PQHS 471
Lecture 13: Good Statistical Practice

The Lady Tasting Tea

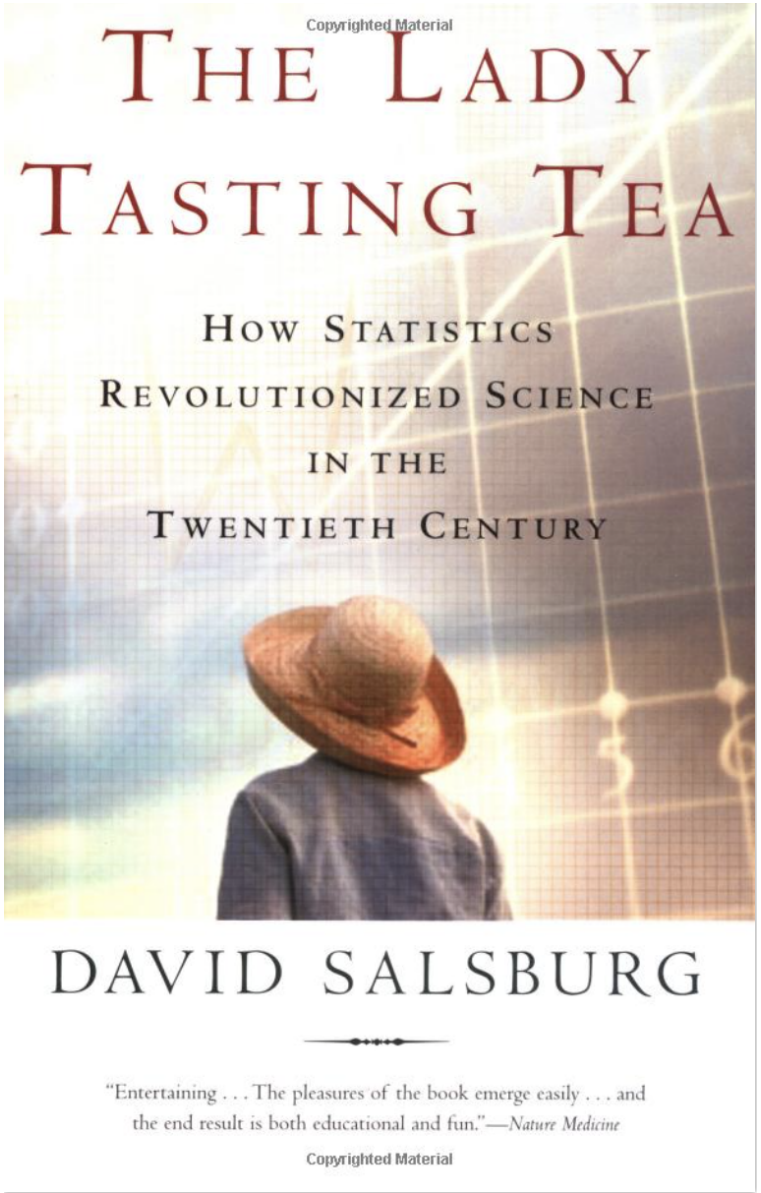
- It was a summer afternoon in Cambridge, England, in the 1920s.
- A groups of university dons, their wifes, and some guests were having afternoon tea.
- A lady was insisting that tea tasted different depending upon whether *the tea was poured into the milk* OR *the milk was poured into the tea*.

R.A. Fisher

Fisher in 1913

- "Sheer nonsense", the scientific minds among the men scoffed at this.
- A thin, short man, with thick glasses, Ronald Fisher, pounced on the problem: "Let us test the proposition!"

