

# Statistical Rituals

## ReproducibiliTea

Matthew J. Vowels

CVSSP, Faculty of Engineering & Physical Sciences, UniS

August 17, 2020

## Today's Paper

## The Paper in More Detail

Introduction

The Ritual!

Empirical Evidence

What To Do?

## Discussion Points

# The Topic/Paper

- ▶ *'Statistical Rituals: The Replication Delusion and How We Got There'* (Gigerenzer, 2018)
- ▶ Relatively strong polemic on current practice
- ▶ Interesting history about Fisher, Neyman and Pearson, and the birth of the “null ritual”
- ▶ Empirical results to support claims
- ▶ Recommendations for improving current practice
- ▶ Assumes familiarity with NHST and what the  $p$ -value (*really*) means



# Crisis

- ▶ Replication crisis across the empirical sciences
- ▶ Some statistics
  - ▶ Irreproducible [irreplicable?] preclinical research costs 28M (USD) annually
  - ▶ 6/53 landmark cancer studies replicated
  - ▶ 14/67 studies in oncology, women's health, CV medicine, only 14 replicated
  - ▶ etc. etc.
- ▶ Well debated reasons for the crisis:
  - ▶ Science as a strategic game (DeDeo, 2020)
- ▶ Novel view...

# Crisis

## Statistical Ritual Hypothesis

“the replacement of good scientific practice by a statistical ritual that researchers perform not simply on the grounds of opportunism but because they have internalized the ritual and genuinely believe in it”

(Gigerenzer, 2018)



# History

## Fisher

- ▶ Testing single hypothesis against the null
- ▶ No power
- ▶ Stat. sig.  $\equiv$  subjective confidence / belief

## Neyman & Pearson

- ▶ Test against a second specified hypothesis
- ▶ Specify power (and therefore  $\alpha$ ,  $\beta$ )
- ▶ Neyman (not Pearson) stat. sig.  $\equiv$  decision not belief

# History

- ▶ Neyman said Fisher's tests were "worse than useless"
- ▶ Fisher said Neyman's theory was childish and "horrifying [for] the intellectual freedom of the west"
- ▶ Avoiding such subjectivity, a hybrid was born (from **neither** Fisher nor Neyman & Pearson).
- ▶ The results is what we know today...

# The Ritual!

1. Set up a null hypothesis of no mean difference or zero correlation. Do not specify the predictions of your own research hypothesis.
2. Use 5% as a convention for rejecting the null. If the test is significant, accept your research hypothesis.
3. Always perform this procedure.



# The Ritual!

- ▶ Step 1 violates Neyman & Pearson's logic and *kind of* violates Fisher's too - need a sensible null. Null grew to mean no difference - i.e. no thought/judgement required!
- ▶ Step 2 contradicts N&P AND Fisher vis the 5%:
  - ▶ No scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas.
- ▶ Step 3 mindless statistics (Gigerenzer, 2004)

# Why the terminology?

- ▶ Ritual consists actions undertaken in prescribed order with the following attributes:
  - ▶ sacred numbers or colors (e.g.,  $p$ -values or MRI images)
  - ▶ repetition (same procedure without adaptation)
  - ▶ fear of being punished when ceasing to perform the ritual (no publication)
  - ▶ wishful thinking (NHST tells you something it doesn't)

# Results - What does a significant $p$ -value mean to Psychologists?

- ▶ Null is false: 1- 66% depending on country
- ▶ Probability that null is true is known: 17-68%
- ▶ Alternative has been shown to be true: 10%
- ▶ Probability of alternative being true is known: 6-33%
- ▶ Probability of incorrectly rejecting null is known: 67-87%

# What do do?

- ▶ Editors no longer accept papers with dichotomous NHST
- ▶ Editors should distinguish between exploratory and confirmatory research
- ▶ Editors should require competitive-hypothesis testing (CHT)
- ▶ Change statistics education in psychology

# Discussion Points

- ▶ General thoughts?
- ▶ Is delusion the right word?
- ▶ Do we think it is a major contributor to the replication crisis?
- ▶ Education in psychology - can/should it be changed? (e.g., include separate research track?)

# Bibliography

DeDeo, S. (2020). When science is a game. *arXiv:2006.05994v2*.

Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*, 33:587–606.

Gigerenzer, G. (2018). Statistical rituals: the replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, 1(2):198–218.