Hypothesis testing, p-values, t-tests (part 1)

With your neighbor, come up with a graph of a cumulative distribution function for a fair 6-sided die.

Kind of distribution: uniform

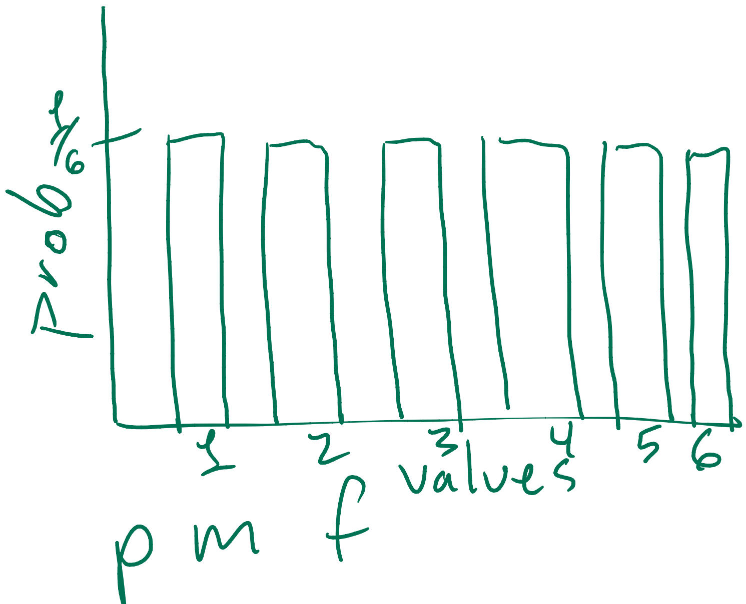
discrete vs. real: discrete

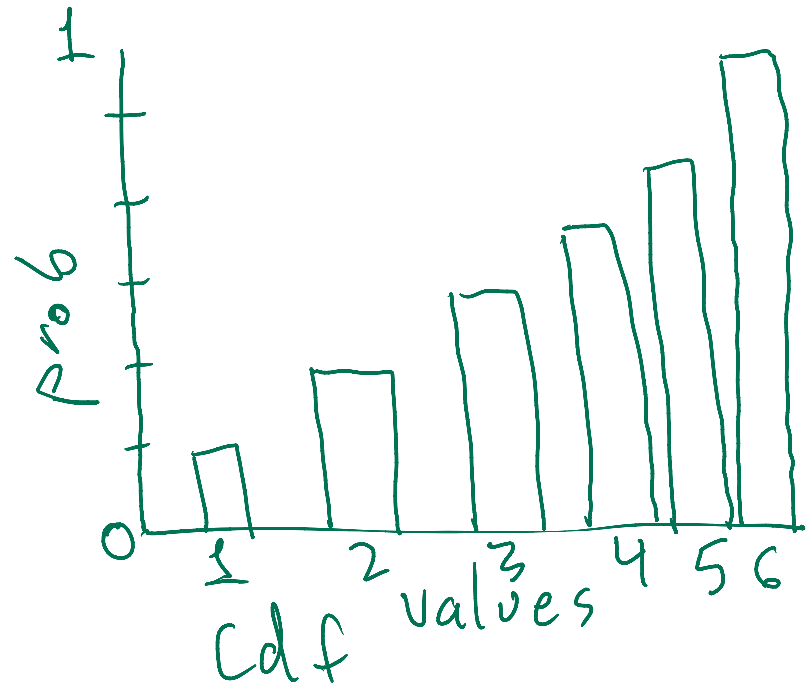
enumerable options $\{1, 2, 3, 4, 5, 6\}$

pmf vs. pdf: pmf (because X is discrete)

cdf for a die

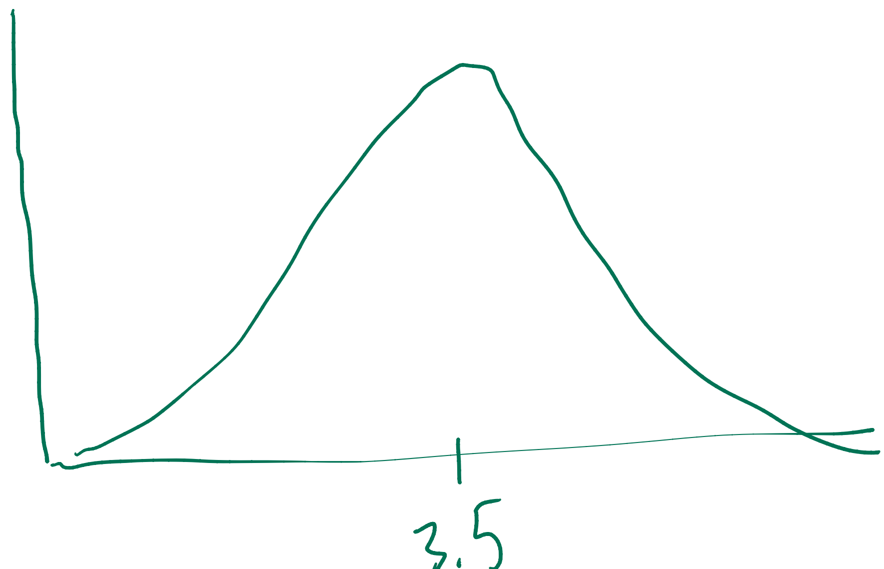
↳ prob that r.v. is ≤ a target value





central limit theorem:

↳ plot the means of samples of ind. r.v.s (same properties) → normal dist.



