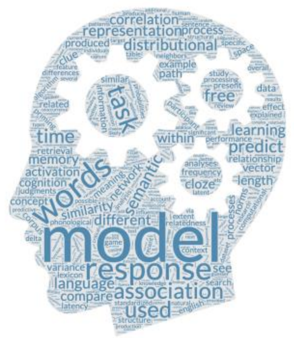
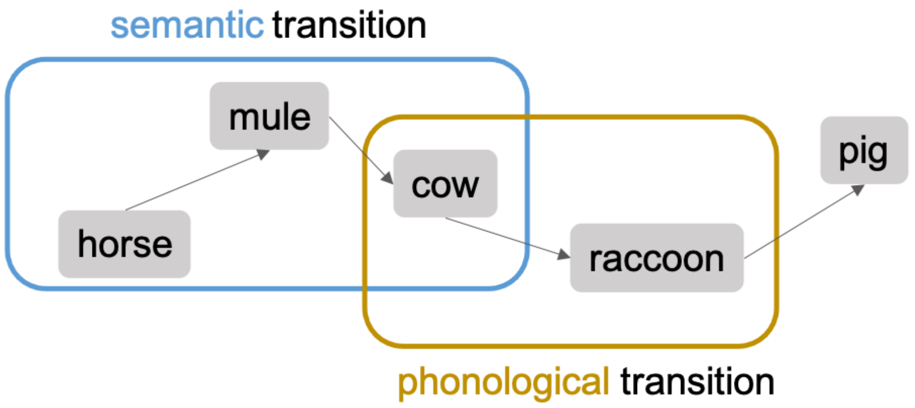
DATA ANALYSIS

Week 7: Hypothesis testing continued…

summer research @ lexicon lab



- exploring individual differences in memory search
- 1-2 students, funded by NSF
- 8-10 weeks (40 hours/week)
- Python coding (pandas, numpy, functions) experience required
- reach out by **March 24** latest
- read lab website and then email with requested info (a.kumar@bowdoin.edu)



students: I'm interested! how do I reach out?

We are especially interested in students who have programming experience (web development, game development, and machine learning), statistical skills (knowledge of R/Python), prior research experience (in psychology, mathematics, neuroscience, or digital/computer science), and a strong desire to learn!

If you are interested in working in the lab, please read the information below carefully and then email the PI, Abhilasha Kumar, with:

1. Which position and project are you applying for? Please review the open positions and SPECIFY which position + project you're interested in and how your skills map onto the required experience.
2. A brief description (1 page) or CV highlighting your relevant coursework, skills, and research experiences
3. Evidence of past research/coding experience (a paper you wrote / analysis you conducted / code you wrote)
4. A short summary (250 words) of any *one* paper from our publications page, with your key takeaways from the paper

today's agenda

- midterm discussion
- errors in hypothesis testing
- assumptions and extensions

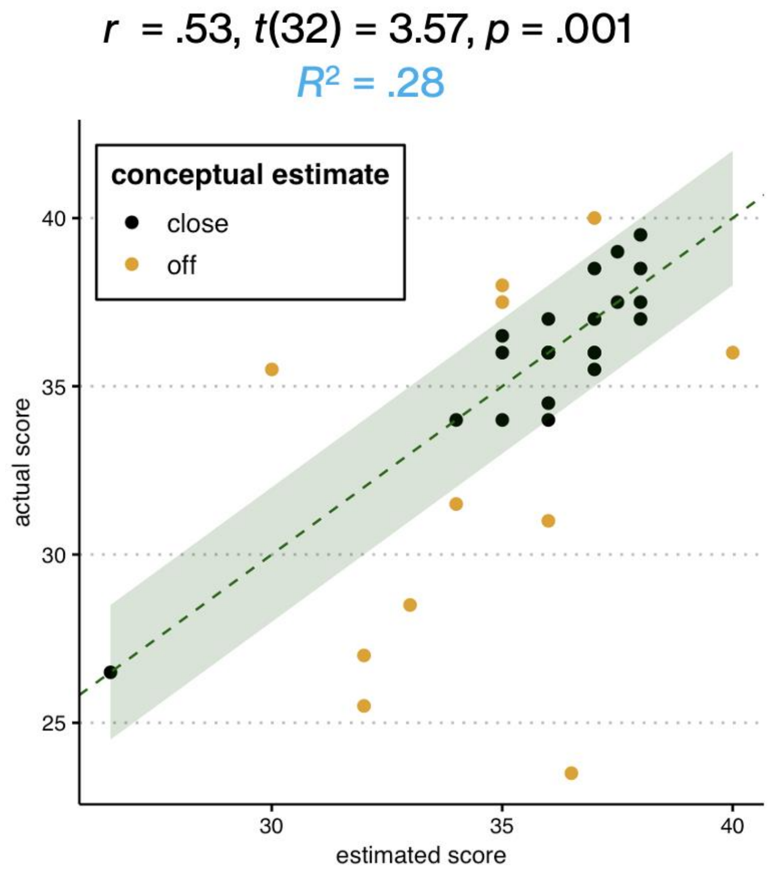
come chat!

