Practical Significance of the T-Test

Practical Significance for the T-Test

After using a t-test to assess statistical significance, it is important to assess practical significance.

> *Your main goal is to explain to your audience why they should or should not care about the effect.*

Three common effect size measures:

1. Difference in means
2. Cohen's d
3. Correlation r

Difference in Means

Difference in means

$$\bar{X}_A - \bar{X}_B$$

- Answers the question "*How different are these groups?*"
- Often makes great headlines and is a good choice if units are familiar
- But lacks context in its calculation
- People who eat chocolate live 1.5 years longer than those who do not each chocolate

Cohen's d

Cohen's d

Cohen's d is a measure of difference of means standardized by the variance in the data.

$$\frac{\bar{X}_A - \bar{X}_B}{s}$$

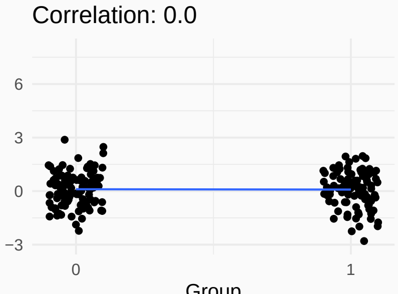
Where s is a pooled standard deviation: $\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2}}$

- Answers the question "*How many standard deviations apart are the groups?*"
- The difference in sarcasm score between frequentists and Bayesians is $d = 0.54$ standard deviations.

Correlation

Correlation answers the question "How strong is the relationship between group identity and the outcome?"

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

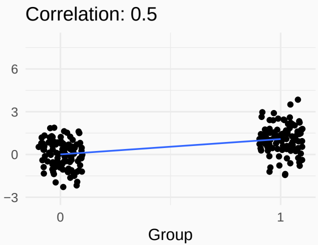


Correlation

Biserial correlation

Correlation answers the question "How strong is the relationship between group identity and the outcome?"

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

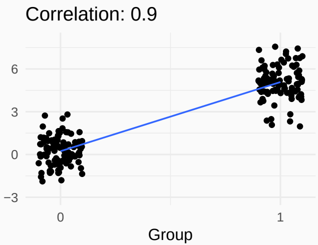


Correlation

Biserial correlation

Correlation answers the question "How strong is the relationship between group identity and the outcome?"

$$\rho = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$



Practical Significance is about Context

- How strong is the same relationship between *different* groups?
- How strong is a *different* relationship between the same group?
- What is the underlying dispersion in the data?
- What is a meaningful anchor or reference point that you can use for context?

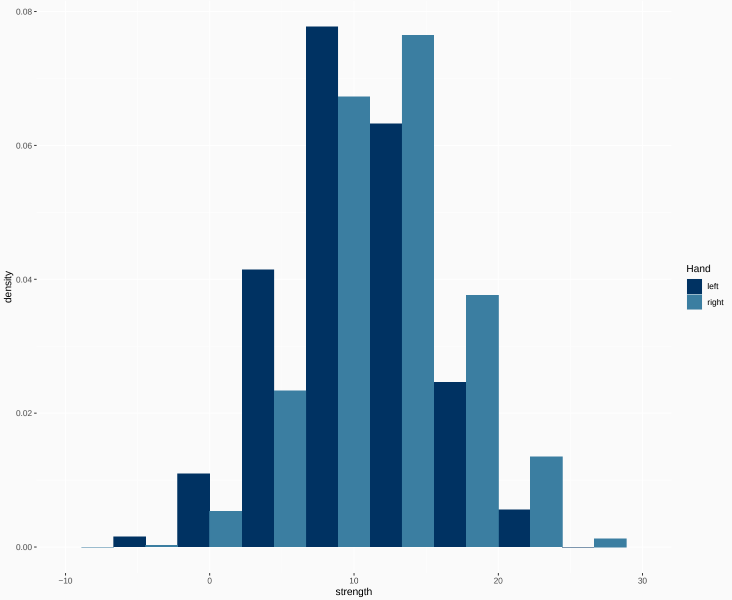
The Paired t-Test

Climbing grip

Suppose you randomly sample 30 Berkeley students. For each student i, you measure right-hand strength (R_i) and left-hand strength (L_i).

