4.3 Determining Statistical Significance and 4.4 A Closer Look at Testing

E. Nordmoe

Math 261

Section 4.3 Determining Statistical Significance

p-value and H_0

- If the p-value is small, then a statistic as extreme as that observed would be unlikely if the null hypothesis were true, providing significant evidence against H_0.
- The *smaller* the p-value, the *stronger* the evidence against the null hypothesis and in favor of the alternative.

Formal Decisions

If the p-value is small:

- Reject H_0
- the sample would be extreme if H_0 were true
- the results are statistically significant
- we have evidence for H_a

If the p-value is not small:

- Do not reject H_0
- the sample would not be too extreme if H_0 were true
- the results are not statistically significant
- the test is inconclusive; either H_0 or H_a may be true

But how small is small?

Significance Level

The significance level α is the threshold *below* which the p-value is deemed small enough to reject the null hypothesis.

- p-value $< \alpha \implies$ Reject H_0
- p-value $\geq \alpha \implies$ Do not reject H_0

The most common value of α is .05

- Values of .10 and .01 are also common.

Significance Level

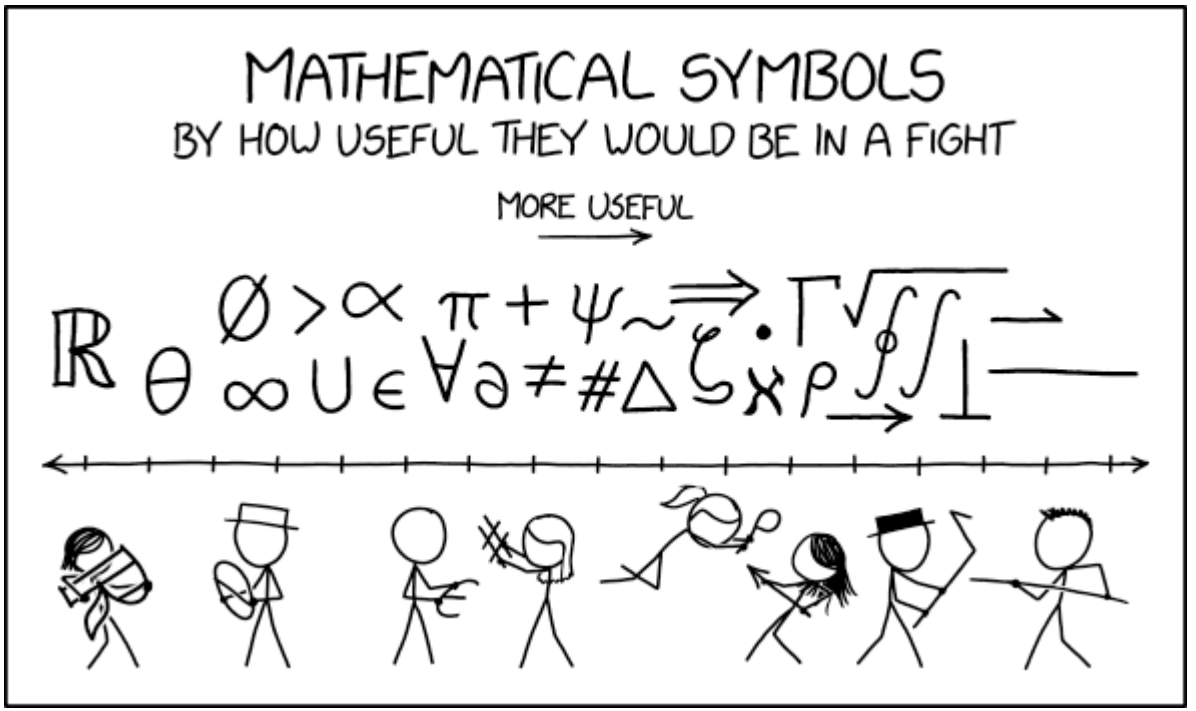
- p-value $< \alpha \implies$ Results are statistically significant
  - Reject H_0 in favor of H_a
- p-value $\geq \alpha \implies$ Results are not statistically significant
  - Test is inconclusive