- midterm grades available
- <u>everyone</u> schedule a 15-minute meeting post spring break / tomorrow [calendly link]
- fill out **anonymous** mid semester survey [opens on Friday]

7	T: March 4, 2025	W7: Sampling and Hypothesis Testing
7	Th: March 6, 2025	W7 continued…
7	F: March 7, 2025	**PS3 revision due**
7	F: March 7, 2025	**Week 7 Quiz due**
8	T: March 11, 2025	**Spring Break!**
8	Th: March 13, 2025	**Spring Break!**
9	T: March 18, 2025	**Spring Break!**
9	Th: March 20, 2025	**Spring Break!**
10	T: March 25, 2025	W10: Modeling Relationships
10	Th: March 27, 2025	W10 continued…
10	Su: March 30, 2025	**Week 10 Quiz due**
11	M: March 31, 2025	**PS4 due / Opt-out Deadline 2**
11	T: April 1, 2025	W11: Special Cases
11	Th: April 3, 2025	W11 continued…
12	M: April 7, 2025	**PS5 + PS4 revision due**
12	T: April 8, 2025	W12: Loose Ends / Exam 2 review
12	Th: April 10, 2025	**Exam (Midterm) 2**

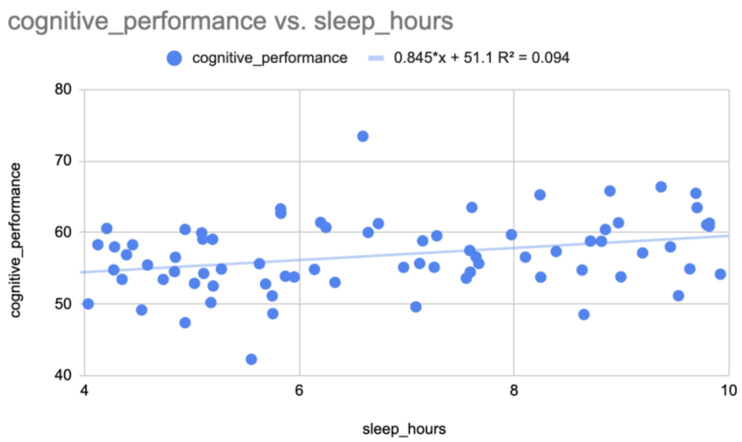
common misconceptions

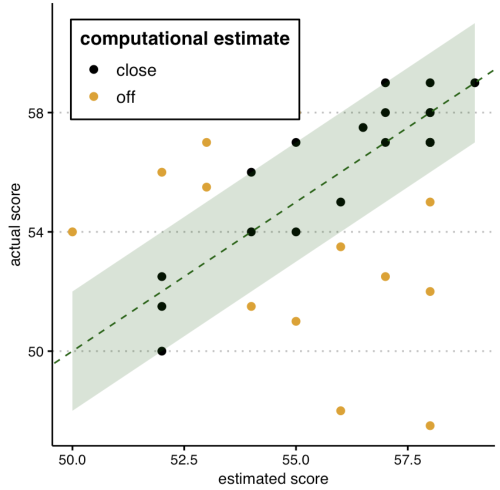
- does a significant p-value of 0.001 mean:
- the probability of the null hypothesis being true is .001?
- the probability that you are making the wrong decision is .001?



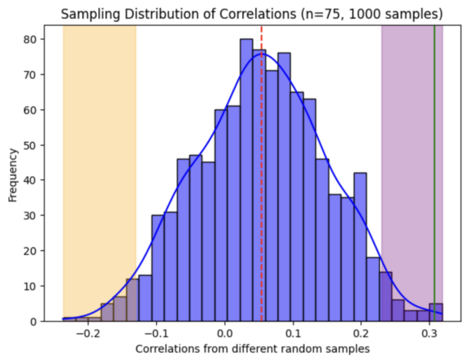
statistical significance ≠ practical significance!

- statistical significance only tells us the probability of the data under the null hypothesis; it does not tell us the *how important* the finding is or how practically significant this effect may be
- what are some reasons that these findings may not be practically significant?
- ultimately, practical significance depends on the **goals** and **drivers** of research

