EDP308: STATISTICAL LITERACY

The University of Texas at Austin, Fall 2020

RAZ: Rebecca A. Zárate, MA

Overview

- Statistical vs. Practical Significance
- Effect Size: Cohen's d
  - Calculations
  - Small, Median, and Large Effect Sizes
- Cohen's d vs. Statistical Tests
- Cohen's d in R

Effect Size

Cohen's d

“The results were significant.”

- Again, statisticians are not good at naming things. We think “significant” means:
  - “sufficiently great or important to be worthy of attention”
- This common definition can confuse what “statistical significance” means…
- Results can be statistically significant but only have a tiny, negligible real world effect.
  - Ex. “The results were statistically significant, there was a .02 point increase in IQ when using a special shampoo.”
    - .02 when the mean is 100 and the SD is 15 is negligible

“The results were significant.”

- Statistical significance DOES NOT equal practical significance
  - *Hypothesis tests only tell us the probability that our finding would occur under the null hypothesis*
  - With a large enough sample size, lots of negligible differences can be “statistically significant”
- Practical significance are differences that are large enough to mean something in everyday life

Statistical vs. Practical Significance

- How different are the two means?
  - Not just are they "statistically significantly different" because as we have seen that can mean very little practically.

How do we quantify practical significance?

Effect Size: Cohen's d

Cohen’s d, Effect Size

- Cohen's d is an effect size used to indicate the <u>standardized</u> difference between two means

sample mean

population mean

effect size in standard units

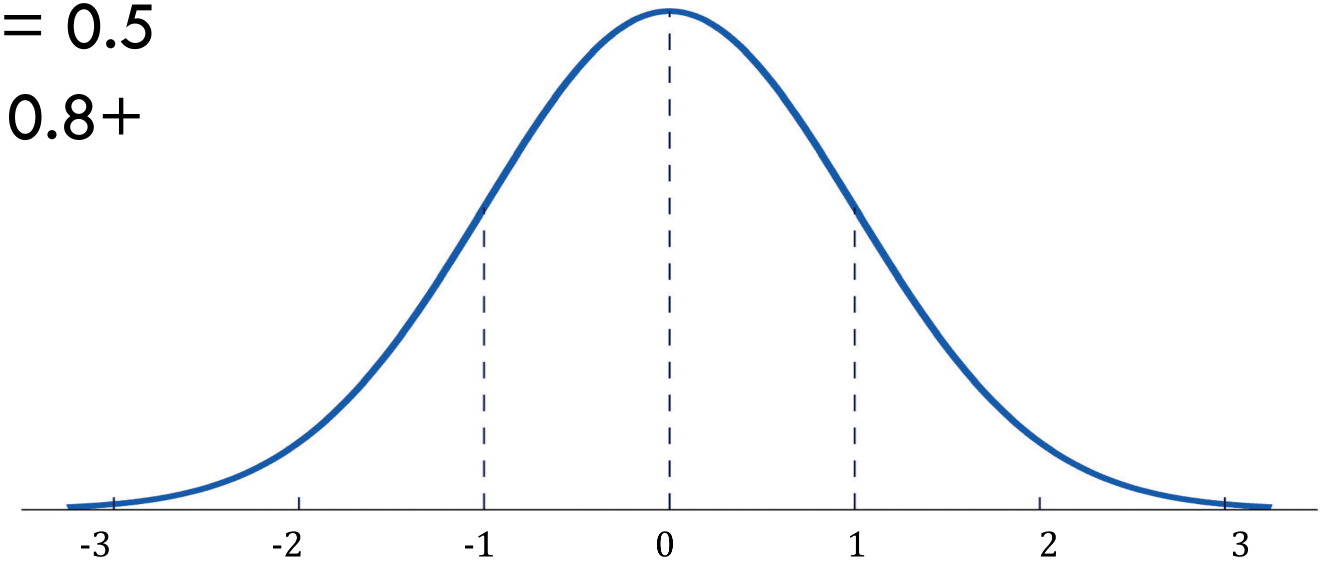
$$d = \frac{\bar{x} - \mu}{s}$$

sample standard deviation

It looks just like our t-test but now we are dividing by standard deviation rather than standard error.

