Statistical Foundations: Effect Sizes

26 February 2020

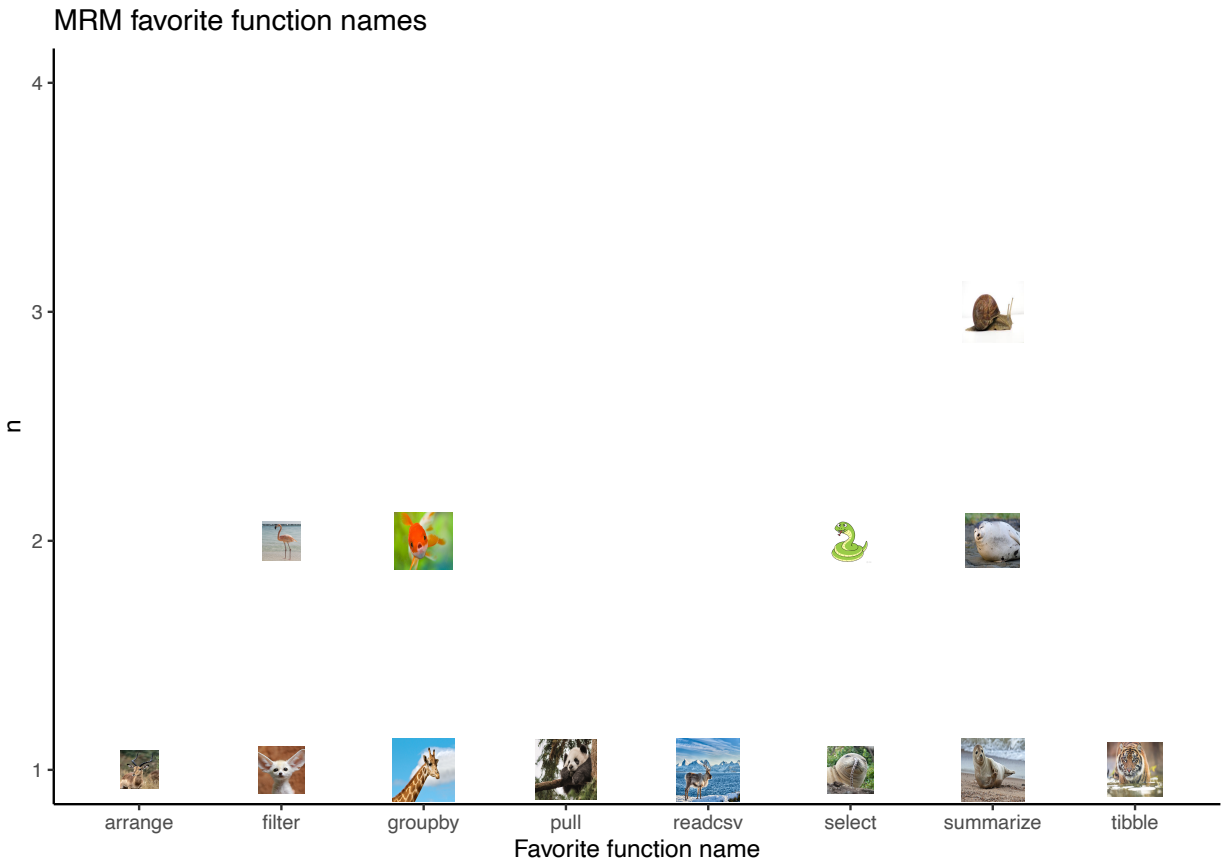
Modern Research Methods

Business

- Assignment 4 graded
- Assignment 5 due tomorrow (Thursday) at noon
- Midterm handed out on Friday at 5pm
- Review session in lab on Friday – Come with your questions.
- Additional review session Friday afternoon – time TBD

