Notes 11/6

- Randomized trials for the null distribution
- Bootstrapping

Central Limit Theorem Recap
- Used to estimate a p-value, but only after obtaining a z-score

Recap of Handout 17
1. Calculate Expected value of a coin flip
2. Calculate variance
    a. Weight expected values by probability
3. Use expected value and variance to calculate central limit theorem
4. Calculate the Z score
5. Plug in to get p value, shade in graph of distribution to highlight the probability of observing something as or more extreme than 3.13
    a. P Value is shaded area (area <-3.13 or >3.13)
    b. Use calculator to find p = .001745
        i. Compare to typical value is .05 (called alpha) to prove coin is “unfair”

Randomized trials for the null distribution

Could we do this in a better way?
- What do we do when we don't have all the data up front? Where is this data coming from?
- If you don't know your null distribution, what do you do?
    - What if we don't know the mean or variance of the null distribution?

The answer is randomized trials
- Randomized trials: general idea
    - Run T trials
    - Record relevant information
    - Count num of times you observe results as or more extreme than your data

In class activity:
- Everyone in the class rolls die 10 times and calculates mean
    - The goal is to see how often the mean is less than or equal to 2.4 (data were given)
- My results
    - 3, 1, 4, 1, 3, 5, 2, 2, 5, 6
    - Mean: (3 + 1 + 4 + 1 + 3 + 5 + 2 + 2 + 5 + 6)/10 = 3.2

- Now, list everybody's results:
    - T= 20
    - Ne = 1
    - P-value = 1/20 = .05
        - This is close
- A larger version of the same experiment:
    - T = 1000
    - Ne = 30
    - P-value = 30/1000 = .03

Handout 18:

- This is an opposite example, where there are a lot of "extreme" examples
- This is two sided, meaning we need to account for values that are proportionally too high and too low
- We have so many extreme values because they original value was not extreme, therefore a lot of values are considered extreme

- 60% in part 3 means "you will get a result this extreme or more extreme 60% of the time"
    - AKA this was not a surprising result, or not statistically significant at alpha = .05 confidence level

Are the means of two samples different?

Difference in means

*Goal of drug is to lower Blood Pressure

example

Before drug: [117, 54, 96, 123, 157, ...] $\bar{X}_n = 112$ n examples

Will giving the drug lower mean blood pressure?

don't need to be same # of tests

After drug: [72, 98, 105, 82, ...] m examples
$\bar{X}_m = 96$

H_0: all #'s drawn from same distribution
H_1: after the drug, blood pressure means down (one-sided)

Permutation testing

There are a lot of things we don't know about the data
Goal: simulate null distribution

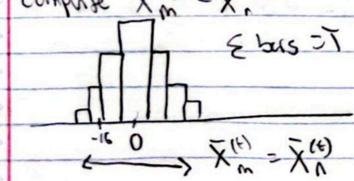
Permute the "labels" of the data (before and after)

1 trial
Before [98, 123, 105, 54, ...], need to ensure there are still n examples
After [82, 72, 117, 157, 96, ...] still m examples

$\bar{X}^{(1)}_m = 105$ → $\bar{X}^{(1)}_n = 101$

Now, compute means of before and after

Next, do for T trials (T = 1000 - 100,000)
compute $\bar{X}^{(t)}_m - \bar{X}^{(t)}_n$



Σ bars = T

$\bar{X}^{(t)}_m = \bar{X}^{(t)}_n$

$\bar{X}_m - \bar{X}_n = 96 - 112 = \boxed{-16}$

Now, how do we get a P-value?
Count # of values that are more extreme

$N_e = \#\left(\bar{X}^{(t)}_m - \bar{X}^{(t)}_n \leq -16\right)$

$\boxed{\text{P-value} = \frac{N_e}{T}}$

t-tests

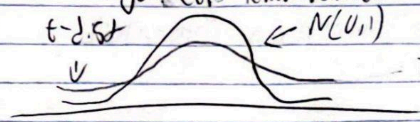
For (when we don't know σ^2)

use sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$t = \frac{\bar{X}_n - \mu}{\sqrt{\frac{s^2}{n}}} \sim \text{t-distribution}$$
(in some cases $N(0,1)$)

↑ like z for when you don't know variance



t-dist is flatter

Difference in Means

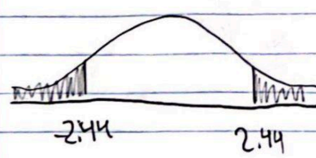
example pops (when analyzing data)

	A	B
$\bar{X}_n$	1.3	1.6
s	.5	.3
n	22	24

sample stddev → s

$H_0: \mu_A = \mu_B$
$H_1: \mu_A \neq \mu_B$
(2-sided)

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \sim \text{t-dist}$$



$$= \frac{1.3 - 1.6}{\sqrt{\frac{.25}{22} + \frac{.09}{24}}} = \boxed{-2.44}$$

p-value = .0236 < .05 So reject null

Bootstraps

Idea:

- Getting something from nothing, a measure of uncertainty

Example:

- Estimate the mean
- Sampling with replacement

Allows us to get a confidence interval
As long as we can resample data, calculate what we want to estimate, and from that get a sense of how good the estimate is.

1. Bootstrap T times

    Run method on test data set

    1. Xtest1 → .82
    2. Xtest2 → .91
    3. Xtest3 → .86
    4. Xtest4 → .95

2. Sort results
3. Take middle 95% for confidence interval