CS2810 Day 17
Mar 25

Admin:

- quiz3 next Friday
    - review session next week with Prof Felix
    - I'll be around Thursday at OH for review too

Content:

Hypothesis Testing

- Forming Hypothesis/Null Hypothesis pair
- P-values
- Errors:
    - type 1 (false alarm)
    - type 2 (missed detection)
- Deciding between hypotheses

What is Hypothesis Testing?

Hypothesis Testing is a use a statistical model to make a decision.
(there are other ways of statistical modelling which make decisions which aren't hypo test too!)

Research example
Self driving cars

DIE ROLL DEMO

Die rolling "prizes"

If student is able to roll a value higher than (or equal to) my own ...

... 1 time, then student gets a high five
... 2 consecutive times, then student gets a fist bump
... 3 consecutive times, then student gets two fist bumps
... 4 consecutive times, then student gets two high fives
... 5 consecutive times, then section will have an additional lowest ICA dropped from final grade
... 6 consecutive times, then student gets three fist bumps
... 7 consecutive times, then some students from front row get a fist bump
... 8 consecutive times, then section will have two lowest ICAs dropped from final grade
(including the prize for 5 consecutive times)

P-value is probability that result is as atypical as observations under the null hypothesis

(in demo, its the prob that Michael gets 8 in a row given Prof Higger is not cheating)

Hypothesis Testing:

A hypothesis describes some possibility of interest

A null hypothesis (of some hypothesis) contains all other possibilities.

Hypothesis:	Null Hypothesis: (example non complementary nu
Earth is heating up due to people	Earth isn't heating up due to people (Earth is not heating up)
The Utah Jazz (basketball) will win NBA champs this season.	The Utah Jazz won't win championship this year. (Boston Celtics will win NBA champs th

Null hypotheses are the default / assumed / natural state as compared to hypothesis.

Appropriate:

Hypothesis: Prof Higger didn't do demo fairly
Null Hypothesis: Prof Higger did demo fairly

Inappropriate:

Hypothesis: Prof Higger did demo fairly
Null Hypothesis: Prof Higger didn't do demo fairly

This is inappropriate as we should assume that Prof Higger does demo fairly by default.

ICA 1:

Wine drinkers claim that they know a good bottle of wine by its taste. Maybe its the case that they're just conditioned some other way in distinguishing good wine from bad (mirroring others preferences, the price of the bottle) and they can't tell the difference.

- Describe an experiment which is able to test whether a person can distinguish good wine from bad wine.
    - You may express your experimental design with a few sentences or as a comic. (comics preferred!)
    - Assume you have access to 8 bottles of ground truth good wine and another 8 bottles of ground truth bad wine
- Write a hypothesis / null hypothesis pair

Null Hypothesis: Proportion of wine drinkers which classify good/bad wine is 50/50
Hypothesis: Proportion of winedrinkers which classify good/bad wine is > 50%

experiment: ask wine drinker to sort wines into two groups (good and bad).