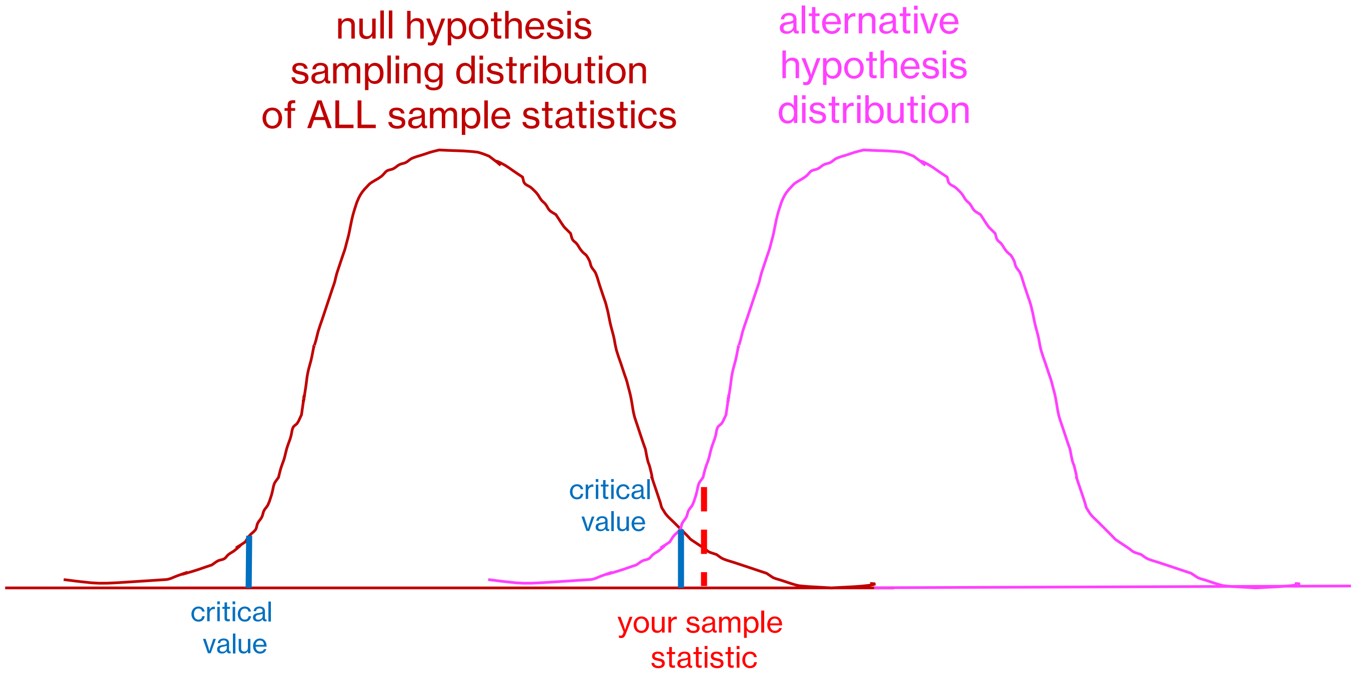


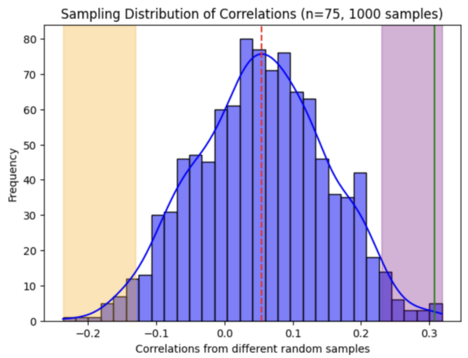


rejecting the null hypothesis



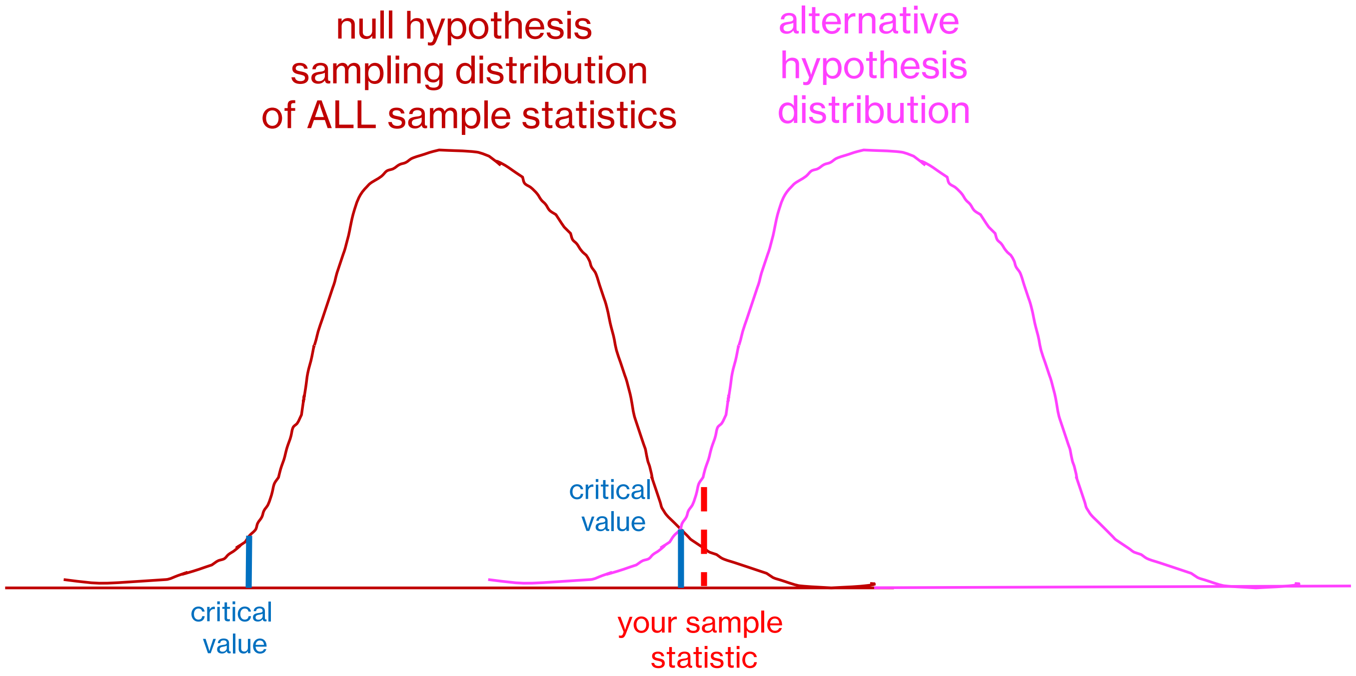
- when a null hypothesis is rejected based on a given α-level, you are making an inference that the sample statistic you have obtained is highly unlikely based on the null hypothesis, i.e., it is not likely to be part of the null distribution
- this implies that the sample statistic you have obtained is consistent with a ***different*** sampling distribution, i.e., the one suggested by the alternative hypothesis