Components of a Hypothesis Test

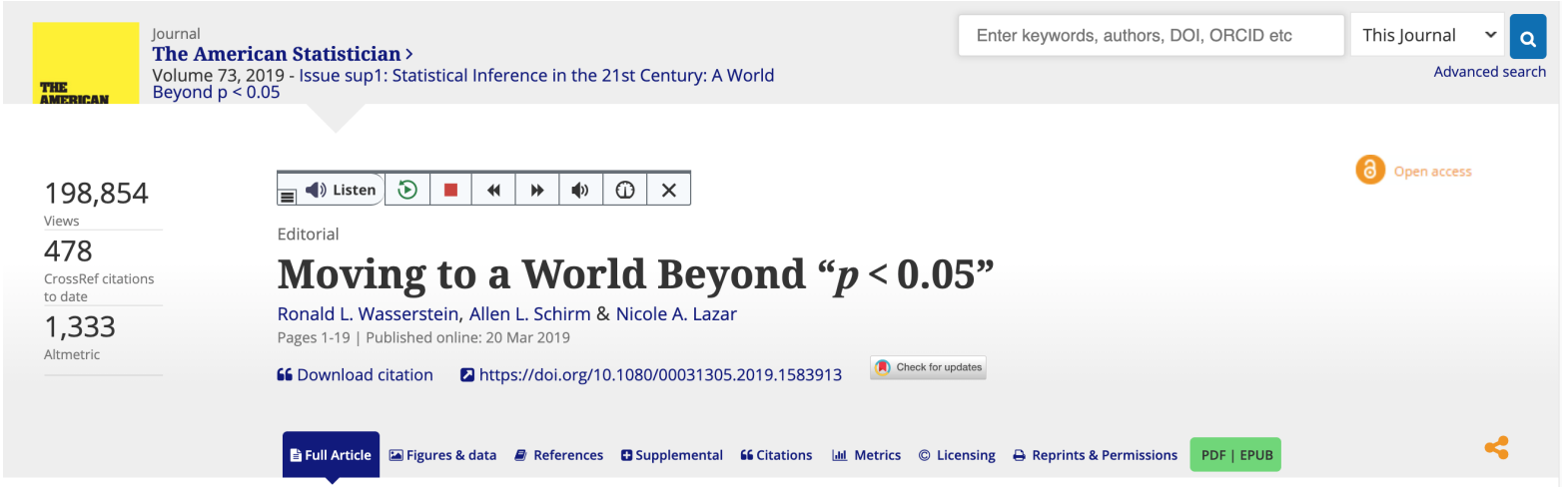
A formal hypothesis test includes the following components:

1. State the null and alternative hypotheses (defining parameters when necessary).
2. Determine the value of the observed sample statistic.
3. Find the p-value.
4. Make a generic decision about H_0 **:** Reject H_0 or do not reject H_0.
5. Write a sentence explaining the conclusion of the test in context, indicating whether or not we have convincing evidence for H_a and referring back to the question of interest.