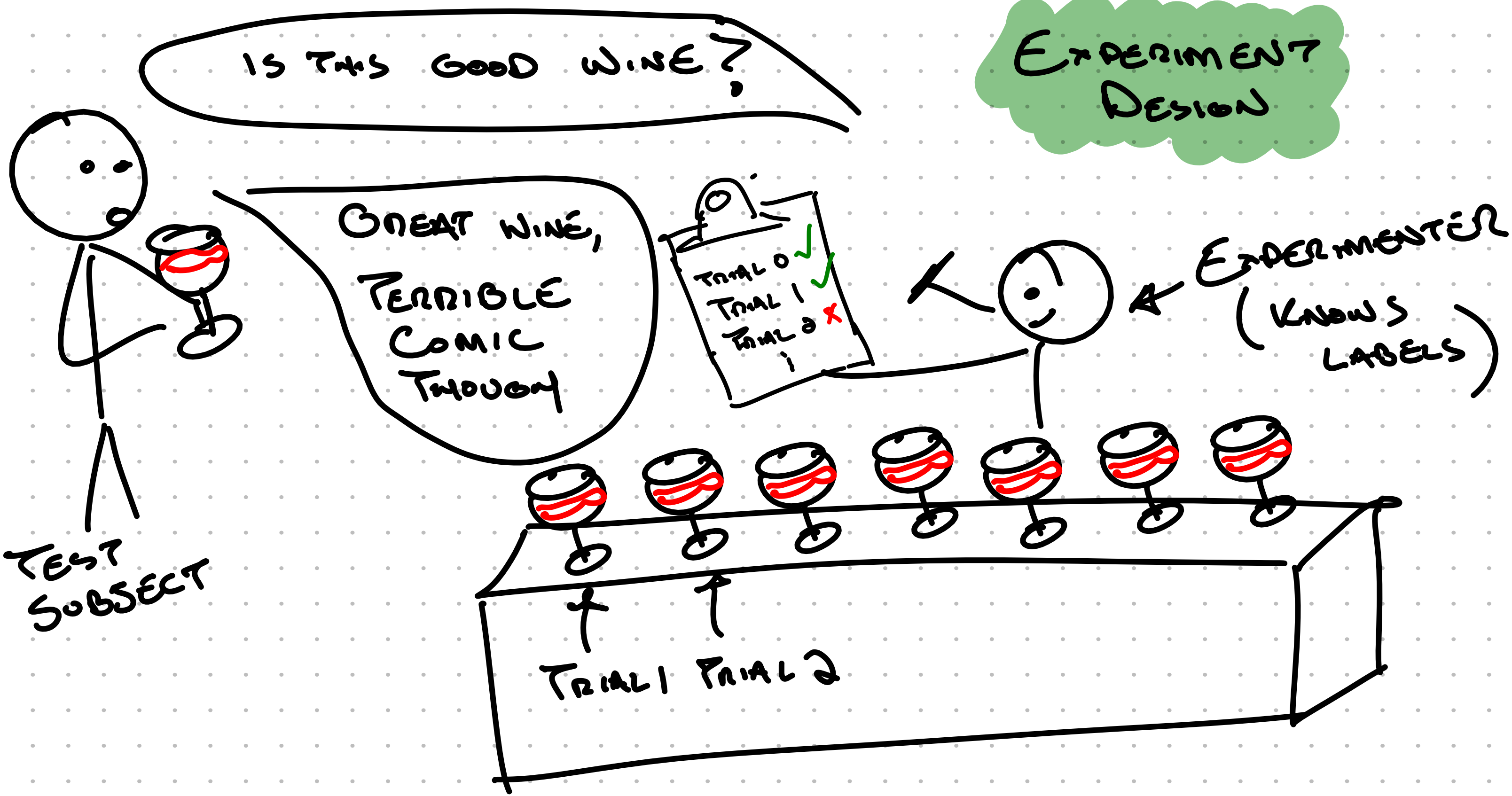
Null Hypothesis: the wine drinker will organize wine correctly
Hypothesis: the wine drinker will organize the wine incorrectly

Hypothesis (H_1):

test subject can distinguish between good and bad wine better than guessing.

Null Hypothesis (H_0):

test subject cannot distinguish between good and bad wine better than guessing.


EXPERIMENT DESIGN
IS THIS GOOD WINE?
GREAT WINE, TERRIBLE COMIC THOUGH
TRIAL 0 ✓
TRIAL 1 ✓
TRIAL 2 ✗
⋮
EXPERIMENTER (KNOWS LABELS)
TEST SUBJECT
TRIAL 1
TRIAL 2

