Learn to live with lack of rejection

Einar Holsbø & Kajsa Møllersen, 21. Oct 2019

Table of contents

Table of contents

Table of contents

Table of contents

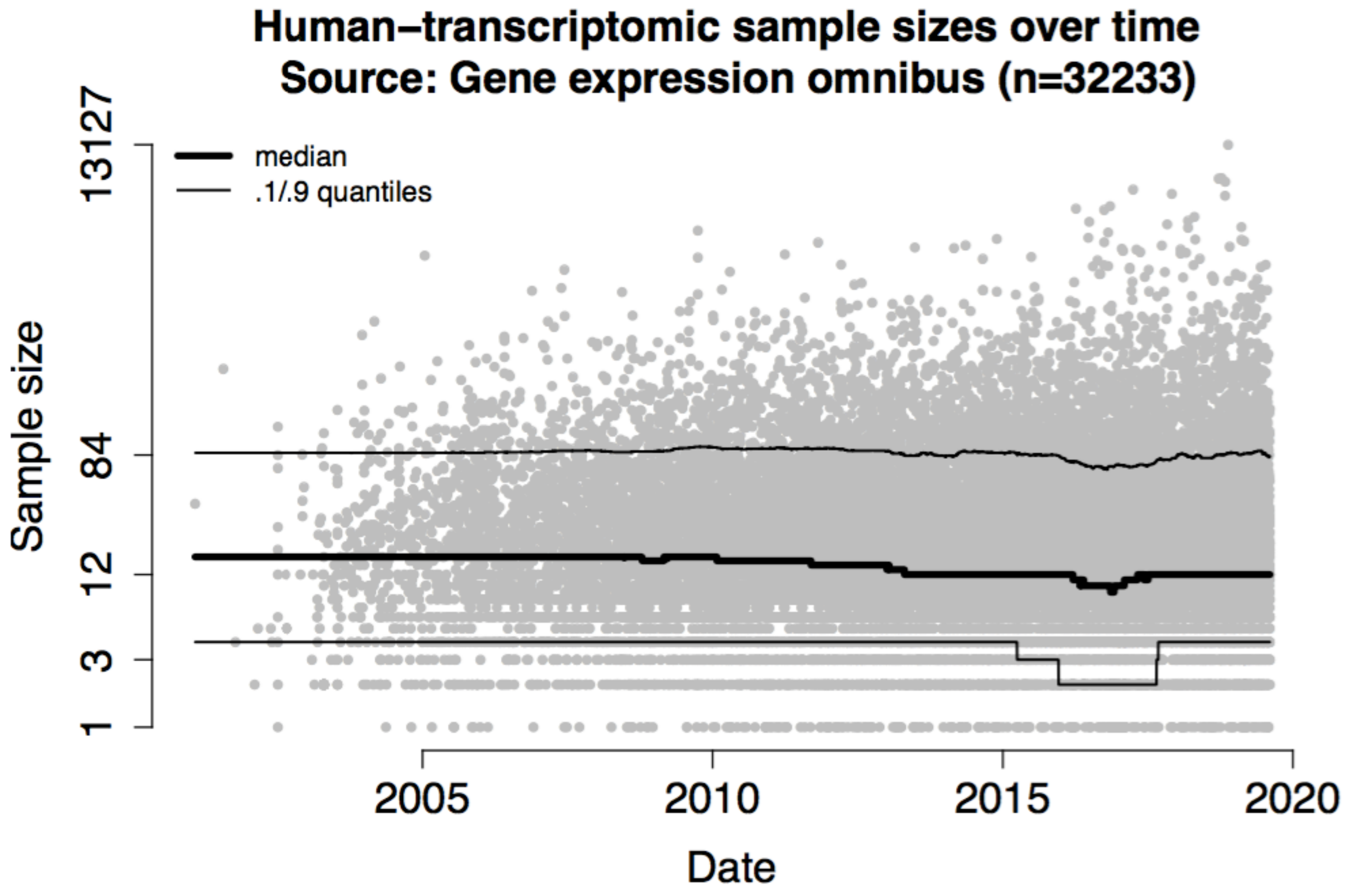
- Your datasets are tiny
- Typical data sizes limit the usefulness of hypothesis tests
- Typical avg. fold changes fall outside those limits

Table of contents

Table of contents

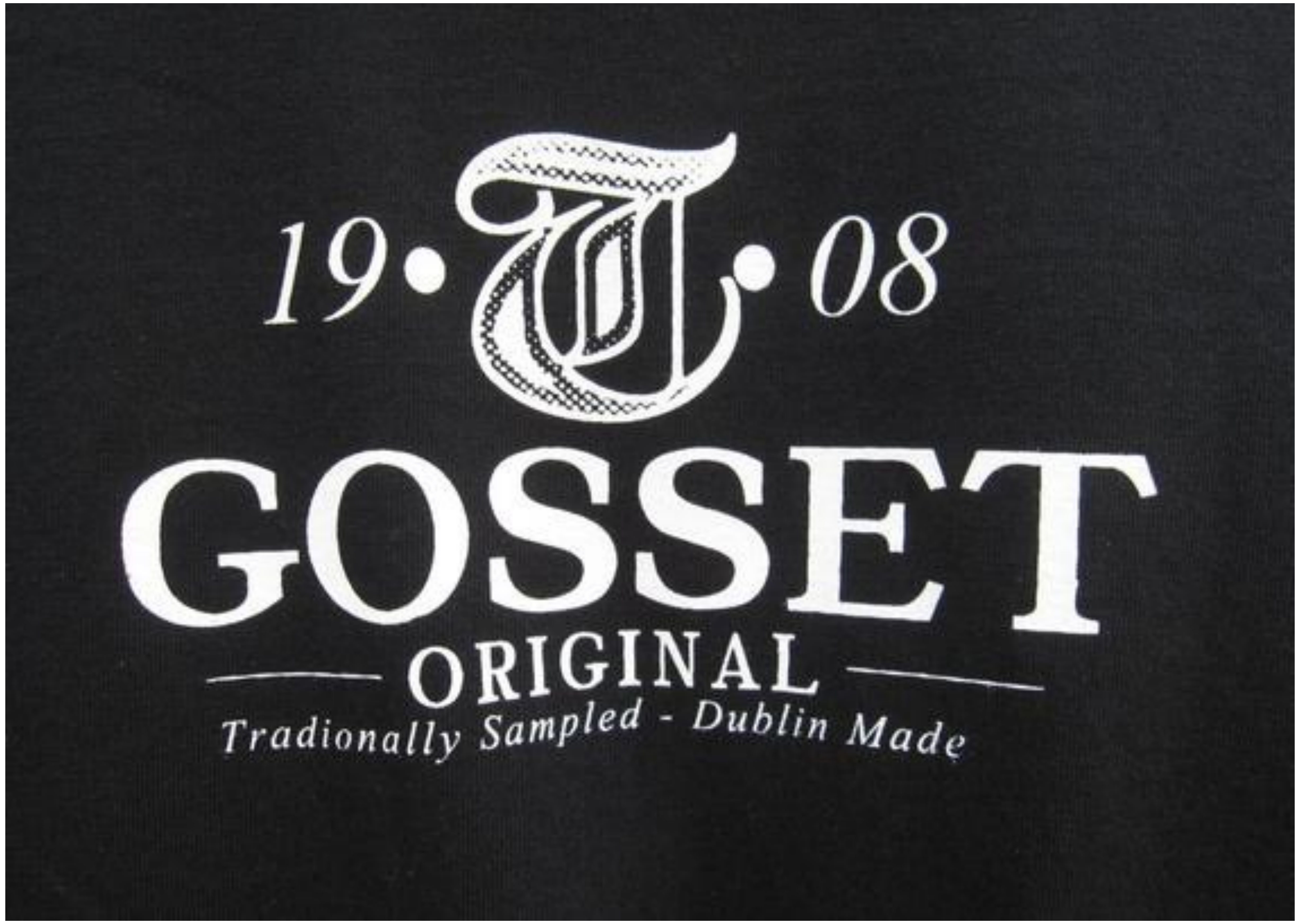
- Your datasets are tiny
- Typical data sizes limit the usefulness of hypothesis tests
- Typical avg. fold changes fall outside those limits
- Leads to all kinds of errors
- There is a better way; Kajsa will explain