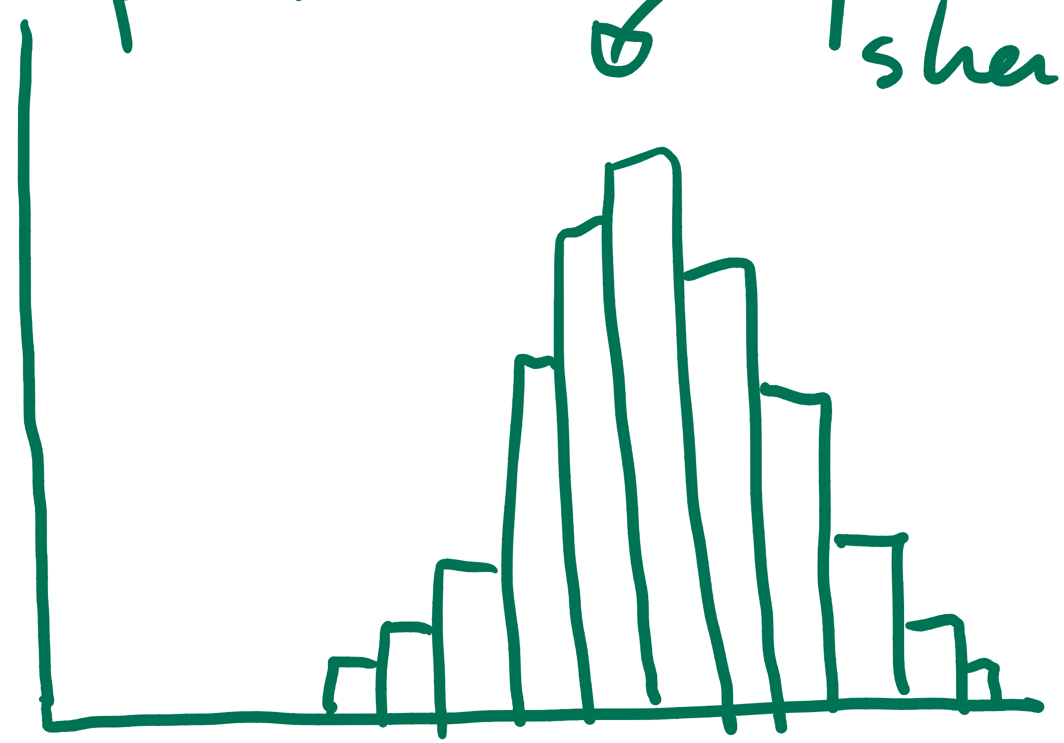
ICA Question 1: central limit theorem

Say that we are trying to determine the word accuracy rate for our newly developed voice assistant technology.

