

STEFANO CORETTA

HOW SIGNIFICANCE TESTING MIGHT BE LIMITING PHONETIC RESEARCH

And what to replace it with

BAAP 2022

Colloquium of the British Association of Academic Phoneticians



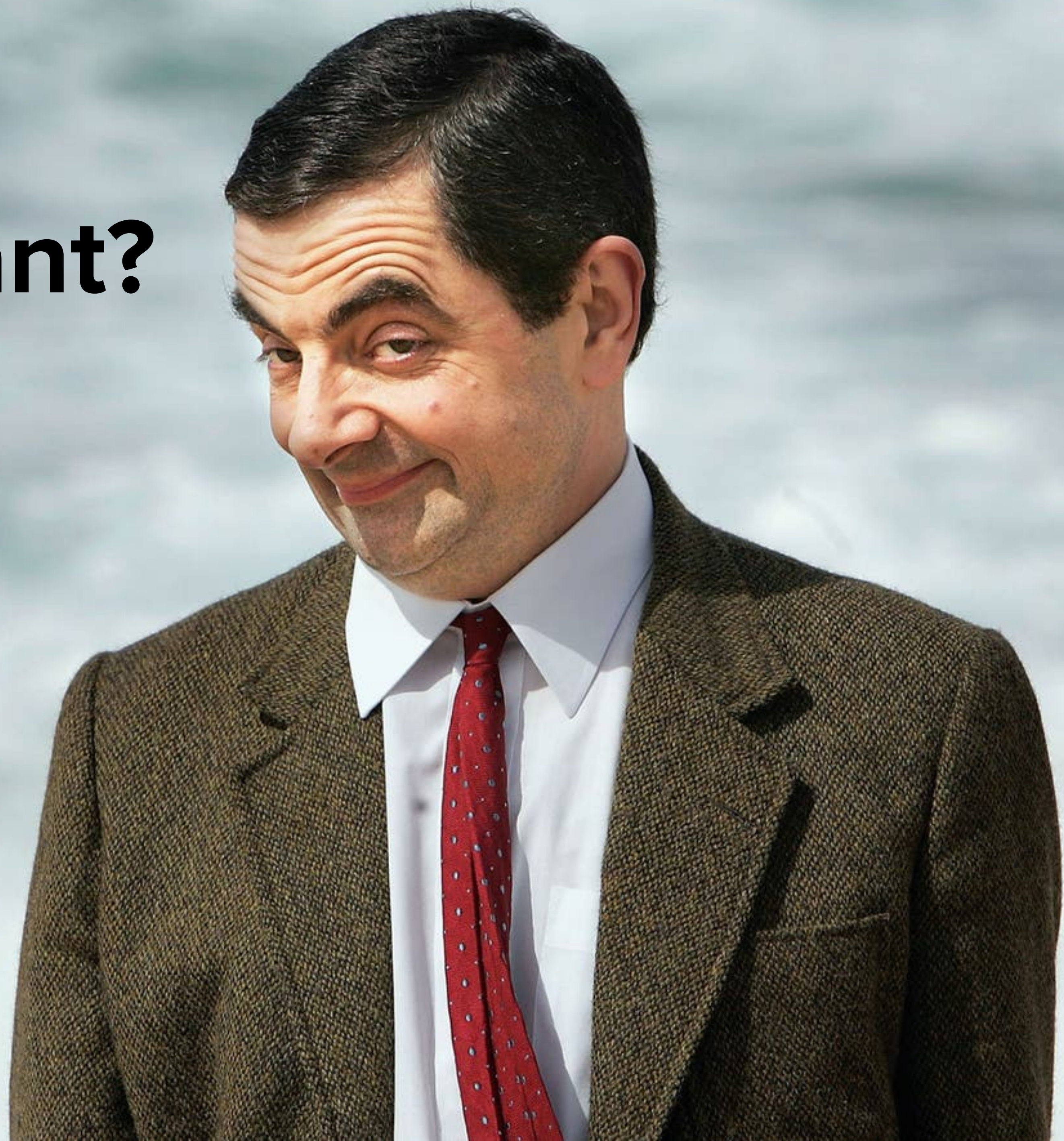
**THE UNIVERSITY
of EDINBURGH**



Design study
Collect data
Start analysis