More than 90% of datasets comprise under 100 observations

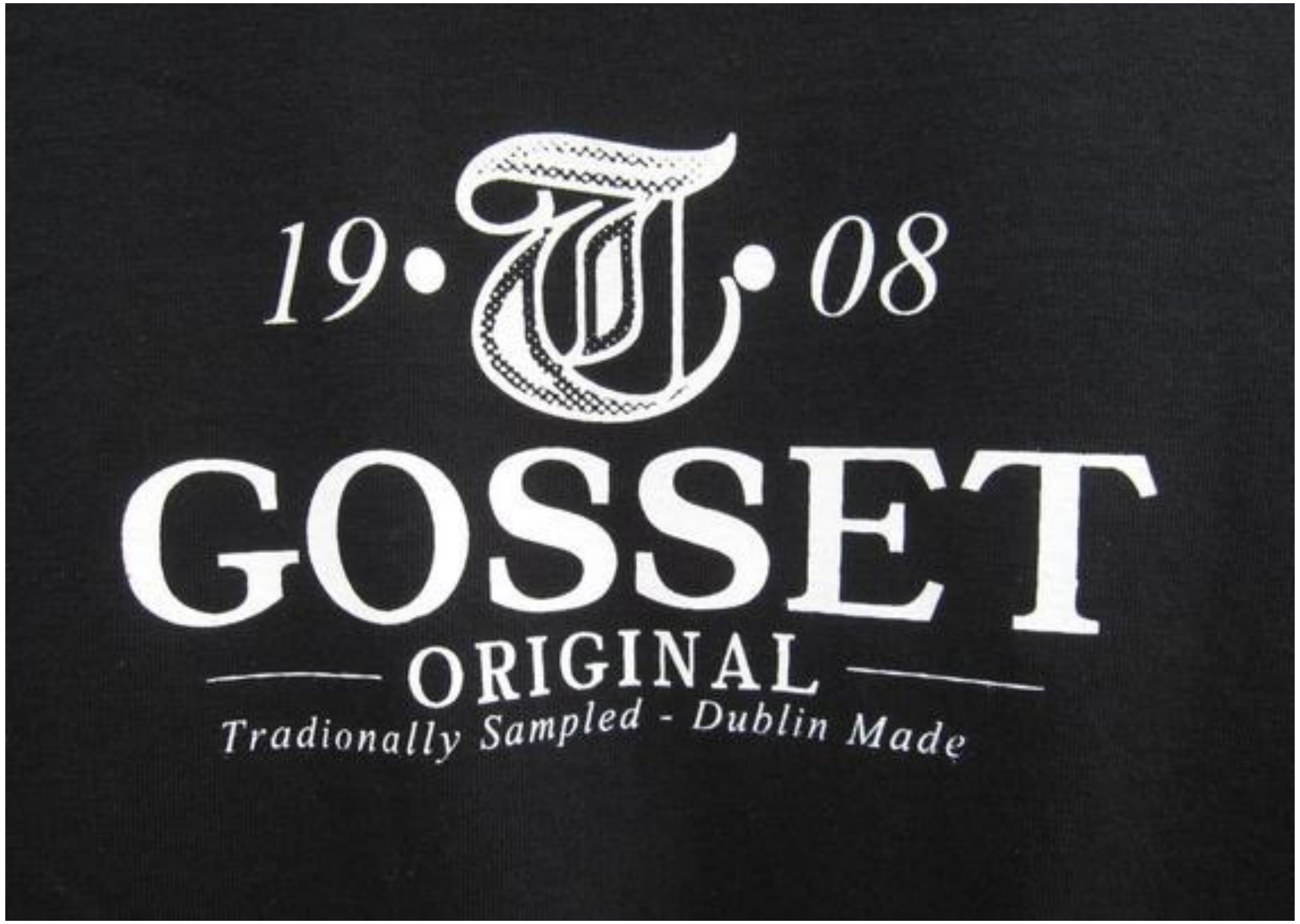


Components of a hypothesis test

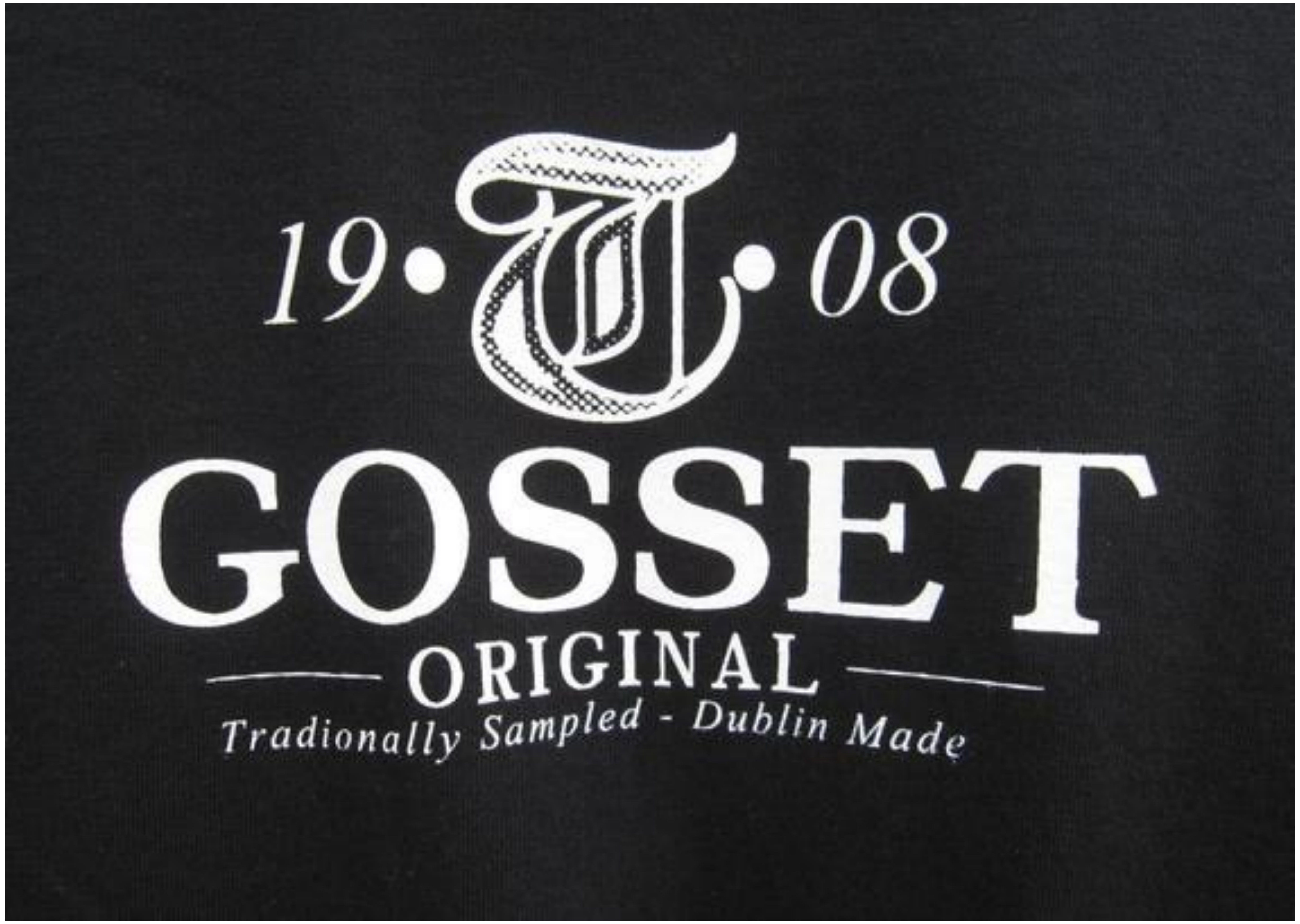
1. Mathematical model of what “nothing” looks like: the null model
2. Mathematical way of comparing data to “nothing”: a statistic
3. A decision rule for what is far enough away from “nothing” to be interesting