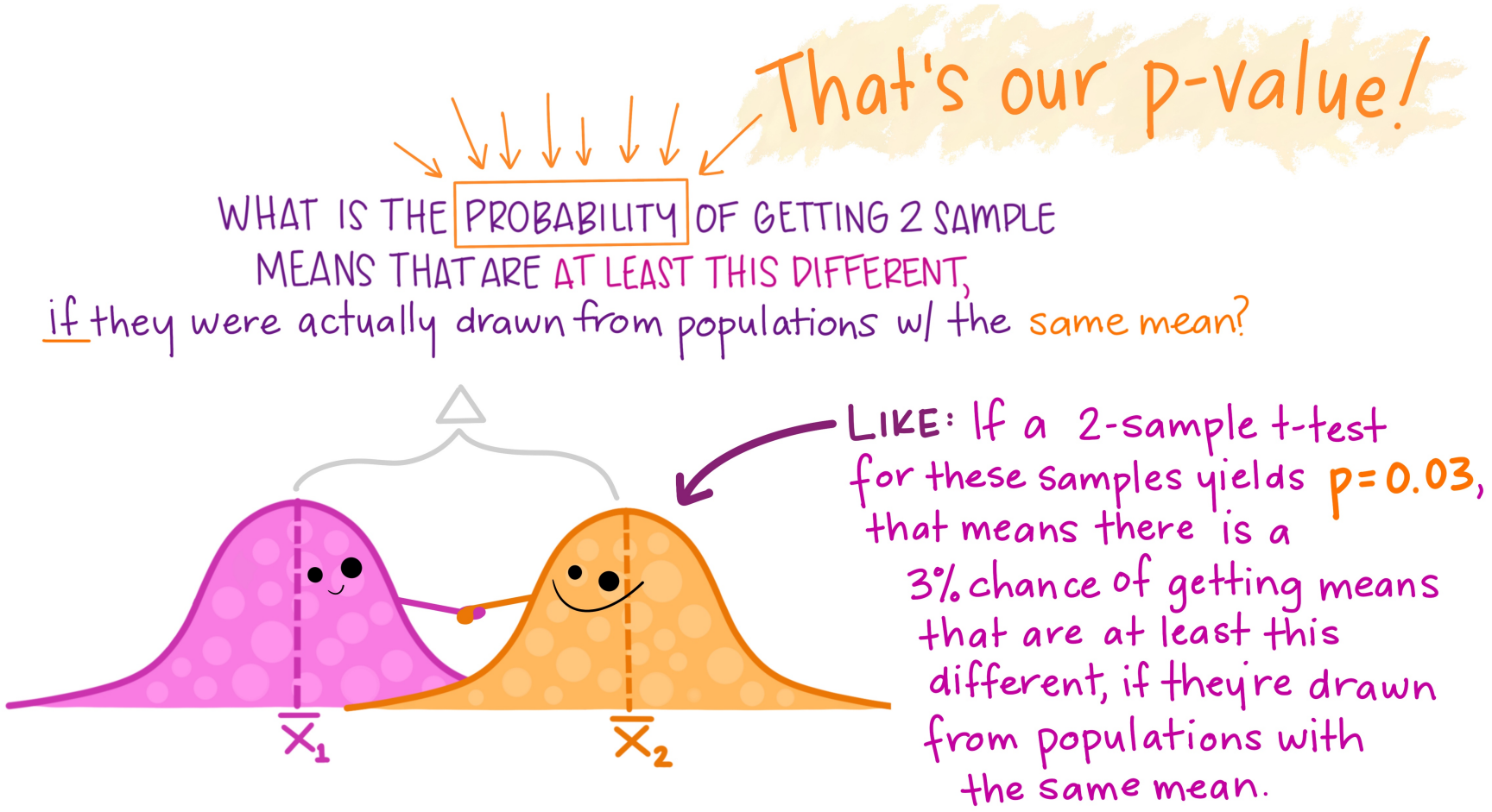
Your favorite tidyverse functions

```
library(tidyverse)
library(ggimage)

picture_files <- list.files("animal_pictures/", full.names = T)

pic_files_df <- tibble(full_file_name = picture_files)

clean_files <- pic_files_df %>%
  mutate(file_name = basename(full_file_name)) %>%
  separate(col = "file_name", sep = "_", into = c("function_name", "animal", "student_name")) 
  separate(col = "student_name", sep = "\\.", into = c("student_name", "temp")) %>%
  select(-temp)

tidy_animals <- clean_files %>%
  group_by(function_name) %>%
  mutate(n = n()) %>%
  ungroup()

ggplot(tidy_animals, aes(x = function_name, y = n)) +
  geom_image(aes(image = full_file_name), by = "width") +
  ggtitle("MRM favorite function names") +
  xlab("Favorite funtion name") +
  ylim(1, 5) +
  theme_classic()
```



