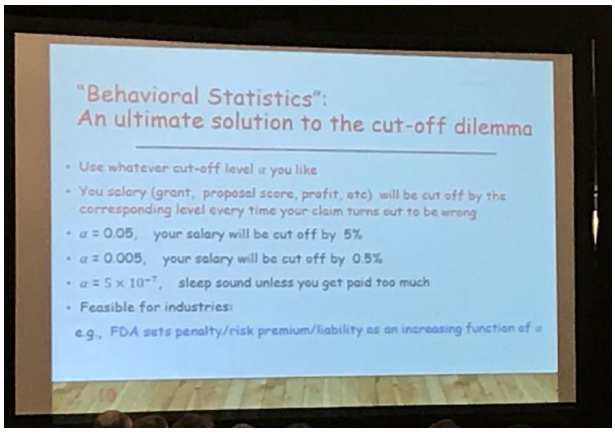


The following slides are taken from, and discussed at, this excellent blog post:
https://scientistseessquirrel.wordpress.com/2018/10/02/presenting-statistical-results-pointers-in-scientific-writing/#foot1

"P < 0.05" or "P = 0.022"?

- Next three issues are closely related. How do you report, discuss *P* value?
- Some papers only indicate $P < 0.05$ or $P > 0.05$
- This is based on a "line in the sand" philosophy of interpreting *P:* one should set alpha, and then care only about whether *P* is smaller than alpha.
- BUT: many statisticians see the magnitude of *P* as indicating strength of evidence; if so, then $P = 0.04$ and $P = 0.000004$ tell us different things…
  … even though both are "$P < 0.05$" (more about this later)

1. Exact *P*-values can be used in meta-analysis
   e.g., Fisher's method for combining P-values (https://wp.me/p5x2kS-qo)
2. A "line in the sand" reader may prefer a different line in the sand (e.g. $\alpha = 0.01$)
3. A "line in the sand" reader can always ignore the exact value of *P*; but a strength-of-evidence reader can't infer the exact value from "$P < 0.05$".

$$P = 0.007 \underset{\text{hard}}{\overset{\text{easy}}{\rightleftharpoons}} P < 0.05$$

"P = 0.037" or "P = 0.022823511"?

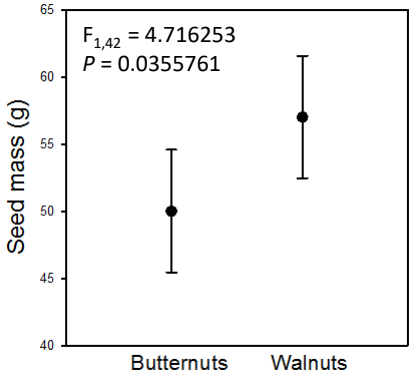
- To what precision should you report *P*? (See https://wp.me/p5x2kS-oN)
- I frequently see manuscripts reporting "$P = 0.022823511$". Do those 1's matter?
- This is a special case of a larger issue with "significant digits" in scientific writing
- For stats: think about
  (1) data-significant digits
  (2) reader-significant digits
- Usually, 2 or at most 3 significant digits

  $F_{1,42} = 4.72$, $P = 0.0356$

  or

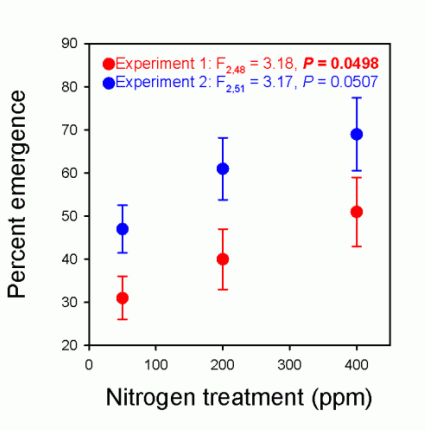
  $F_{1,42} = 4.7$, $P = 0.036$

- *OK, but what if* P = *0.049999999999?*



What about P = 0.051?

- What do you do with *P* = 0.051?
  - "nearly significant"?
  - "marginally significant"?
  - "trending toward significance"?
- Many people have strong, but poorly justified, opinions on this…
- Two alternative views of the *P*-value:
  - line-in-the-sand, or "absolutist"
  - strength-of-evidence, of "continualist"



What about P = 0.051?

- In the 1930s, Fisher and Neyman feuded, in part over line-in-the-sand vs. strength-of-evidence interpretation of *P.*
- Result was a tangled literature full of very strong statements…
- Strength-of-evidence view comes from Fisher's original insight that P(apparent pattern|no real effect) indicates evidence against the null *in a single experiment.*

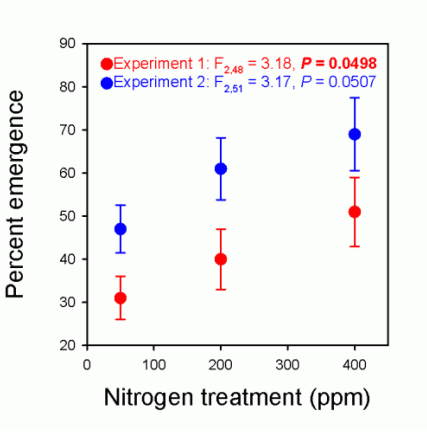
- Line-in-the-sand view comes from the application of stats to "process control": batches of manufactured product must be accepted or rejected as they come off an assembly line.
- Line-in-the-sand view happens to line up easily with superficial readings of Popper.

What about P = 0.051?

- Here's the thing: *P* = 0.049 and *P* = 0.051 aren't meaningfully different outcomes
  - statistically
  - logically
- So go ahead:

  "There were fewer loopers in sprayed plots, and the difference was nearly significant/ marginally insignificant/etc. (P = 0.051)"

- But expect to get pushback.



(More discussion: https://wp.me/p5x2kS-cR)

Stats first or pattern first?

Awful:

"The Welch's t-test produced significant results (t = 2.77, df=9, P = 0.022); see Figure 1"

Bad:

"The Welch's t-test produced significant results (t = 2.77, df=9, P = 0.022), with a difference in looper density (Figure 1) between sprayed and unsprayed plots"

Better:

"There was a significant difference in looper density (Figure 1) between sprayed and unsprayed plots (Welch's t-test, t = 2.77, df=9, P = 0.022)"

Good:

"Spraying reduced looper densities by about 23% (Figure 1; $t_{(9)}$ = 2.77, P = 0.022)"

Whose job is it to tell the story - the writer's or the reader's?
(Do "the data speak for themselves"?)