**But is it
significant?**



SIGNIFICANCE

NULL HYPOTHESIS SIGNIFICANCE TESTING

P-VALUES

STATISTICAL

SIGNIFICANCE

STATISTICS \neq

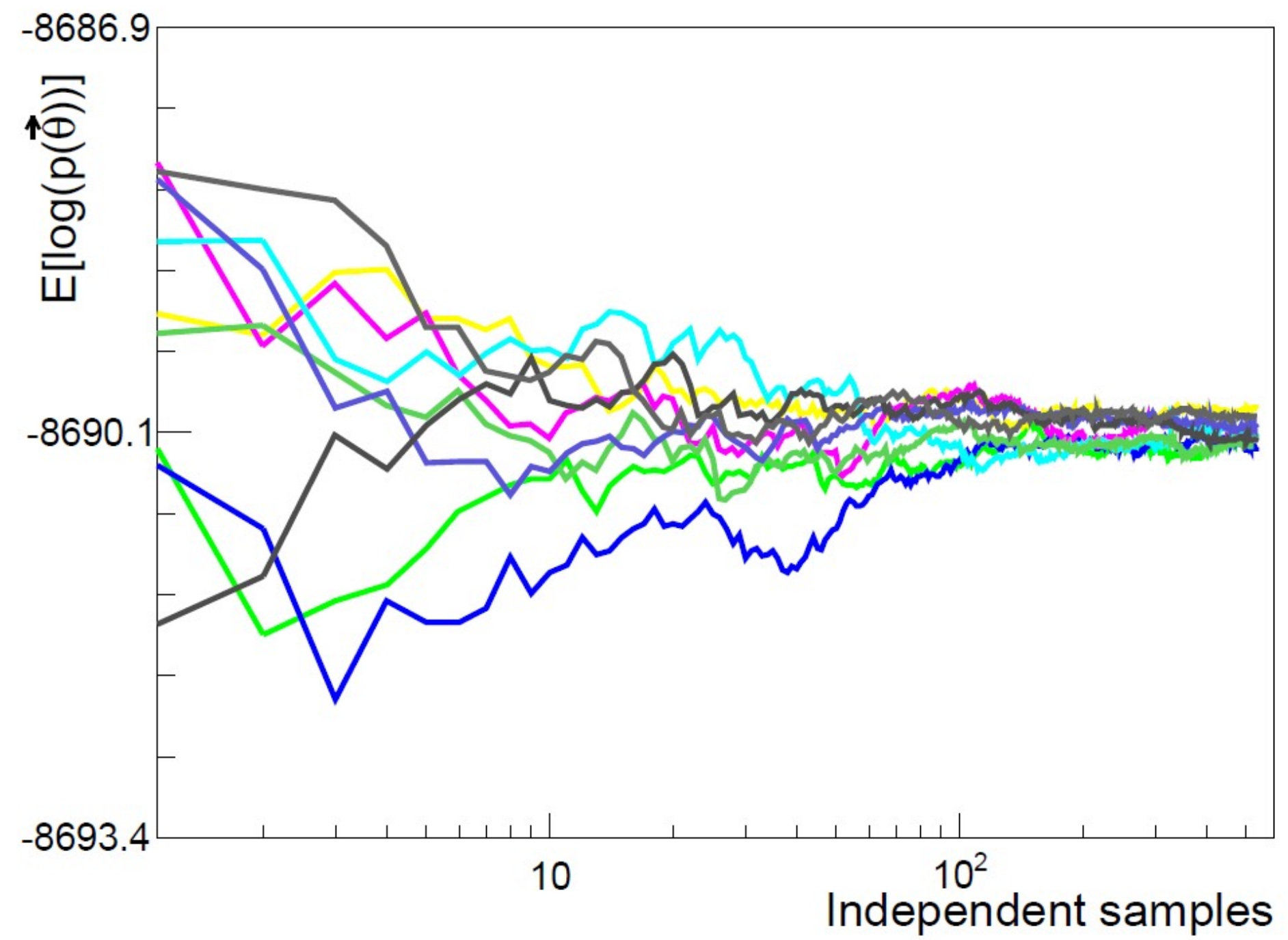
SIGNIFICANCE

**What's the
alternative?**



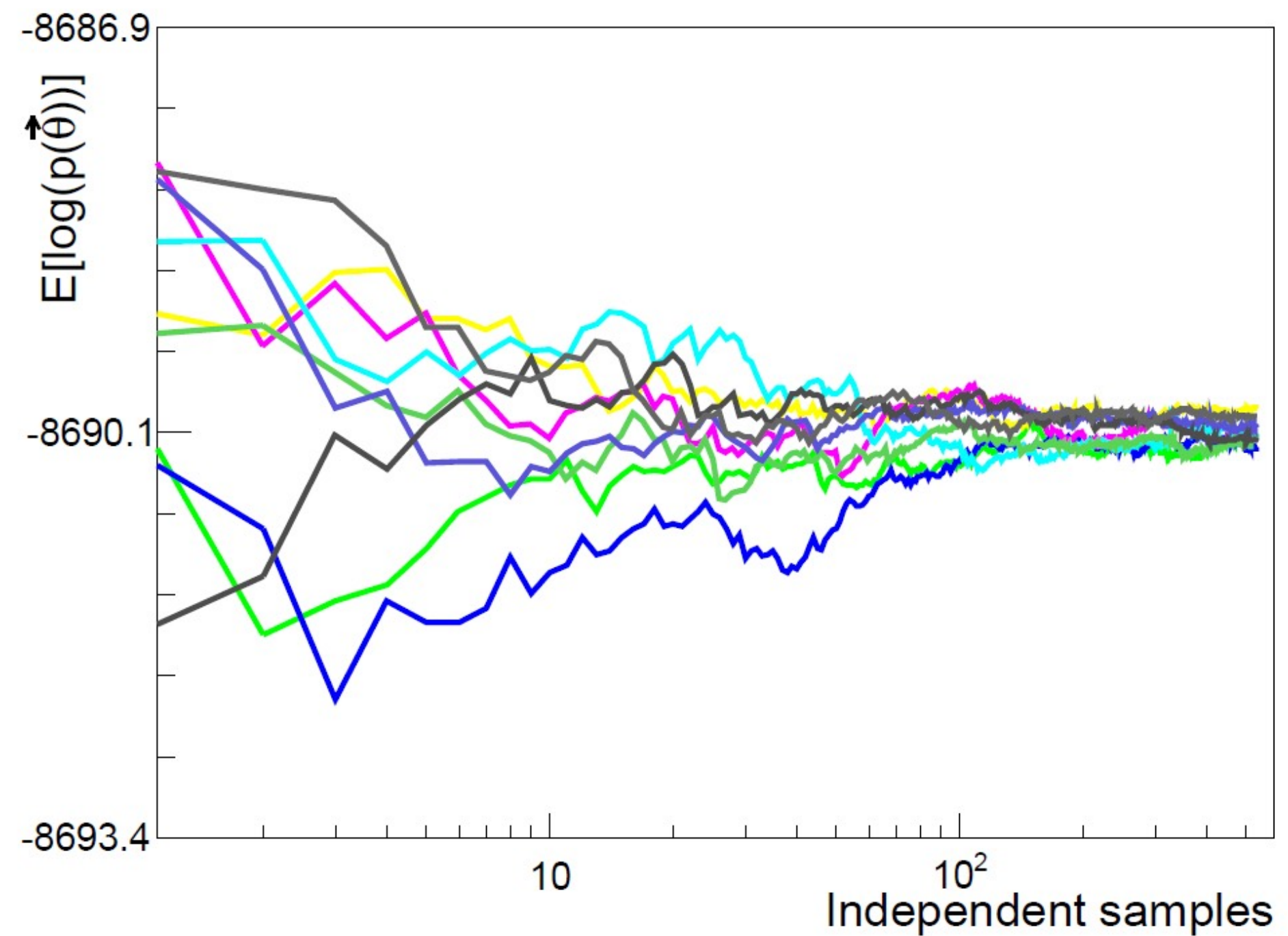
PARAMETER ESTIMATION
QUANTIFICATION OF UNCERTAINTY
