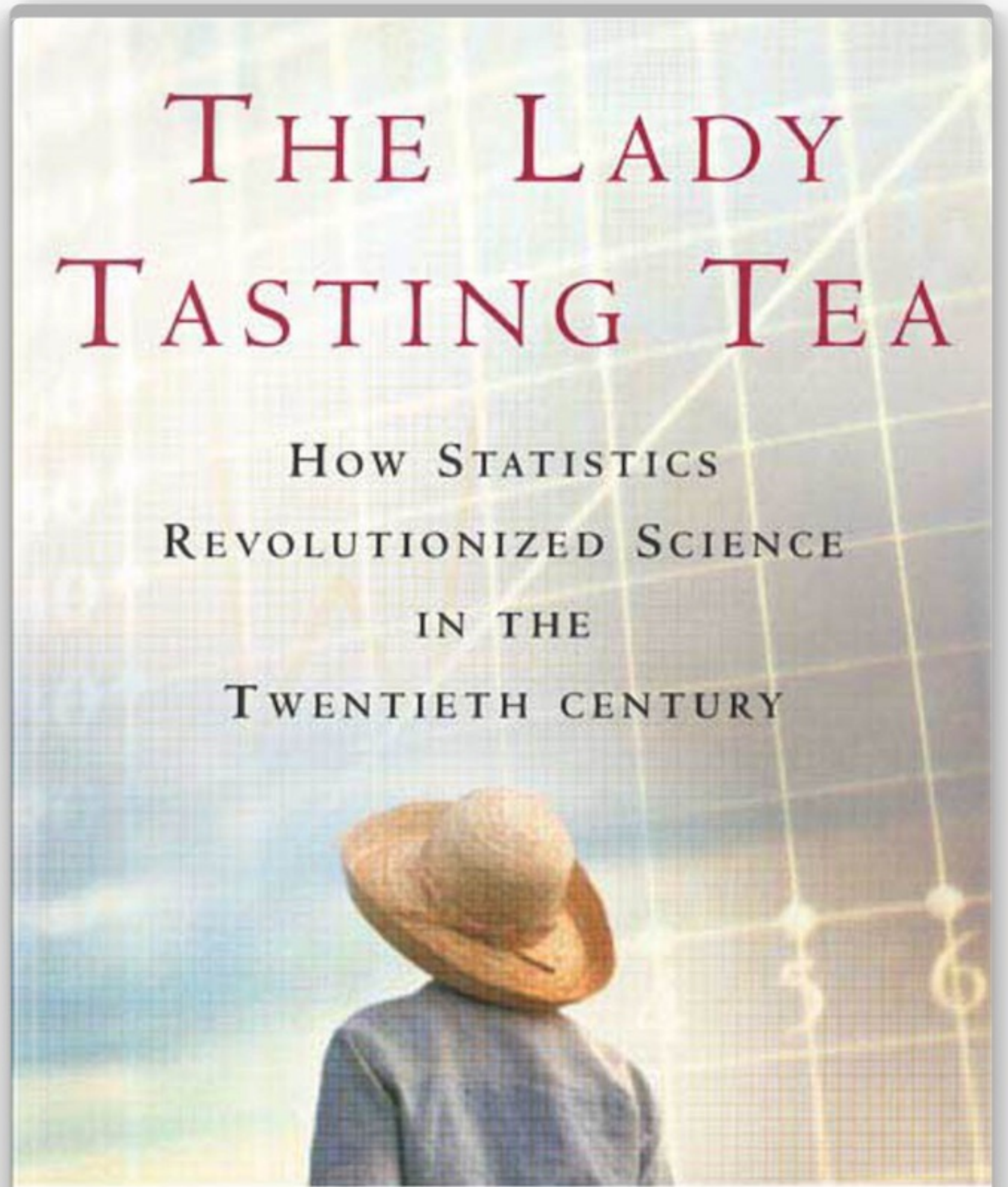
The botanist keeps tasting tea

A gentle introduction to e-values and sequential statistical inference

Ian Waudby-Smith

Department of Statistics, UC Berkeley

Miller retreat, 2024



DAVID SALSBURG

"A fascinating description of the kinds of people who interacted, collaborated, disagreed, and were brilliant in the development of statistics."
—Barbara A. Bailar, National Opinion research Center

The botanist tasting tea (1920s)

Ronald Fisher

Muriel Bristol

The botanist tasting tea (1920s)

Ronald Fisher

Would you like some tea? →

Muriel Bristol

The botanist tasting tea (1920s)

Ronald Fisher

Would you like some tea?

No, T in M $\neq$ M in T

Muriel Bristol

The botanist tasting tea (1920s)

Ronald Fisher

Would you like some tea? →

← No, T in M $\neq$ M in T

Can you *really* tell them apart? →

Muriel Bristol

The botanist tasting tea (1920s)

Ronald Fisher

Would you like some tea? →

← No, T in M $\neq$ M in T

Can you *really* tell them apart? →

← Indeed, yes!

