p value = .10
- 95% confidence has z = 1.96
  - p value = .05
- 99% confidence has z = 2.58
  - p value = .01

Many times we do not know σ and sample sizes can be small sometimes. We usually need to use the t-distribution …

Unlikeliness with Unknown σ

- When we are working with the t-distribution, the critical t-values ("cut offs) associated with those chances need to be looked up on a t-table and vary depending on the sample size (degrees of freedom) and level of confidence (level of significance) you want to be able to have in your statistical test.

All these confidence intervals and levels of confidence will enable us to finally move into…

Hypothesis Testing

- We may want to know if there are patterns or differences between groups (ex. control vs. exp groups) and if so, whether those differences are due to *true* differences or just due to random chance…
  - Again we know how finicky samples can be… But! Using distributions and confidence intervals, we have a much better chance of getting closer to the truth.
- "If we assume everything is fine, everyone is the same, what is the likelihood of getting this outcome by random chance?"
  - We've been doing this with our z and t-tables to determine the probability of observing a certain value
- Now, we are going to start formally testing our hypotheses!

Hypothesis Testing

Going Beyond the Mean

- So far, we've been calculating confidence intervals for statistic of interest like the mean
- But rarely do we just want to know the mean of one group…

What are some more interesting research questions?

Research Question Examples

- Do Females have more Empathy than Males?
- Are Men better at spatial tasks than Women?
- Does exercise plus diet work better than just exercise or just diet?
- Do more confident people make more eye contact than less confident people?
- Are certain racial or ethnic groups more likely to be accepted to a particular university?
- Are certain racial or ethnic groups more likely to experience police violence?
- Does this drug work better than a placebo?

Research Question Examples

- Some of those examples may seem fairly easy to answer, but we know now just from the way samples work, all samples are going to vary a bit from one to the other. And, how much actually counts as "more"?
  - Do Females have more Empathy than Males?
    - Maybe some females have more empathy, but what about those more empathetic males? How much more empathy does one group have to have over the other for us to say there is a difference? Would just half a point be considered a "real difference"? Or do we need 10 points? How much is enough?
- We must find a way to quantify…

Assuming that…

- To conduct hypothesis testing, certain assumptions need to be met in order to run certain statistics.
  - The data was collected using randomized sampling
- Other assumptions may include:
  - A certain sample size, n
  - A certain shape of the population distribution
    - Ex. Normally distributed
  - Assumption may differ depending on what population parameter we're interested in

Significance Test: Step by Step

A significance test about a hypothesis has six steps.

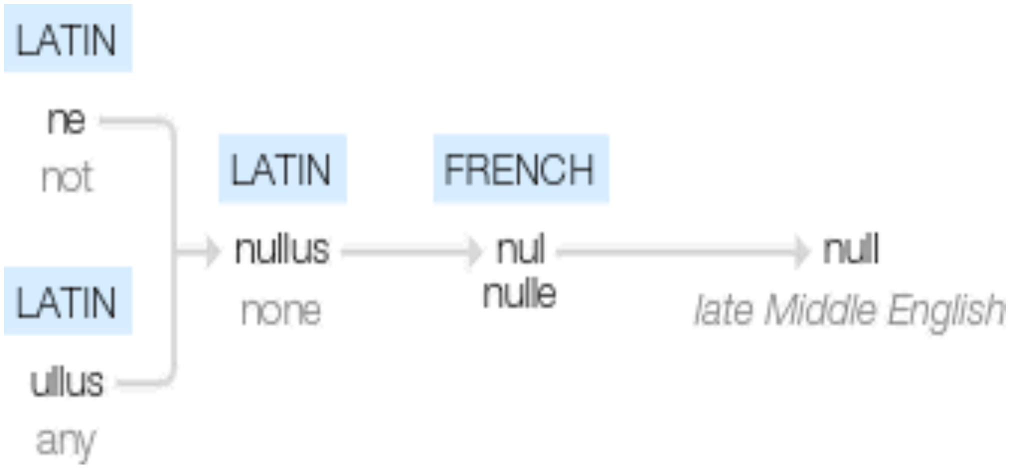
1. State the null and alternative hypotheses
   - And draw the picture!
2. Select a level of significance (α)
3. Choose a statistical test
4. Find the critical values
5. Compute the test statistic and the p-value
6. Formulate a decision (reject or fail to reject the null hypothesis)

Step 1: State the Hypotheses

- There are always two hypotheses – they always come in pairs!
  - The NULL Hypothesis (H_0)
    - Usually this is the norm, status quo, the assumption that there is no difference, no change, no effect
      - As a researcher, many times this is the thing you want to reject
        - Ex. "There is NO difference between the lead content in Flint, MI's water and the water for the rest of the US."
  - The ALTERNATIVE hypothesis (H_1 or H_A)
    - Usually the idea you are trying to provide evidence for, the new idea, the difference
      - As a researcher, it's usually your big idea or thing you believe
        - Ex. "There IS a difference between the lead content in Flint, MI's water and the water for the rest of the US."