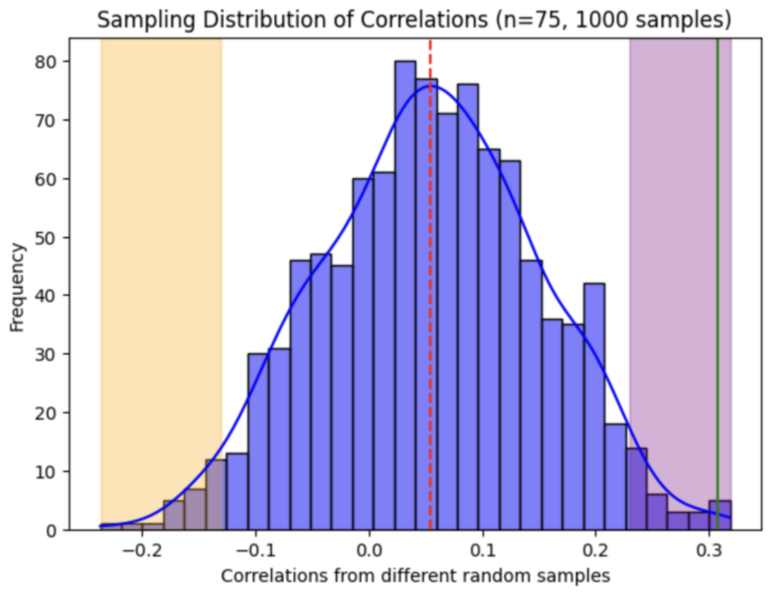
rejecting the null hypothesis

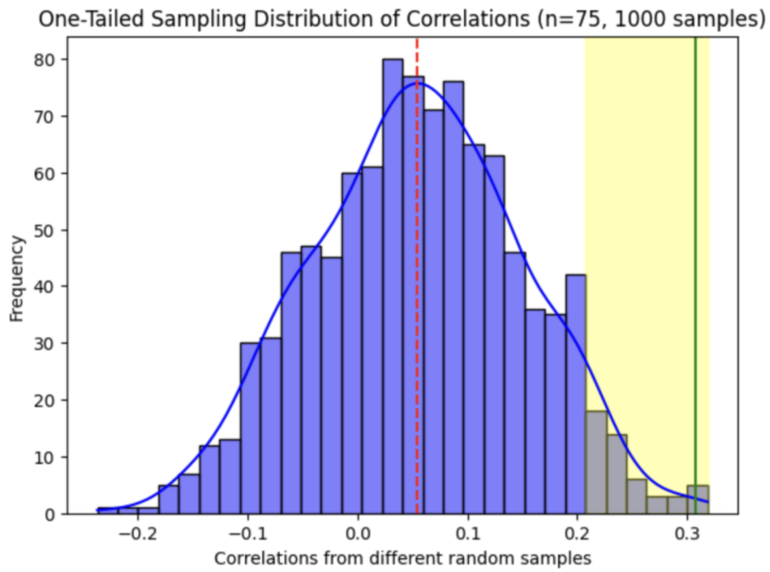
- BUT highly unlikely ≠ impossible!
- rare events are still possible events, i.e., your sample statistic could STILL be part of the null distribution
- For α=.05, there is a 5% chance of obtaining a rare sample that lies in the tails of the null distribution
- there is a 5% chance that you may be making an error when rejecting the null hypothesis in favor of the alternative hypothesis
- this is called a Type I error



one vs. two tailed test

- in a two tailed test, your α-level is split across two sides and your chance of making a Type I error is lower, compared to a one-tailed test
- lower values of α make it harder to reject the null hypothesis!
  - you are less likely to reject a null hypothesis in a two-tailed test, i.e., it is more stringent than a one-tailed test
- in practice, you are expected to report the two-tailed test even if you have a directional hypothesis





signal detection and hypothesis testing

- the process of evaluating the effect of a manipulation in experimental research is quite similar to differentiating a signal from noise, an idea with roots in signal detection theory with broad applications

	world truth	
your data	true effect	noise
effect found	hit	false alarm
effect not found	miss	correct rejection

activity: signal vs. noise

- groups of 2
- you will be presented with a research situation and you need to come up with a signal vs. noise formulation of the situation