Cohen's d

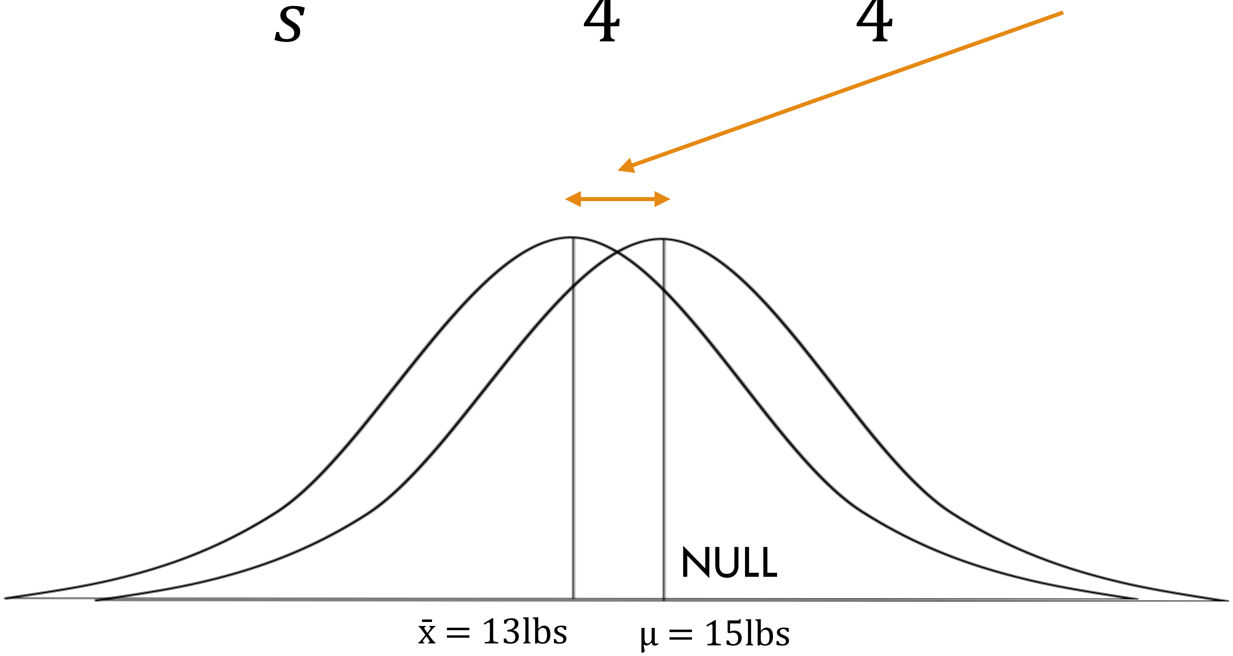
- The effect size (d) is in standard units, aka standard deviations, think back to our simple z-distribution
- So, an effect size of $d = 1.0$ says that the means are a whole one standard deviation apart
  - Small = 0.2
  - Medium = 0.5
  - Large = 0.8+



Standardized Differences

Freshman 15 Example

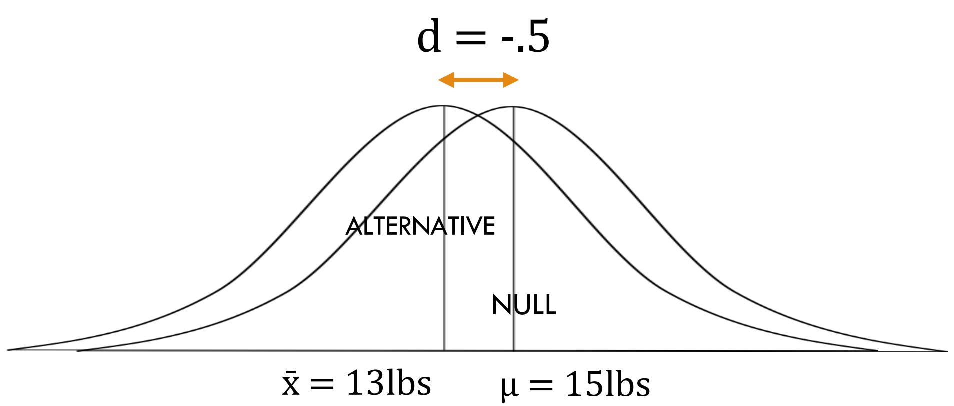
$$d = \frac{\bar{x} - \mu}{s} = \frac{13 - 15}{4} = \frac{-2}{4} = -0.5$$

