Lecture 14: Effect Size

Efffect Size

If there’s an effect, how big is it?
How different is \( \mu \) from \( \mu_0 \), or \( \mu_A \) from \( \mu_B \), etc.?

Separate from reliability
Inferential statistics measure effect relative to standard error
Tiny effects can be reliable, with enough power
Danger of forgetting about practical importance

Estimation vs. inference
Inferential statistics convey confidence
Estimation conveys actual, physical values

Ways of estimating effect size
Raw difference in means
Relative to standard deviation of raw scores

Direct Estimates of Effect Size
Goal: estimate difference in population means
One sample: \( \mu - \mu_0 \)
Independent samples: \( \mu_A - \mu_B \)
Paired samples: \( \mu_{diff} \)
Solution: use \( M \) as estimate of \( \mu \)
One sample: \( M - \mu_0 \)
Independent samples: \( M_A - M_B \)
Paired samples: \( M_{diff} \)

Point vs. interval estimates
We don’t know exact effect size; samples just provide an estimate
Better to report a range that reflects our uncertainty

Confidence Interval
Range of effect sizes that are consistent with the data
Values that would not be rejected as \( H_0 \)

Computing Confidence Intervals
CI is range of values for \( \mu \) or \( \mu_A - \mu_B \) consistent with data
Values that, if chosen as null hypothesis, would lead to \( |t| < t_{crit} \)

One-sample t-test (or paired samples):

\[
\text{Retain } \mu_0 \text{ if } |t| = \left| \frac{M - \mu_0}{SE} \right| < t_{crit}, \text{ i.e. } |M - \mu_0| < t_{crit} \cdot SE
\]

Therefore any value of \( \mu_0 \) within \( t_{crit} \cdot SE \) of \( M \) would not be rejected
Formulas for Confidence Intervals

Mean of a single population (or of difference scores)

\[ M \pm t_{\text{crit}} \cdot SE \]

Difference between two means

\[ (M_A - M_B) \pm t_{\text{crit}} \cdot SE \]

Always use two-tailed critical value

\[ p(t_{\text{df} > t_{\text{crit}}} = \alpha/2 \]

Confidence interval has upper and lower bounds

Need \( \alpha/2 \) probability of falling outside either end

Effect of sample size

Increasing \( n \) decreases standard error

Confidence interval becomes narrower

More data means more precise estimate

Interpretation of Confidence Interval

Pick any possible value for \( \mu \) (or \( \mu_A - \mu_B \))

**IF** this were true population value

5% chance of getting data that would lead us to falsely reject that value

95% chance we don’t reject that value

For 95% of experiments, CI will contain true population value

"95% confidence"

Other levels of confidence

Can calculate 90% CI, 99% CI, etc.

Correspond to different alpha levels: \( \text{confidence} = 1 - \alpha \)

Leads to different \( t_{\text{crit}} \): \( t_{\text{crit}} = qt(\alpha/2, df) \)

Higher confidence requires wider intervals (\( t_{\text{crit}} \) increases)

Relationship to hypothesis testing

If \( \mu_0 \) (or 0) is not in the confidence interval, then we reject \( H_0 \)

Standardized Effect Size

Interpreting effect size depends on variable being measured

Improving digit span by 2 more important than for IQ

Solution: measure effect size relative to variability in raw scores

**Cohen’s \( d \)**

Effect size divided by standard deviation of raw scores

Like a \( z \)-score for means
### Samples

<table>
<thead>
<tr>
<th>$d$</th>
<th>One Independent Paired</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>$\frac{\mu - \mu_0}{\sigma}$</td>
</tr>
<tr>
<td>Estimated</td>
<td>$\frac{M - \mu_0}{\sqrt{MS}}$</td>
</tr>
</tbody>
</table>

**Meaning of Cohen’s $d$**

How many standard deviations does the mean change by?

- Gives $z$-score of one mean within the other population
- (negative) $z$-score of $\mu_0$ within population
- $z$-score of $\mu_A$ within Population B

$\text{pnorm}(d)$ tells how many scores are above other mean (if population is Normal)

- Fraction of scores in population that are greater than $\mu_0$
- Fraction of scores in Population A that are greater than $\mu_B$

**Cohen’s $d$ vs. $t$**

- $t$ depends on $n$; $d$ does not
- Bigger $n$ makes you more confident in the effect, but it doesn’t change size of the effect

<table>
<thead>
<tr>
<th>Statistic</th>
<th>One</th>
<th>Independent</th>
<th>Paired</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>$\frac{M - \mu_0}{\sqrt{MS}}$</td>
<td>$\frac{M_A - M_B}{\sqrt{MS}}$</td>
<td>$\frac{M_{\text{diff}}}{\sqrt{MS}}$</td>
</tr>
<tr>
<td>$t$</td>
<td>$\frac{M - \mu_0}{\sqrt{MS/n}}$</td>
<td>$\sqrt{MS\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$</td>
<td>$\frac{M_{\text{diff}}}{\sqrt{MS/n}}$</td>
</tr>
</tbody>
</table>