Samples

w_1	- - -	- - -	w_n	
✓	✓	✓	✓	1
✓	x	✓	✓	.75
✓	x	x	✓	.5
✓	x	✓	x	.5
✓	✓	✓	✓	1

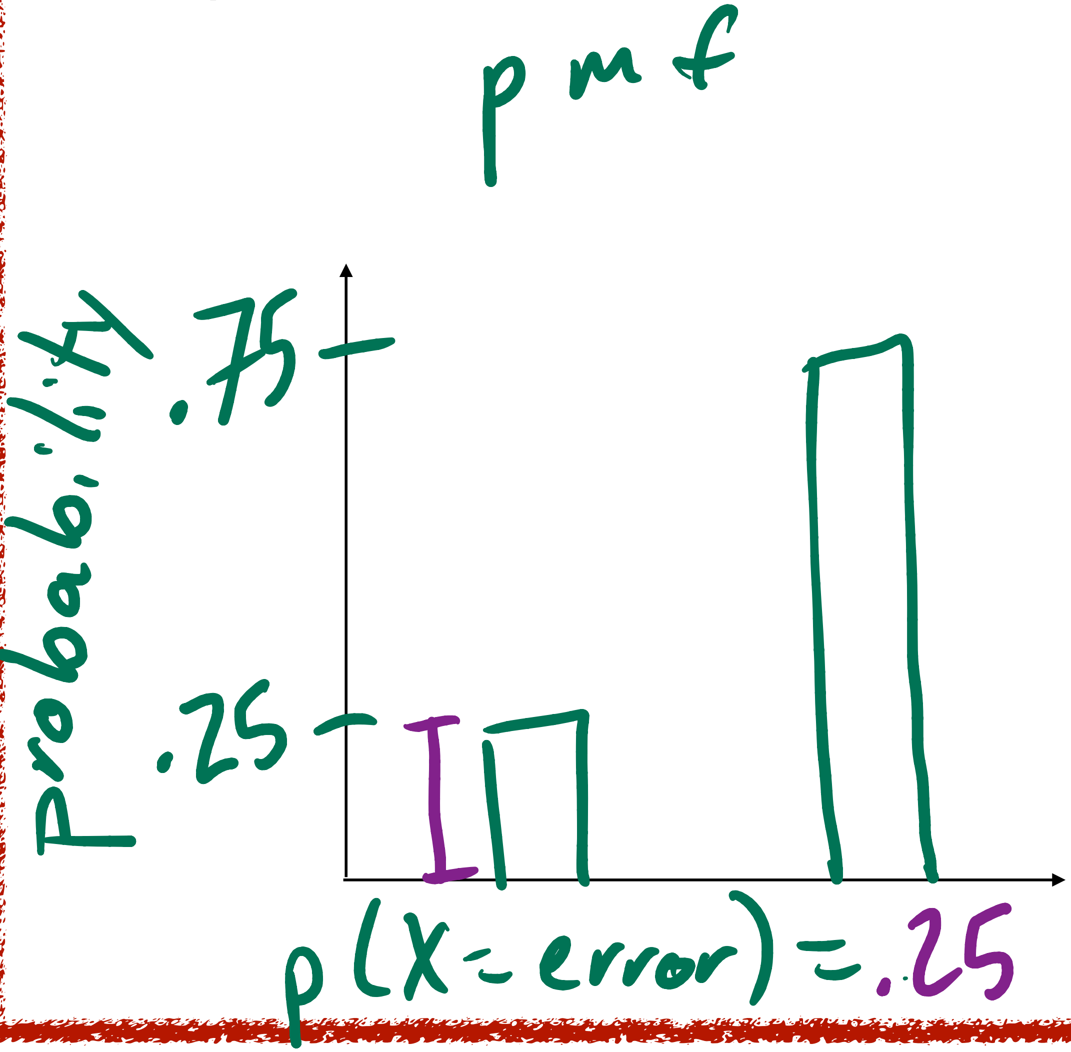
$p(w_{correct}) = .75$

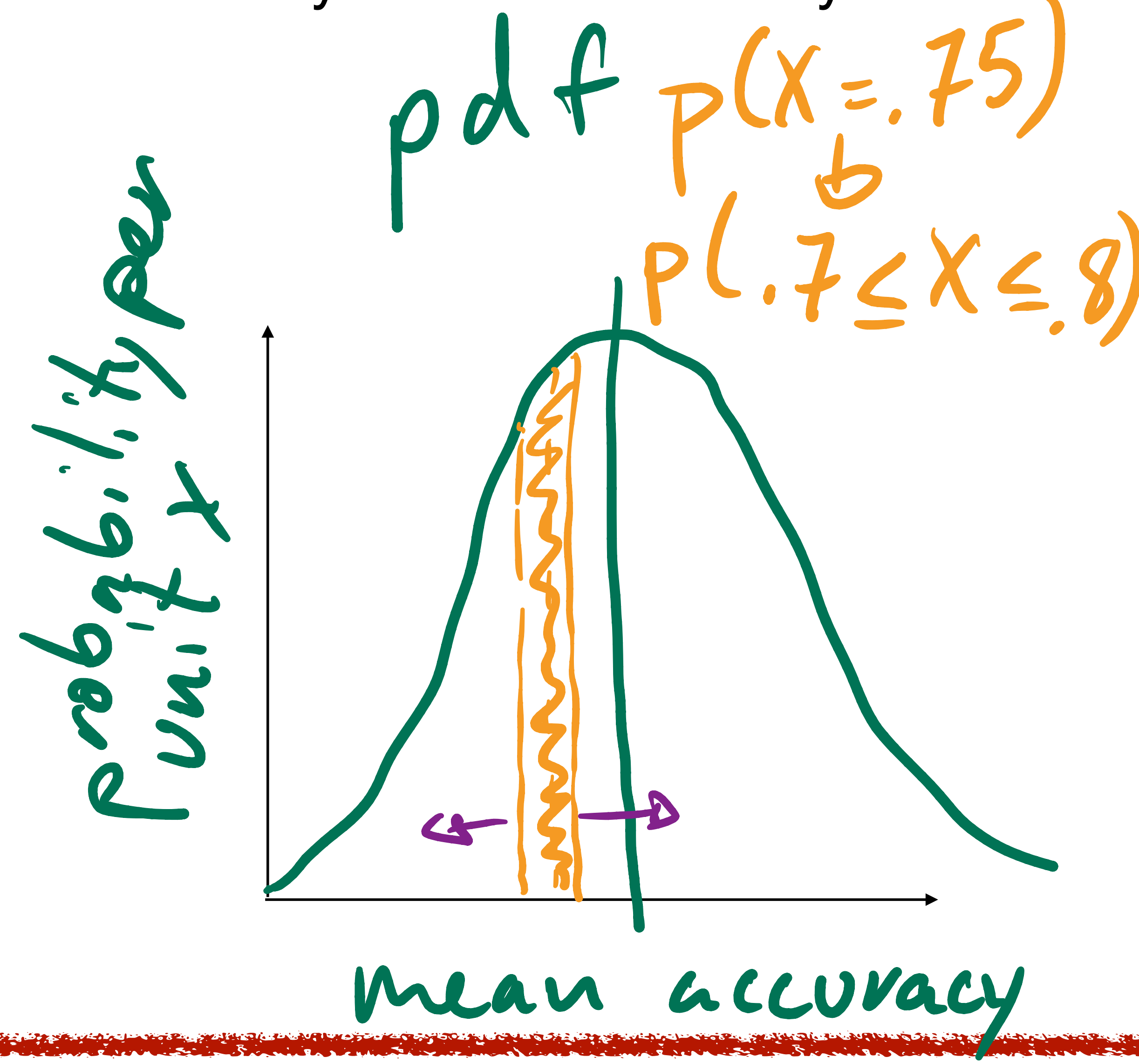


$p(0 \le X \le 1) = 1$

ICA Question 2: pmf vs. pdf

Say that we are trying to determine the word accuracy rate for our newly developed voice assistant technology.





