Safe Summer Stats Workshop 2020

The era of bad stuff, but also free virtual meetings.

Visit Erin Nishimura's lab at: onishlab.colostate.edu

Writing workshop – July 16th

Science communication workshop – August 6th

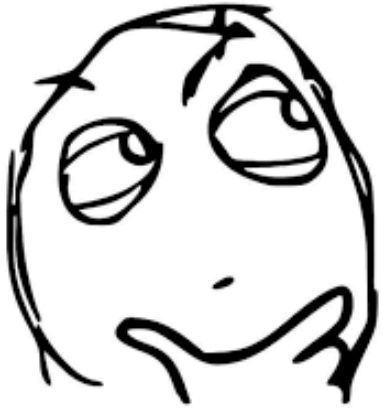
Show and Tell Style. Informal! Ask questions! Discuss!

Program:

1. Mikaela Elder – Regression
2. Rob Williams – Permutation Test
3. Adeline Williams – Chi-square
4. Adeline and David King – Equivalence Testing
5. Ben Prytherch – Effect Size and Power Analysis

Equivalence testing: Statistical test for NO difference

We learned how to do this in undergrad stats…. right?

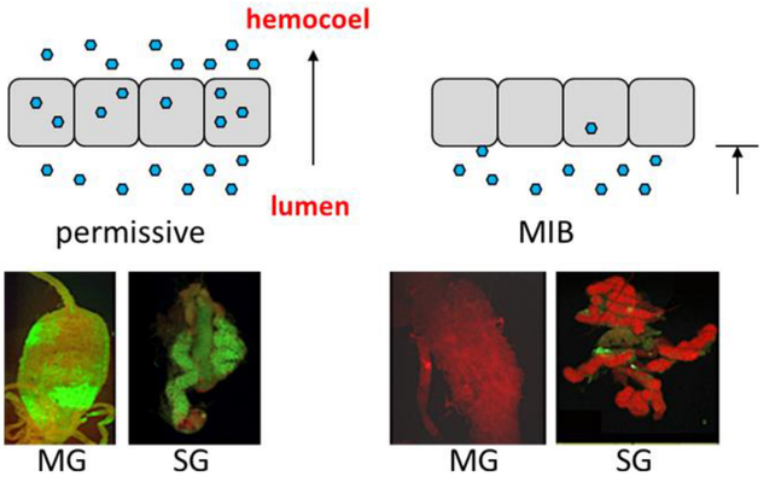


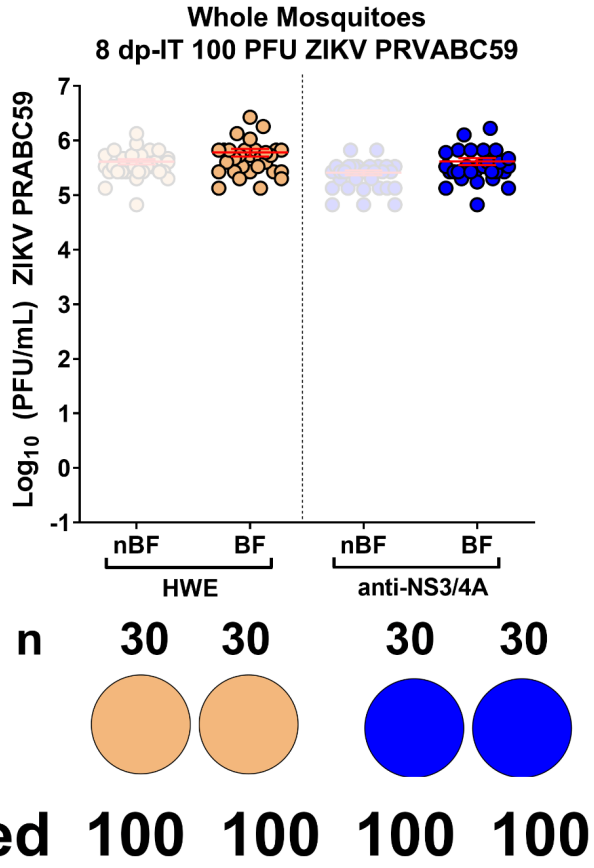
Lakens, Daniël, et al. *"Equivalence Testing for Psychological Research: A Tutorial."* Advances in Methods and Practices in Psychological Science, vol. 1, no. 2, 2018 *(not really a tutorial)*