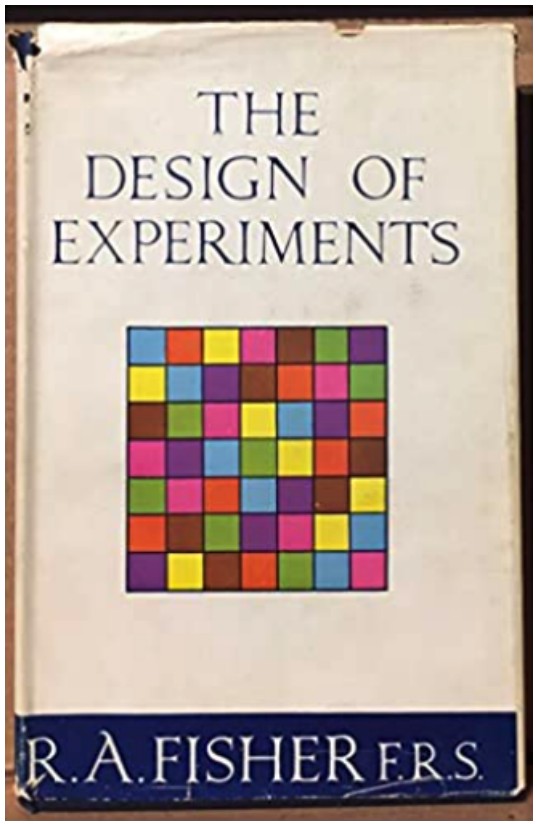
The Lady Tasting Tea



Hypothesis Testing

- Fisher's notion of a *null hypothesis*
  — Null hypothesis
  — Popularize p-value
- Neyman-Pearson Lemma
  — Error of the 2nd kind
  — Alternative/competing hypothesis
  — Power function

Most influential books on statistical methods



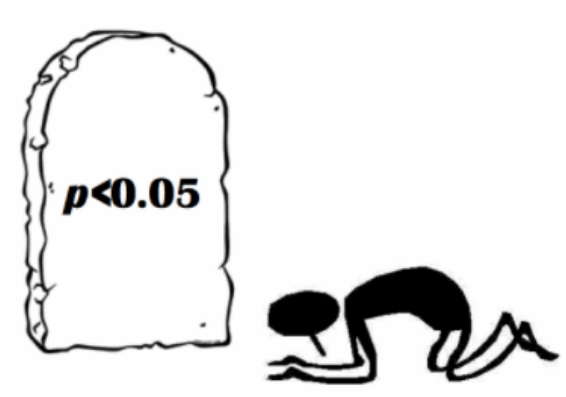
- **Statistical Methods for Research Workers**
- **The Design of Experiments**

“...the best thing about being a statistician...”

John Wilder Tukey

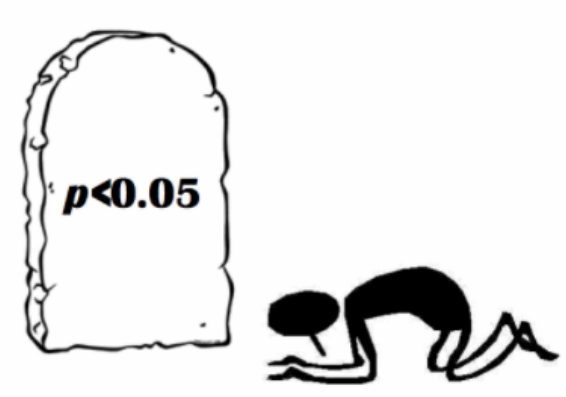
“... is that you get to play in everyone’s backyard.”

Misuse of p-value



- Q: Why do so many colleges and grad schools teach p = 0.05?
- A: Because that's still what the scientific community and journal editors use.
- Q: Why do so many people still use p = 0.05?
- A: Because that's what they were taught in college or grad school.

Misuse of p-value



- Q: Why do so many colleges and grad schools teach p = 0.05?
- A: Because that's still what the scientific community and journal editors use.
- Q: Why do so many people still use p = 0.05?
- A: Because that's what they were taught in college or grad school.

"We teach it because it's what we do; we do it because it's what we teach."

Fisher's words in SMRW

"Personally, the writer prefers to set a low standard of significance at 5 percentage point... A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance."