Wait, why is the probability of a value for for a real-valued random variable 0?

- The practical part:

$P(x=.75) \neq 0$

$P(x=.751) \neq 0$

$P(x=.7513) \neq 0$

because infinite #s,

$\sum$ #s = infinity

non-zero, positive

$P(.745 \leq x \leq .755)$

- For probability **density** functions, we care about **area** for probability, and that for probability **mass** functions, we care about **height** for probability (y-axis)

- for pdfs, y-axis is probability per x-unit ("probability density")

for a non-zero prob. in practice

ICA Question 3: cdf vs. ppf

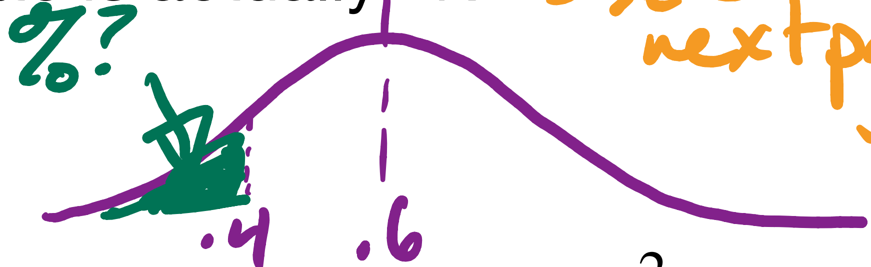
cdf → in: real val
out: %

ppf: in: %
out: real val

Say that we are trying to determine the word accuracy rate for our newly developed voice assistant technology.

↳ used the cent. lim. thm → normal dist.

Given a μ of .6 and a σ^2 of 0.05, what is the chance that the true word accuracy rate is actually .4? *→ see next pg

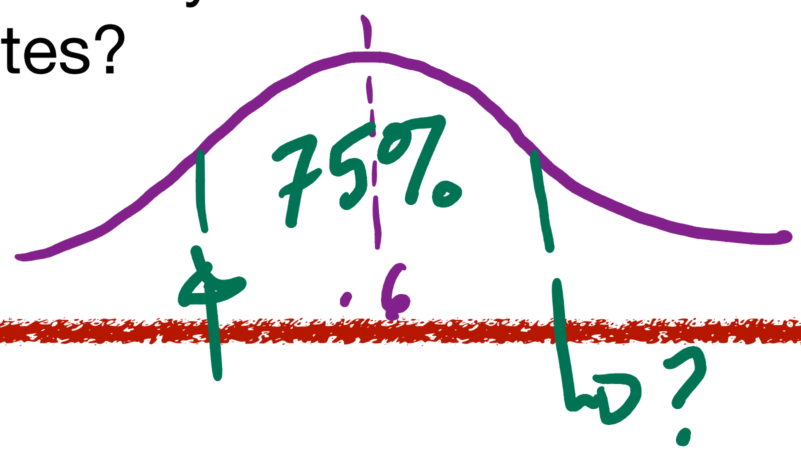


cdf
in: .4

$\sigma = .224$

norm.cdf(.4, μ, σ)

Given a μ of .6 and a σ^2 of 0.05, what is the **lower bound** for the true word accuracy rate if we want to claim that we are in the top 25% of possible accuracy rates?



norm.ppf(1-.25, μ, σ)

"exactly" .4

$\text{cdf}(.45, \mu, \sigma) - \text{cdf}(.35, \mu, \sigma)$

pdf(.4) → theoretically, yes

Hypothesis

- A **hypothesis** is a tentative assumption made in order to draw out and test its logical or empirical consequences
  - (Merriam-Webster)

→ bring this to math land, testing assumptions about distributions.

Hypothesis testing

- We'll be starting with a **question**
  - Is there a change in student test scores based on whether or not they listen to music beforehand?
- Next, we'll need to describe some **observations**

M - students who listened to music + test scores

S - students who didn't + test scores

- Then, we'll write down the **hypothesis** being tested

$H_1: \mu_M \neq \mu_S$

The null hypothesis

- The **null hypothesis**—H_0 is the hypothesis that there is no difference between the observed groups

- For example, given the question:

    - Is there a change in student test scores based on whether or not they listen to music beforehand?

    - with the hypothesis: $H_1 : \mu_{music} \neq \mu_{nomusic}$

    - the null hypothesis is $H_0 : \mu_{music} = \mu_{nomusic}$

The null hypothesis

- The **null hypothesis**—H_0 is the hypothesis that there is no difference between the observed groups

- For example, given the question:

    - Do students who eat strawberries for breakfast have higher test scores than students who don't?

    - with the hypothesis: $H_1: \mu_{strawbs} > \mu_{no\ strawbs}$

    - the null hypothesis is: $H_0: \mu_{strawbs} = \mu_{no\ strawbs}$

ICA Question 4: hypotheses

Come up with a **question**, what **observations** you'd need to answer it, and what your **hypothesis** is.

questions	observations	hypos
med. cure headache?	headaches w/ med w/o med	$\mu_{med} < \mu_{no\,med}$
gas mileage w/ gas types?	cars w/ prem cars w/ reg	$\mu_{prem} > \mu_{reg}$

→ issues to think about: diff. car types, gas source

P-values

- A **p-value** is the probability of observing test results that are at least as extreme as the results that were actually observed.

↳ ⚠ we don't use p-values to describe individuals!