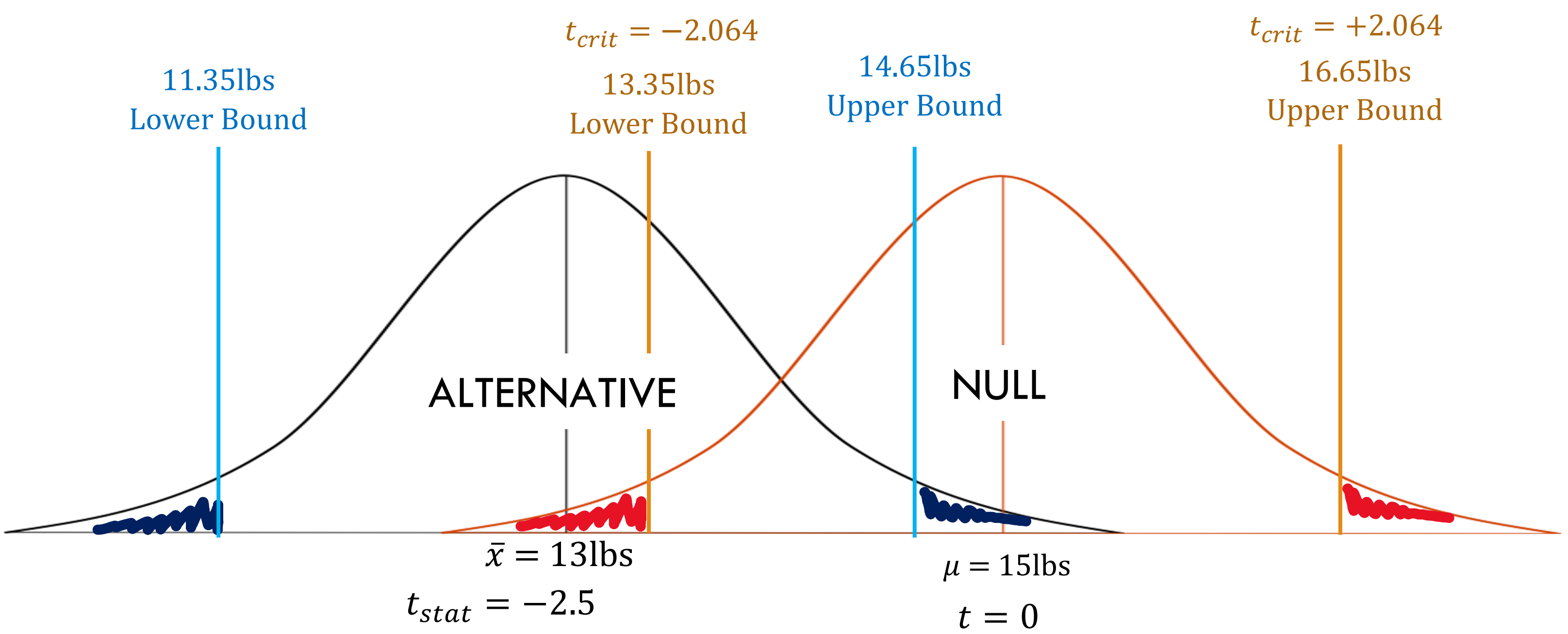


There is a difference of half a standard deviation (d = -0.5) between the average college student and UT college students when it comes to weight difference after their first year of college.

