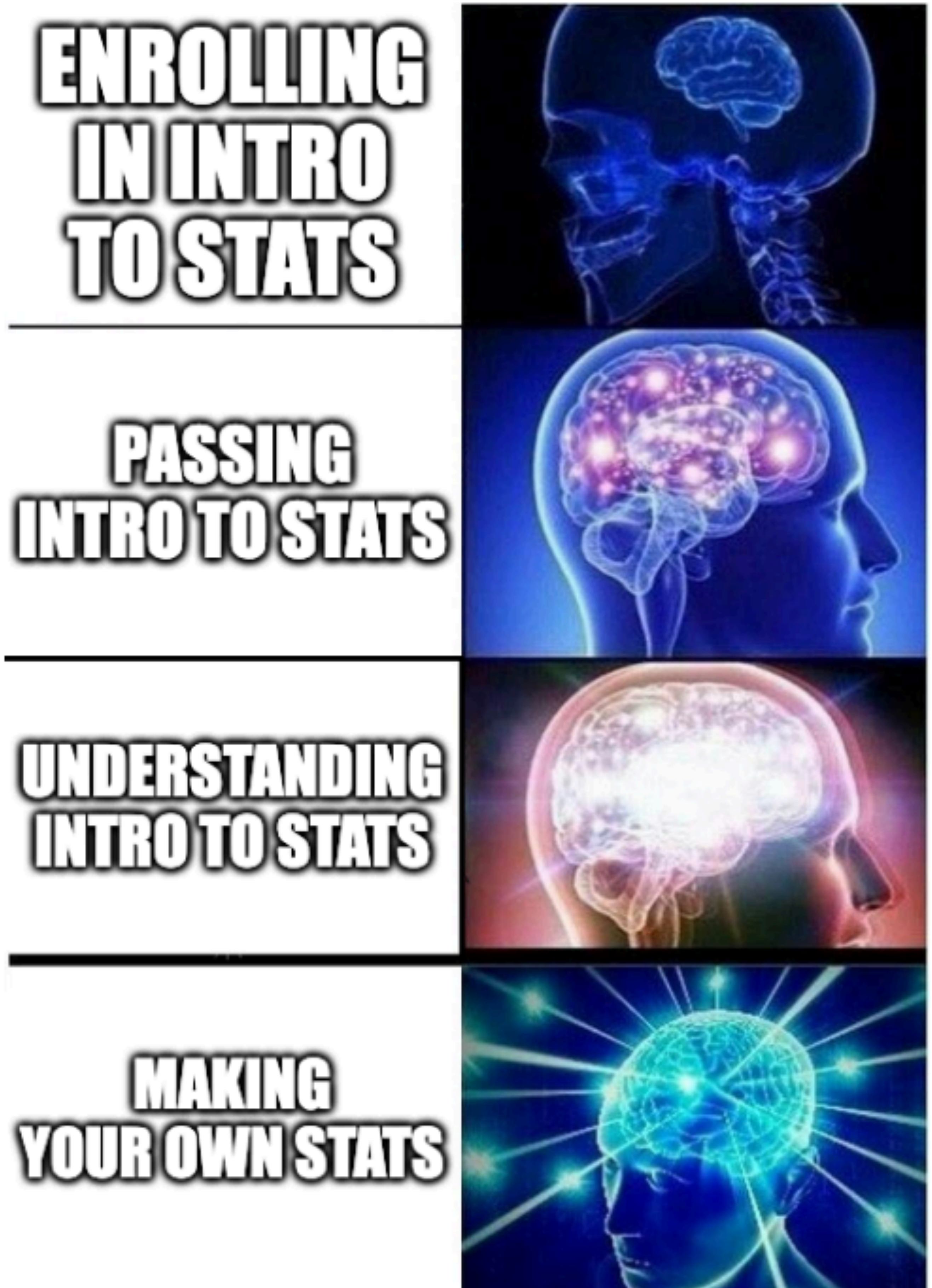
Statistical building blocks

PSYC 11: Laboratory in Psychological Science
March 27, 2024

Jeremy Manning
jeremy@dartmouth.edu


ENROLLING IN INTRO TO STATS
PASSING INTRO TO STATS
UNDERSTANDING INTRO TO STATS
MAKING YOUR OWN STATS

The "intro to stats" view of stats

- You have several options to choose from:
  - Want to compare the means of distributions? Use t-tests or ANOVAs
  - Want to compare trends? Use correlations or regressions
  - Etc.