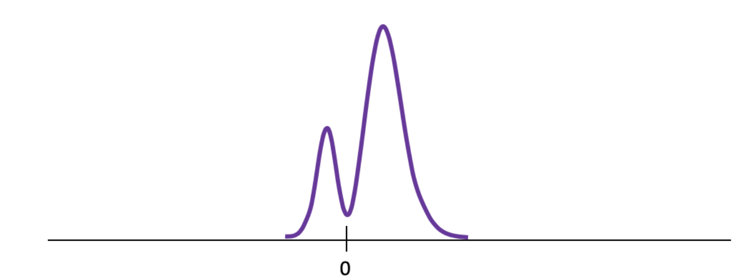
- You conduct a t-test with $H_0 : \mathrm{E}[R] = \mathrm{E}[L]$
- **Problem**: Grip strength varies a lot person-to-person, $\Rightarrow$ t-test has low power.

Paired t-test

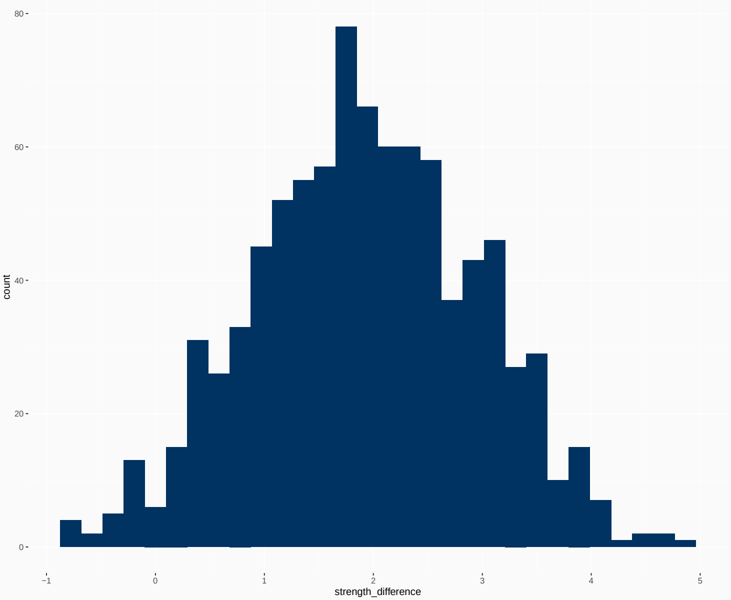


Paired t-Test

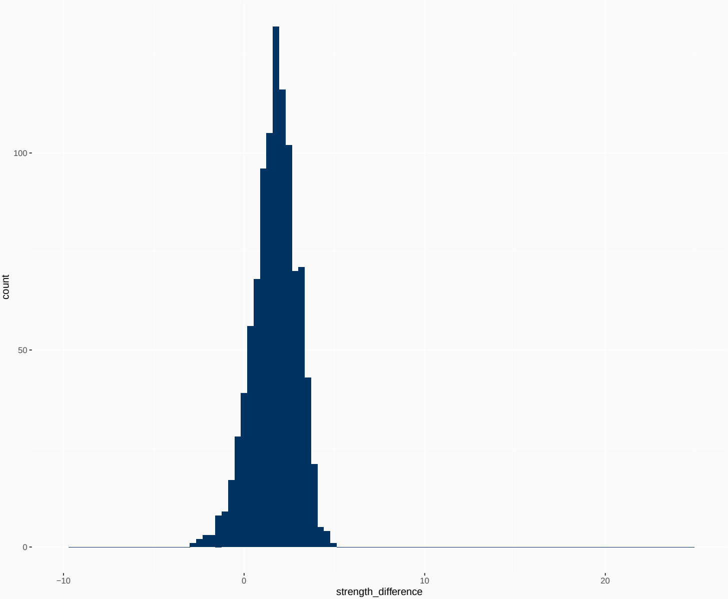
- **Idea:** For any *particular* subject i, the difference between right-hand strength and left-hand stregth, $R_i - L_i$, will usually be small.
- Within-person variation is small.



Paired t-Test



Paired t-Test



Paired t-Test

Paired t-test

A *paired t-test*, sometimes called a *dependent t-test*, builds an explicit dependency between data. Instead, perform a one-sample t-test with $H_0 : \mathrm{E}[R_i - L_i] = 0$.

- This dependency must actually exist
- Cannot simply change the test

Unpaired vs. Paired t-Test

Unpaired

- $t = \frac{\bar{A} - \bar{B}}{\sigma_{A\&B}}$

Paired

- $t = \frac{\bar{A} - \bar{B}}{\sigma_{(A-B)}}$

Paired t-Test Assumptions

- A and B have a metric scale with the same units.
- There is a natural pairing between observations for A and for B.
    - pre-test and post-test for same individual
    - response to two types of stimulus for same mouse
    - responses for a pair of spouses
- Each pair (A_i, B_i) is drawn i.i.d.
- The distribution of $A - B$ is sufficiently normal given the sample size.