ASA Statement on p-values

Taylor & Francis

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

The ASA Statement on *p*-Values: Context, Process, and Purpose

Ronald L. Wasserstein & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein & Nicole A. Lazar (2016) The ASA Statement on *p*-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108

To link to this article: https://doi.org/10.1080/00031305.2016.1154108

pop quiz

Which(s) of the following statements is/are reasonable?

- p-value is a probability.
- $p > 0.05$ is the probability that the null hypothesis is true.
- 1 minus the p-value is the probability that the alternative hypothesis is true.
- A statistically significant test result ($p \leq 0.05$) means that the test hypothesis is false or should be rejected.
- A p-value greater than 0.05 means that no effect was observed.

The status quo

Informally, a p-value is the probability **under a specified statistical model** that a statistical summary of the data (e.g., the sample mean difference between two compared groups) would be *equal to or more extreme* than its observed value.

Six principles of p-value

- 1. P-values can indicate how incompatible the data are with a specified statistical model.
  — The most common context is a model (under a set of assumptions): H_0
  — Often H_0 postulates the absence of an effect (e.g. no difference between two groups)
  — The smaller the p-value, the greater the incompatibility of the data with H_0
  — Incompatibility casting doubt on H_0

Six principles of p-value

- 1. P-values can indicate how incompatible the data are with a specified statistical model.
  — The most common context is a model (under a set of assumptions): H_0
  — Often H_0 postulates the absence of an effect (e.g. no difference between two groups)
  — The smaller the p-value, the greater the incompatibility of the data with H_0
  — Incompatibility casting doubt on H_0
- 2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
  — Never turn a p-value into a statement about the truth of H_0
  — p-value is a statement about the **relationship** between the data and H_0, NOT about the **explanation** (H_0) itself.

Six principles of p-value (cont'd)

- 3. Scientific conclusions and business or policy decisions should NOT be based only on whether a p-value passes a specific threshold.
  — "bright-line" rule (e.g. $p < 0.05$ alone) can lead to erroneous beliefs and poor decision making.
  — A conclusion does not immediately become "true" on one side of the divide and "false" on the other.
  — Researchers should bring many contextual factors into play to derive scientific inferences, including the design of a study, the quality of the measurements, the external evidence for the phenomenon under study, and the validity of assumptions that underlie the data analysis.
  — Using $p < 0.05$ alone as a license for making a claim of a scientific finding leads to considerable distortion of the scientific process.

Six principles of p-value (cont'd)

- 3. Scientific conclusions and business or policy decisions should NOT be based only on whether a p-value passes a specific threshold.
  — "bright-line" rule (e.g. $p < 0.05$ alone) can lead to erroneous beliefs and poor decision making.
  — A conclusion does not immediately become "true" on one side of the divide and "false" on the other.
  — Researchers should bring many contextual factors into play to derive scientific inferences, including the design of a study, the quality of the measurements, the external evidence for the phenomenon under study, and the validity of assumptions that underlie the data analysis.
  — Using $p < 0.05$ alone as a license for making a claim of a scientific finding leads to considerable distortion of the scientific process.
- 4. Proper inference requires full reporting and transparency
  — number of hypotheses explored, all data collection decisions, all statistical analyses conducted
  — No "cherry-picking"

Six principles of p-value (cont'd)

- 5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
  — $pval \neq$effect size
  — Statistical sig. vs. biological sig.

Six principles of p-value (cont'd)

- 5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
  — $pval \neq$effect size
  — Statistical sig. vs. biological sig.
- 6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

Usage of p-value

- **Good statistical practice** is an integral part of **good scientific practice**.
  — study design and conduct, summaries of data, understanding of the phenomenon under study, interpretation of results in context, complete reporting, proper logical understanding of results.

Usage of p-value

- **Good statistical practice** is an integral part of **good scientific practice**.
  — study design and conduct, summaries of data, understanding of the phenomenon under study, interpretation of results in context, complete reporting, proper logical understanding of results.
- **No single index should substitute for scientific reasoning.**