Where do those tests come from?

- Consider what sorts of values your observations can take on:
  - Real numbers?
  - Counts?
  - Probabilities?
  - Sets of numbers that sum to 1?

Where do those tests come from?

- Then we can ask: under different (typically very simple) assumptions about where the numbers came from, how likely would it be to see something like our actual data?

Where do those tests come from?

- Example: [-1.2, 2.04, 0.087, -0.1, …]
- How unexpected would these numbers be if:
  - We thought the numbers came from a Normal distribution with mean = 0, var = 1
  - We thought the numbers came from a Normal distribution with mean = 100, var = 1

Where do those tests come from?

- To create the different tests you learn about in introductory stats courses, people have solved out the probabilities of observing different (sets of) values under different assumptions
- The p-value we get out tells us how unlikely it was that the observed data came from some "null" distribution

Making your own statistical tests

- Different distributions can produce different types of draws— Real numbers, counts, etc.
- Each distribution is typically controlled by one or more parameters— mean/ variance, probabilities, etc.

Probability distributions

Distribution name	Parameter(s)	What you get out	Example draws
Gaussian (Normal)	Mean, variance	Real numbers	-0.2, -10.923, 45.08, -6.4545
Bernoulli	Probability that x = 1	Results of "coin flips"	0, 1, 1, 0, 1, 0, 0, 0
Binomial	Number of observations, probability that each observation is 1	The number of observations where x = 1	10, 5, 38, 0, 267
Multinomial	Number of observations, probability that each *feature* in each observation is 1	The per-feature counts showing how many times each feature was 1	[3, 10, 2, 27], [46, 5, 4, 0]
Uniform	Start and end points (Real numbers)	A number between the start and end points	0.2, 0.7532, 0.00000123
Von Mises	Circular mean and concentration	Angles	3.6°, 186°, 240°, 359.98°

Making your own statistical tests

1. Pick an appropriate distribution
2. Pick parameters that correspond to your “null hypothesis” (e.g., that the distribution has a mean of 0, that the values are equally likely, etc.)
3. Take a bunch of samples from your distribution
4. Compare the values of those samples to your actual data

Example: is a coin fair?

- Suppose you observe some coin flips: [0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, …]
- How could you figure out whether the coin is fair (e.g., explainable by 0 and 1 being equally likely)?

Example: is a coin fair?

- Suppose you observe some coin flips: [0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, …]
- First, we need an appropriate distribution

Probability distributions

Distribution name	Parameter(s)	What you get out	Example draws
Gaussian (Normal)	Mean, variance	Real numbers	-0.2, -10.923, 45.08, -6.4545
Bernoulli	Probability that x = 1	Results of “coin flips”	0, 1, 1, 0, 1, 0, 0, 0
Binomial	Number of observations, probability that each observation is 1	The number of observations where x = 1	10, 5, 38, 0, 267
Multinomial	Number of observations, probability that each *feature* in each observation is 1	The per-feature counts showing how many times each feature was 1	[3, 10, 2, 27], [46, 5, 4, 0]
Uniform	Start and end points (Real numbers)	A number between the start and end points	0.2, 0.7532, 0.00000123
Von Mises	Circular mean and concentration	Angles	3.6°, 186°, 240°, 359.98°

Example: is a coin fair?

- Suppose you observe 12 coin flips: [0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0]
- For a fair coin, p(1) = 0.5
- We know that the number of "events" is 12 (i.e., the number of flips)
- Now we can ask: what's the probability of observing 3 or fewer 1s if the coin is fair?

Example: is a coin fair?

- Suppose you observe 12 coin flips: [0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0]
- Parameters: $N = 12$, $p(x = 1) = 0.5$
- Now take a bunch of draws from the binomial distribution. Let's take 1,000,000 draws and ask: what proportion of those draws have a count less than or equal to 3?
- That's our p-value!

Demo