activity: signal vs. noise

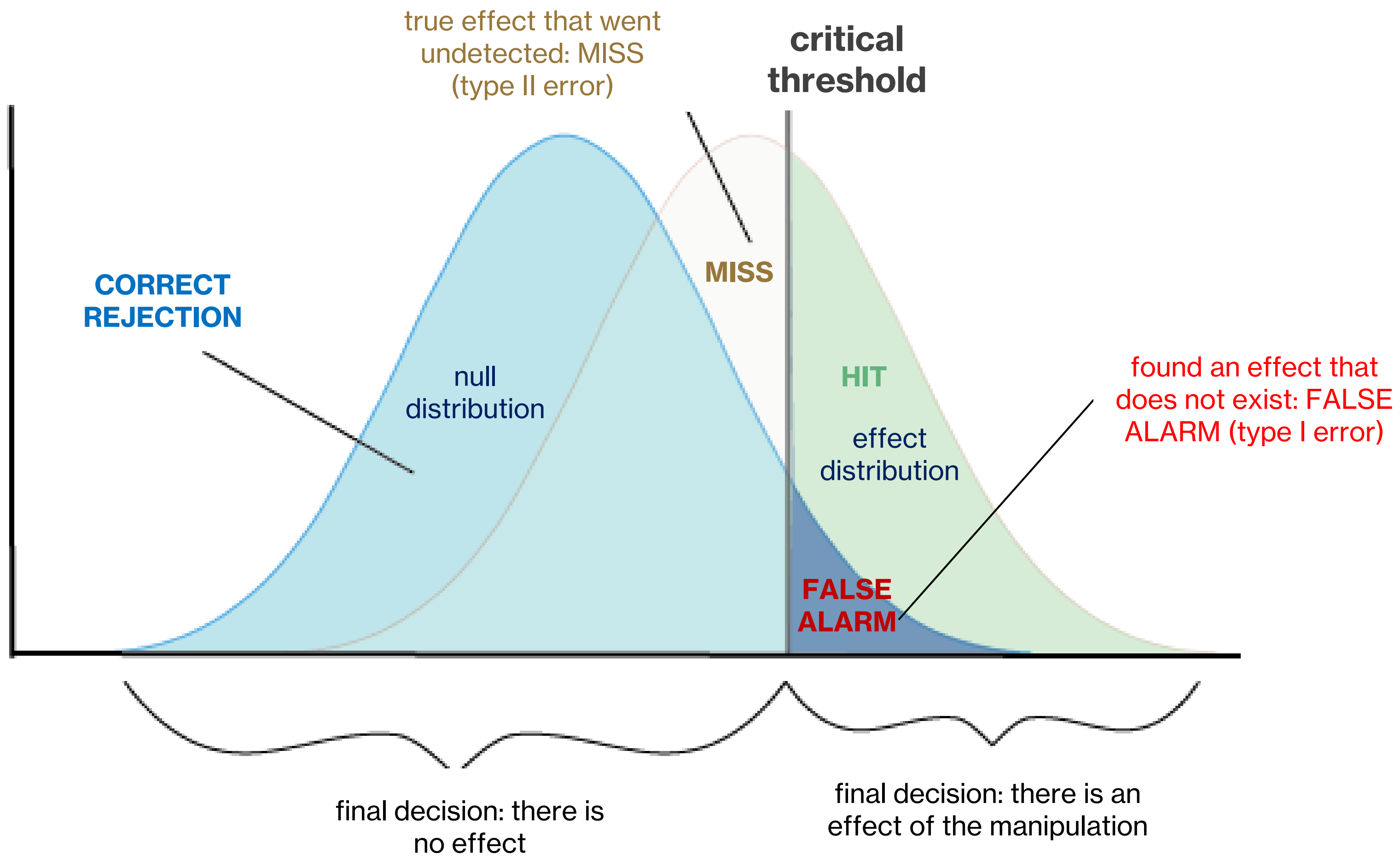
- a researcher is trying to understand if there are differences in the ability to manage money between freshmen and seniors. They conduct an experiment where they record the amount of money saved by both groups in a given month. Their hypothesis is that freshman and seniors save different amounts of money.

activity: signal vs. noise

- construct the signal vs. noise table for this experiment design

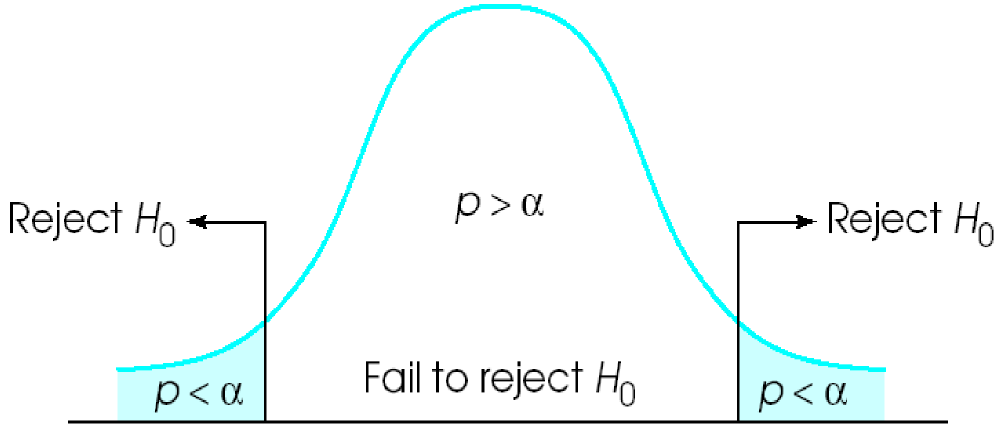
	world truth	
your data	true effect ???	noise ????
effect found ???	hit	false alarm
effect not found ???	miss	correct rejection