MODEL SPECIFICATION

PARAMETER ESTIMATION



PARAMETER ESTIMATION

most phonetic frameworks



present



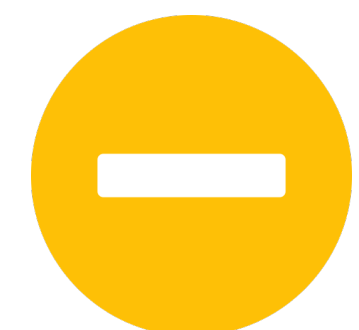
absent



increase



decrease



PARAMETER ESTIMATION

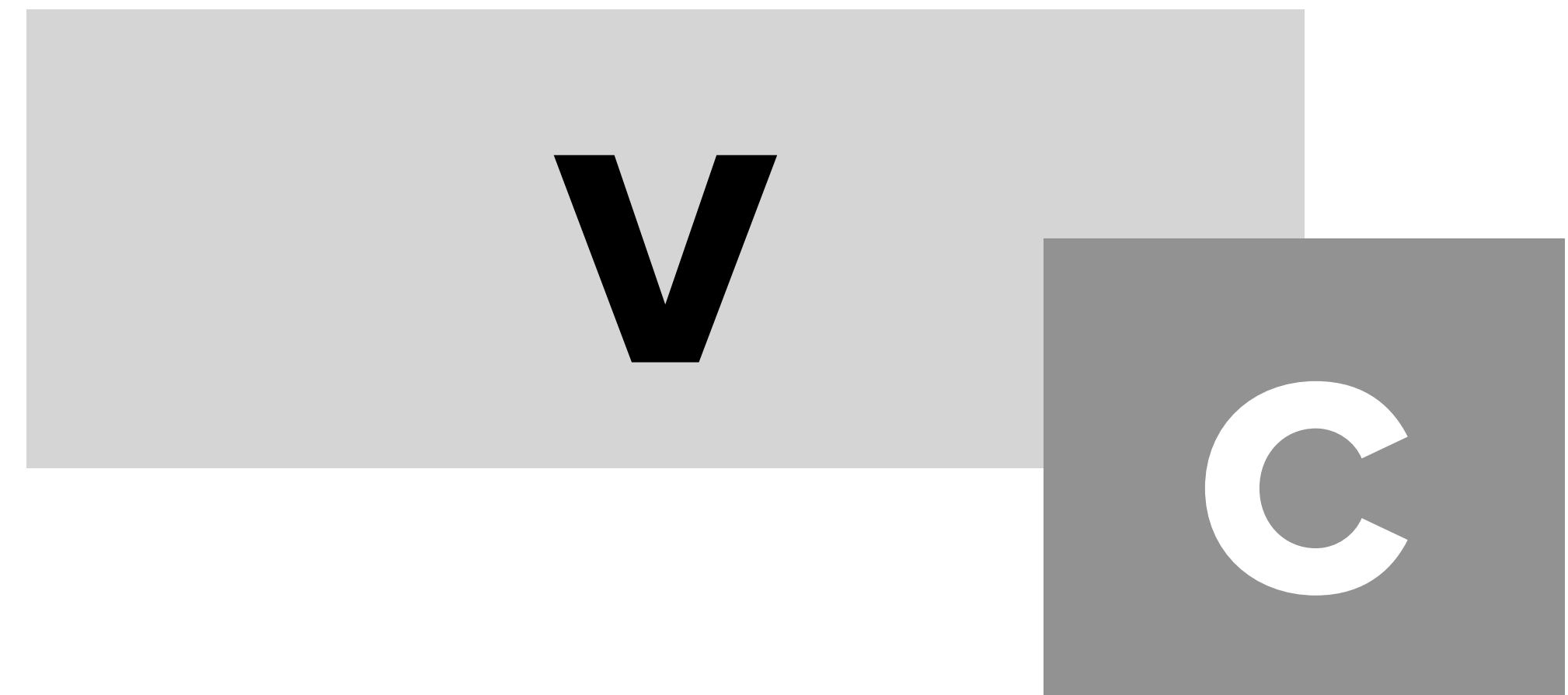
In-phase and anti-phase coordination of gestures