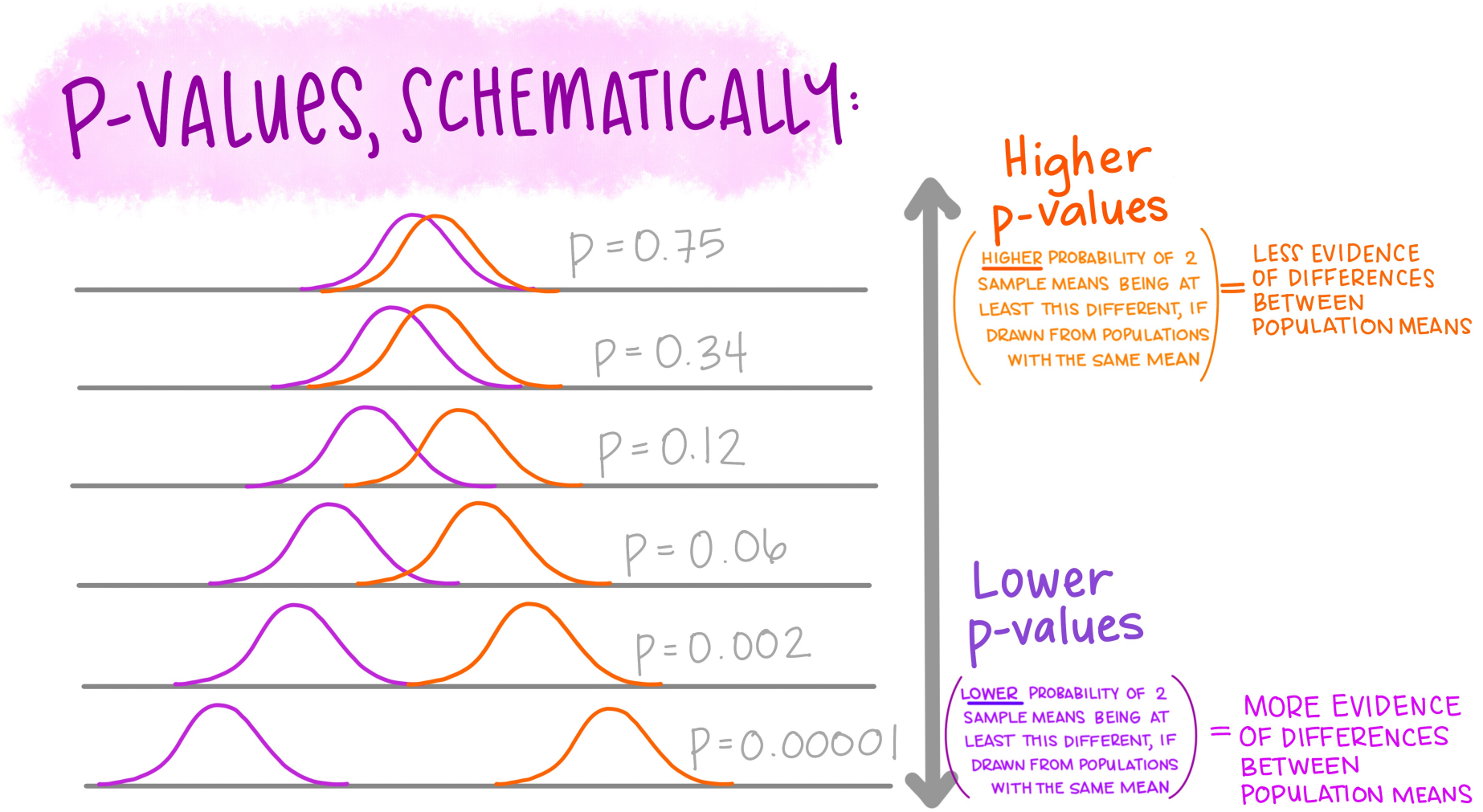
* Code and pictures on R studio cloud

Last time: Null Hypothesis Testing and *p*-values



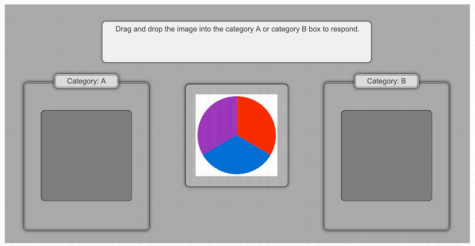
Artwork by @Allison_horst

Last time: Null Hypothesis Testing and *p*-values



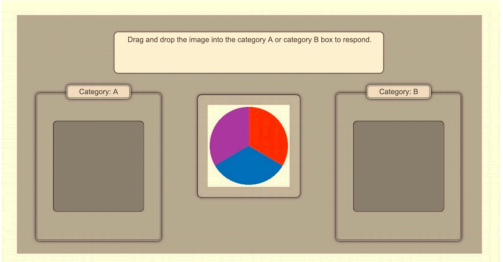
Our “replication”

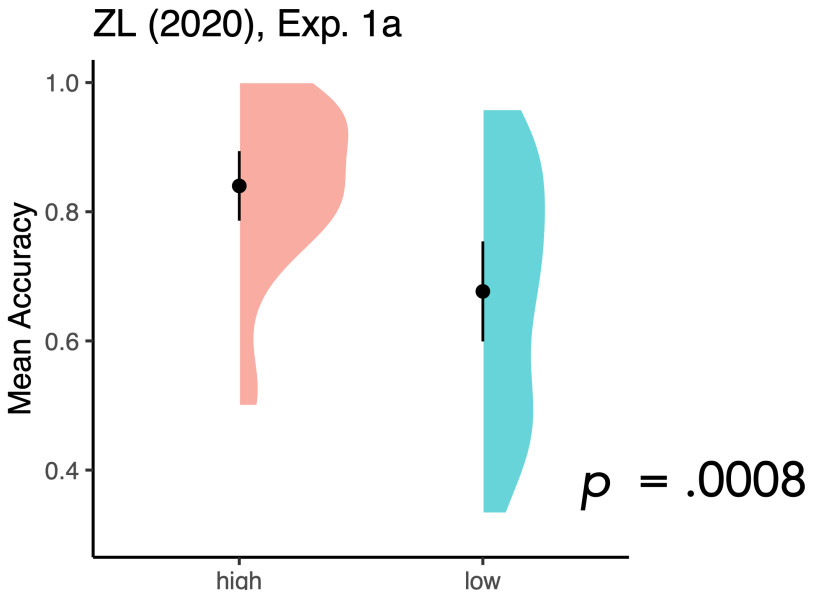
Original

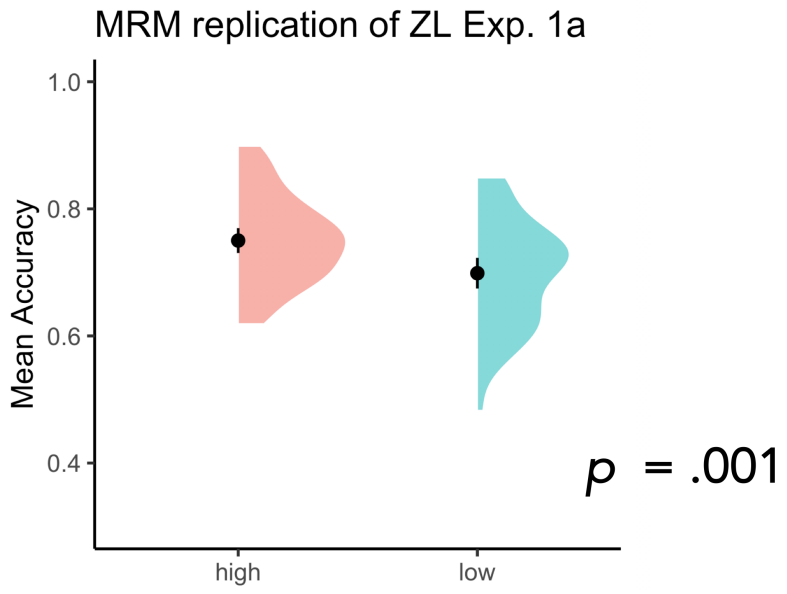


Replication

[Us]

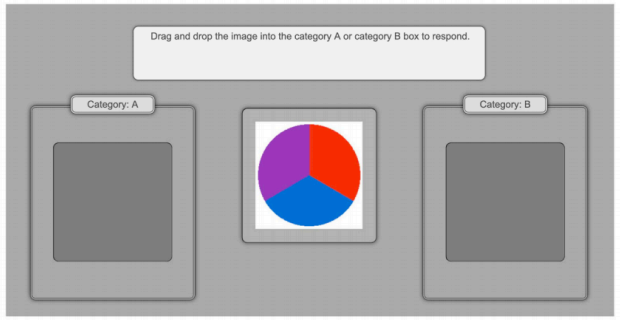






Replicating Zettersten and Lupyan (2020)

Original

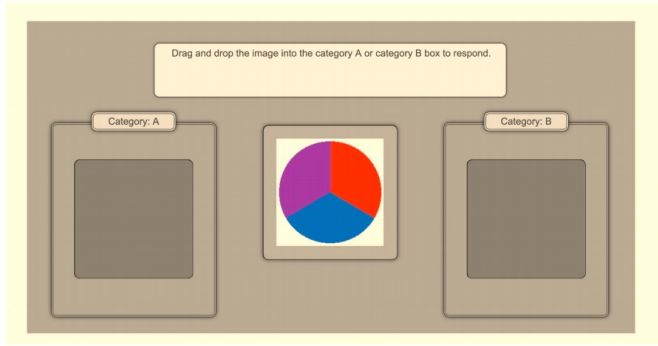


CLAIM: It's easier to learn a category when the colors are nameable.