Standardized Differences

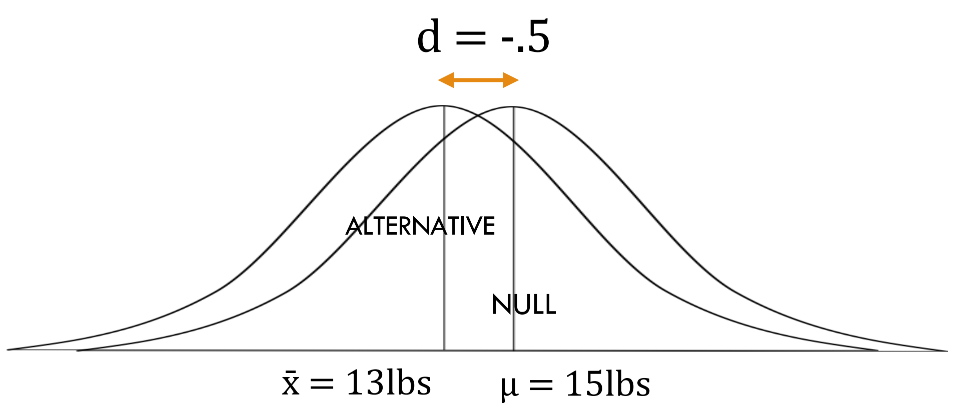
How is this different from what we calculated in the Freshman 15 t-test?

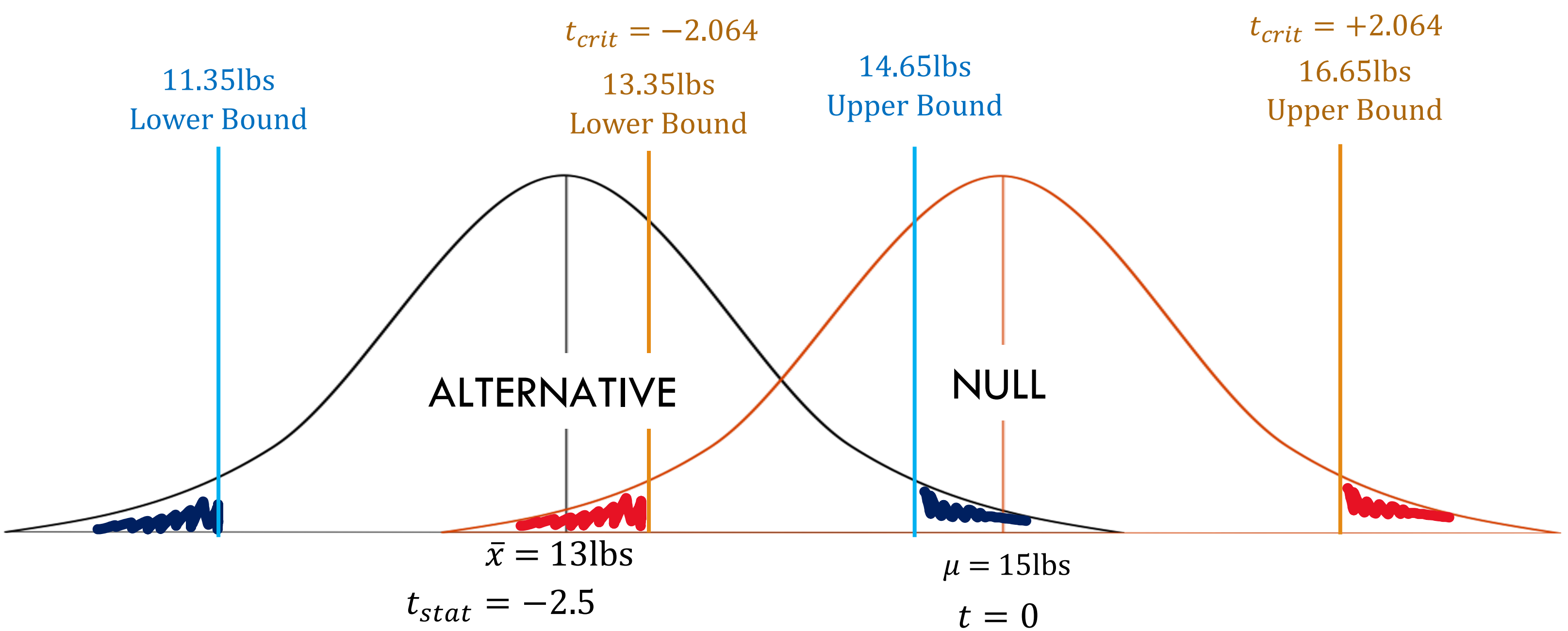




Standardized Differences

Cohen's d is a standardized difference between two means, while the t-test is using a *sampling distribution* to test a hypothesis.