"NULL"

- To nullify (verb) something means to invalidate or cancel it out
- When something is null (noun) it means zero
- In hypothesis testing, we want to (usually) nullify the null hypothesis, get rid of it, cancel it out…



We (typically) want to REJECT the premise that the difference is zero

Step 1: State the Hypotheses

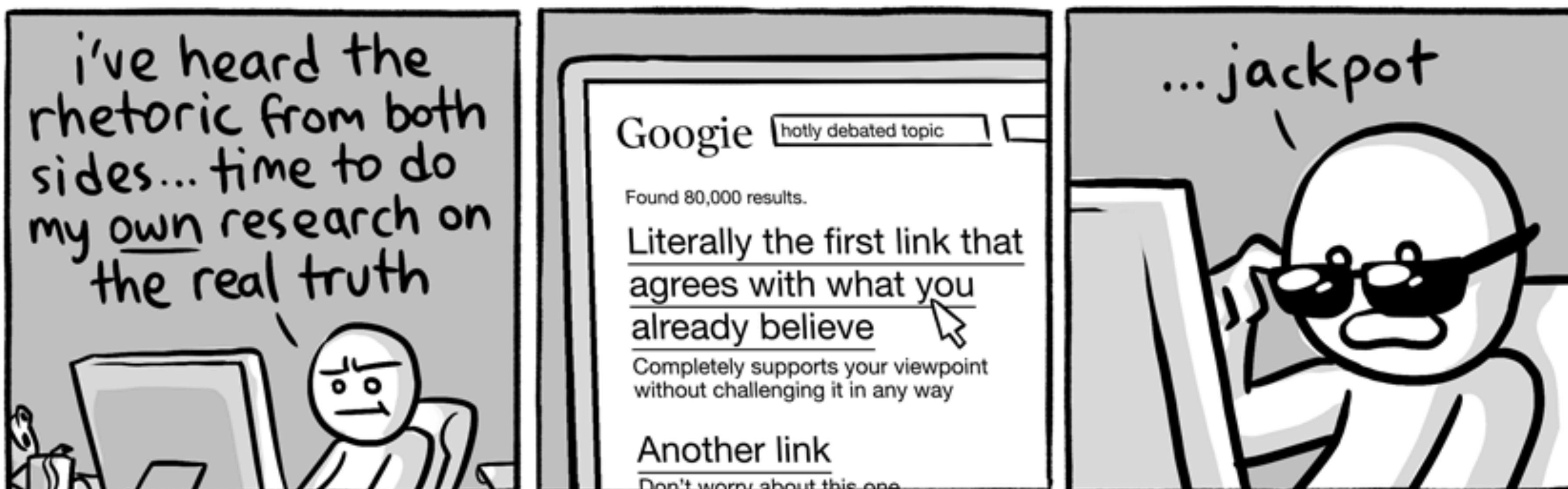
- Null Hypothesis: H_0
  - General assumption, "There is nothing there."
    - "There is zero difference."
    - "There is no effect."
    - "There is no difference between these groups."
    - "There is no relationship between these two things."
    - "Everything is cool."
- Alterative Hypothesis: H_a
  - Novel assertion, "There IS something there."
    - "There is a non-zero difference."
    - "There is an effect."
    - "There is a difference between these groups."
    - "There is a relationship between these two things."
    - "Everything is NOT cool."

The Logic of Falsification

- We *never* actually PROVE anything with statistics… Rather we "provide evidence for" some things, like the alterative hypothesis, or against other things like the null hypothesis
  - This is how science works… by disproving other explanation for a phenomenon
- To support and hold onto your belief, you need to be able to reject the opposing argument
- We assume the NULL is true and test how likely a certain outcome is based on that assumption

Confirmation Bias

- Unfortunately, we tend to look for and accept things that agree with and confirm our beliefs, while disregarding things that don't.



Chainsawsuit by Kris Straub

Confirmation Bias

“All swans are white.”