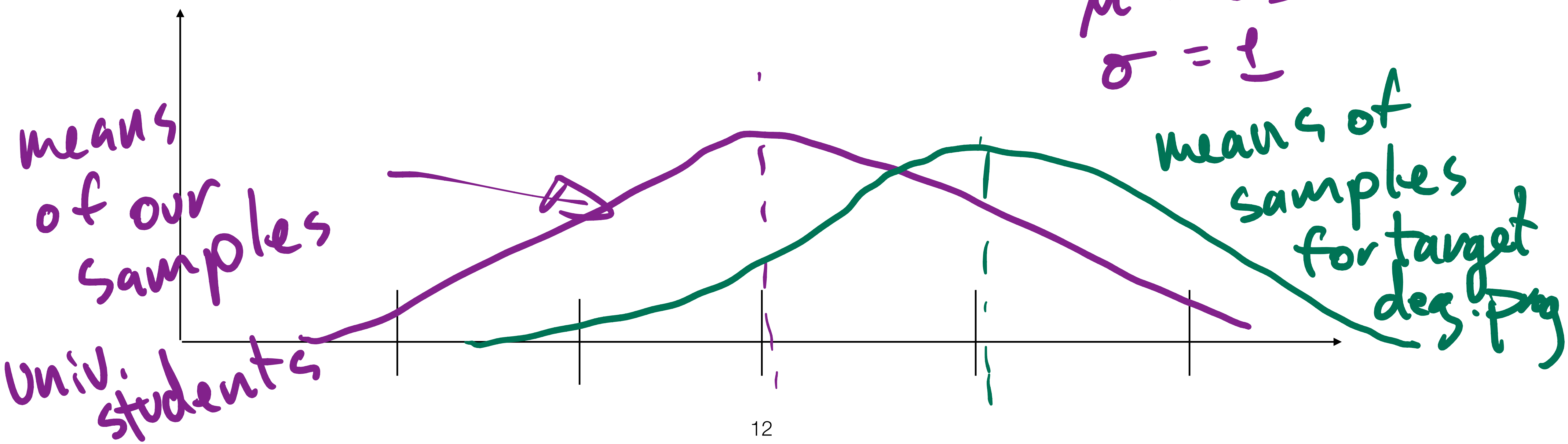
more extreme values

P-values

- Say that I want to know if a population of students in a certain degree program has a mean age that is significantly different than the mean ages of students in the university as a whole.

- First, we'll rely on the central lim. th., to build a distribution of mean ages of students in the university as a whole.

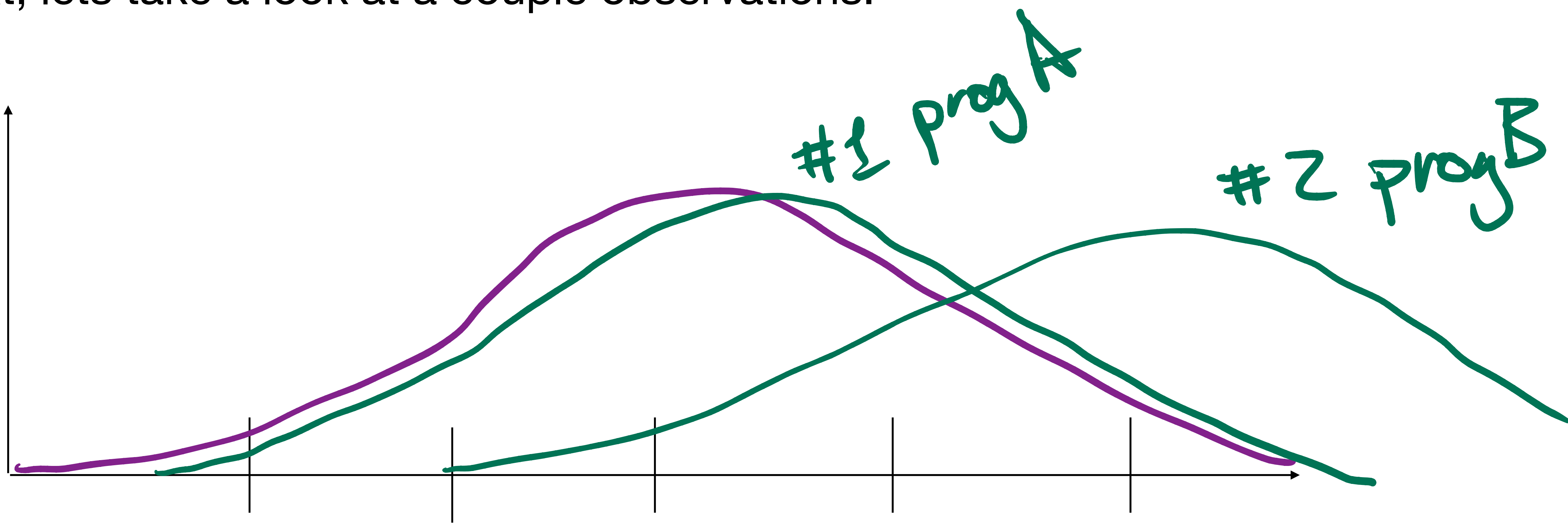
$\mu = 21$
$\sigma = 1$



P-values

- Say that I want to know if a population of students in a certain degree program has a mean age that is significantly different than the mean ages of students in the university as a whole.