Statistical test	Null hypothesis (H_0)	p-value	p-value < 0.05	p-value > .05
Two-sample t-test	No difference between the **means of the samples**	Probability of seeing observed Δ (or greater) if H_0 is true	Reject H_0. Support H_a: **means** are unequal.	Fail To Reject H_0

- How do we test the inverse? that the means are *not different?*
- Can't we just use a high p-value to *accept H_0? No?* Why the H naught?
- What if $p(H_a)$ is also high? We're only describing $p(H_0)$
- Other methods compare probabilities, i.e., $p(H_1)/p(H_2)$ to see which is higher; Bayesian modeling, mixture models…
- How do we test, and get a p-value, for two things being equal?
- Example from Adeline's data.

Anti-NS3/4A are resistant to ZIKV due to midgut infection barrier (MIB)

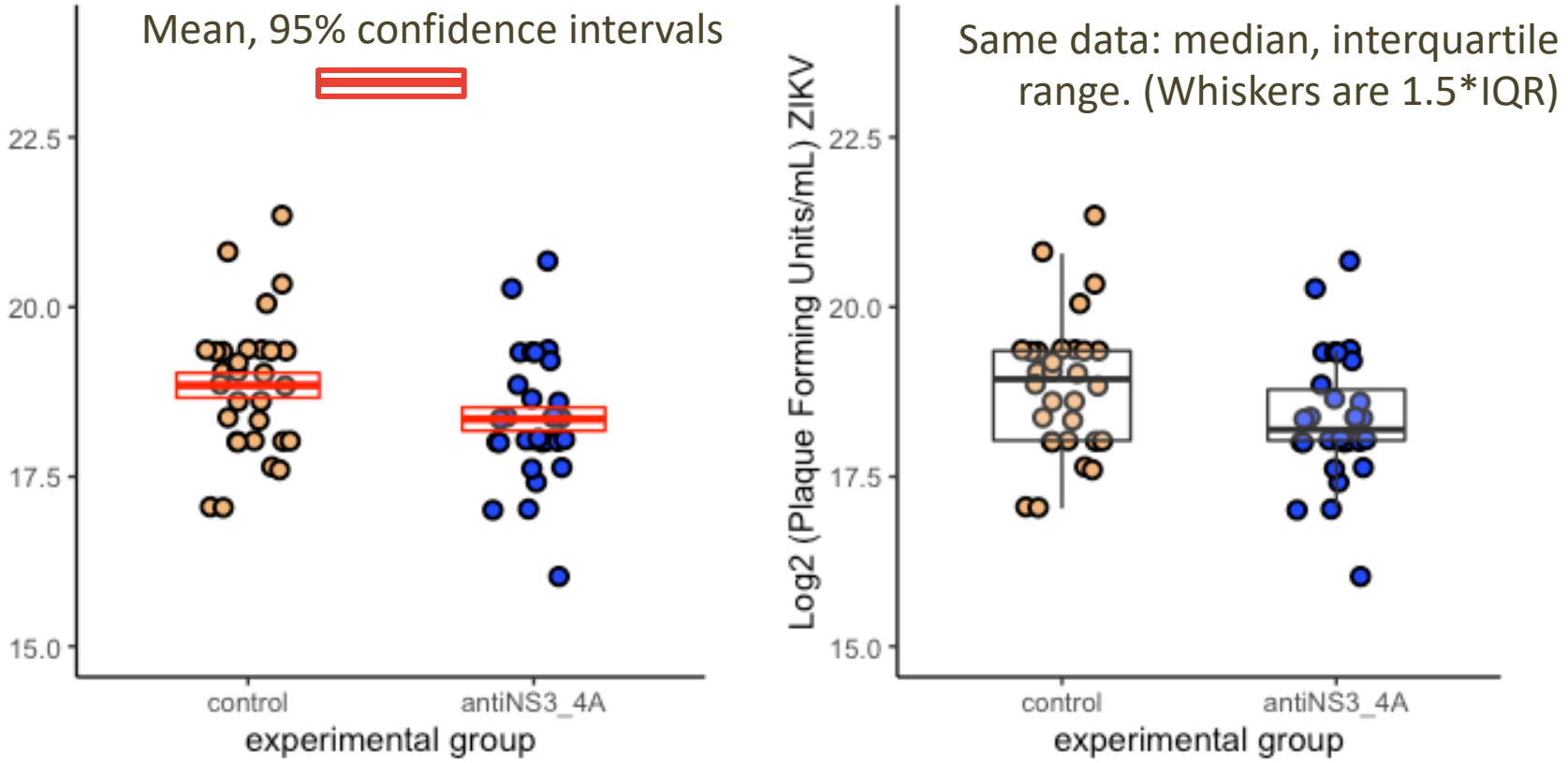




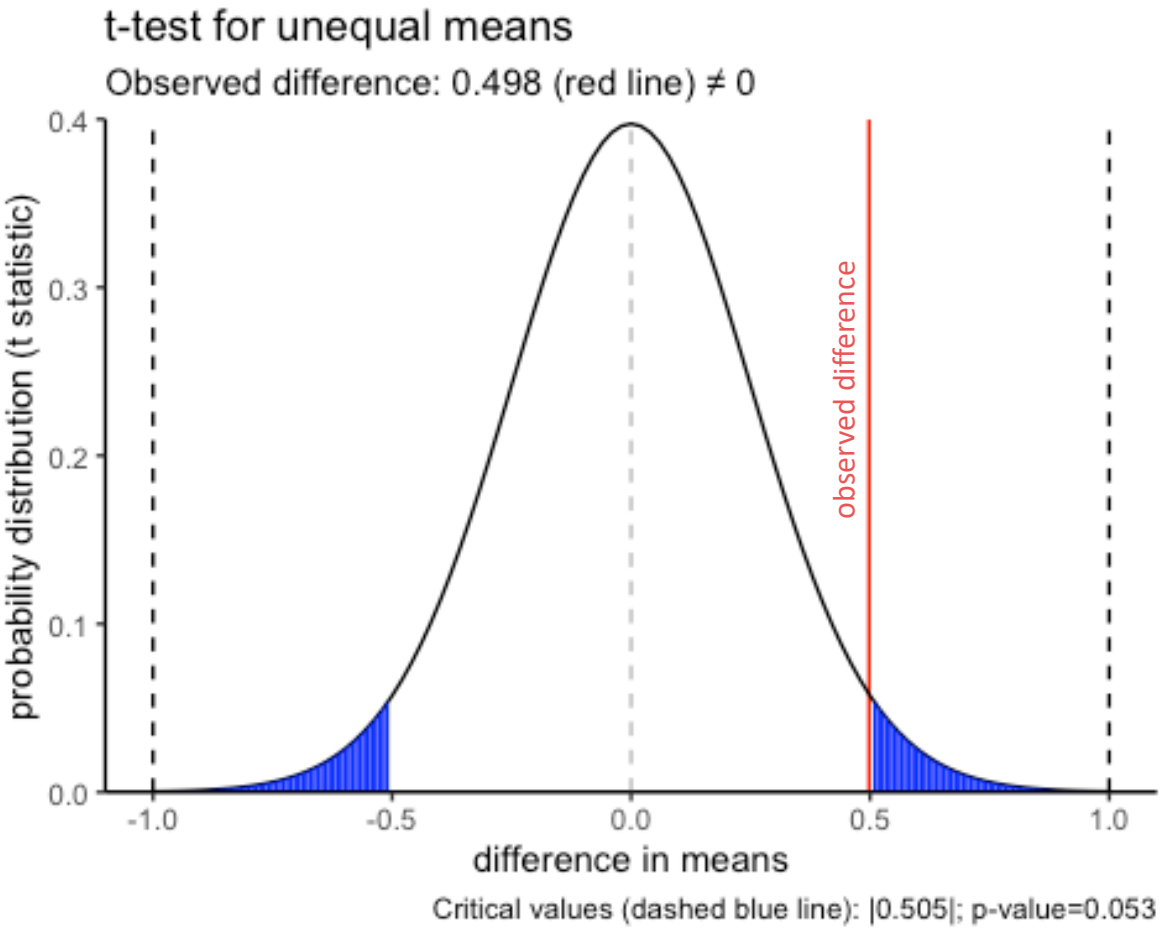
Franz et al., 2015

Zooming in on the data: is there a difference between antiNS3/4A and the control?