Muriel Bristol

The botanist tasting tea (1920s)

Ronald Fisher

Would you like some tea? →

← No, T in M ≠ M in T

Can you *really* tell them apart? →

← Indeed, yes!

Muriel Bristol

T M T T M M M T

The botanist tasting tea (1920s)

Ronald Fisher

Would you like some tea? →

← No, T in M $\neq$ M in T

Can you *really* tell them apart? →

← Indeed, yes!

Muriel Bristol

T M T T M M M T

What's the probability that a *chance guess* would be *perfect*?

$1/70 \approx 0.014$

The botanist tasting tea (1920s)

Ronald Fisher

Would you like some tea? →

← No, T in M ≠ M in T

Can you *really* tell them apart? →

← Indeed, yes!

Muriel Bristol

T M T T M M M T

What's the probability that a *chance guess* would be *perfect*?

$1/70 \approx 0.014$

This is a p-value for H_0 : Muriel cannot distinguish btwn MT and TM.

The botanist tasting tea (1920s)

The botanist tasting tea (1920s)

However, the odds were stacked against Muriel from the start!

The botanist tasting tea (1920s)

However, the odds were stacked against Muriel from the start!

The probability that a *chance guess* would yield *at most one error* is 17/70 ≈ 0.24, which is not so impressive.

The botanist tasting tea (1920s)

However, the odds were stacked against Muriel from the start!

The probability that a *chance guess* would yield *at most one error* is 17/70 ≈ 0.24, which is not so impressive.

If she had made exactly one mistake, could we just let her keep tasting tea?
No, this is blatant p-hacking.

The botanist tasting tea (1920s)

However, the odds were stacked against Muriel from the start!

The probability that a *chance guess* would yield *at most one error* is 17/70 ≈ 0.24, which is not so impressive.

If she had made exactly one mistake, could we just let her keep tasting tea?
No, this is blatant p-hacking.

That is, the type-I error is not controlled: $\mathbb{P}_{H_0}(P_\tau \leq 0.05) \nleq 0.05.$

I have done some work on “e-values”.

I have done some work on "e-values".

An *e-value* is a function E of the data (MT or TM) so that

$$E \geq 0 \quad \text{and} \quad \mathbb{E}_{H_0}(E) \leq 1.$$

Suppose Muriel keeps tasting tea:

I have done some work on "e-values".

An e-value is a function E of the data (MT or TM) so that

$$E \geq 0 \quad \text{and} \quad \mathbb{E}_{H_0}(E) \leq 1.$$

Suppose Muriel keeps tasting tea:

I have done some work on "e-values".

An e-value is a function E of the data (MT or TM) so that

$$E \geq 0 \quad \text{and} \quad \mathbb{E}_{H_0}(E) \leq 1.$$

Suppose Muriel keeps tasting tea:

E_1 E_2 E_3 E_4 E_5 E_6 E_7 E_8

I have done some work on "*e*-values".

An *e*-value is a function E of the data (MT or TM) so that

$$E \geq 0 \quad \text{and} \quad \mathbb{E}_{H_0}(E) \leq 1.$$

Suppose Muriel keeps tasting tea:

I have done some work on "e-values".

An e-value is a function E of the data (MT or TM) so that

$$E \geq 0 \quad \text{and} \quad \mathbb{E}_{H_0}(E) \leq 1.$$

Suppose Muriel keeps tasting tea:

E_1 E_2 E_3 E_4 E_5 E_6 E_7 E_8 E_9 E_{10}

I have done some work on "*e*-values".

An *e*-value is a function E of the data (MT or TM) so that

$$E \geq 0 \quad \text{and} \quad \mathbb{E}_{H_0}(E) \leq 1.$$

Suppose Muriel keeps tasting tea:

$E_1 \quad E_2 \quad E_3 \quad E_4 \quad E_5 \quad E_6 \quad E_7 \quad E_8 \quad E_9 \quad E_{10} \quad \ldots$

I have done some work on “e-values”.

An e-value is a function E of the data (MT or TM) so that

$$E \geq 0 \quad \text{and} \quad \mathbb{E}_{H_0}(E) \leq 1.$$

Suppose Muriel keeps tasting tea:

E_1 E_2 E_3 E_4 E_5 E_6 E_7 E_8 E_9 E_{10} …

Then, $\mathbb{P}_{H_0}\left(P^\star_\tau \leq 0.05\right) \leq 0.05$, at any data-dependent sample size τ!

where $P^\star_n := \left(E_1 \cdot E_2 \ \cdots \ E_n\right)^{-1}$.

Thank you!

ianws.com

Ian Waudby-Smith

Department of Statistics, UC Berkeley

Miller retreat, 2024