IN-PHASE



no time lag

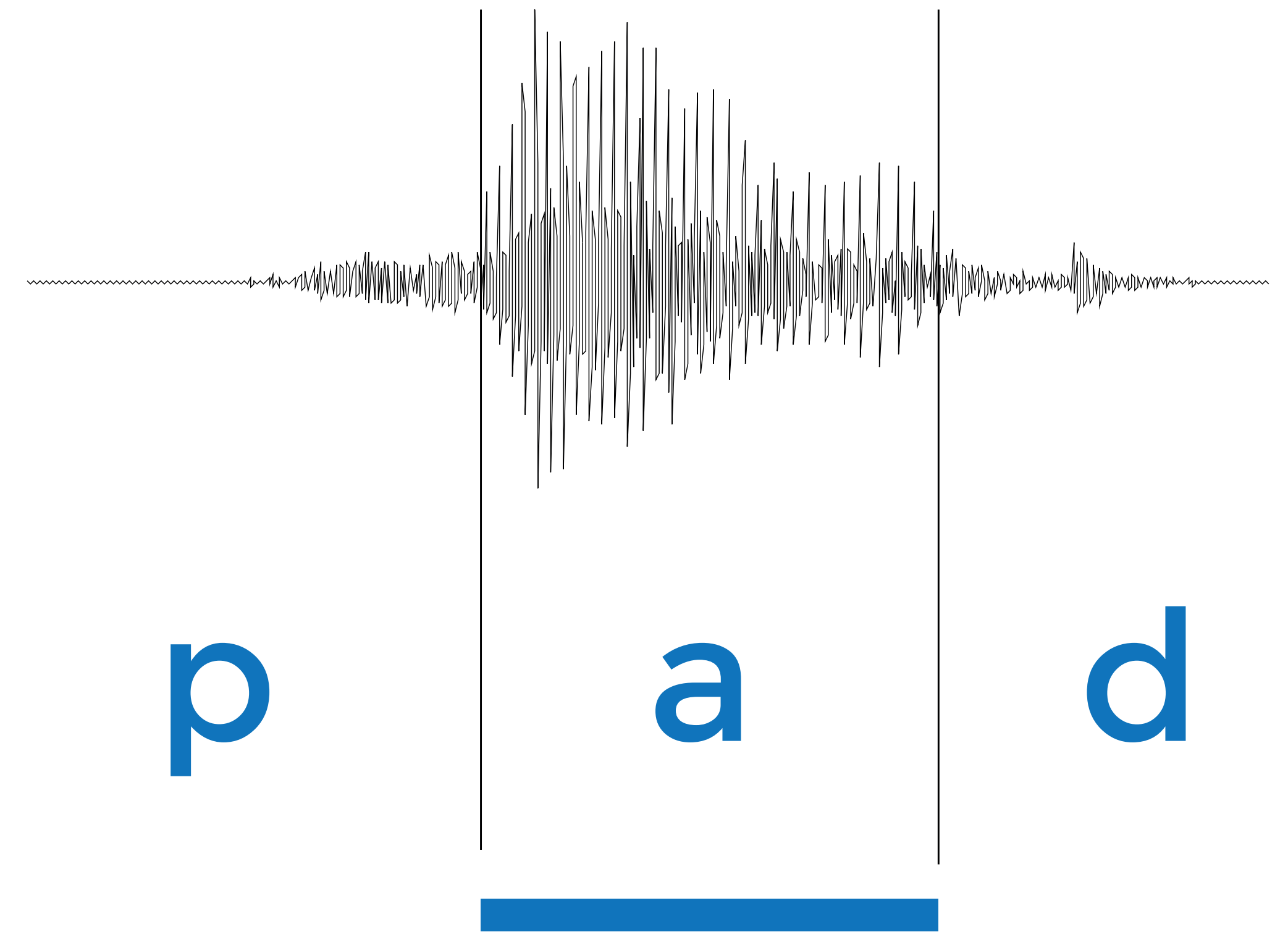
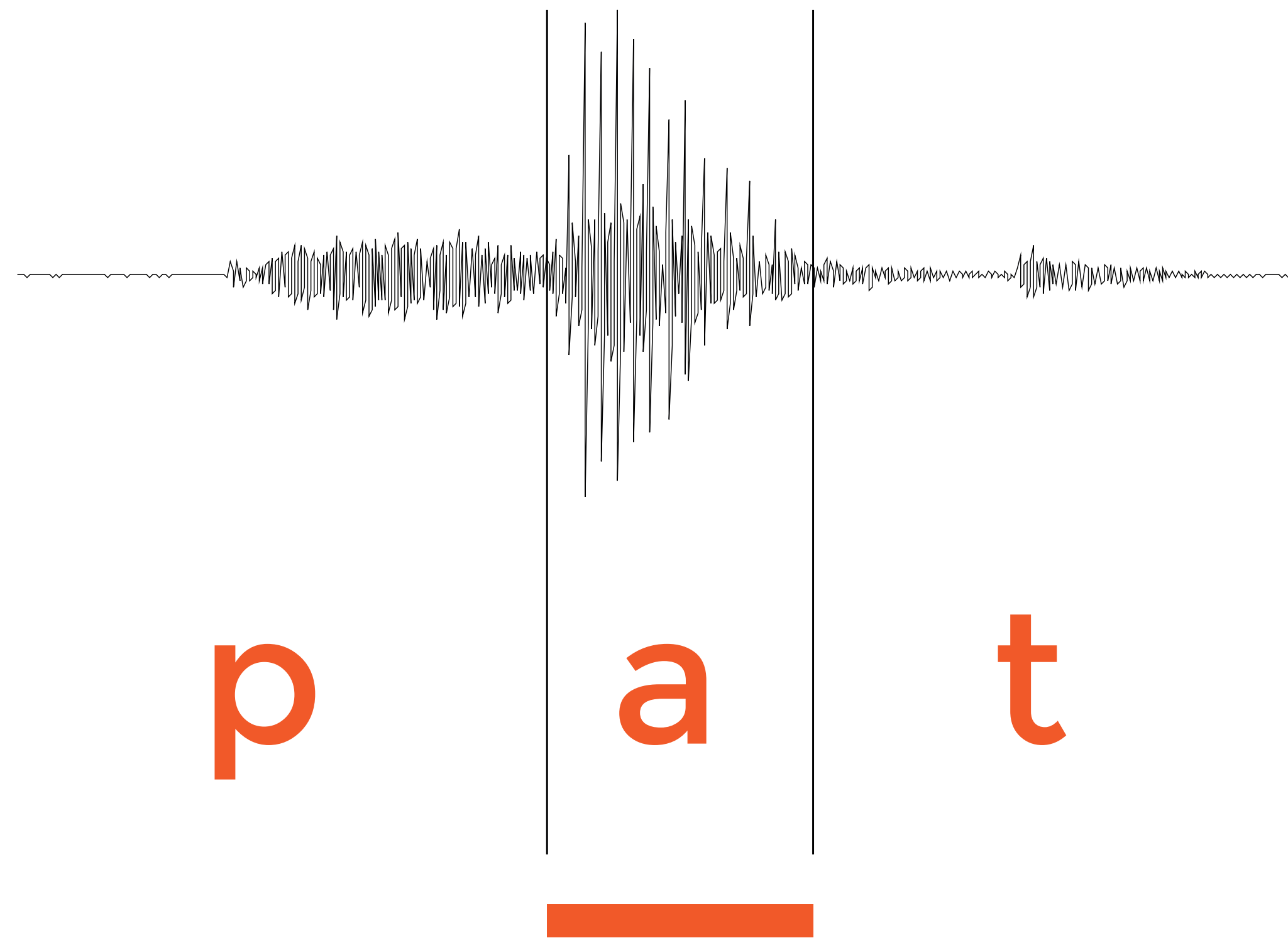
ANTI-PHASE