$$\frac{\bar{x} - \mu}{\text{s.e.}(\bar{x})}$$



$$\frac{\text{“Signal”}}{\text{s.e.}(\bar{x})}$$

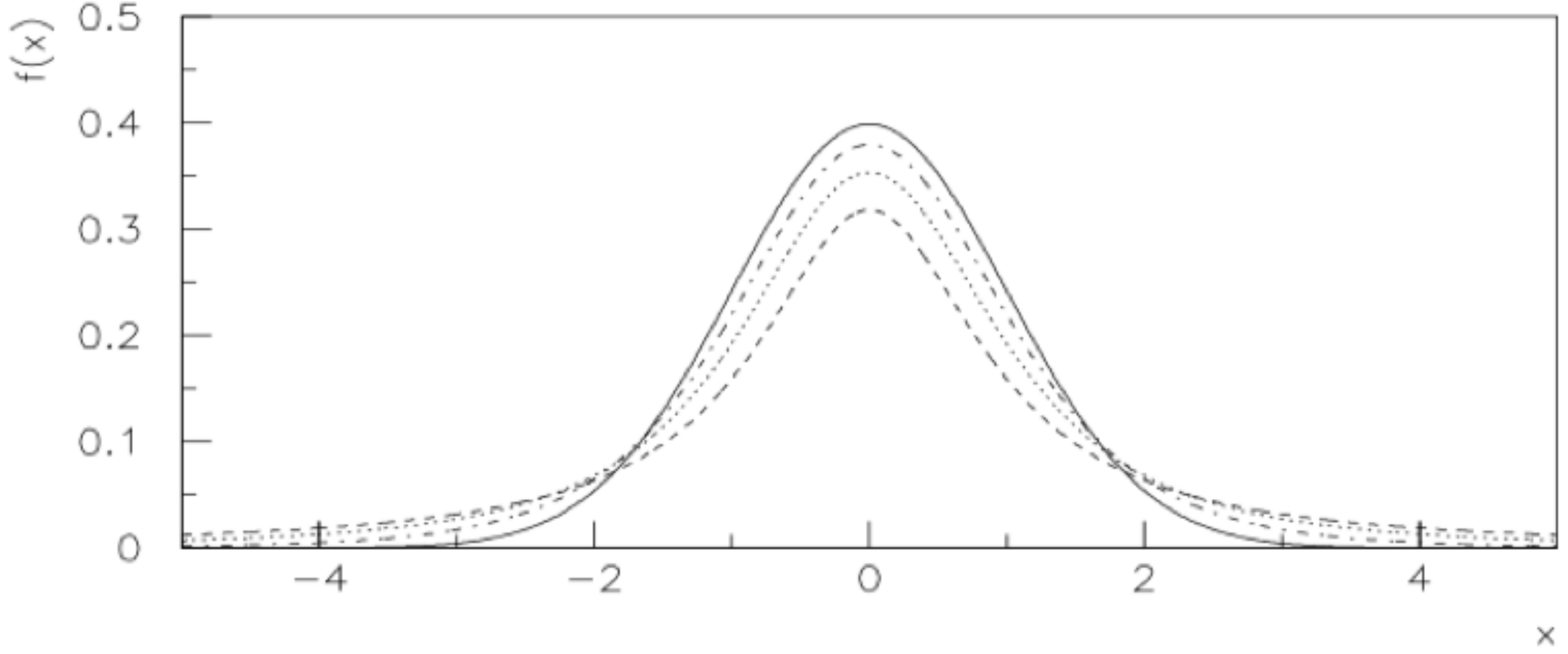


$$\frac{\text{“Signal”}}{\text{“Noise”}}$$

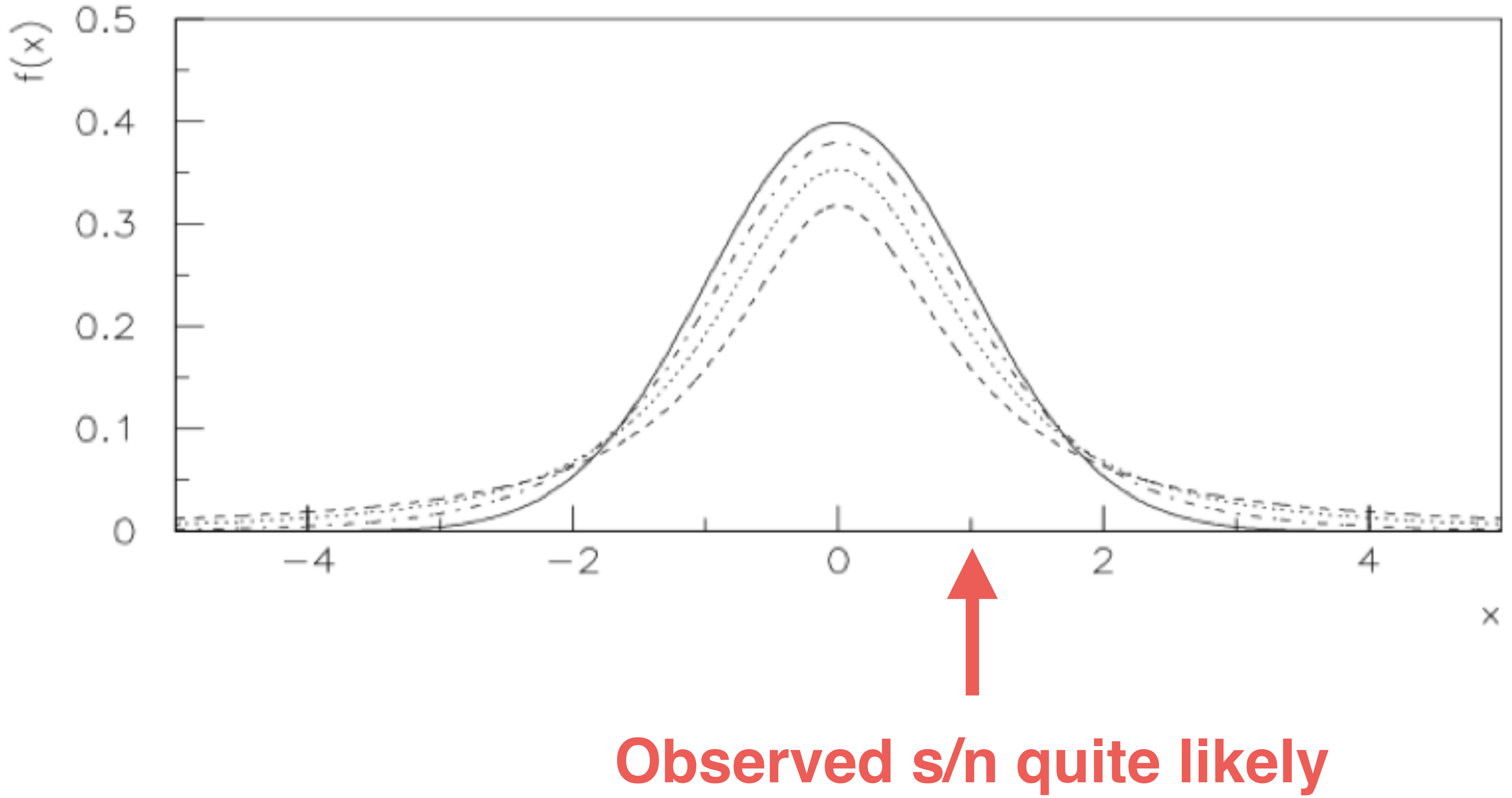
How many times larger is the signal than the noise?

$$\frac{\text{“Signal”}}{\text{“Noise”}}$$

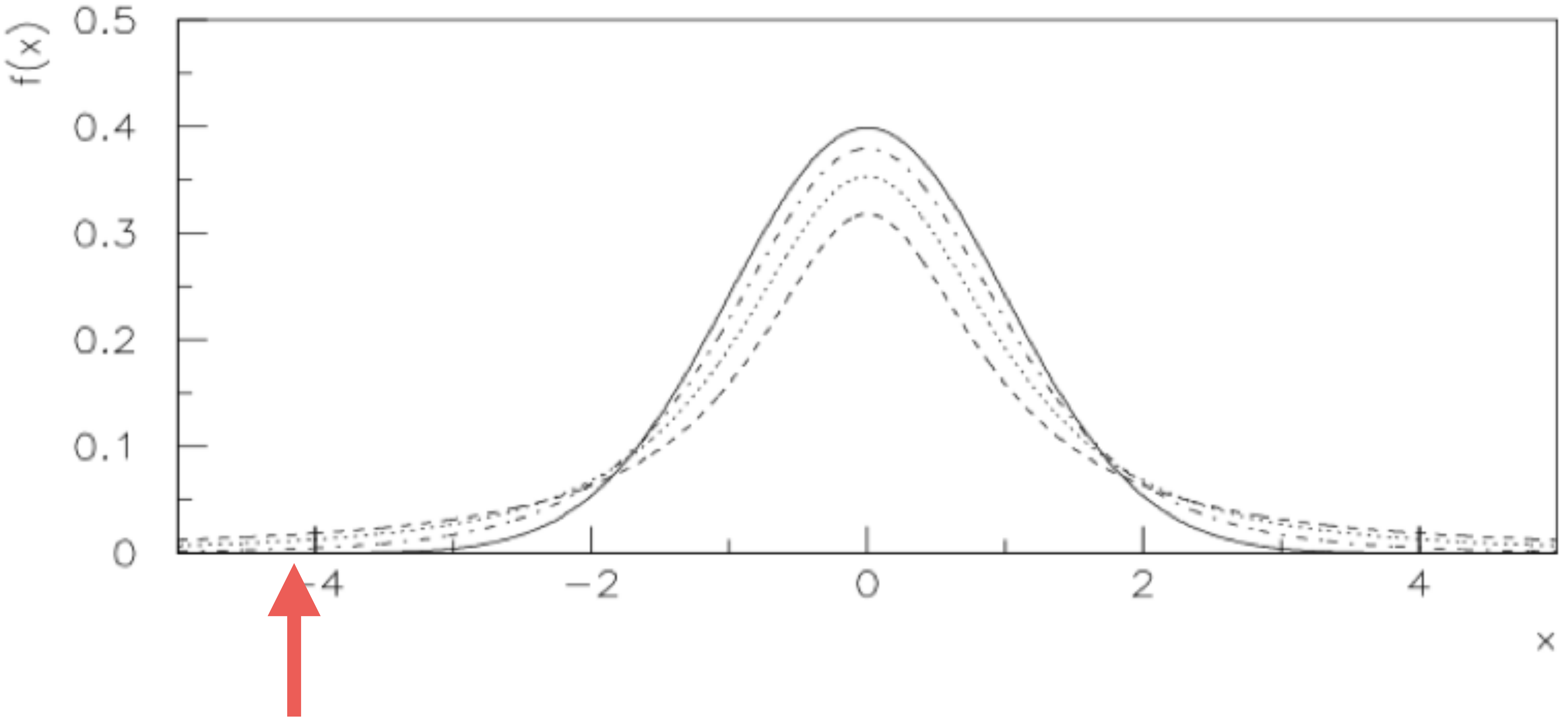
Schematic representation of “nothing”



Schematic representation of “nothing”