predicting participants' trial-by-trial accuracy on training trials from condition, including a by-subject random intercept.[3] We used the lme4 package version 1.1-21 in R (version 3.6.1) to fit all models (D. Bates & Maechler, 2009; R Development Core Team, 2019). Participants in the High Nameability condition (M = 84.0%, 95% CI = [78.6%, 89.4%]) were more accurate than participants in the Low Nameability Condition (M = 67.7%, 95% CI = [59.9%, 75.4%]), b = 1.02, 95% Wald

Replication

[Us]



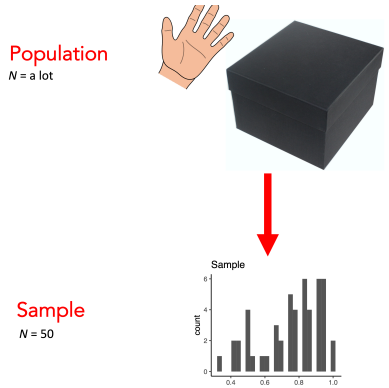
High Nameability Condition = 75%
Low Nameability Condition = 69%

Did we replicate it? YES!

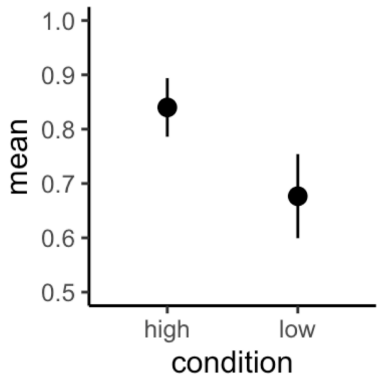
Today: Effect Sizes

Last week:

Sampling



Confidence Intervals (CI)



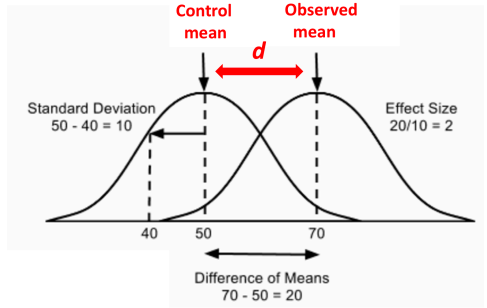
This week:

Null Hypothesis Testing

Are the means different? (yes/no)

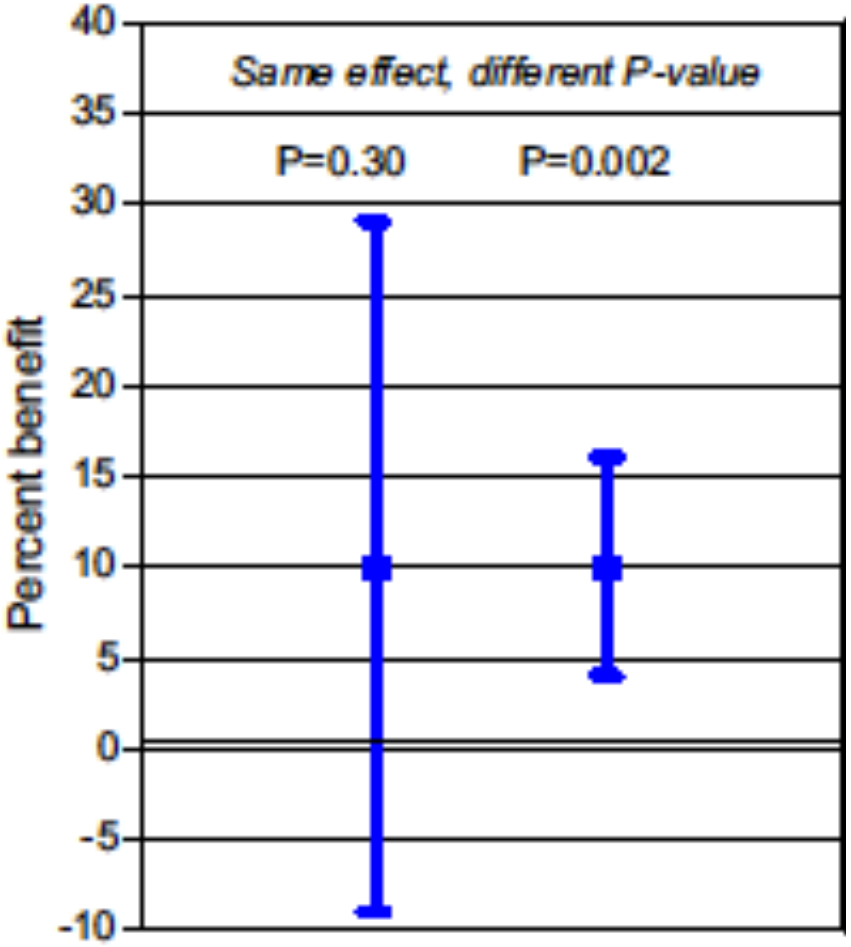
Effect Sizes

How different are the means?