time lag

PARAMETER ESTIMATION

Effect of consonant voicing on vowel duration



QUANTIFY UNCERTAINTY

**With great
power comes
great
responsibility**

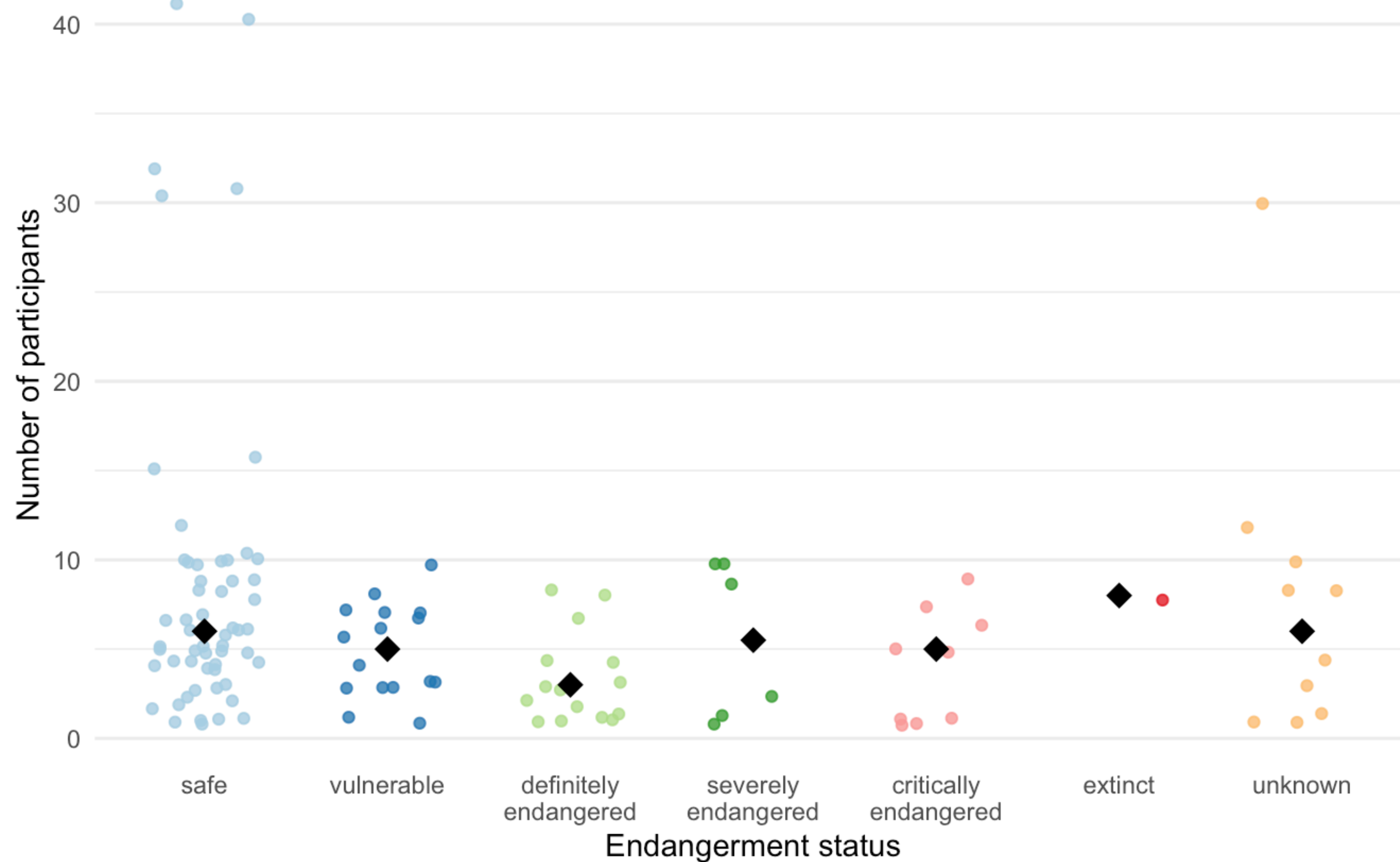
QUANTIFY UNCERTAINTY

**With low
statistical power