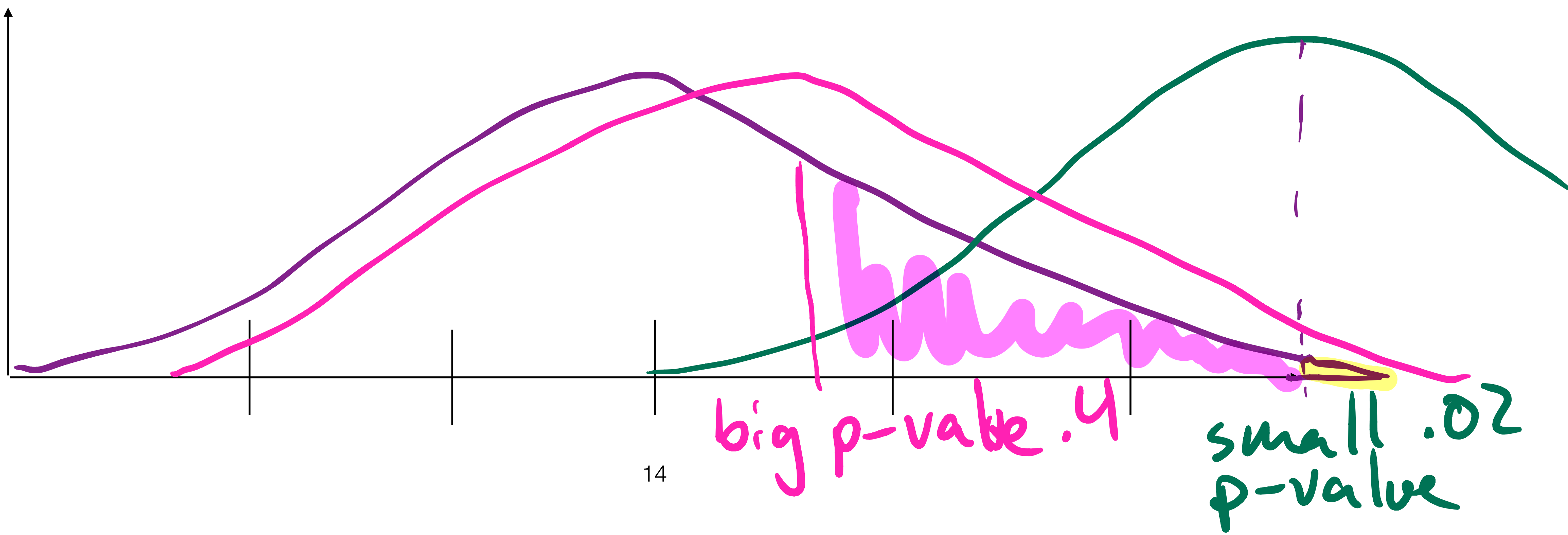
- Next, lets take a look at a couple observations:



P-values

- Say that I want to know if a population of students in a certain degree program has a mean age that is significantly different than the mean ages of students in the university as a whole.

- A larger p-value means that we are **more likely** to observe something that is **at least as extreme** as what we have observed.



P-values

- Say that I want to know if a population of students in a certain degree program has a mean age that is significantly different than the mean ages of students in the university as a whole.

- What do we need to calculate a p-value?

  - null hypothesis

  - test statistic

  - data /observations

test statistics

- Remember: overall goal is to be able to answer the question "is what I have observed meaningfully different than what I expect?" (vs. just due to random chance)

- We want to know if a coin is fair.
  - null hypothesis H_0: $P(\text{heads}) = 0.5 = P(\text{tails})$
  - test statistic count the heads
  - data - counts of heads in the sample

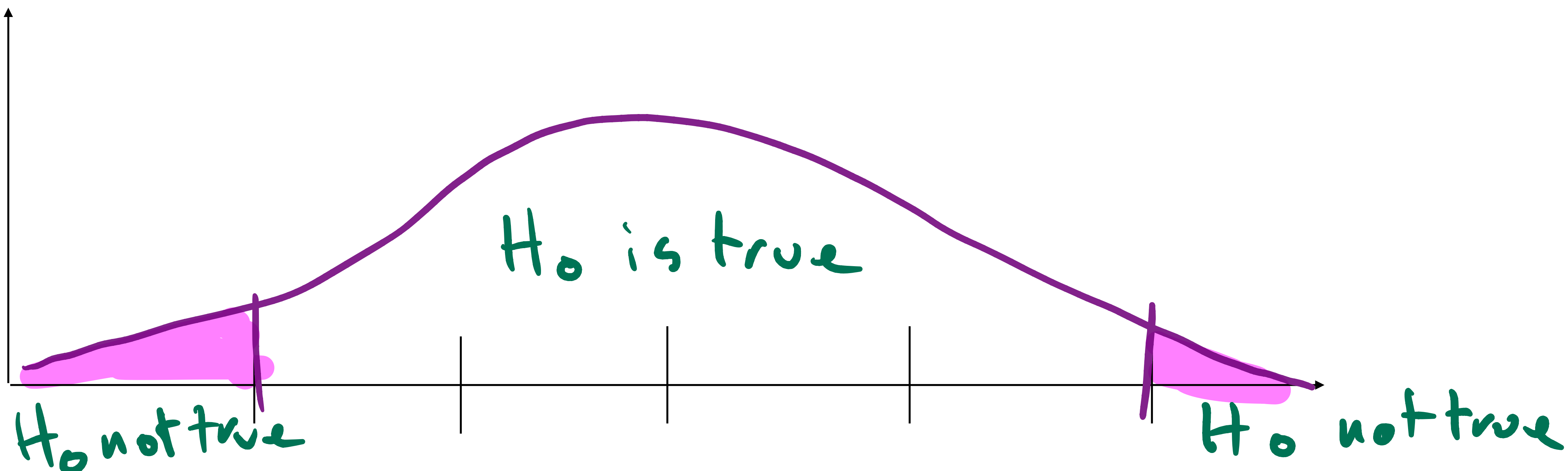
test statistics

- Remember: overall goal is to be able to answer the question "is what I have observed meaningfully different than what I expect?" (vs. just due to random chance)