Let me prove it to you…

Confirmation Bias

See?
So many white swans. There's another and another and another.

"Obviously, all swans are white."
What is this "researcher" missing?

Confirmation Bias

- It only takes ONE black swan to disprove that theory that all swans are white… Rather than looking for more white swans to amass your white swan theory and collection, you should be looking for the black sw

Hypothesis Language

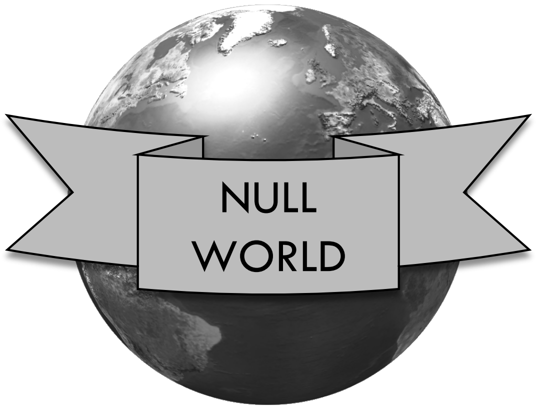
- We only "reject" or "fail to reject" NULLs, not the alternative hypothesis.
- We never "accept" any hypothesis, just "retain"
- At the end of your test, you will make one of two possible conclusions:
  - "Fail to reject the NULL hypothesis."
  - "Reject the NULL hypothesis."

You will state one of these two things for any hypothesis test we do.

We never reject or fail to reject the alternative.
We only makes statements about the NULL.

The Point of the Hypotheses

- In statistical testing, we ALWAYS start from the assumption, the belief that the NULL hypothesis is true.
  - Usually that means holding the OPPOSITE of your actual research belief, your alternative hypothesis.
  - We will ALWAYS assume we live in a NULL world until we test it and can reject it…

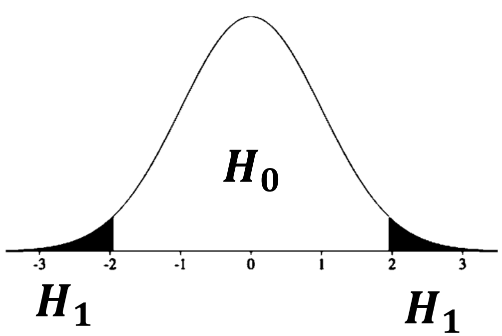


Hypothesis Notation

Draw the picture!

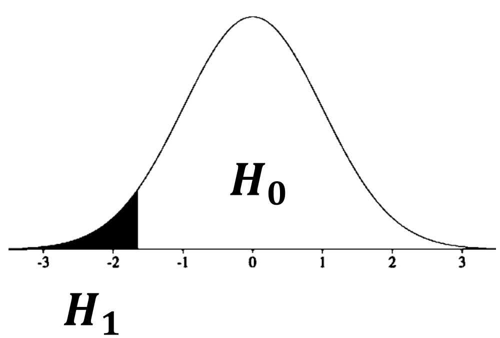
- Whenever we do hypothesis testing you will ALWAYS be making assertions about the POPULATION (the parameter μ). You want to go beyond your sample and say something about the population
- So you will <u>always</u> use $\boldsymbol{\mu}$, not $\bar{x}$, in your hypothesis notation
  - Hypotheses always come in pairs.
    - Equality is always a part of the null, H_0 (i.e. $=, \leq, \geq$)
    - Inequality is always a part of the alterative H_1 (i.e. $\neq, <, >$)

$\boldsymbol{H_0}$: $\boldsymbol{\mu} =$ some value
$\boldsymbol{H_1}$: $\boldsymbol{\mu} \neq$ some value



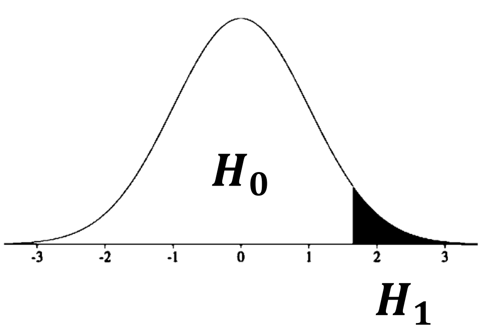
Two Tailed

$\boldsymbol{H_0}$: $\boldsymbol{\mu} \geq$ some value
$\boldsymbol{H_1}$: $\boldsymbol{\mu} <$ some value