Viral titre in infected mosquitos



T-test (unequal means)



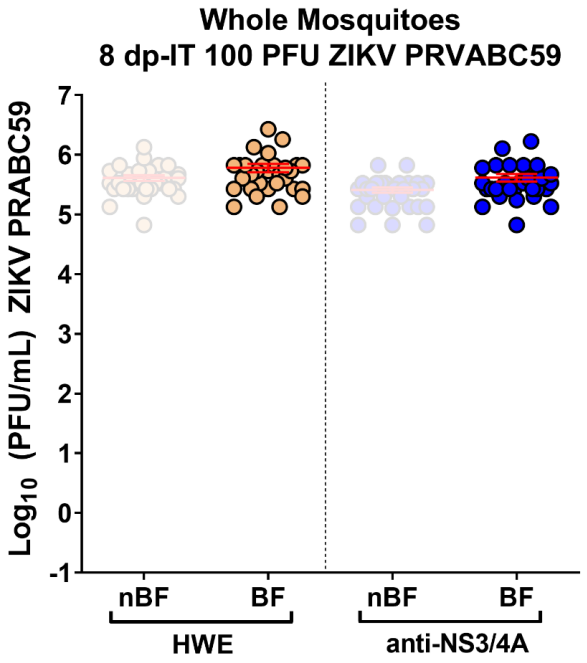
Define an equivalence boundary

Question: When you present this to your peers, how far apart do they need to be before you say they are different?