- We want to know if a population has a different mean age than another population
    - null hypothesis $H_0 : \mu_{p_1} = \mu_{p_2}$
    - test statistic t-statistic (from t-test)
    - data ages of people in p_1
      ages of people in p_2

test statistics & tails

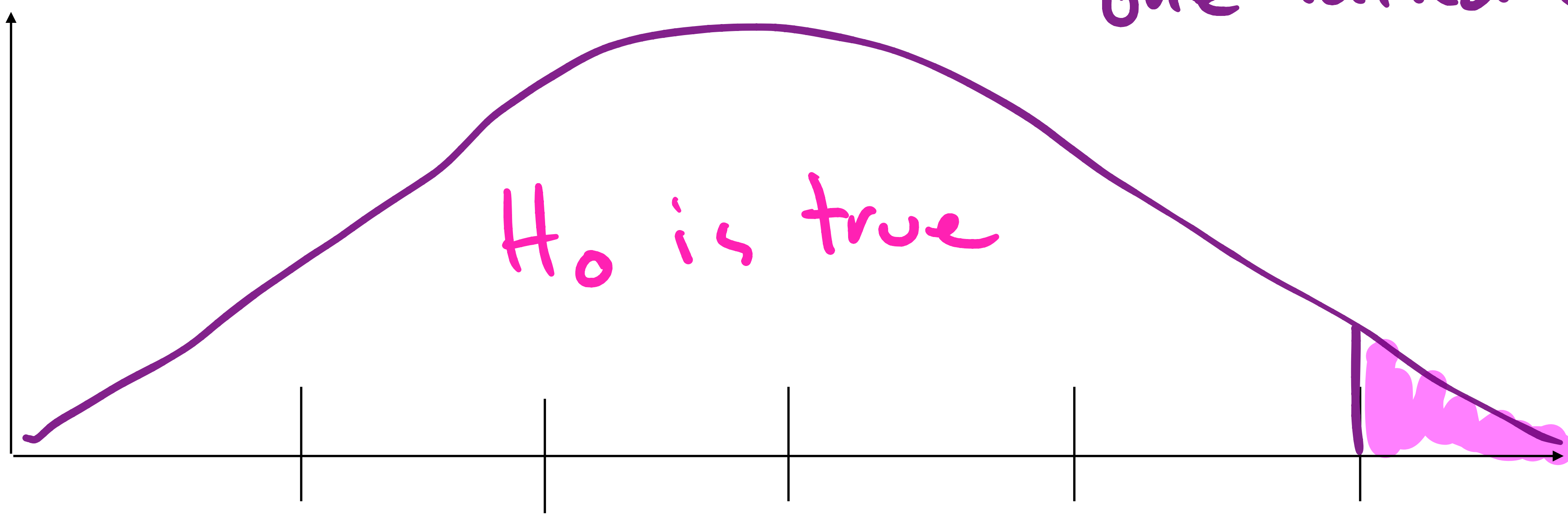
- We want to know if a population has a **<u>different</u>** mean age than another population
- **one** vs. **two** tailed tests refer to which part of the distribution we care about when performing significance testing



test statistics & tails

- We want to know if a population has a <u>**larger**</u> mean age than another population
- **one** vs. **two** tailed tests refer to which part of the distribution we care about when performing significance testing

one-tailed (right)