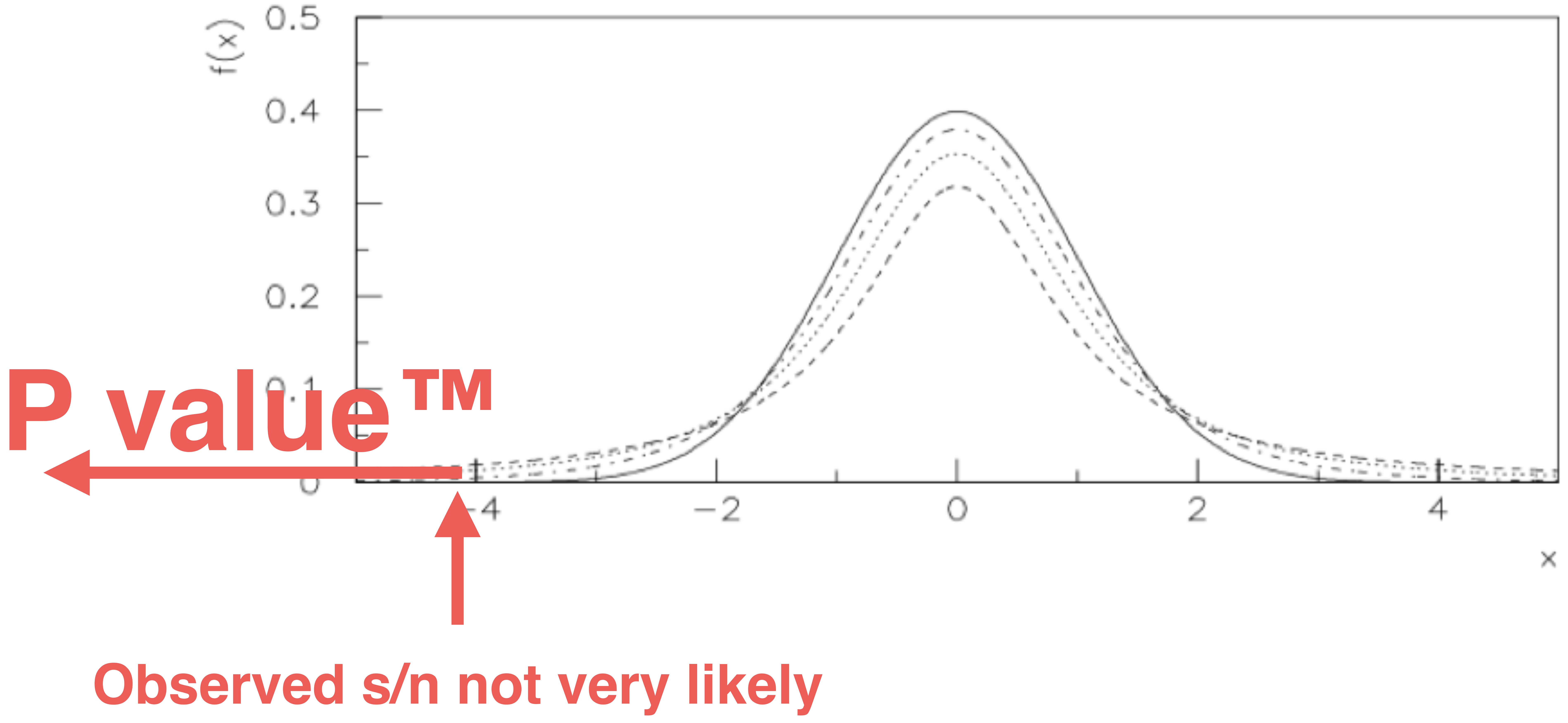


Schematic representation of "nothing"



Observed s/n not very likely

Schematic representation of "nothing"



What is a good p-value?

What is a good p-value?

The [number of standard deviations] for which

What is a good p-value?

The [number of standard deviations] for which P=.05, or 1 in 20, is 1.96 or nearly 2; it is

What is a good p-value?

The [number of standard deviations] for which P=.05, or 1 in 20, is 1.96 or nearly 2; it is convenient to take this point as a limit in judging

What is a good p-value?

The [number of standard deviations] for which P=.05, or 1 in 20, is 1.96 or nearly 2; it is convenient to take this point as a limit in judging whether a deviation is to be considered

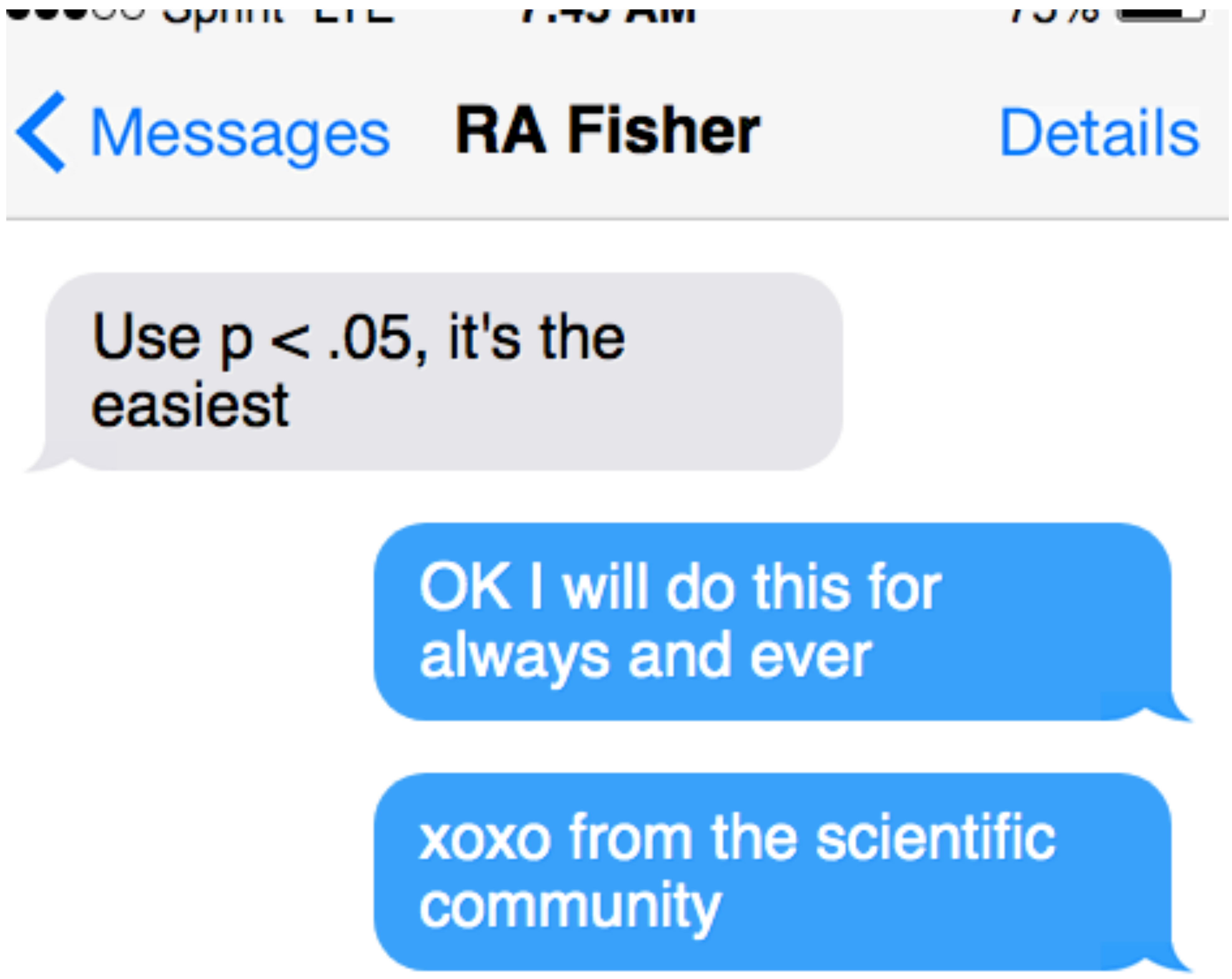
What is a good p-value?

The [number of standard deviations] for which P=.05, or 1 in 20, is 1.96 or nearly 2; it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not.

What is a good p-value?

The [number of standard deviations] for which P=.05, or 1 in 20, is 1.96 or nearly 2; it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not.

– RA Fisher
(Statistical Methods for Research Workers)


Messages
RA Fisher
Details
Use p < .05, it's the easiest
OK I will do this for always and ever
xoxo from the scientific community

Back to the 100 observations

- If we observe fold change so that it is 2 times the noise, we “win”
- Implies that true fold change should at least be 4 times the noise if we want to almost surely win
- Noise shrinks with the inverse of square root of number of observations
- Details in **The Book**