Significance Testing is Controversial

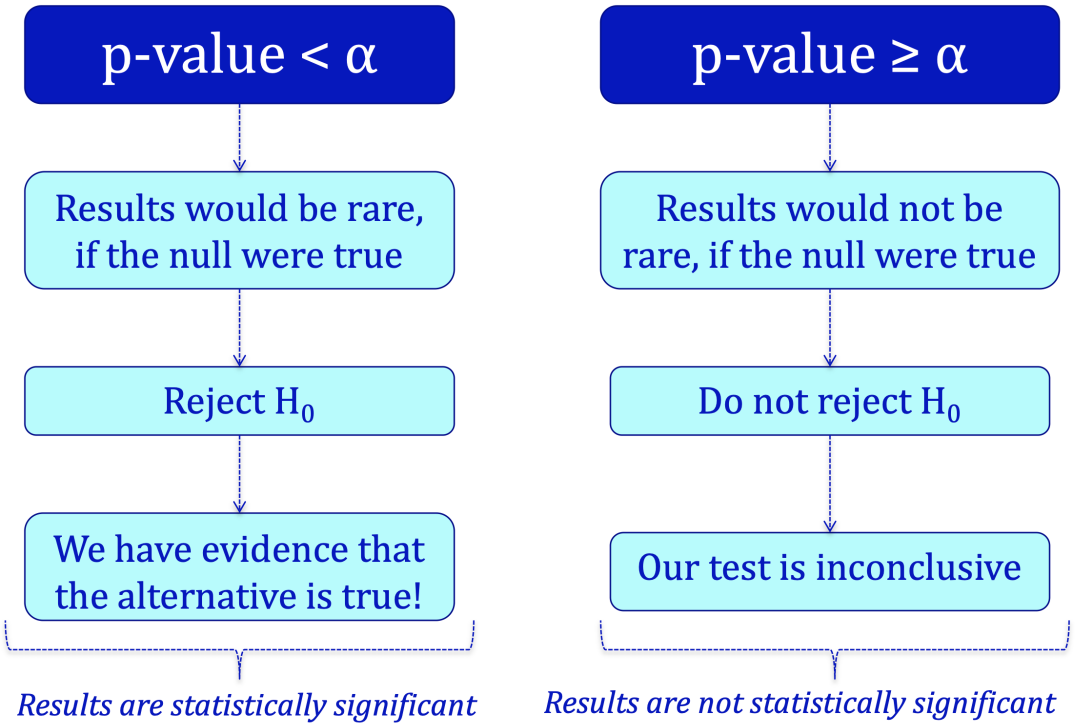

MATHEMATICAL SYMBOLS
BY HOW USEFUL THEY WOULD BE IN A FIGHT
MORE USEFUL

Moving to a World Beyond $p < 0.05$



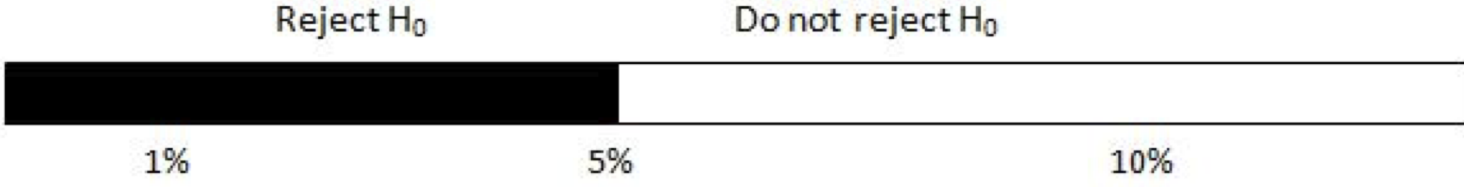
- Link to the article

P-value, Significance Levels, and Decisions

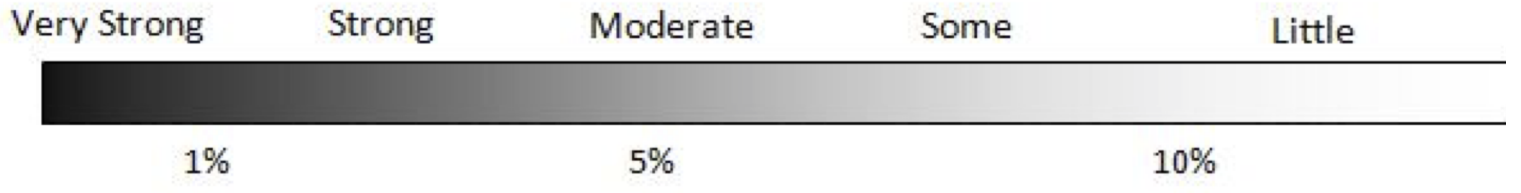


Statistical Conclusions

Formal decision of hypothesis test, based on $\alpha = 0.05$:



Informal strength of evidence against H_0:



Never Accept the Null

- "Do not reject H_0" is not the same as "accept H_0"!
- Lack of evidence against H_0 is not the same as evidence for H_0!