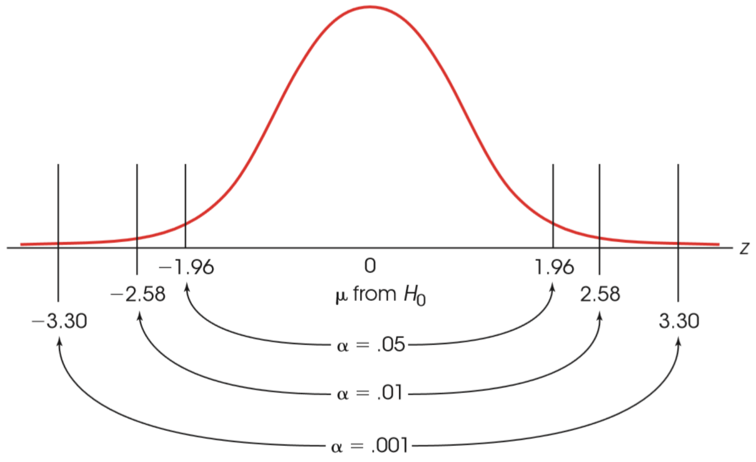
summary of errors



type I error: false alarms

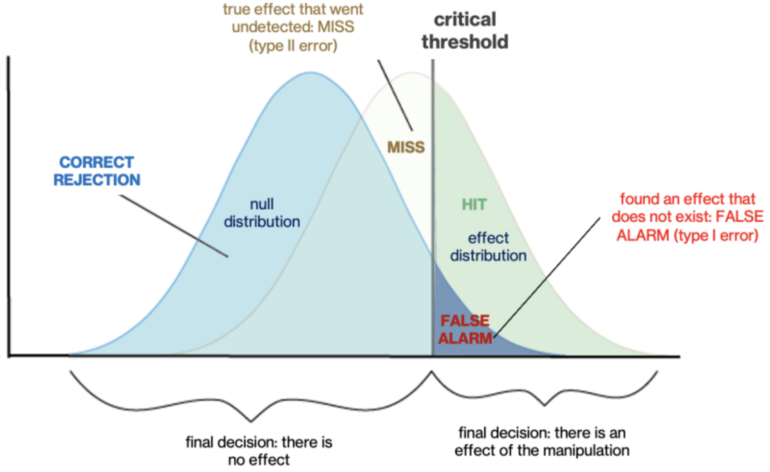
- type I errors occur when H_0 (null hypothesis) is rejected when it should not have been, i.e., your sample is simply a rare one within the null distribution
- what determines the rejection of H_0?
  - α (critical threshold): probability of making a type I error
  - lower the threshold (typically $\alpha < .05$) lower your chance of making a type I error
  - but not too low, else it will be impossible to find evidence to reject H_0
- sample size and type I error?