Back to the 100 observations

$$\frac{\mu}{\text{s.e.}(\bar{x})} > 4 \iff \mu > 4\text{s.e.}(\bar{x}),$$

$$\text{s.e.}(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

Back to the 100 observations

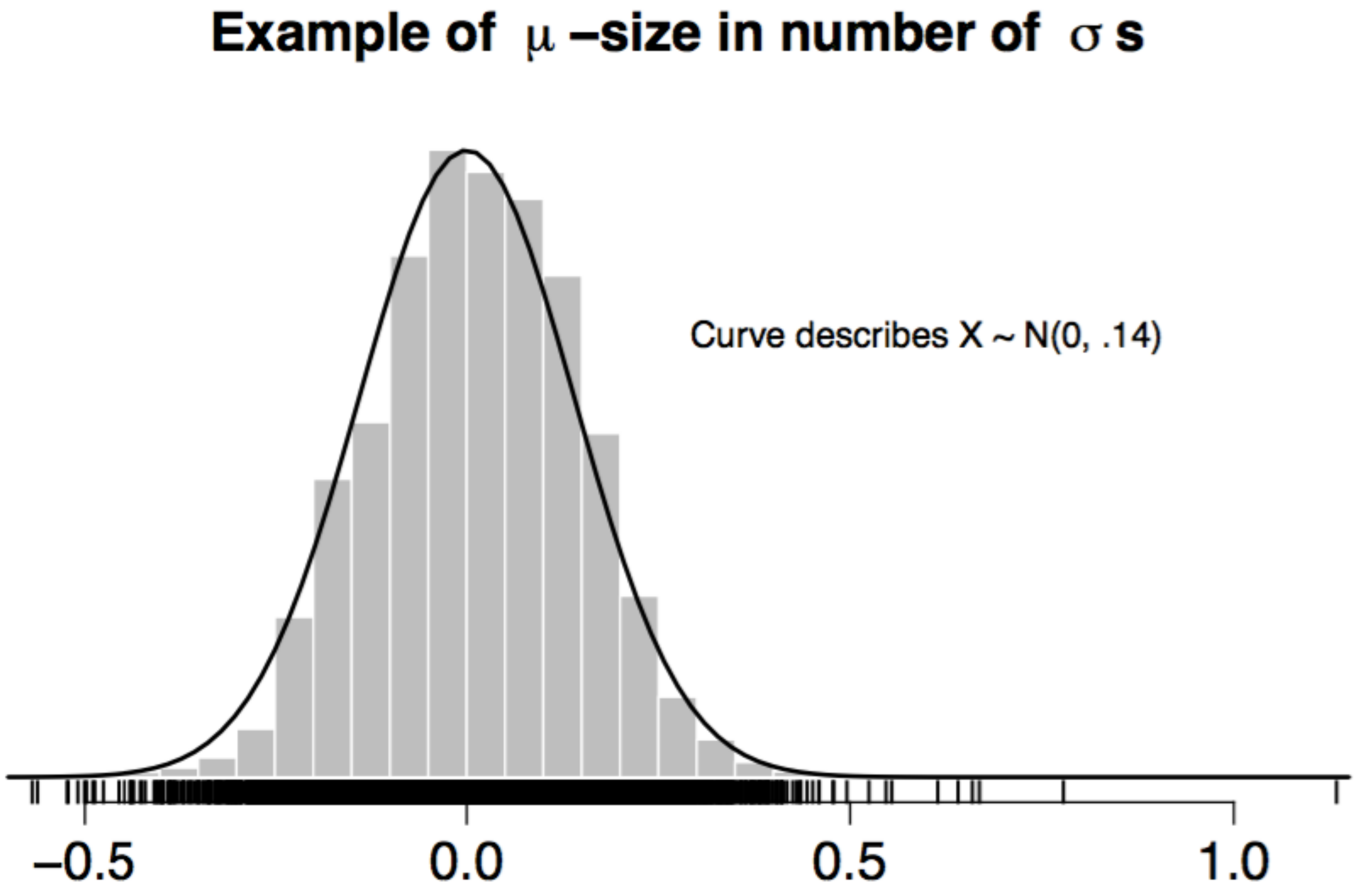
$$\frac{\mu}{\text{s.e.}(\bar{x})} > 4 \iff \mu > 4\text{s.e.}(\bar{x}),$$

$$\text{s.e.}(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

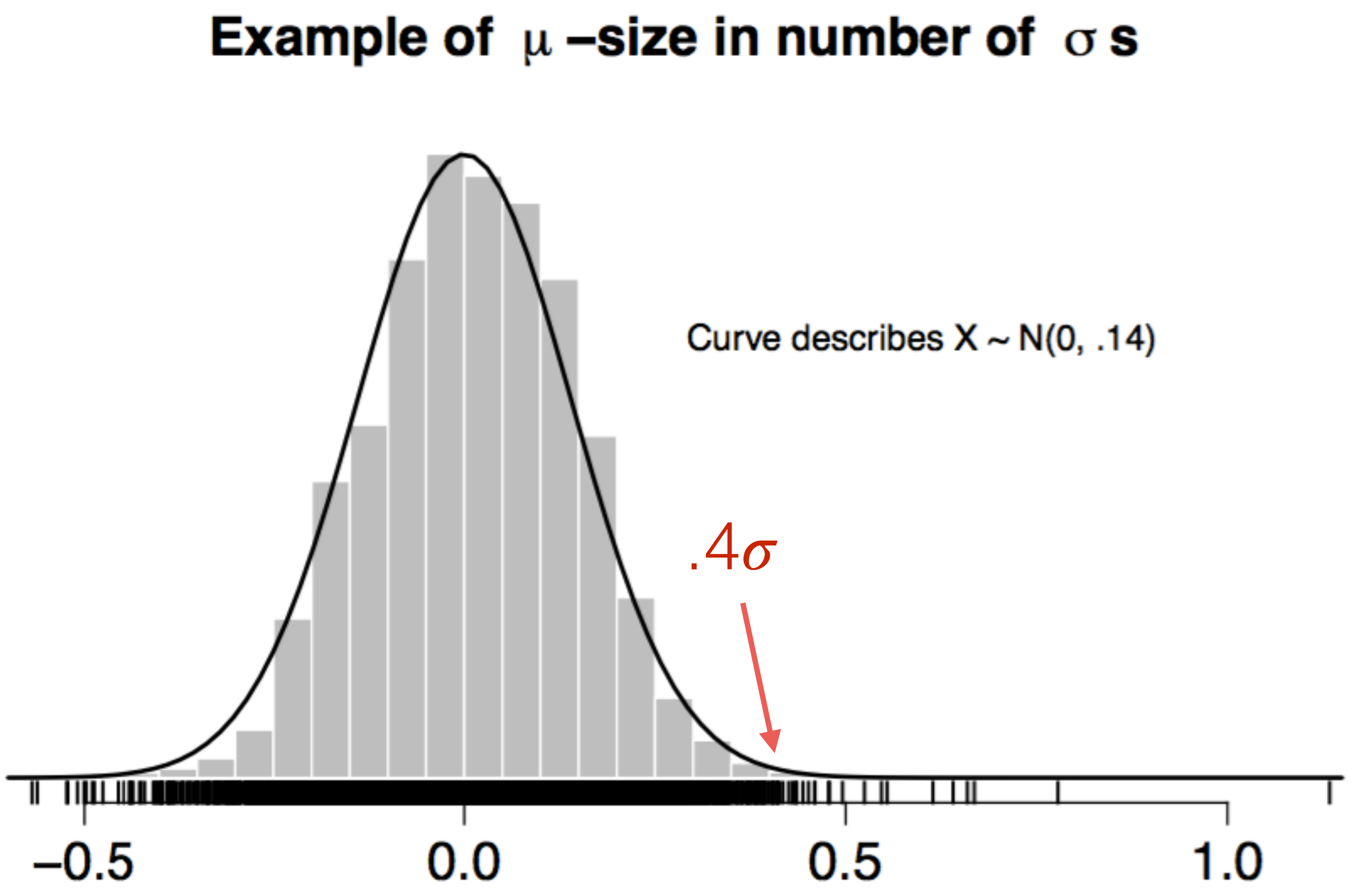
Smallest detectable means in number of standard deviations

Difference from zero:	$.4\sigma$
Difference between two groups of 50:	$.8\sigma$
Subgroups of 25, diff from zero:	1.1σ
Difference between subgroups:	1.6σ