comes great
uncertainty**

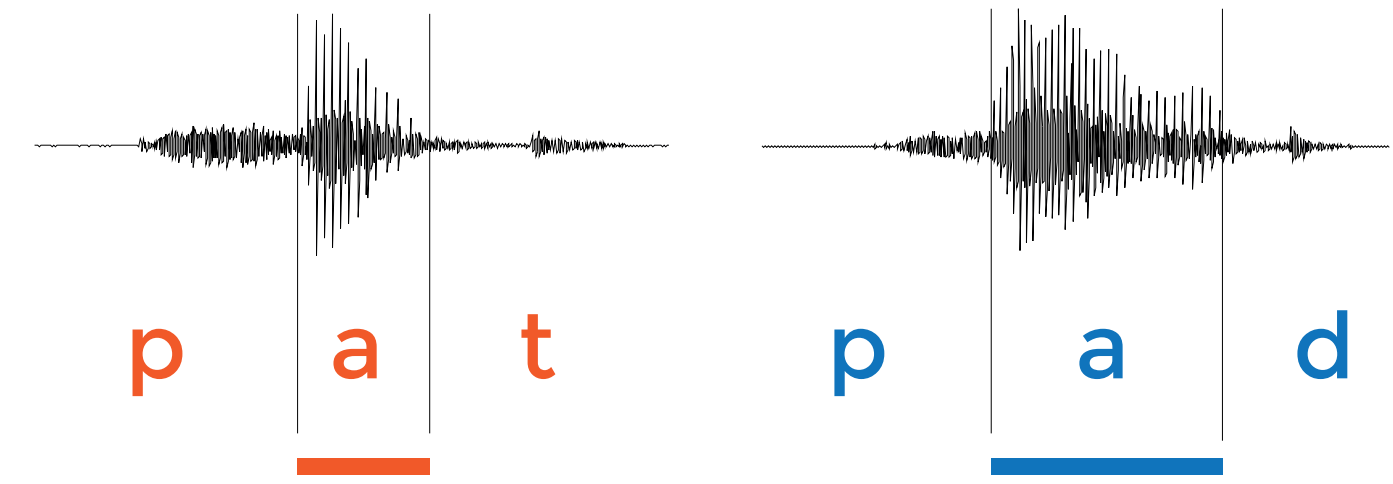
QUANTIFY UNCERTAINTY

Number of participants per study by language endangerment status

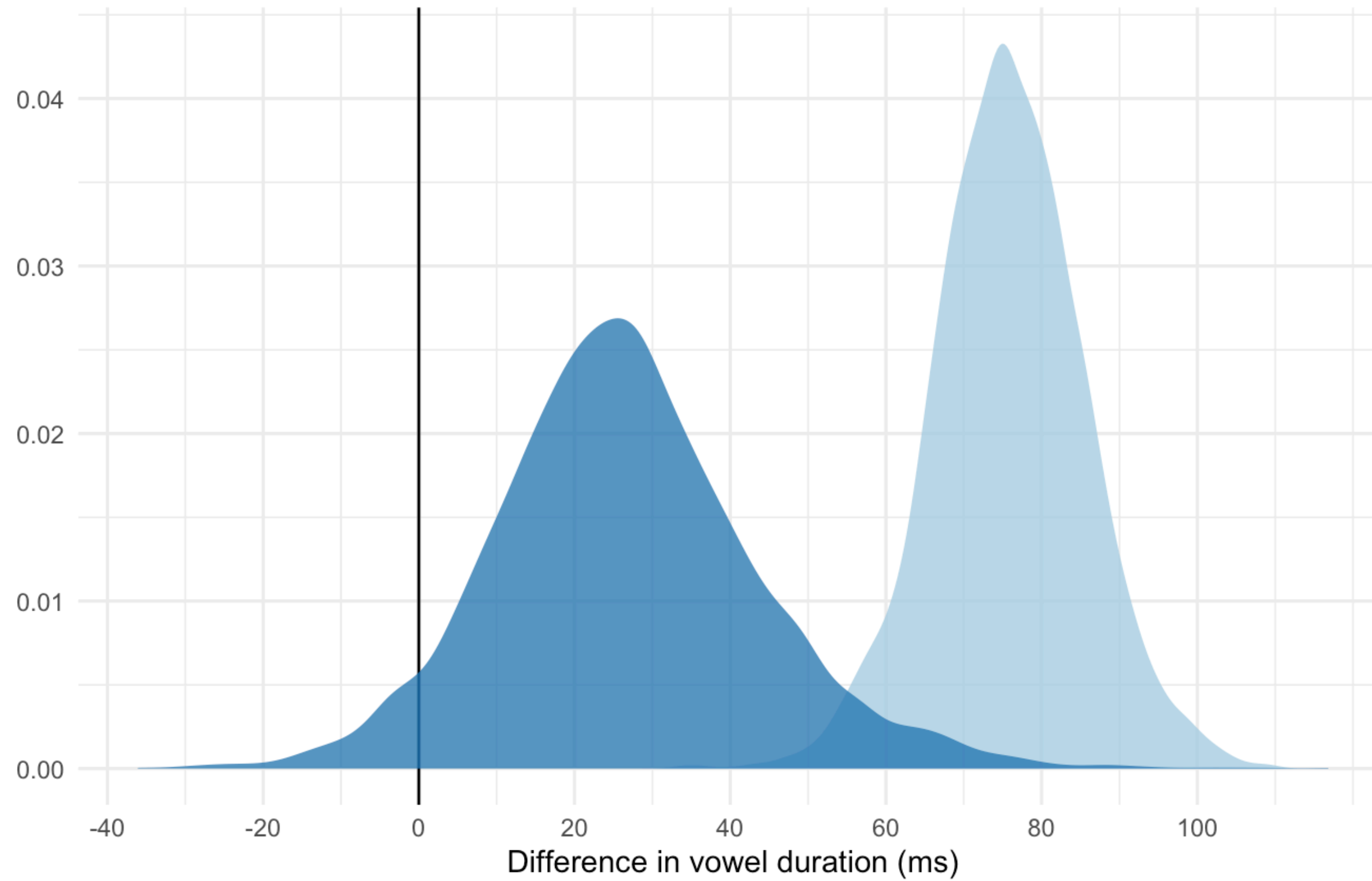
The diamonds indicate the median.