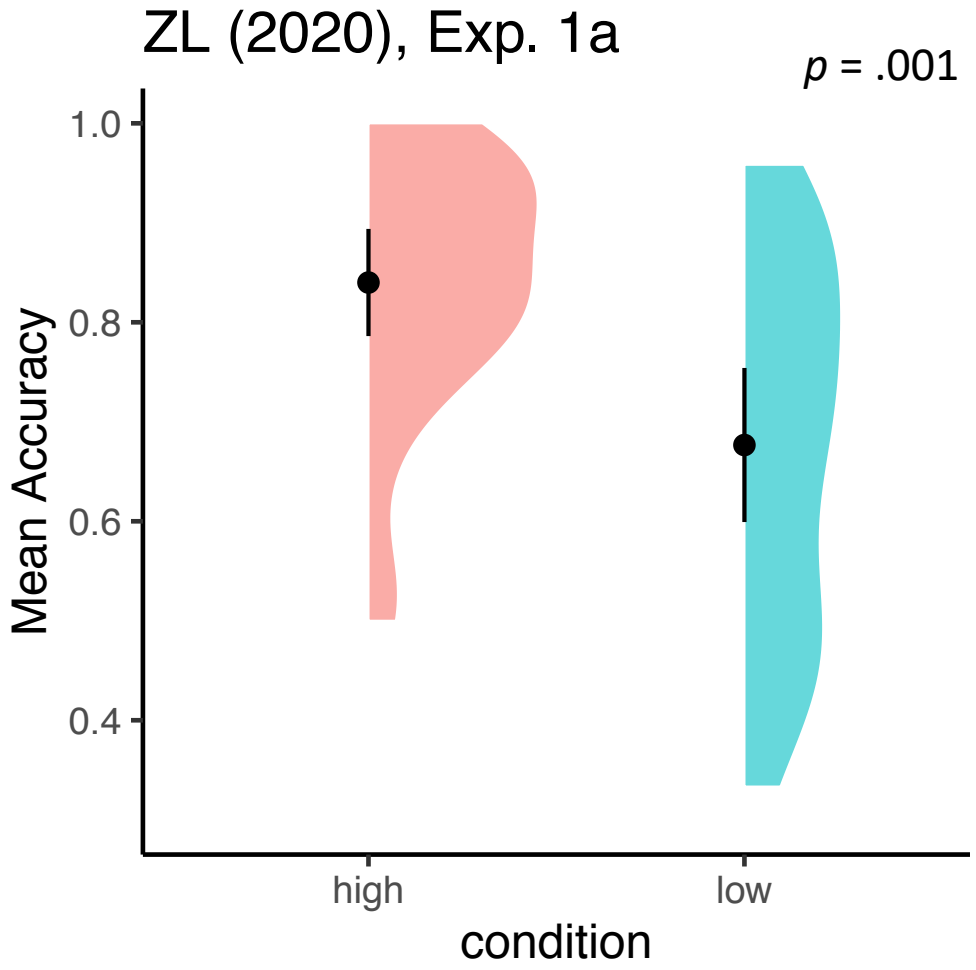


Interpreting Confidence Intervals

- 95% confident that true value is in the range?
- **No!**
- "95% of confidence intervals contain the true value of the parameter in the population"
- Plausible values for estimate
- Related to *p*-values:
  - Bigger CI -> bigger *p*-value
  - Smaller CI -> smaller *p*-value



Interpreting *p*-values



- Probability H_0 is true?
- Strength of evidence for H_1?
- Size of effect?

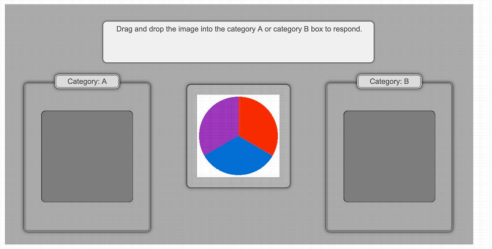
- **No!**

- *p*-values just give us yes/no answer about significance.
- Size of *p*-value related to sample size
- "significant" ≠ practically meaningful

- To understand size of effect, use effect size measures.

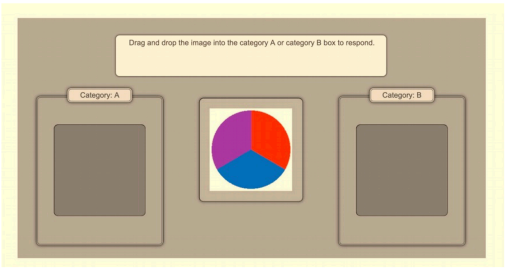
In which experiment is the effect bigger?

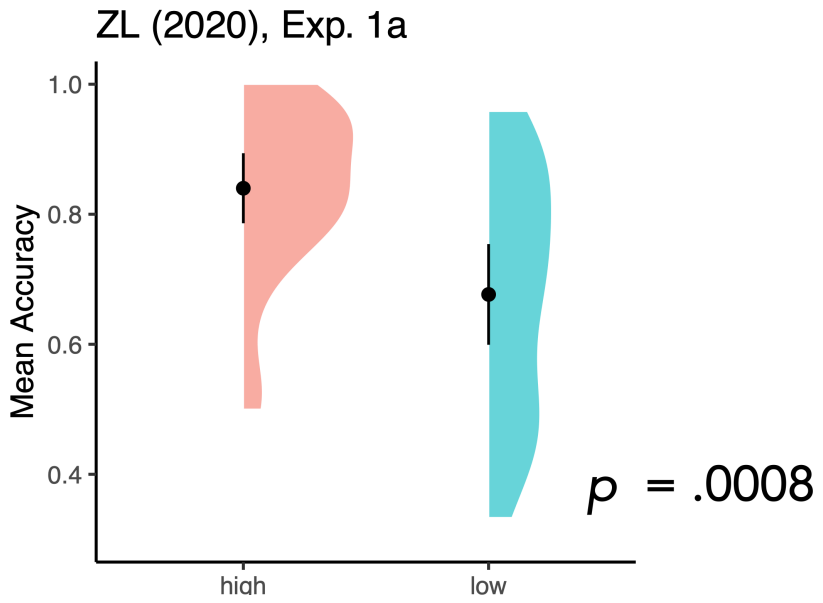
Original

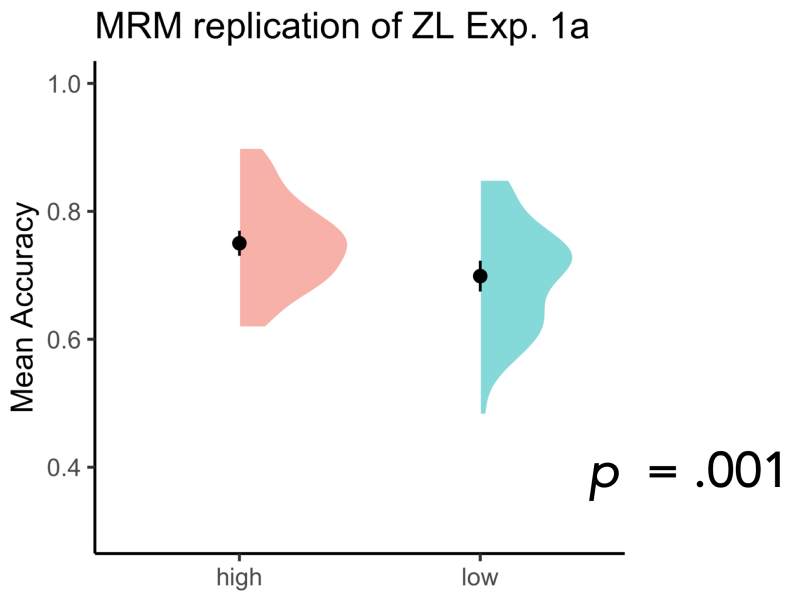


Replication

[Us]