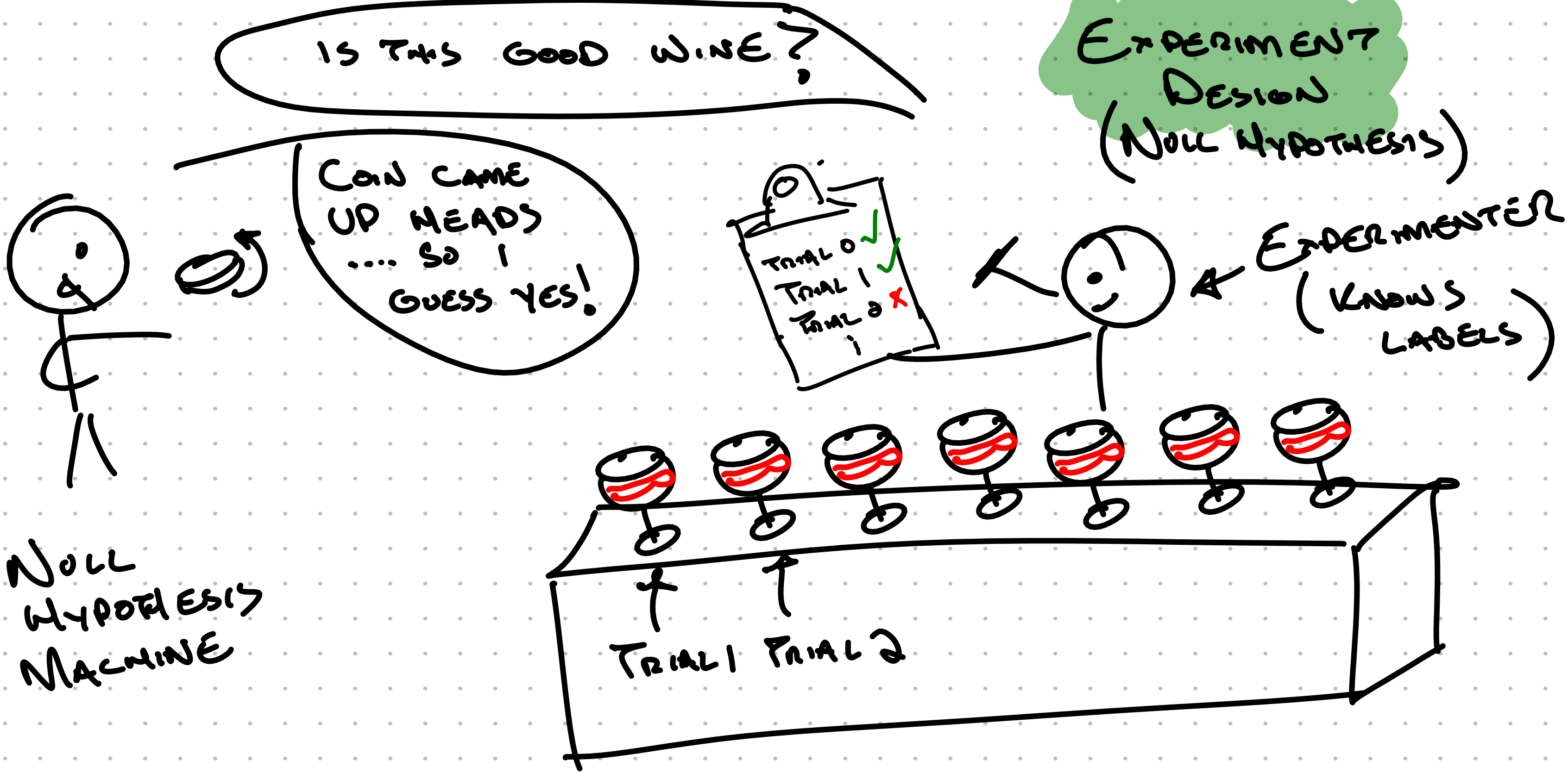
Result: 13 of the 16 trials were labelled correctly

How do we evalute this result to decide between the null hypothesis and hypothesis?

Compute the probability, given the null hypothesis, subject would have gotten 13 (or more) trials labelled correctly

... if this is very likely, then the null hypothesis may be true.
... if this is very unlikely, then the null hypothesis is probably not true.

Experiment Design (Null Hypothesis)



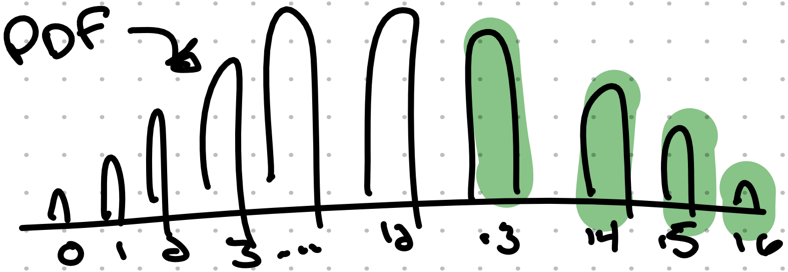
Result: 13 of the 16 trials were labelled correctly

How do we evalute this result to decide between the null hypothesis and hypothesis?

Compute the probability, given the null hypothesis, subject would have gotten 13 (or more) trials labelled correctly.
... if this is very likely, then the null hypothesis may be true.
... if this is very unlikely, then the null hypothesis is probably not true

LET X BE R.V. OF NUMBER OF GUESSES ONE GETS CORRECT IN EXPERIMENT.