Section 4.4 A Closer Look at Testing

Type I and Type II Errors

	Reject H_0	Do not reject H_0
H_0 is true	Type I Error (α)	No error
H_0 is false	No error	Type II Error (β)

Type I Error $==>$ False positive

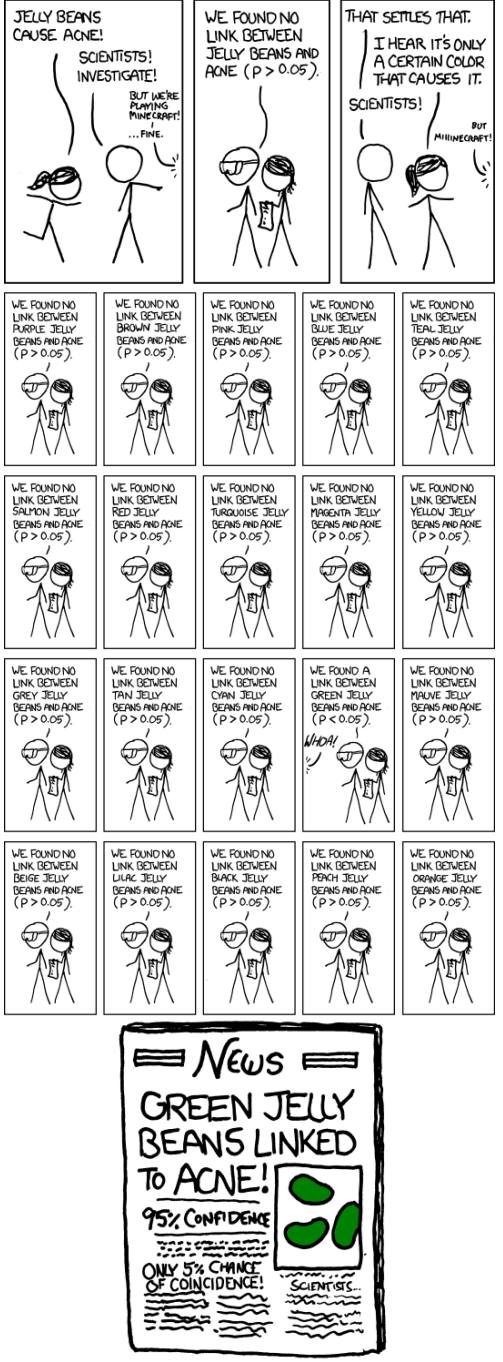
Type II Error $==>$ False negative

Significance Level

The significance level, α, represents the tolerable probability of making a Type I error.

Multiple Testing

- When multiple hypothesis tests are conducted, the chance that at least one test incorrectly rejects a true null hypothesis increases with the number of tests.
- If the null hypotheses are all true, α of the tests will yield statistically significant results just by random chance.


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
... FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
News
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Publication Bias

- Publication bias refers to the fact that usually only the significant results get published.
- The one study that turns out significant gets published, and no one knows about all the insignificant results.
- This combined with the problem of multiple comparisons, can yield very misleading results.

Replicating Results

By attempting to *replicate* significant results with another study, this second study can either:

- Reject H_0, providing further confirmation that H_a is true

OR

- Fail to reject H_0, suggesting that the first study may have yielded a Type I error.

Practical vs Statistical Significance

- With small sample sizes, even large differences or effects may not be significant.
- With large sample sizes, even a very small difference or effect can be significant.
- A statistically significant result is not always practically significant, especially with large sample sizes.