QUANTIFY UNCERTAINTY



Syllable position ■ final ■ penultimate

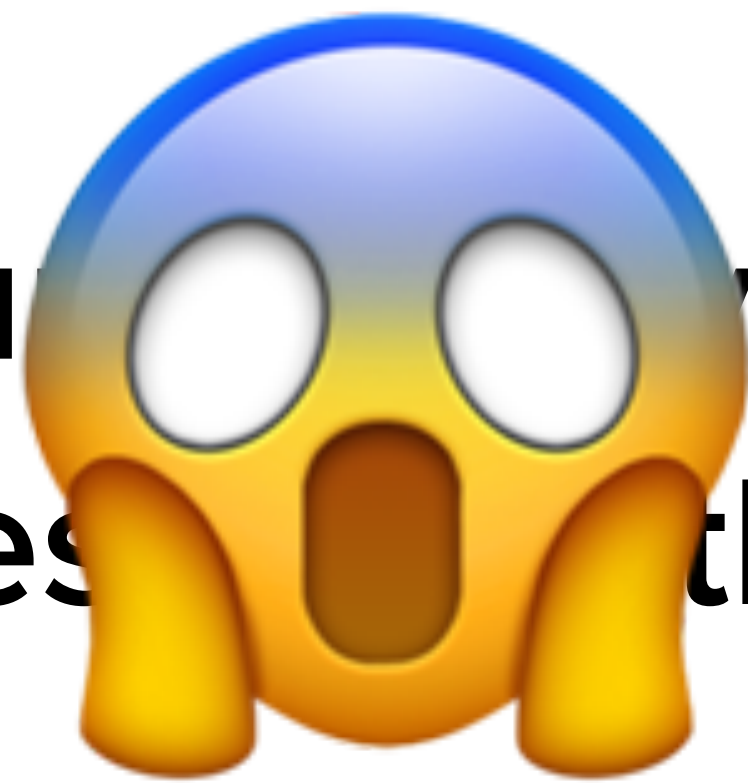


MODEL SPECIFICATION

4 genders × 3 age-groups × 2 vowels × 3 varieties × 2
modalities × 3 ethnic × ...

MODEL SPECIFICATION

4 genders × 3 age-groups × 2 modalities × 3 ethnic × ...
vowels × 3 varieties × 2



Warning messages:

```
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
  Model failed to converge with max|grad| = 0.131568 (tol = 0.001, component 1)  
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
  Model is nearly unidentifiable: very large eigenvalue
```

PARAMETER ESTIMATION
QUANTIFICATION OF UNCERTAINTY
MODEL SPECIFICATION

Bayesian statistics



PARAMETER ESTIMATION



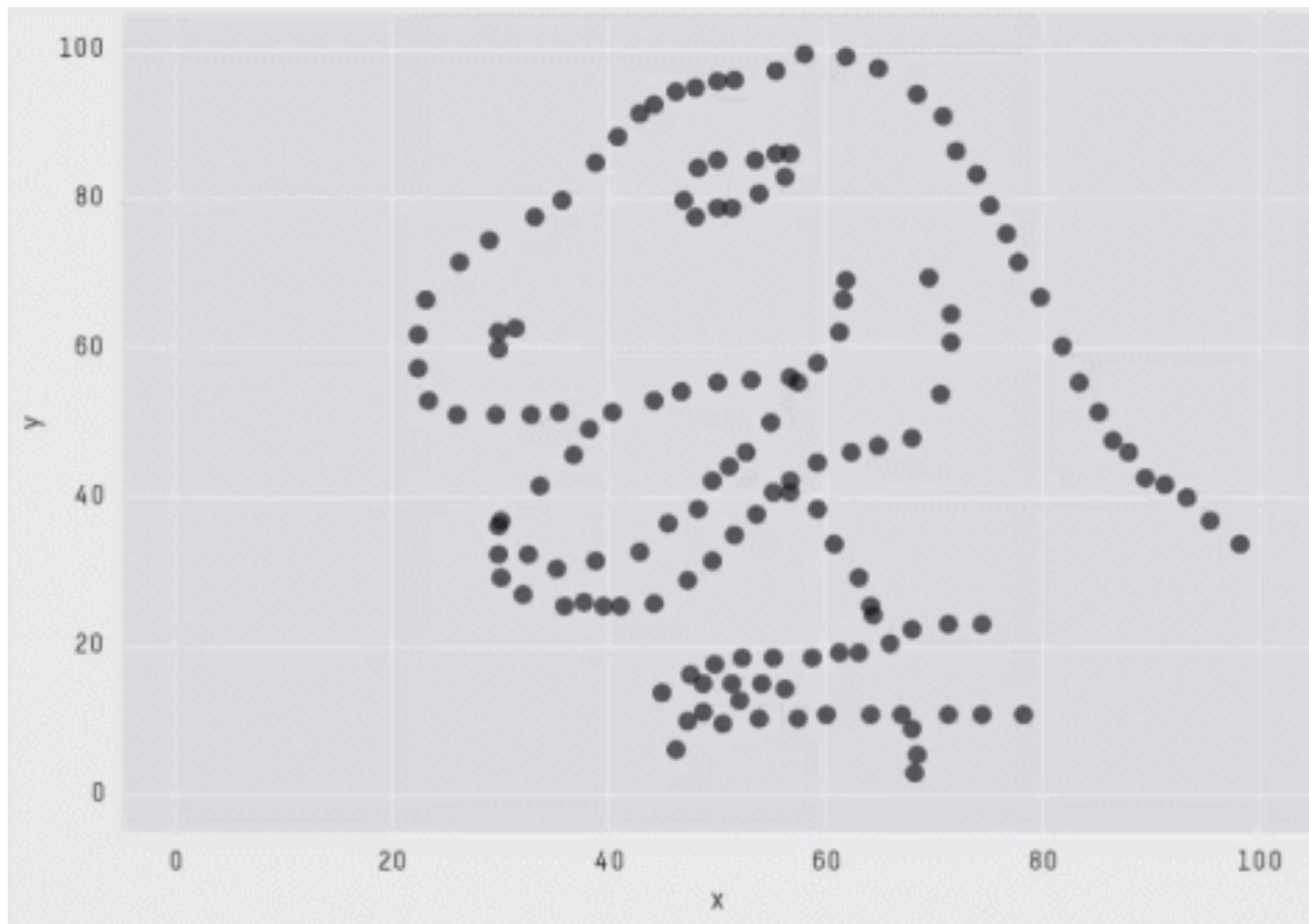
QUANTIFICATION OF UNCERTAINTY



MODEL SPECIFICATION

REFERENCES

<https://stefanocoretta.github.io/biblio/>



X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

FAQ

WHAT IF WE ARE STILL INTERESTED IN PRESENCE VS ABSENCE OF AN EFFECT?