Cohen's d and Test Statistics

- Test statistics determine how likely these findings would be if the NULL hypothesis were true
  - This is why it takes the variance AND the sample size into account (i.e. Standard Error)
  - It does not determine the **magnitude** of the difference
- Cohen's d does determines the magnitude of difference, but cannot determine the likelihood of seeing these findings

Cohen's d and t-statistic Examples

How would you interpret the following?

- $t(24) = +2.45, p < .05, d = .15$
- $t(24) = +2.05, p > .05, d = .15$
- $t(4) = +2.45, p > .05, d = .80$
- $t(4) = +2.78, p < .05, d = .80$

Cohen's d and t-statistic Examples

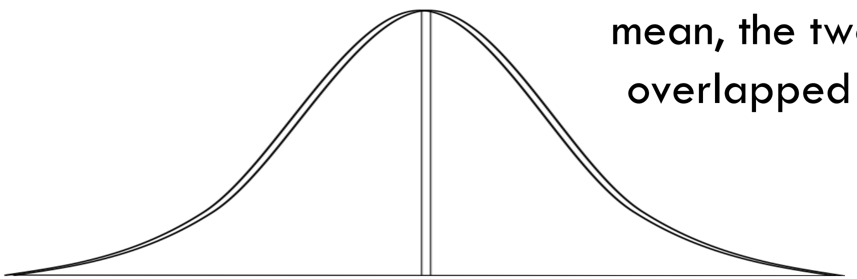
- How would you interpret the following?
  - $t(24) = +2.45, p < .05, d = .15$
    - Statistically significant, but very small effect
      - We can be confident that there is only small difference
  - $t(24) = +2.05, p > .05, d = .15$
    - NOT statistically significant, and very small effect
      - We can't be confident…
  - $t(4) = +2.45, p > .05, d = .80$
    - NOT statistically significant, BUT very large effect size
      - This suggests we might be UNDERPOWERED, not a big enough n
  - $t(4) = +2.78, p < .05, d = .80$
    - Statistically significant, and a very large effect size
      - We can be confidence that there is a large effect

Cohen's d Interpretation

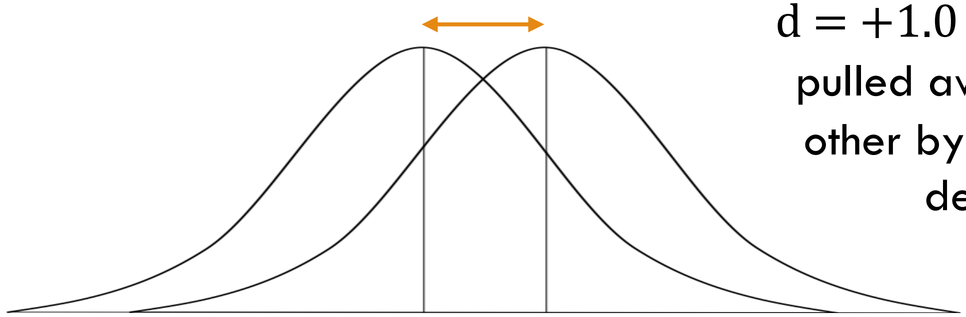
- In a one sample t-test example, a Cohen's d value of $\mathrm{d} = +1.0$ would suggest that your sample of people ($\bar{x}$) are 1.0 standard deviations above the average of the population (μ)
  - We may not be familiar with a certain scale and whether or not a difference of 10 points means anything, but standard units are helpful because we are familiar with them and know what they imply
    - If I said the Cohen's d was +1.5, we know there was a large difference between the two means
    - If I said Cohen's d was .06, you'd know there was little practical difference between the means

Overlaps and Shifts

- With Cohen's d we know that a $d = +1.0$ value implies that two distributions (ex. UT vs non-UT students) are shifted away from each other by one full standard deviation.



If there was no difference between the sample mean and population mean, the two curves would be overlapped nearly perfectly.



But if they have a have a $d = +1.0$, the curves are pulled away from each other by one standard deviation.

Up Next…

- Ok, now on comparing more than one group to one known mean…

Independent Samples t-tests

Cohen’s d in R

Cohen's d in R

7.2-4 Effect Size Cohen's d

```{r}
sample_mean <- 13
pop_mean <- 15
sd <- 4

cohens_d <- (sample_mean - pop_mean)/sd
```