$X \sim \text{Binom}(p=.5, n=16)$



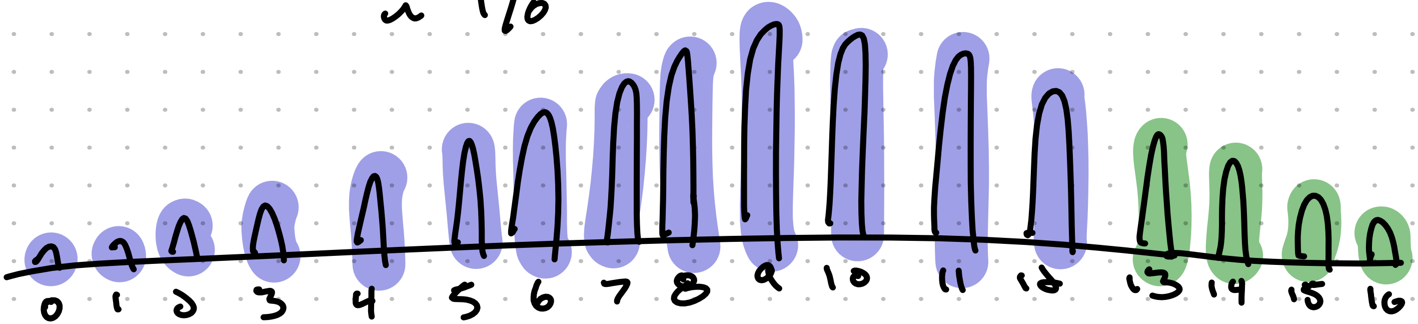
Python Discrete CDF Quirk

Notice: CDF includes boundary for Discrete

$$P(X \geq 13) = 1 - P(X \leq 12)$$

$$= 1 - \text{Binom.CDF}(x=12, n=16, p=.5)$$

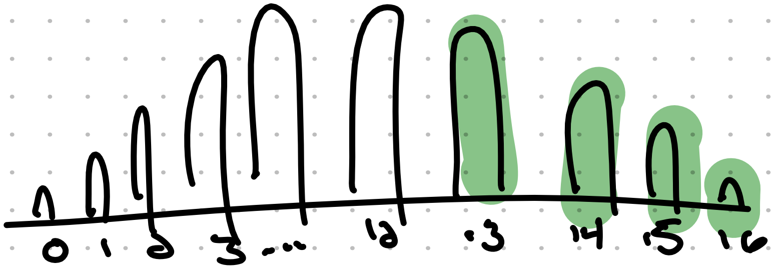
$$\approx 1\%$$



Result: 13 of the 16 trials were labelled correctly

How do we evalute this result to decide between the null hypothesis and hypothesis?

Compute the probability, given the null hypothesis, subject would have gotten 13 (or more) trials labelled correctly.
... if this is very likely, then the null hypothesis may be true.
... if this is very unlikely, then the null hypothesis is probably not true



$P($Result as atypical as our observation assuming null hypothesis$) = .01$

NULL HYPOTHESIS

HYPOTHESIS

HOW DO WE DECIDE BETWEEN H_0 / H_1?