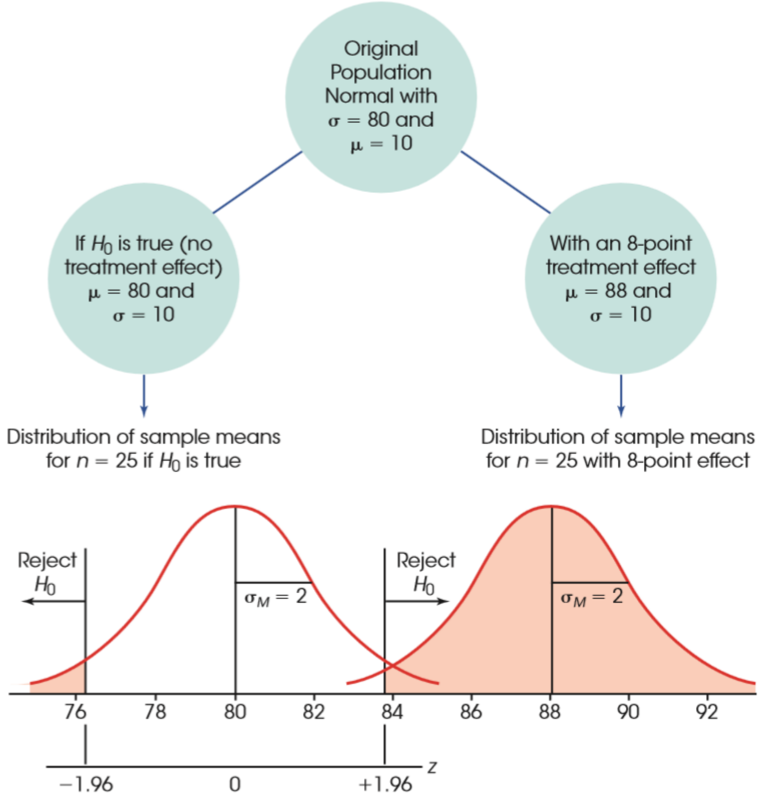




type II error: miss

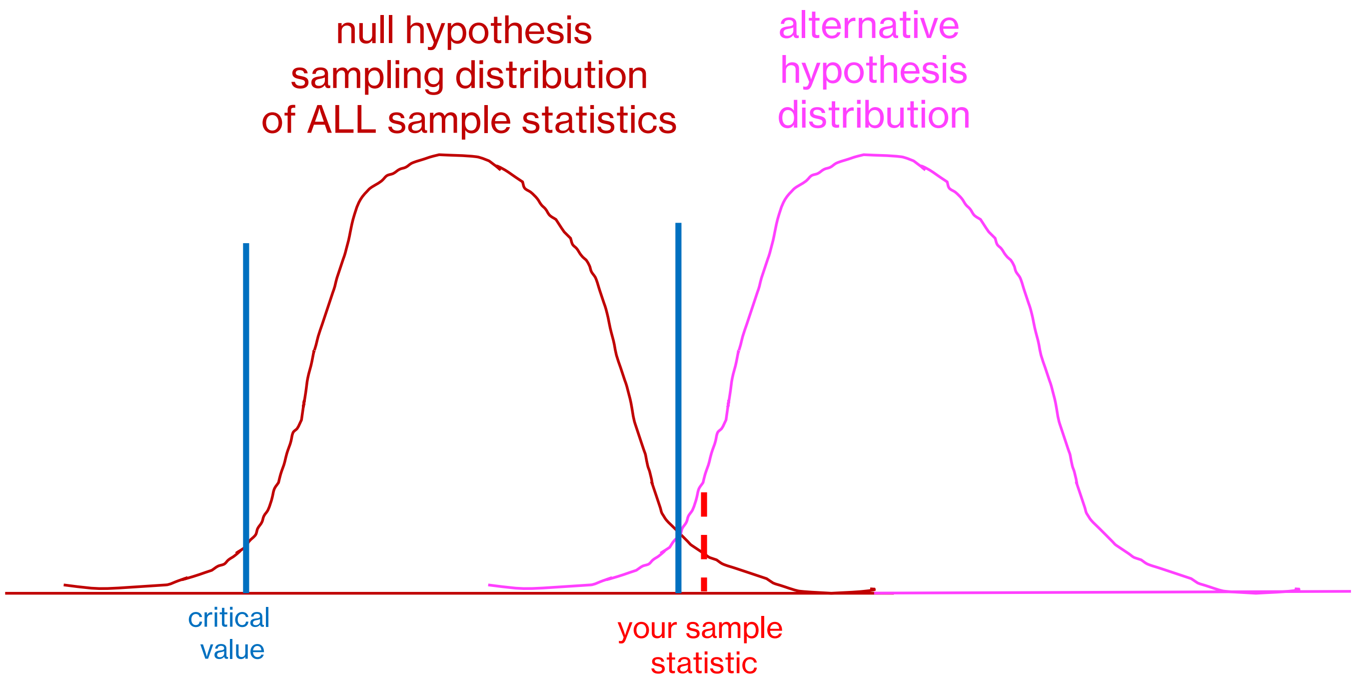
- type II errors occur when you fail to reject H_0 (null hypothesis) while there was an effect, i.e., the sample actually belonged to the alternative hypothesis distribution but was in the overlapping area with the null distribution
- β: probability of making a type II error
- power: $1 - \beta$, probability that a test will correctly reject the null hypothesis
  - typically calculated ***before*** data collection
  - depends on the (1) α level, (2) one vs. two-tailed test, (3) sample size, (4) effect size,





statistical power: α-level / tails

- increasing the α-level OR using a one-tailed test increases the power because we are more likely to reject the null hypothesis in both cases
- you are more likely to find the "atypical" case of the alternative hypothesis if you increase the α-level BUT also more likely to make a type I error!

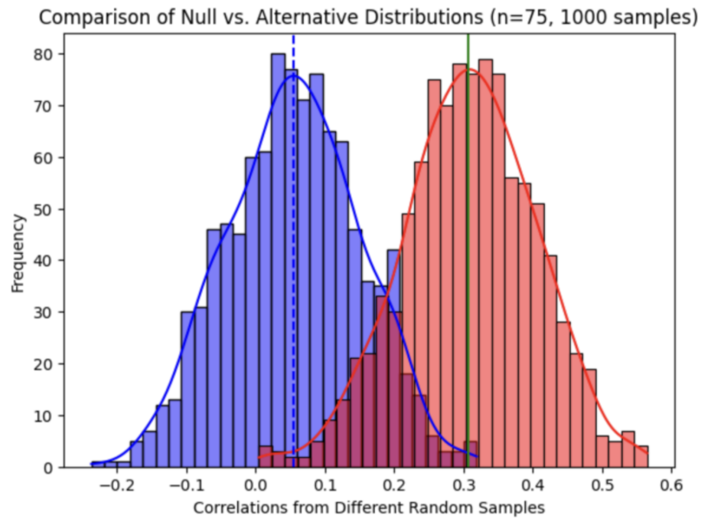


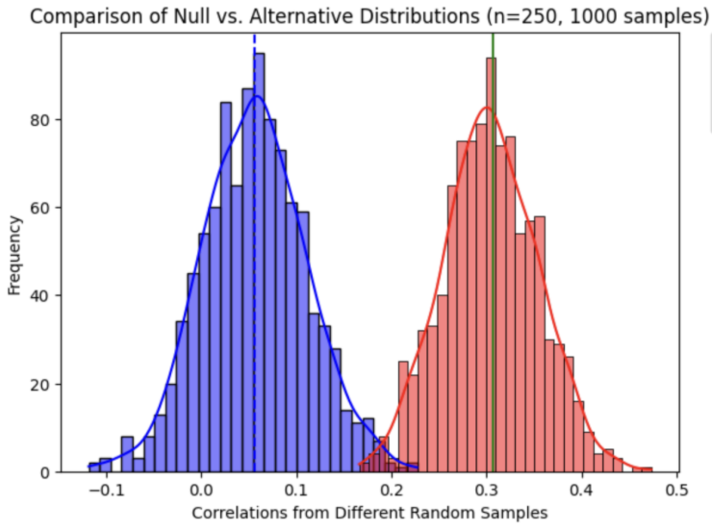
statistical power: sample size

- higher sample size decreases the standard deviation of the sampling distribution