A simple measure of the size of an effect

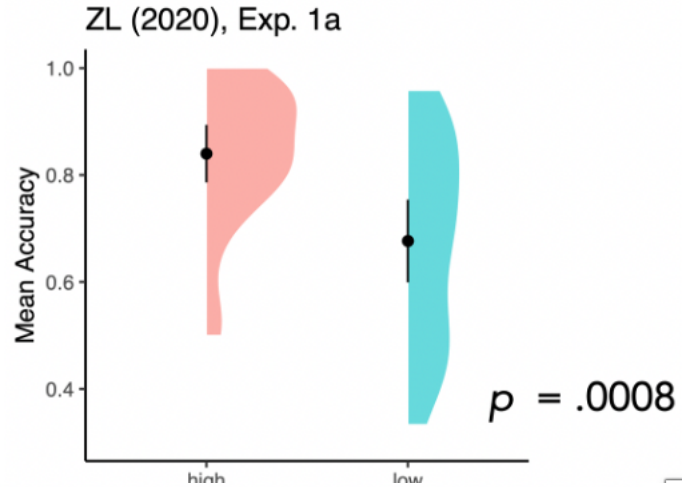
Simple Effect Size = diff. between means

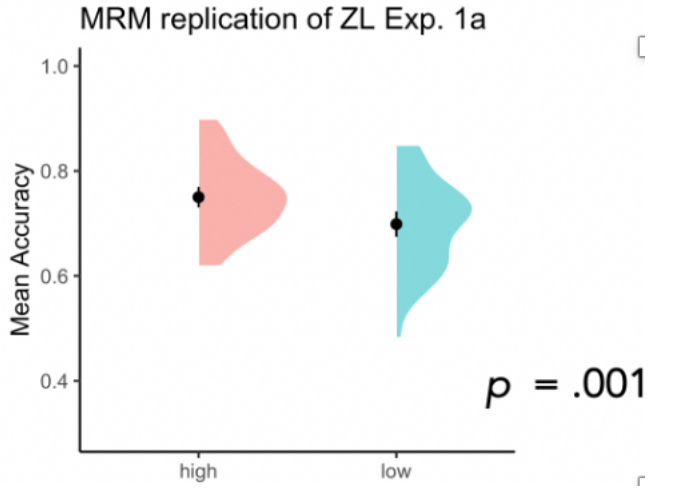
- Size of the effect in Original =

  $Mean_{High} - Mean_{Low} = .84 - .67 = 0.17$

- Size of the effect in Replication =

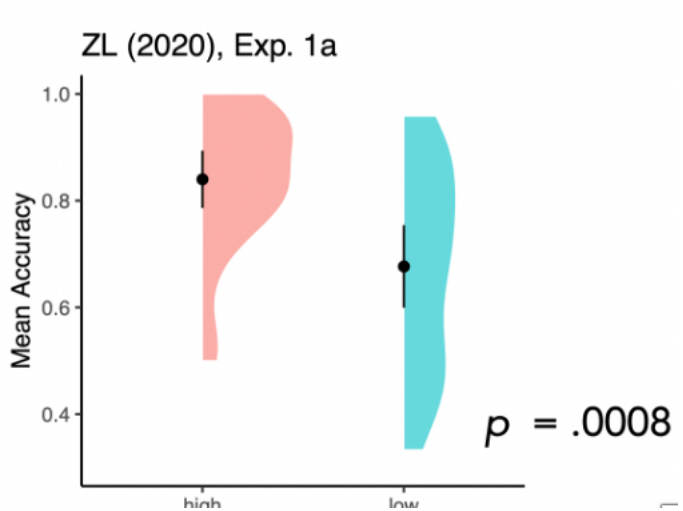
  $Mean_{High} - Mean_{Low} = .76 - .69 = 0.07$

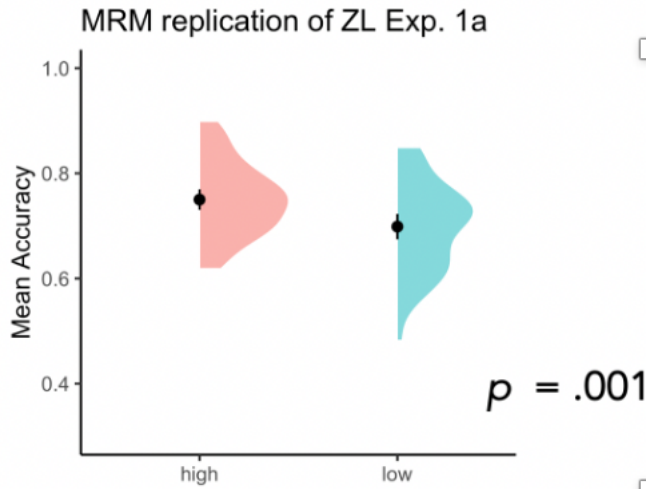




But what if they have very different dispersions?