One Tailed, Left

$\boldsymbol{H_0}$: $\boldsymbol{\mu} \leq$ some value
$\boldsymbol{H_1}$: $\boldsymbol{\mu} >$ some value



One Tailed, Right

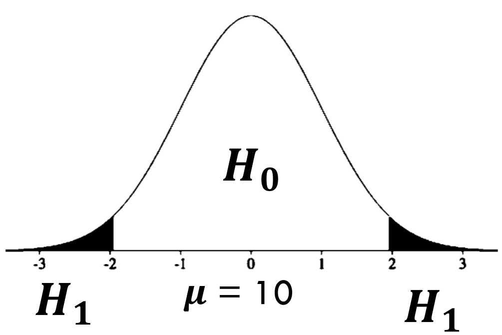
Example Hypothesis Formulation

Two Tailed

Different Than

- H_0: There is no difference in this school's performance on the STAR test compared to national average (10).
- H_1: There IS a difference in this school's performance on the STAR test compared to national average (10).

$$H_0: \mu = 10$$
$$H_1: \mu \neq 10$$

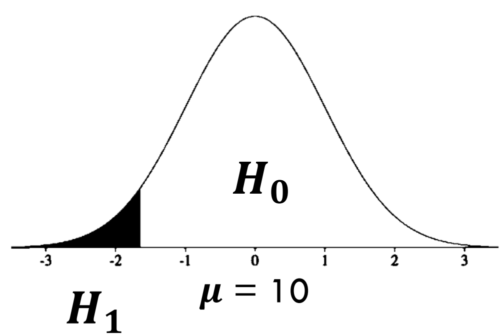


One Tailed (Left)

Less Than

- H_0: This school's performance on the STAR test is greater than or equal to the national average (10).
- H_1: This school's performance on the STAR test is less the national average (10).

$$H_0: \mu \geq 10$$
$$H_1: \mu < 10$$

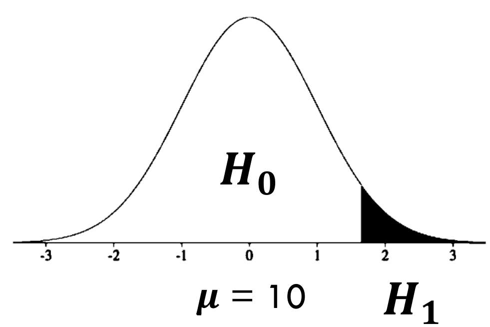


One Tailed (Right)

Greater Than

- H_0: This school's performance on the STAR test is less than or equal to the national average (10).
- H_1: This school's performance on the STAR test is greater the national average (10).

$$H_0: \mu \leq 10$$
$$H_1: \mu > 10$$



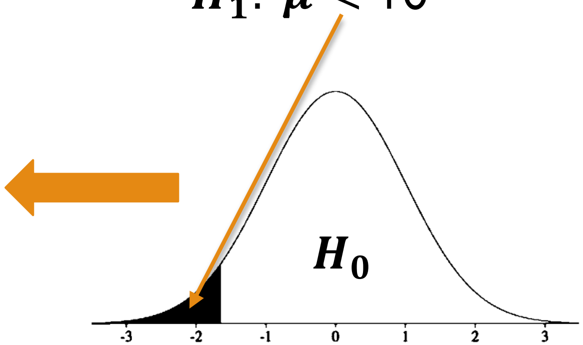
Example Hypothesis Formulation

One Tailed

Less Than

- H_0: This school's performance on the STAR test is greater than or equal to the national average (10).
- H_1: This school's performance on the STAR test is less the national average (10).

H_0: $\mu \geq 10$
H_1: $\mu < 10$



H_1: $\mu < 10$

Notice which way the "gator" is chomping. It always wants to eat the bigger piece, H_0.
(It also "points" in the direction of the alternative H_1.)

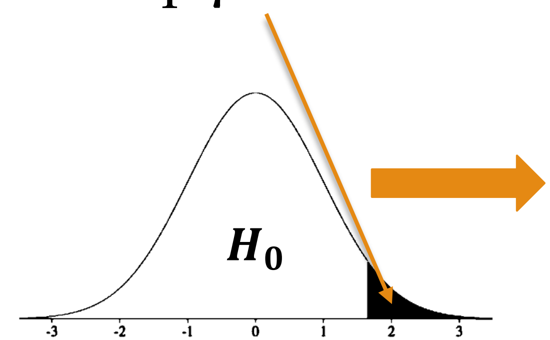
Draw the picture!