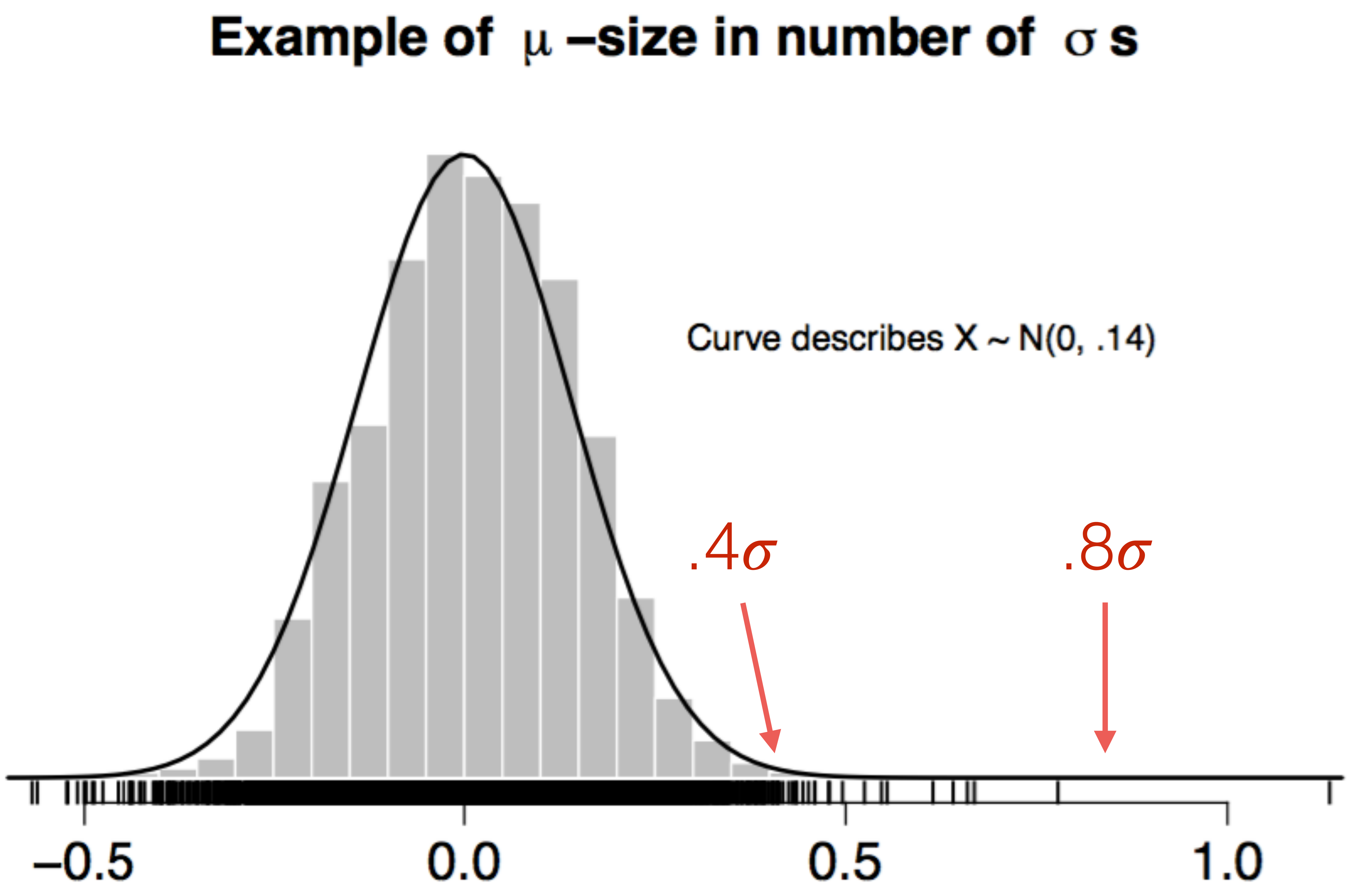
Reality check: smokers vs nonsmokers



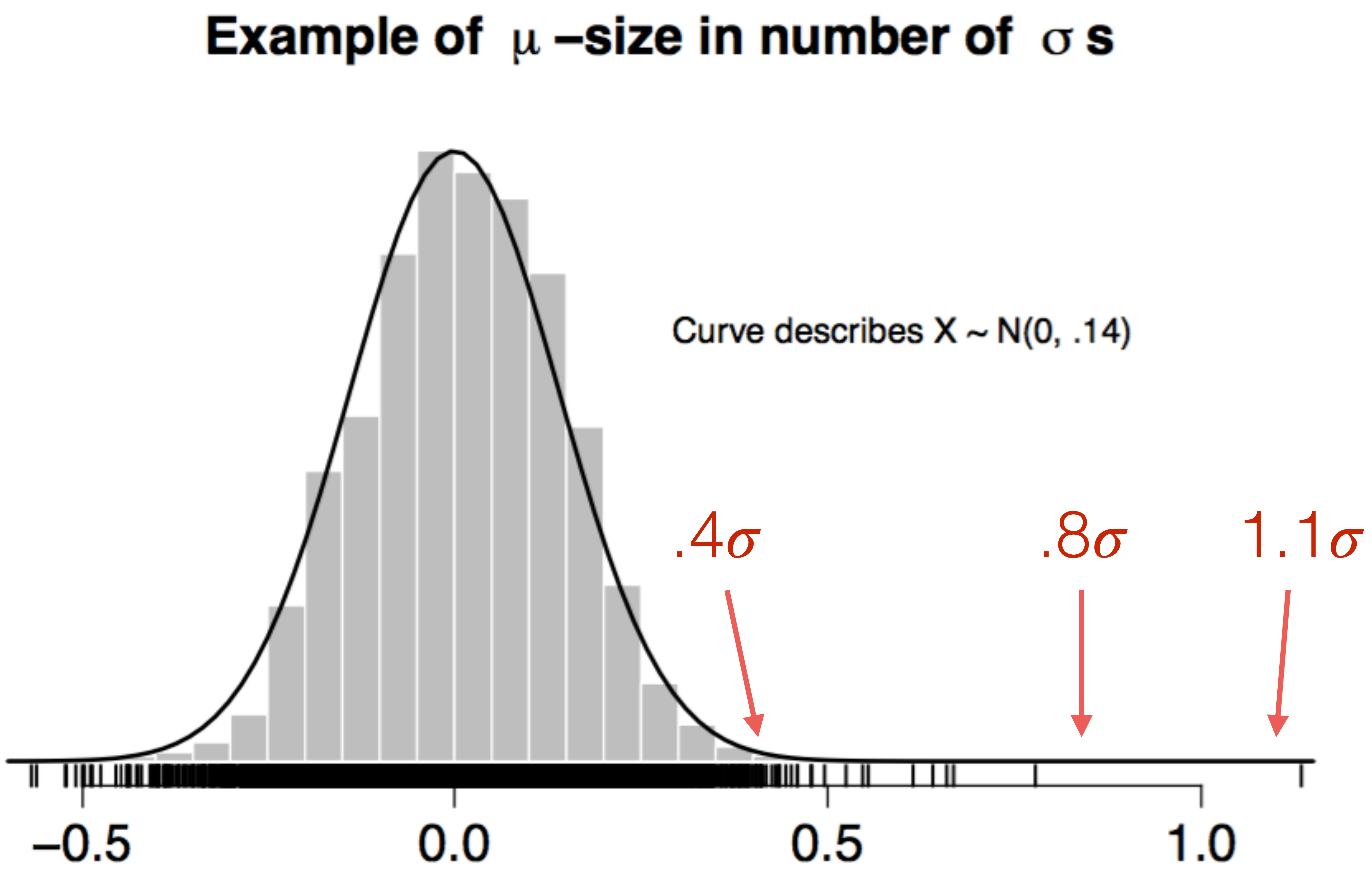
Reality check: smokers vs nonsmokers



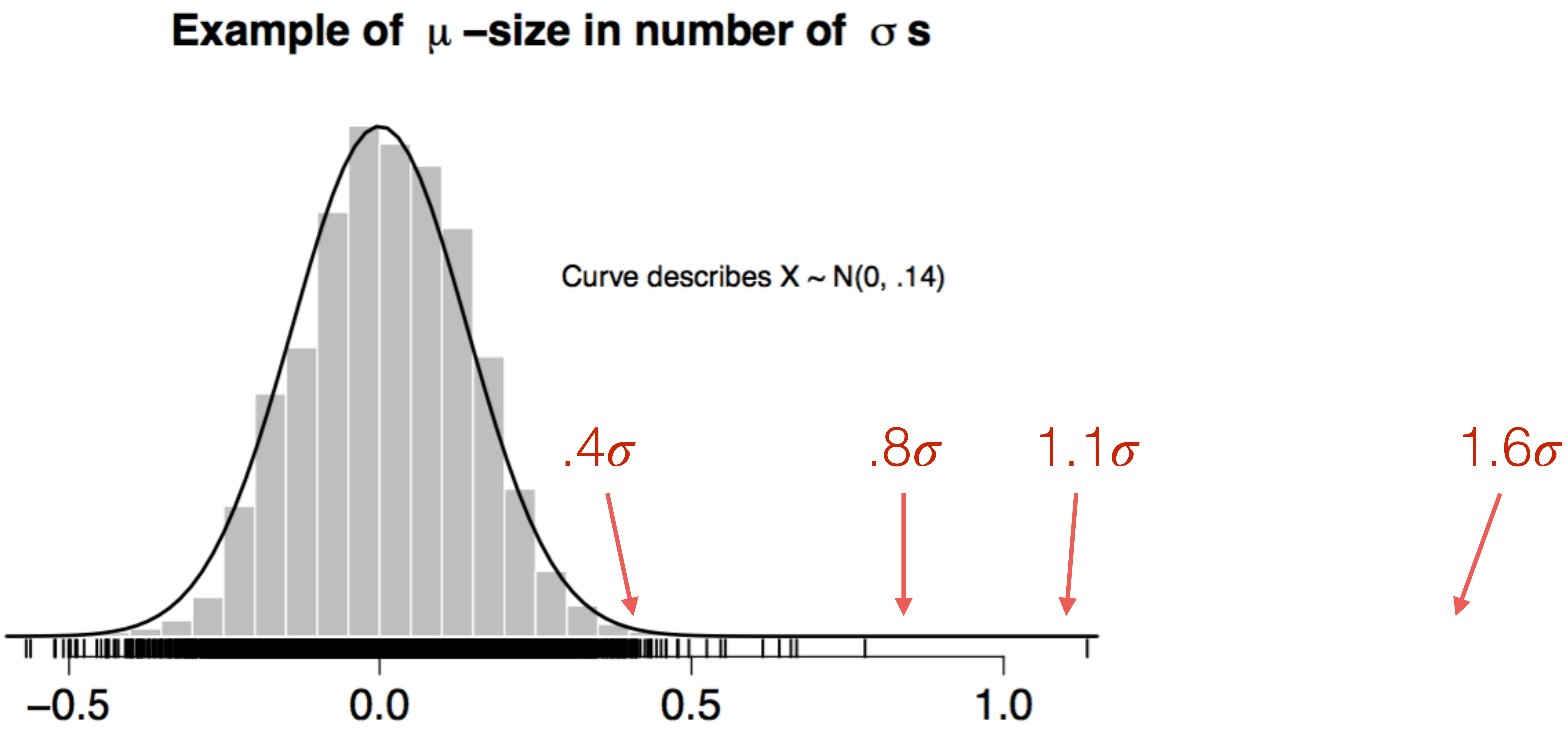
Reality check: smokers vs nonsmokers



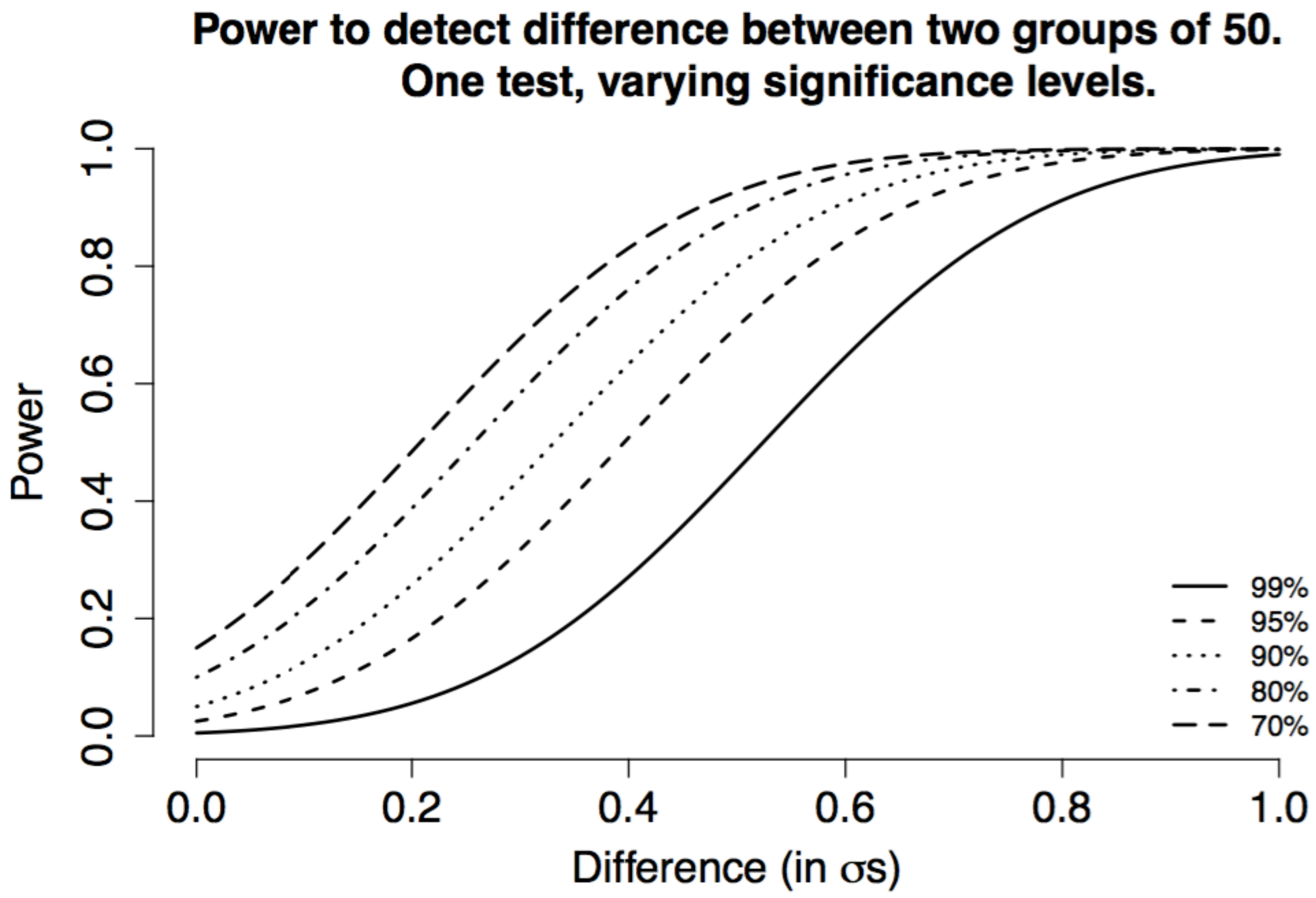
Reality check: smokers vs nonsmokers