- Our simple measure ignores this information.
- We want a measure that takes into account the amount of dispersion in the two conditions.
- More dispersion -> smaller effect;
- Less dispersion -> bigger effect





Cohen's *d*

Standardized measure of the size of an effect

Encodes magnitude *and* direction of effect

Cohen's *d*:

Effect Size = $\frac{\text{diff. between means}}{\text{standard dev.}}$

$$d = \frac{M_{group1} - M_{group2}}{SD_{pooled}}$$

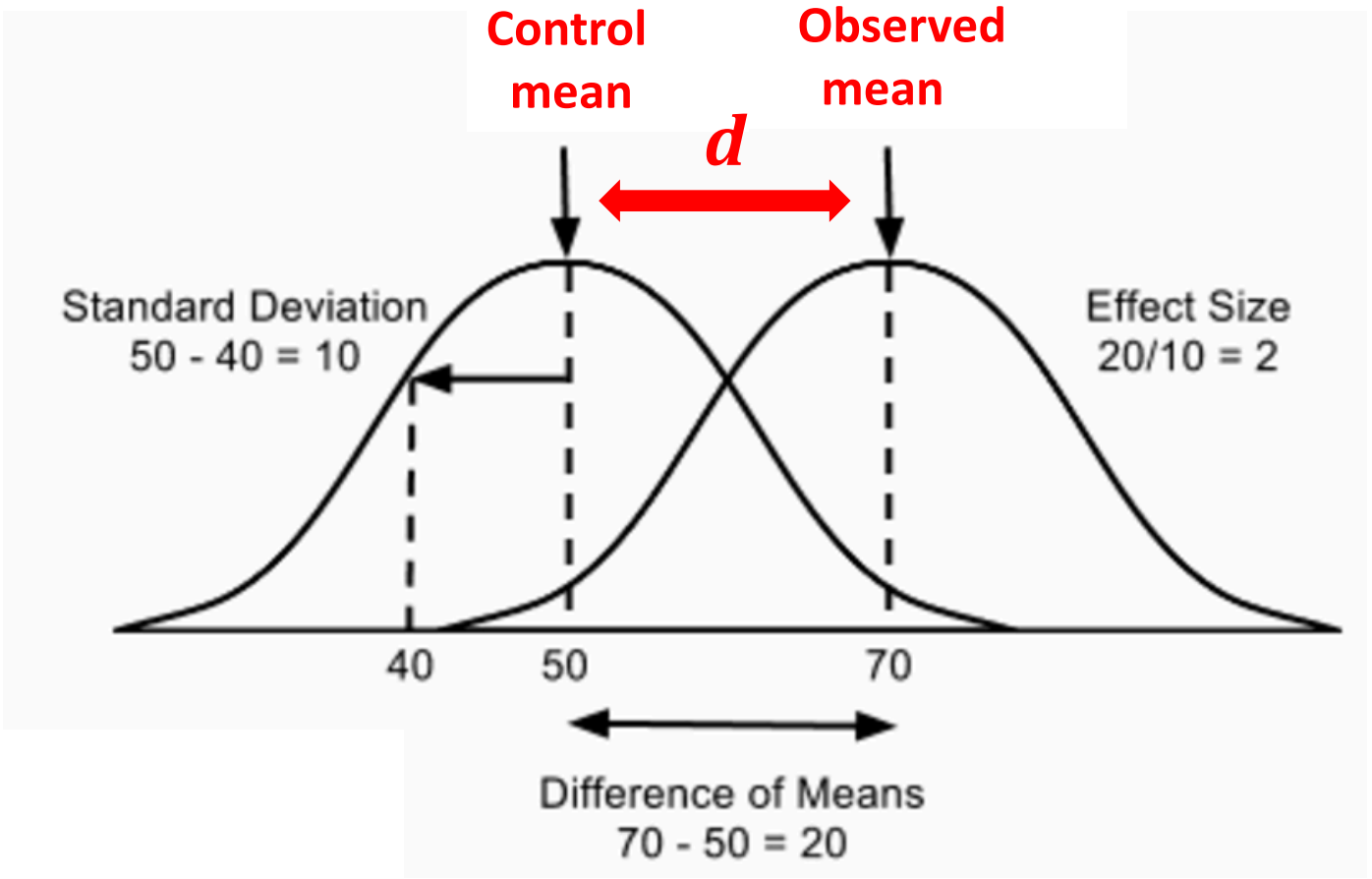
$$SD_{pooled} = \sqrt{(SD^2_{group1} + SD^2_{group2})/2}$$