One Tailed

Greater Than

- H_0: This school's performance on the STAR test is less than or equal to the national average (10).
- H_1: This school's performance on the STAR test is greater the national average (10).

H_0: $\mu \leq 10$
H_1: $\mu > 10$



H_1: $\mu > 10$

How do you know?

- How do you know if it is a two-tailed or a one-tailed test?
  - Ask yourself, is there any suggestion about how the two groups might differ (which direction, "more" or "less") or is it only suggesting they differ (in general)?
    - Two-Tailed (non directional)
      - "Differ"
    - One-Tailed (directional)
      - "Greater than"
      - "More than"
      - "Less than"
      - "Fewer than"

And draw the picture!

Hypothesize it…

Write the null (H_0) and alternative (H_1) hypotheses for each scenario. Indicate whether it is a one (left or right tailed) or a two-tailed test.

- 1) An emergency room advertises a wait time of 10 minutes, but you believe that it is longer.
    - $\boldsymbol{H_0}$: $\boldsymbol{\mu}$
    - $\boldsymbol{H_1}$: $\boldsymbol{\mu}$
- 2) A psychologist believes that watching 9 or more hours of football a week reduces men's self esteem. The population of men score an average of 40 points on a self-esteem questionnaire.
    - $\boldsymbol{H_0}$: $\boldsymbol{\mu}$
    - $\boldsymbol{H_1}$: $\boldsymbol{\mu}$
- 3) Researchers report that infants can listen for an average of 8 seconds before becoming distracted. A psychologist believes that with training, infants can improve their attention span.
    - $\boldsymbol{H_0}$: $\boldsymbol{\mu}$
    - $\boldsymbol{H_1}$: $\boldsymbol{\mu}$

Hypothesize it…

Write the null (H_0) and alternative (H_1) hypotheses for each scenario. Indicate whether it is a one (left or right tailed) or a two-tailed test.

- 1) An emergency room advertises a wait time of 10 minutes, but you believe that it is longer.
  - $\boldsymbol{H_0}: \mu \leq 10$
  - $\boldsymbol{H_1}: \mu > 10$
- 2) A psychologist believes that watching 9 or more hours of football a week reduces men's self esteem. The population of men score an average of 40 points on a self-esteem questionnaire.
  - $\boldsymbol{H_0}: \mu \geq 40$
  - $\boldsymbol{H_1}: \mu < 40$
- 3) Researchers report that infants can listen for an average of 8 seconds before becoming distracted. A psychologist believes that with training, infants can improve their attention span.
  - $\boldsymbol{H_0}: \mu \leq 8$
  - $\boldsymbol{H_1}: \mu > 8$

Hypothesize it…

Write the null (H_0) and alternative (H_1) hypotheses for each scenario. Indicate whether it is a one (left or right tailed) or a two-tailed test.

- 4) The average amount of lead in water is 2.8 ppb, but you think that one city's lead water levels may be higher than 2.8 ppb.
  - H_0: μ
  - H_1: μ
- 5) It is suspected behavior modification will have an effect on the average number of sodas a person drinks (8) in a week.
  - H_0: μ
  - H_1: μ
- 6) You want to test whether a new smoking program actually reduces the average number of cigarettes smoked in a day (9).
  - H_0: μ
  - H_1: μ

Hypothesize it…

Write the null (H_0) and alternative (H_1) hypotheses for each scenario. Indicate whether it is a one (left or right tailed) or a two-tailed test.

- 4) The average amount of lead in water is 2.8 ppb, but you think that one city's lead water levels may be higher than 2.8 ppb.
  - $\boldsymbol{H_0}: \mu \leq 2.8$
  - $\boldsymbol{H_1}: \mu > 2.8$
- 5) It is suspected behavior modification will have an effect on the average number of sodas a person drinks (8) in a week.
  - $\boldsymbol{H_0}: \mu = 8$
  - $\boldsymbol{H_1}: \mu \neq 8$
- 6) You want to test whether a new smoking program actually reduces the average number of cigarettes smoked in a day (9).
  - $\boldsymbol{H_0}: \mu \geq 9$
  - $\boldsymbol{H_1}: \mu < 9$

Next Up…

- The rest of the steps to hypothesis testing…
  1. State the null and alternative hypotheses ✓
  2. Select a level of significance (α)
  3. Choose a statistical test
  4. Find the critical values
  5. Compute the test statistic and the p-value
  6. Formulate a decision (reject or fail to reject the null hypothesis)