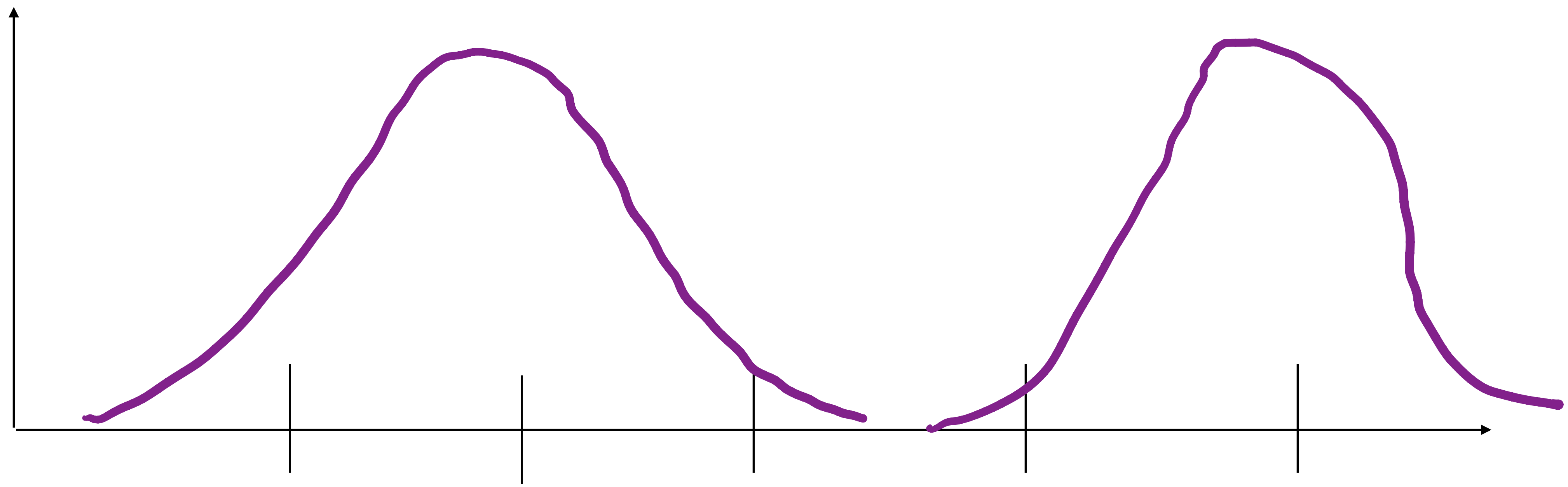
test statistics & tails

- We want to know if a population has a **smaller** mean age than another population

- **one** vs. **two** tailed tests refer to which part of the distribution we care about when performing significance testing

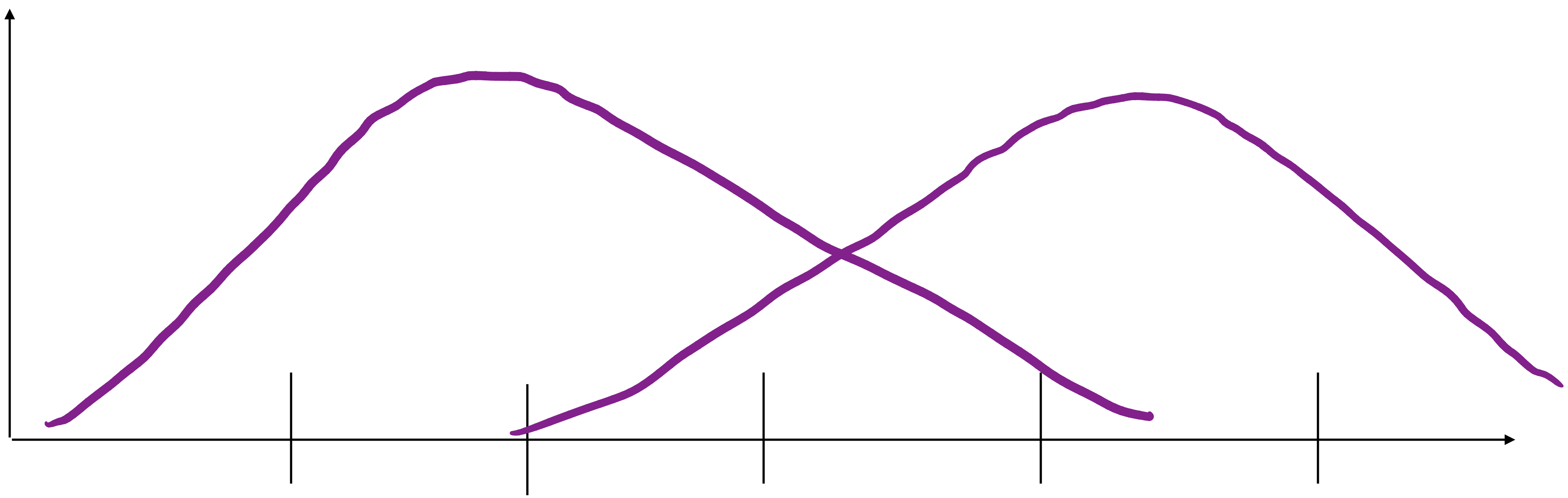
t-tests

- **Student's t-test** is the name of the test statistic that we'll use when we're trying to compare two continuous probability distributions that are normally distributed.



t-tests

- **Student's t-test** is the name of the test statistic that we'll use when we're trying to compare two continuous probability distributions that are normally distributed.



t-tests

- Before we go wild with t-tests on everything, there are a few requirements!
  - distributions should be normal
  - the two populations should have the same variance
  - data should be sampled independently

Future-you

- On Monday:
    - Actually calculating t-tests (and p-values)
      ↳ one ques on HW 7
    - errors
    - bias
      one ques on HW 7
    - mis-using p-values "harking"

Schedule

ICA : passcode : "cookie"

Turn in ICA 17 on Canvas (make sure that this is submitted by 2pm!)

HW 7 - available on the course website/canvas now. Due April 3rd. You will need some material from lecture on Monday!

Mon	Tue	Wed	Thu	Fri	Sat	Sun
March 21st Lecture 16 - normal distributions	**Felix OH Calendly** **HW 6 due @ 11:59pm**	**Felix OH Calendly**	**Felix OH Calendly** Lecture 17 - hypothesis testing HW 7 out			
March 28th Lecture 18 - t-tests, errors, experimental bias	**Felix OH Calendly** ~~HW 7 due @ 11:59~~	**Felix OH Calendly** 2pm Review (Zoom)	**Felix OH Calendly** Test 3			HW 7 due

More recommended resources on these topics

- p-values: YouTube, StatQuest: P values, clearly explained
- p-values: Wikipedia: https://en.wikipedia.org/wiki/P-value#Calculation
- Student's t-test, assumptions: https://en.wikipedia.org/wiki/Student%27s_t-test#Assumptions
- Student's t-test (we'll go over this in more depth on Monday): Youtube, Bozeman Science, Student's t-test