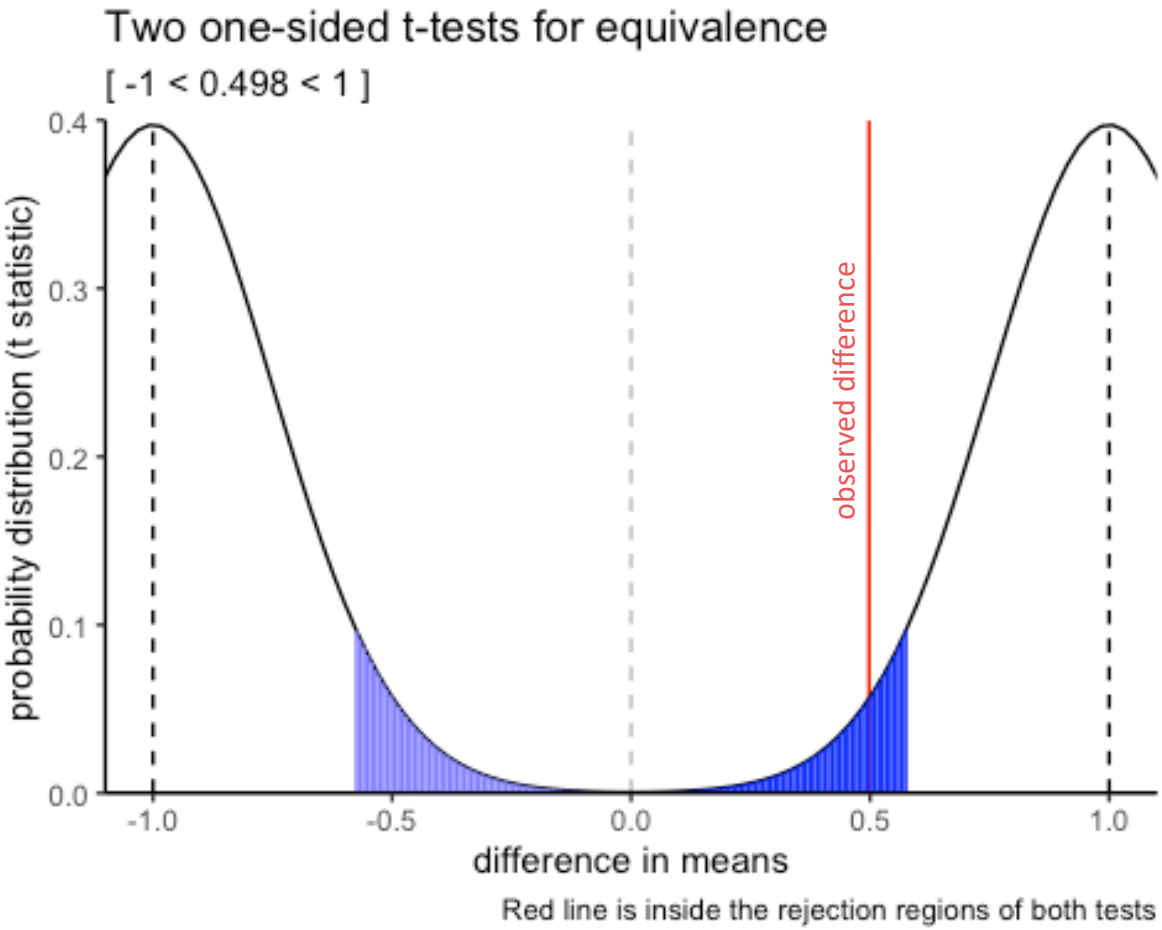
Answer: A 2-fold difference

Equivalence bounds: ±1 in log2 space.

We'll discuss different choices for this at the end.



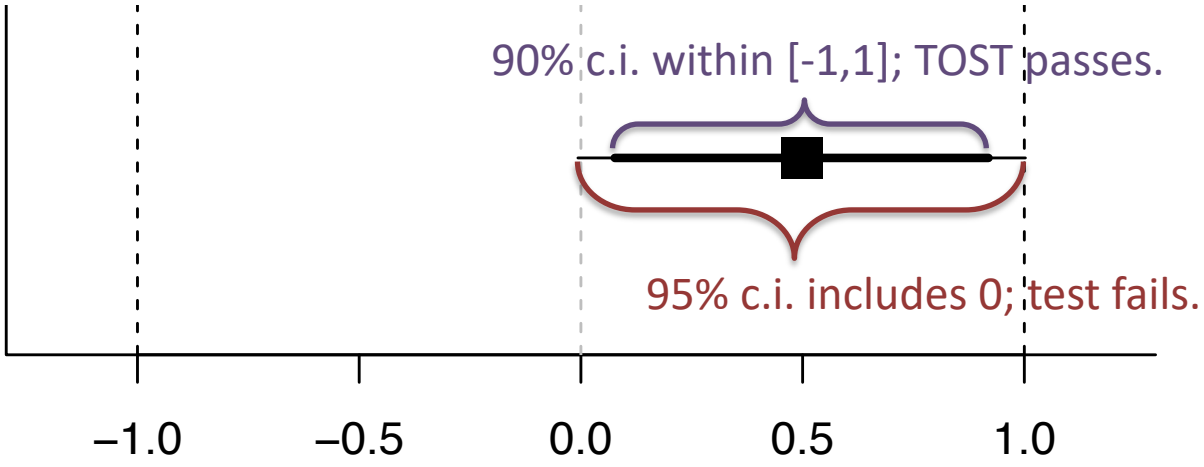
Equivalence test: Two One-Sided Tests (TOST).