- $SE_r = s_r = \sqrt{\frac{1-r^2}{n-2}}$

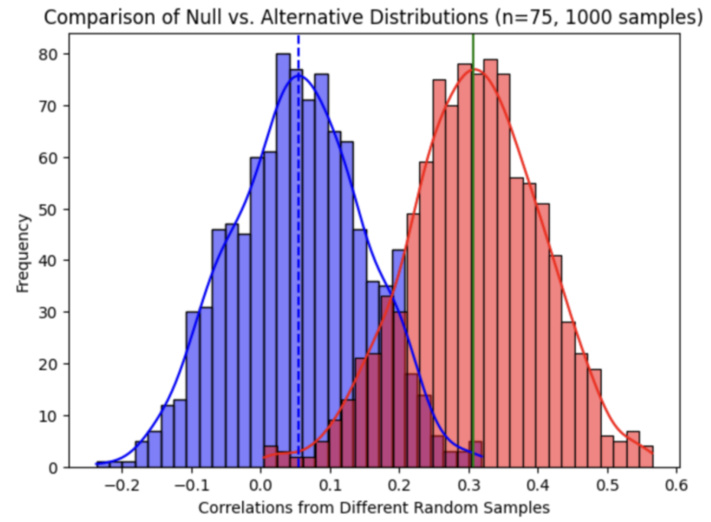
- this leads to narrower distributions, and decreasing the overlap between the two distributions, again leading to higher power to detect significant effects

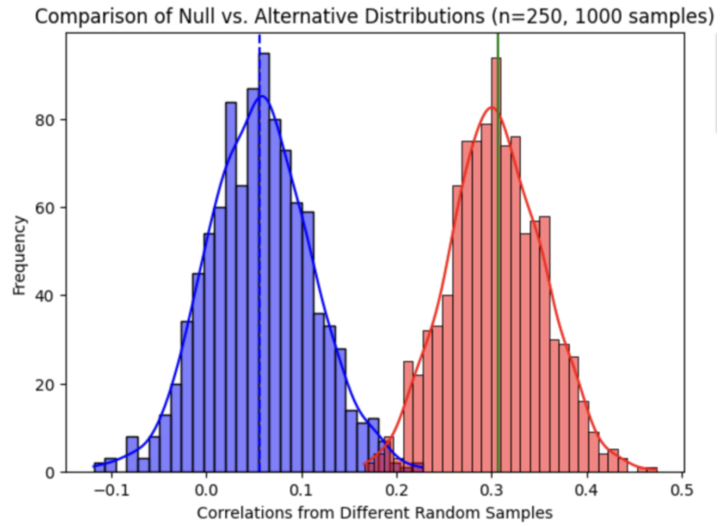




statistical power: effect size

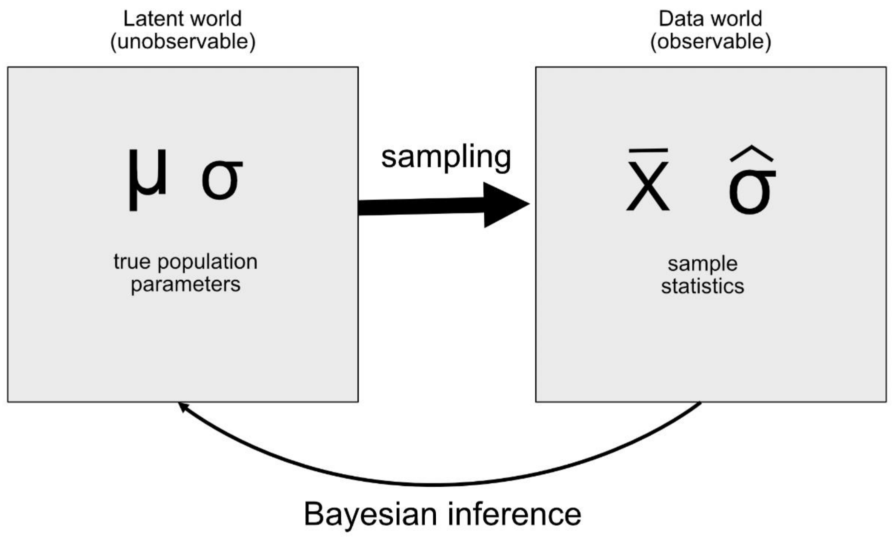
- if the separation between the two distributions is greater to begin with, power is higher
  - this is called effect size!
- effect size is an estimate of the difference between what would be expected by chance (null) vs. what is observed due to our manipulation (effect), irrespective of the sample size
  - for correlations, effect size = coefficient of determination!





other issues

- under NHST, we never actually test the likelihood of our hypothesis!
  - we obtain P (data | null hypothesis)
  - we want P (alternative hypothesis | data)
- NHST is limited because you cannot make actual inferences about the hypothesis you want to test
- alternative framework: Bayesian statistics!
  - Bayesian statistics allows you to evaluate P(alternative hypothesis | data)



terminology review: W7 Activity 3

- Connector game!

where are we going next?

data = model + error

thus far

- data = mean + error
- data = X (interval/ratio) + error

after break

- data = X (interval/ratio/nominal) + error
- data = X + Y + error
- data (NOIR) = model (NOIR) + error