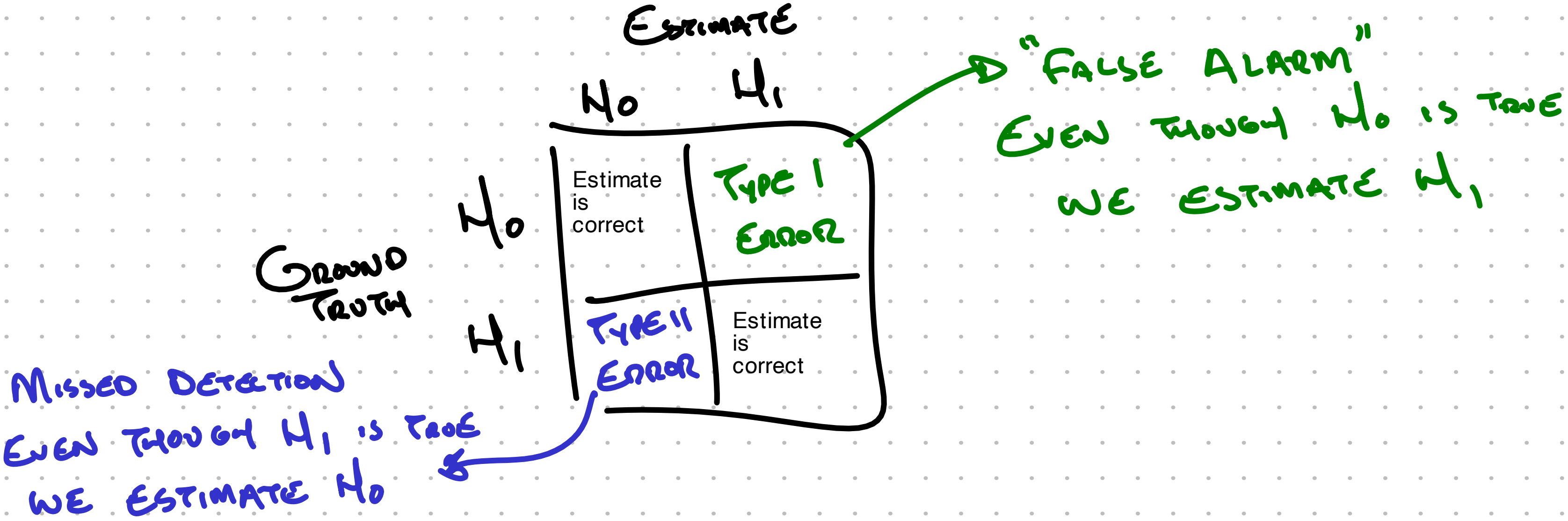
Error Types

Hypothesis (H_1):

test subject can distinguish between good and bad wine better than guessing.

Null Hypothesis (H_0):

test subject cannot distinguish between good and bad wine better than guessing.



		ESTIMATE	
		H_0	H_1
GROUND TRUTH	H_0	Estimate is correct	TYPE I ERROR
	H_1	TYPE II ERROR	Estimate is correct

"FALSE ALARM"
EVEN THOUGH H_0 IS TRUE
WE ESTIMATE H_1

MISSED DETECTION
EVEN THOUGH H_1 IS TRUE
WE ESTIMATE H_0

P-value: The probability that an outcome as atypical as our observation would occur under the null hypothesis.

(In our example, this is the prob that 13 or more trials are correct given the person is guessing if wine is good or bad ... we computed it as .01)

A high P-value means ...

RESULT IS CONSISTENT W/ NULL HYPOTHESIS

A low P-value means ...

RESULT IS INCONSISTENT W/ NULL HYPOTHESIS

Deciding between our hypotheses:

If p-value < .05 then we reject the null hypothesis in favor of our hypothesis.
(e.g. we believe a person can distinguish good / bad wine)

If p-value >= .05 then we do not reject the null hypothesis.
(e.g. we don't have enough evidence to claim a person can distinguish good / bad wine)

Using this rule:
- 5% of the time we'll reject the null hypothesis even though its true (type 1 / false alarm error)
- no promises about how often we'll have type 2 errors (missed detections)

alpha = .05 is the most common p-value rejection threshold, but others may be used too.

WE NEVER ACCEPT
THE NULL HYPOTHESIS

ICA 2

Playing a board game, one scores a "hit" if a 6-sided die roll yields a value of 4 or higher. After 20 rolls, a player has scored 17 hits. Another player is concerned that the die being rolled is not a fair (uniform) 6 sided die.

To determine if the die is fair, perform a hypothesis test:

- define hypothesis / null hypothesis
- compute p-value
    (remember: p-value is the probability an outcome as astypical as observations has occured)
- say if you reject / do not reject the null hypothesis
    - use alpha = .06 as your rejection threshold
    - in one sentence, tell what (if anything) the alpha threshold implies about the probability of type 1 and type 2 errors
- interpret the results in one sentence which may be understood by one who doesn't study statistics

P-value: The probability that an outcome as atypical as our observation would occur under the null hypothesis.

H_1 Hypothesis: die is not uniform
H_0 Null hypothesis: die is uniform

17 hits out of 20

Assuming the null hypothesis, then total number of hits friend gets is:

Binomial(n=20, p=.5)

$$p - P(X \geq 17) = 1 - \text{BINOM.CDF}(x=16, n=20, p=.5)$$

$$= .00129$$

REJECT H_0, DIE IS NOT UNIFORM

One more tip to distinguish null hypothesis from hypothesis:

If $\text{pval} < .05$ we reject the null hypothesis (claiming hypothesis is true)

if $\text{pval} >= .05$ we don't reject the null hypothesis (no claims made)

choose the null hypothesis so that we are able to make the claim we're interested in.