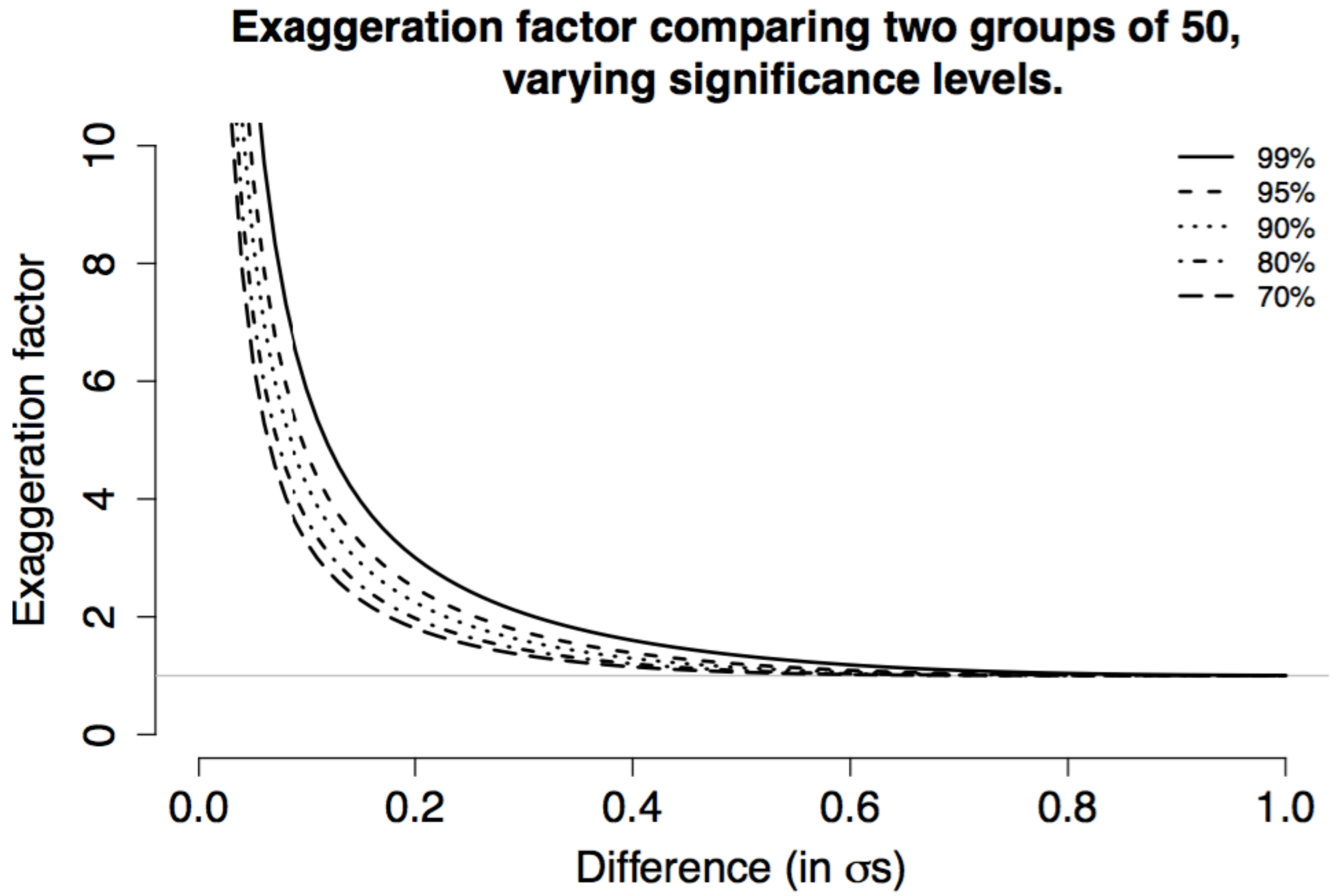
Reality check: smokers vs nonsmokers



Statistical power is tragic in data such as these



The “wins” are overestimated by like 200%



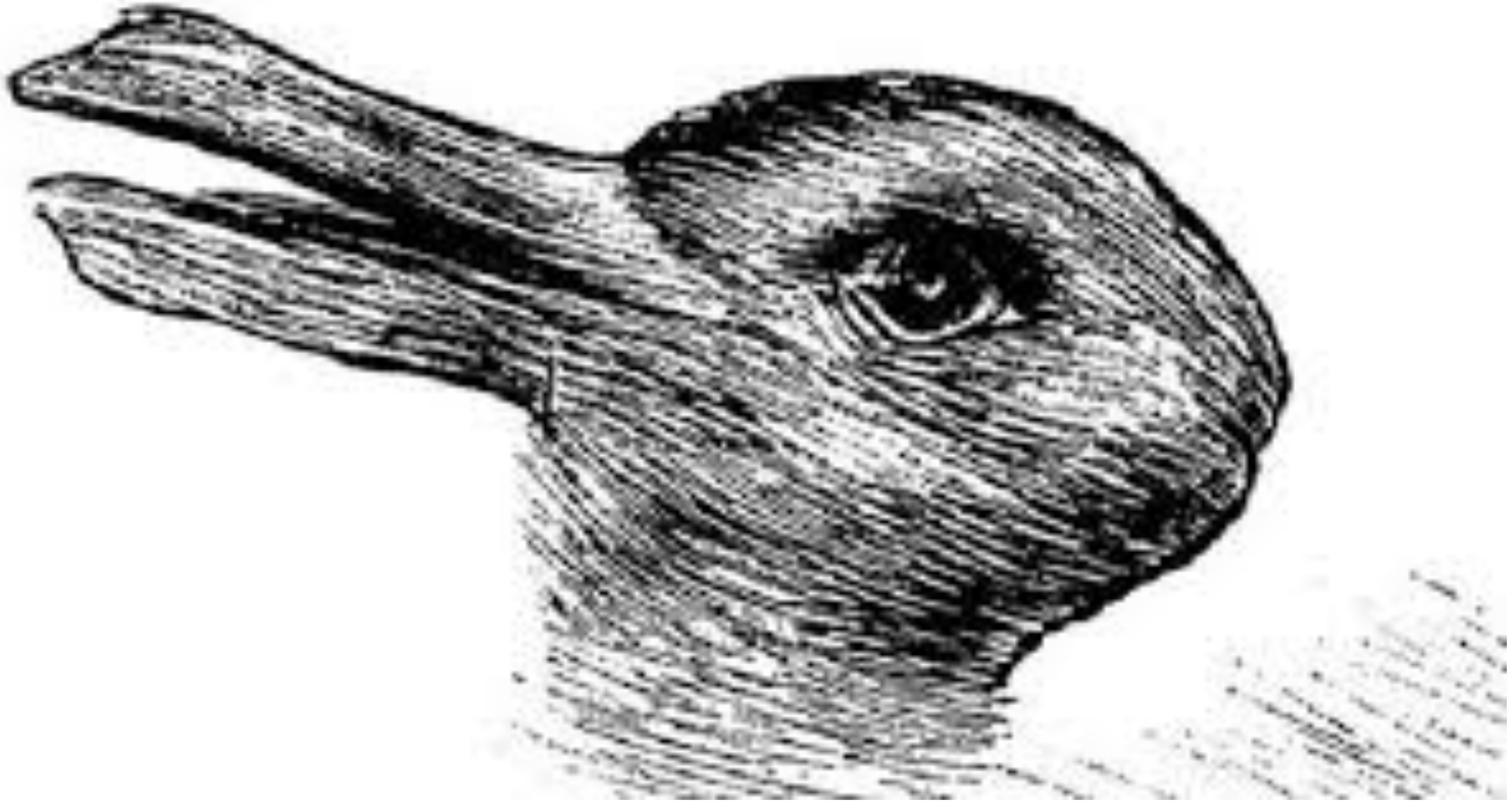
What to expect

Confidence	99%	95%	90%	80%	70%
Magnitide error	2.3	2.0	1.8	1.6	1.5
Power	12%	28%	39%	53%	63%