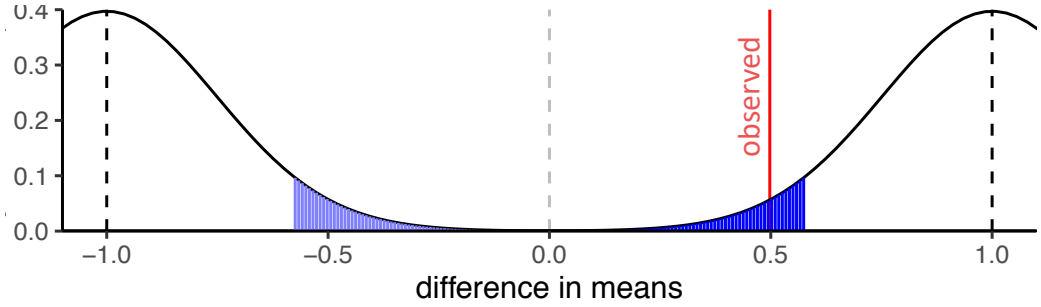
Confused about the switch from 90% confidence intervals to conclusions with 95% certainty? Good. That means you are paying attention. It **is** confusing!

Help doc from Prism/GraphPad

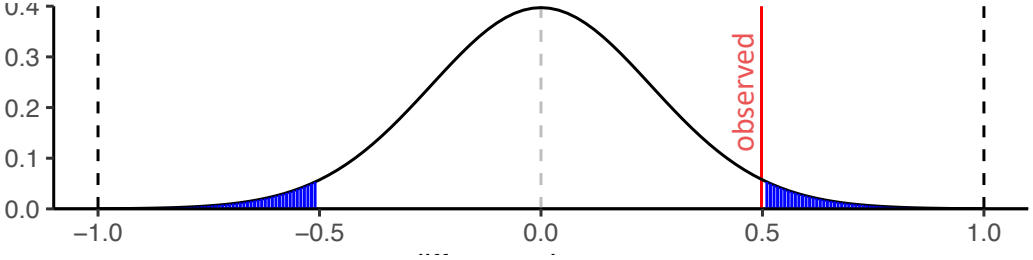
TOSTER package (R) output: confidence intervals



Two one-sided t-test distributions



Original two-sided t-test against 0: distribution



```
## TOST results:
## t-value lower bound: 5.94    p-value lower bound: 0.00000009
## t-value upper bound: -1.99   p-value upper bound: 0.026
## degrees of freedom : 57.8
##
## Equivalence bounds (raw scores):
## low eqbound: -1
## high eqbound: 1
##
## TOST confidence interval:
## lower bound 90% CI: 0.076
## upper bound 90% CI:  0.919
##
## NHST confidence interval:
## lower bound 95% CI: -0.007
## upper bound 95% CI:  1.002
##
## Equivalence Test Result:
## The equivalence test was significant, t(57.8) = -1.992, p = 0.0255,
given equivalence bounds of -1.000 and 1.000 (on a raw scale) and an
alpha of 0.05.
##
## Null Hypothesis Test Result:
## The null hypothesis test was non-significant, t(57.8) = 1.975, p =
0.053, given an alpha of 0.05.
##
## Based on the equivalence test and the null-hypothesis test combined,
we can conclude that the observed effect is statistically not different
from zero and statistically equivalent to zero.
```

Possible outcomes

t-test

TOST

	t-test for difference Reject null	t-test for difference Fail to Reject null
Two one-sided tests: both tests reject null	Different, but within equivalence bounds.	Equivalent (what we saw).
Two one-sided tests either test Fails To Reject	Different and exceeding equivalence bounds.	Inconclusive.

Choice of smallest effect size of interest (SESOI)

Tutorial (Lakëns et al., 2018) discussion.

Just-noticeable difference. Burriss et al., 2015:

Research question: Do ovulating women signal men by becoming more flush in the face?

- NIH funded research?
- Maybelline?

Just-noticeable difference = min. Δ in face redness perceptible by men.

Authors' conclusions: women do become more flush, but not enough to be discernable by the naked eye. (Maybelline cancels rouge ad project)

Statistical Power, no laughing matter

- What **effect size** can we detect at alpha=.05, sample size n, power=.8?
- **Effect size** in units of *Cohen's d.*
- Run TOST using **effect size**, rather than raw value.

Inform your choice from prior research.

One more thing

- I will forever look at the 95% confidence interval as the TOST equivalence boundary at α = .025
- Predefining an equivalence bound *a priori* is a specific test of a hypothesis. A succinct framing of the question.
- TOSTER has functions for:
  - one-sample t-test
  - two-sample t-test
  - proportion test
  - correlation (r)