An example effect size calculation

Effect Size = $\frac{\text{diff. between means}}{\text{standard dev.}}$

$$d = \frac{M_{group1} - M_{group2}}{SD_{pooled}}$$

$$SD_{pooled} = \sqrt{(SD^2_{group1} + SD^2_{group2})/2}$$



Plot from: http://www.realeducationreform.com/questions.html

Explore Cohen's *d*

https://rpsychologist.com/d3/cohend/

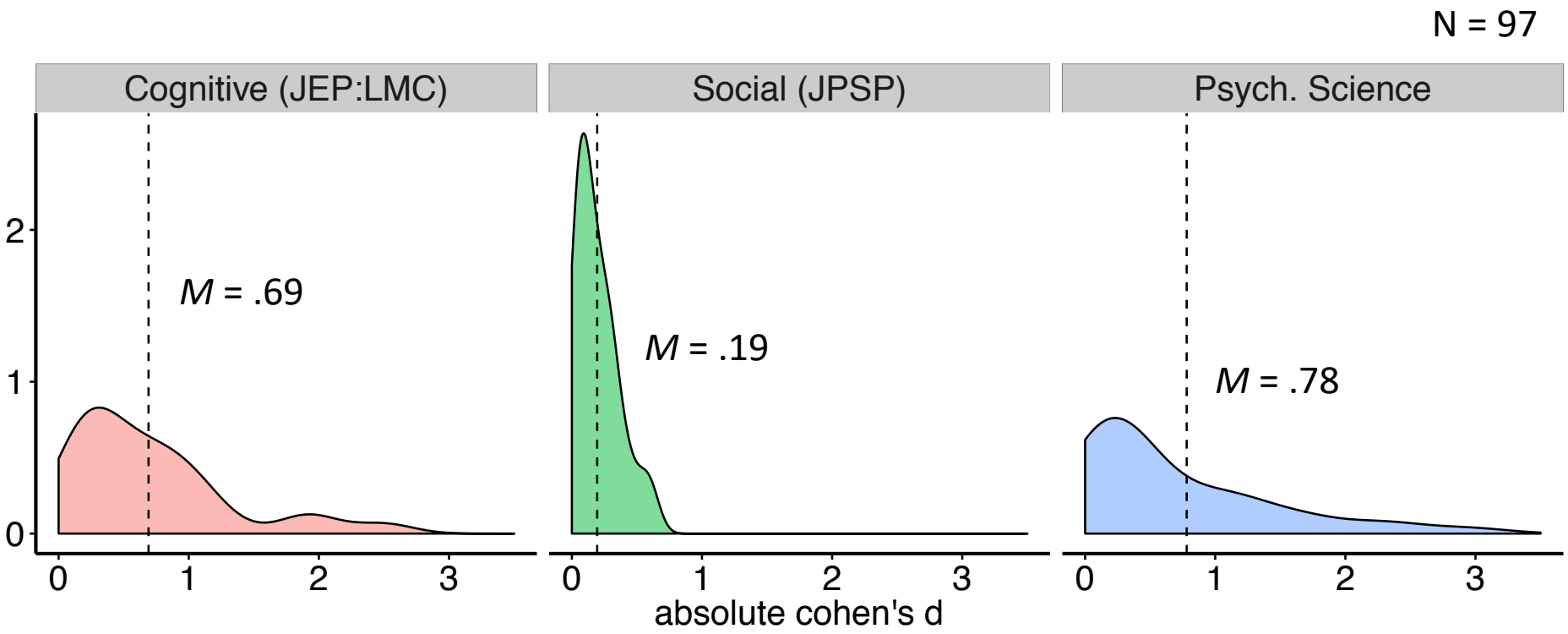
Interpreting Cohen's *d*

Size	Description	Cohen's Intuition	Psychological Example
.2	"small"	Diff. between the heights of 15 yo and 16 yo girls in the US	Bouba-kiki effect in kids (~.15; Lammertink, et al. 2016)
.5	"medium"	Diff. between the heights of 14 yo and 18 yo girls.	Cognitive behavioral therapy on anxiety (~ .4; (Belleville, et al., 2004) Sex difference in implicit math attitudes (~.5; Klein, et al., 2013)
.8	"large"	Diff. between the heights of 13 yo and 18 yo girls.	Syntactic Priming (~.9; Mahowald, et al., under review) Mutual exclusivity (~1.0; Lewis & Frank, in prep)

(Cohen, 1969)

Interpreting Cohen's *d*

Estimating the Replicability of Psychological Science (OSF, *Science*, 2015)



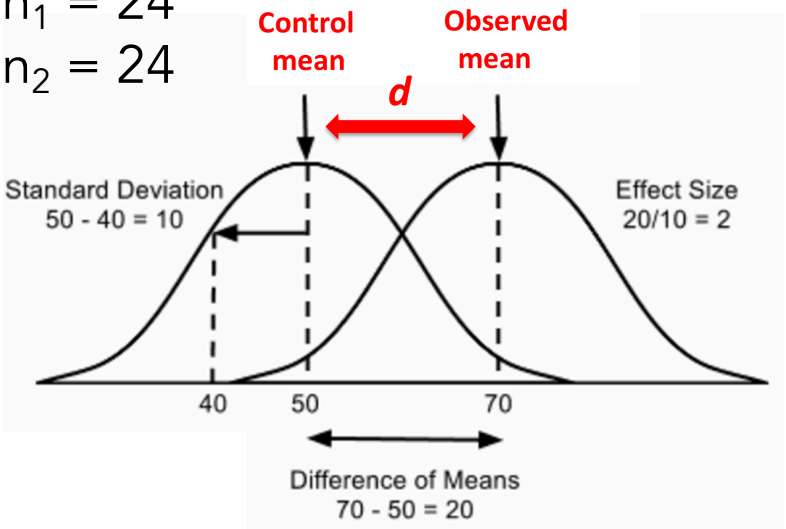
Relatively "large" effects reported in cognitive psychology

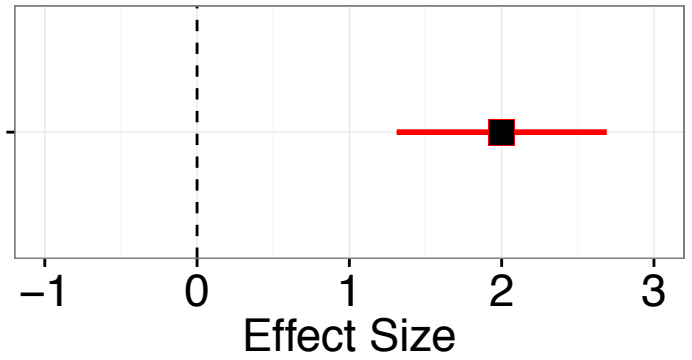
Effect size measures

- Cohen's *d* is just one (prototype)
- Appropriate effect size measure depends on aspect of design (e.g., within vs. between subject), and types of variables (e.g. qualitative vs. quantitative).
- For any statistical test you conduct, can compute effect size (in principle)
  - the difference is between groups (t-test, *d*)
  - the relationship between variables (correlation, *r)*
  - the amount of variance accounted for by a factor (ANOVA, regression, *f)*
  - …
- Can convert between ES metrics

Effect size confidence interval

$n_1 = 24$
$n_2 = 24$





$$
\begin{aligned}
var_d &= \frac{n_1 + n_2}{n_1 * n_2} + \frac{d^2}{2(n_1 + n_2)} \\
&= \frac{24 + 24}{24 * 24} + \frac{2^2}{2(24 + 24)} \\
&= .125
\end{aligned}
$$

$$
\begin{aligned}
CI(d) &= Est(d) \pm z_{(\alpha/2)} * \sqrt{var(d)} \\
&= 2 \pm 1.96 * .35 \\
&= 2 \pm .69
\end{aligned}
$$

Calculating effect sizes in R

- Compute.es package

After Spring Break: Actually doing our own replication study

- Replicate a published experiment online using Amazon Mechanical Turk
- Will collect and analyze are own data