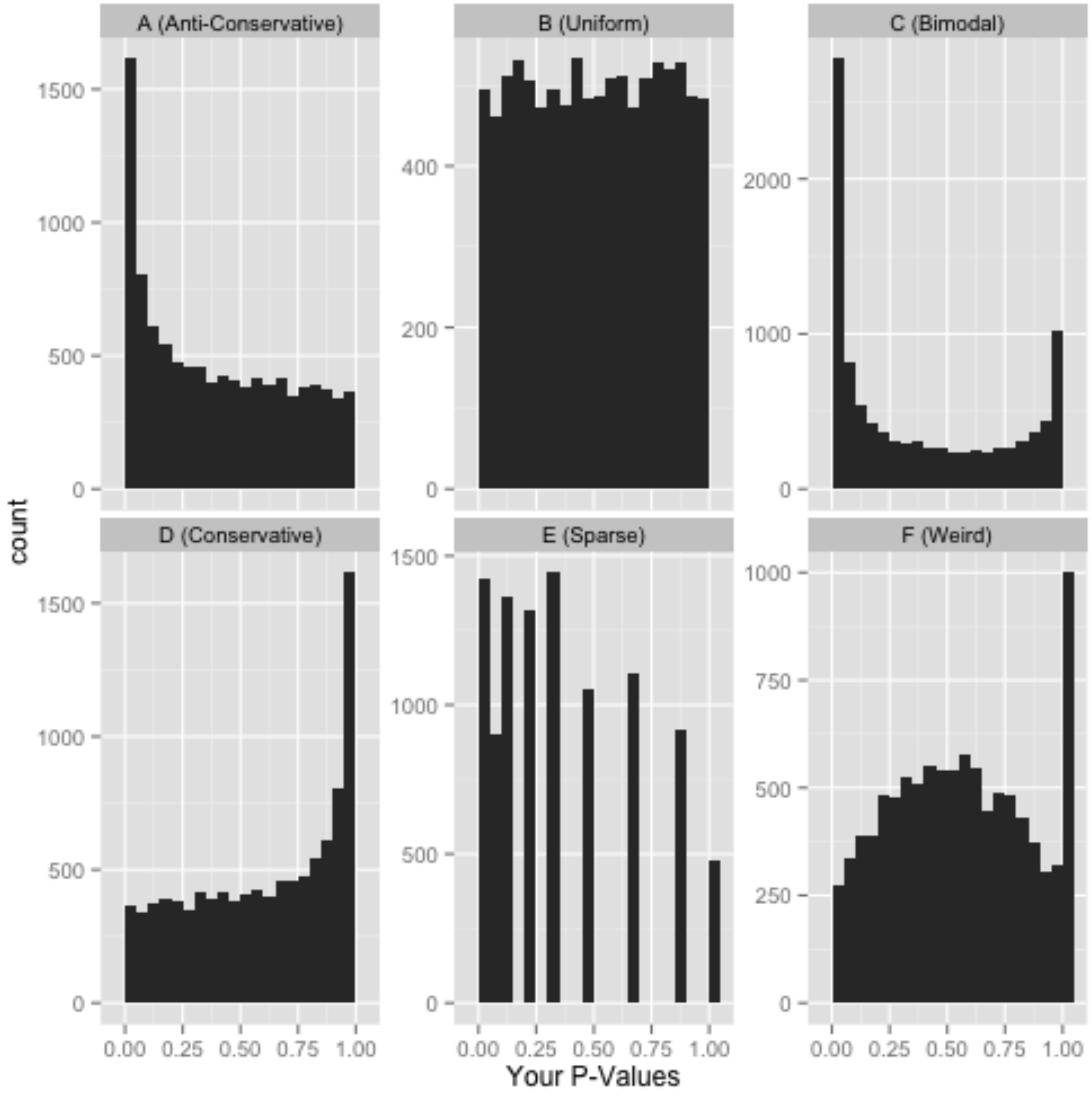
What to expect

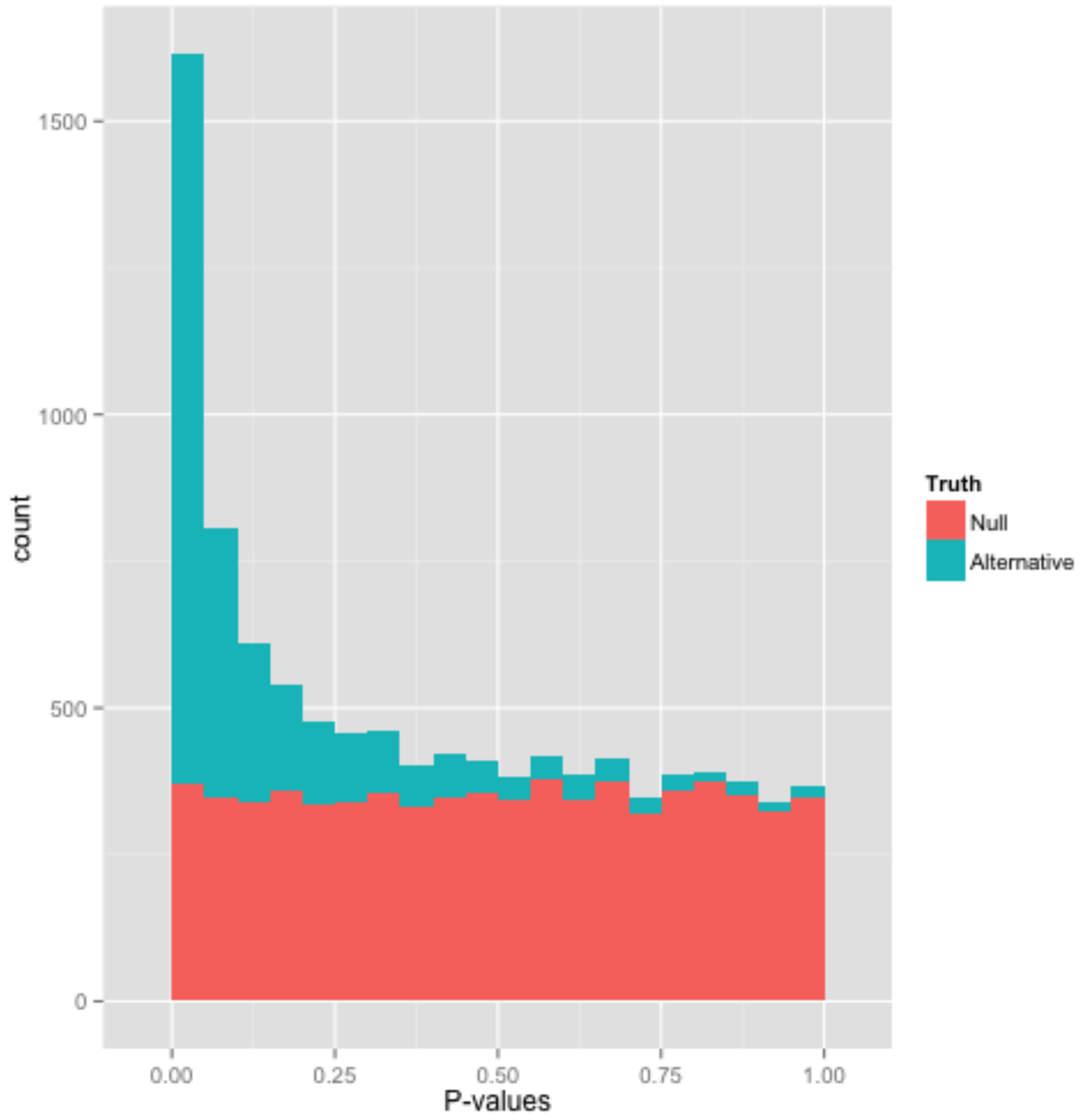
Confidence	99%	95%	90%	80%	70%
Magnitide error	2.3	2.0	1.8	1.6	1.5
Power	12%	28%	39%	53%	63%

Food for thought: are you sure hypothesis testing is for you?



Humble scholars of uncertainty


A (Anti-Conservative)
B (Uniform)
C (Bimodal)
D (Conservative)
E (Sparse)
F (Weird)
count
Your P-Values
0.00 0.25 0.50 0.75 1.00


1500
1000
500
0
count
0.00
0.25
0.50
0.75
1.00
P-values
Truth
Null
Alternative

Be a biologist!

- Ask a biology question
- Let the data answer
- Use a statistical tool to quantify the uncertainty

Be a biologist - use familiar statistics tool

- Choose an α.
- Calculate the mean difference - which ones are interesting? Sort accordingly.
- Confidence intervals for those.
- Pick those with a low p-value, not necessarily lower than α.


Ordre