- Statistical significance is no guarantee that the effect in fact exists.
- Null effects (exactly 0) do not exist.
- Which is the **SMALLEST MEANINGFUL EFFECT SIZE?**

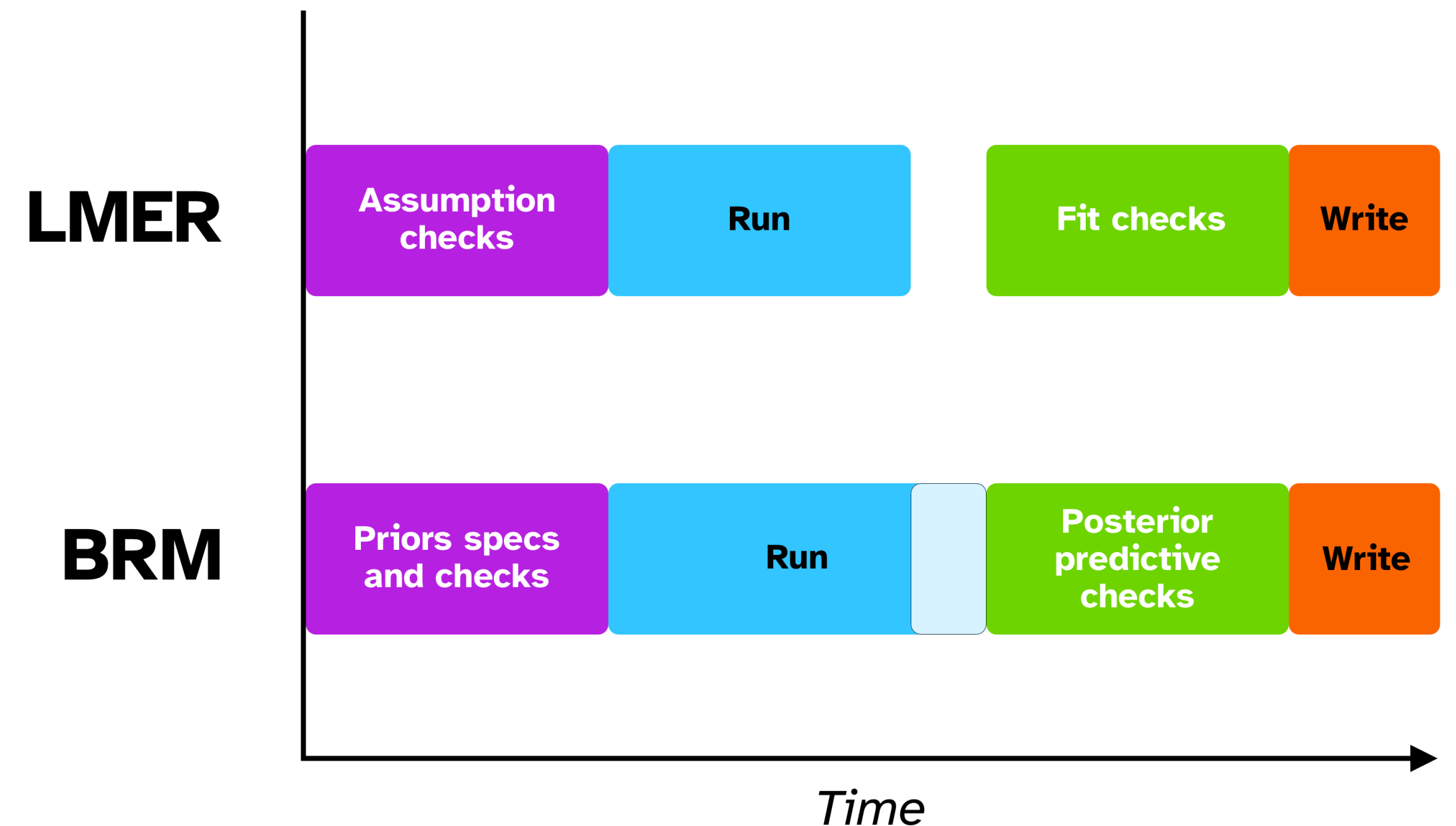
PARAMETER ESTIMATION

WHAT ABOUT ENDANGERED/MINORITISED LANGUAGES

- Bayesian stats works even better!
- You can quantify uncertainty and guard yourself from false results.
- You can still fit complex models, even with small sample sizes.
- (Fun fact: What's the minimal sample size for Bayesian analysis? **ONE OBSERVATION**)

WHAT ABOUT TIME/COMPUTATIONAL COST OR LIMITED RESOURCES?

- Factoring out sampling time, you need more or less the same time.
- While the model is sampling, you can attend to other tasks.
- Use model parallelisation and discretisation on servers (geeky solution).
- If you have research money, invest on a powerful computer and/or a server.



WHAT ABOUT TIME CONSTRAINTS?

- Science is a slow process.
- Fast science is bad for science.
- (By science I mean research).
- Just a philosophical take, the modern western world lives under the assumption that faster is better. But at what cost?
- What about **QUALITY**?

WHAT ABOUT RESEARCH WHICH IS IMPLEMENTING MATHEMATICAL MODELS?

- Yes, those do exist! But they are still the minority.
- For some, see work by Tilsen, Tomaschek, Sóskuthy, Turk (in progress), Beguš (in progress), and others.

WHAT'S YOUR EVIDENCE?

- Personal experience with reading the literature and collaborating with researchers for different subfields.
- Common question: "I see, but is there a way to say that the effect is significant?"
- Reply: "No because statistical significance makes sense only within NHST"
- Follow-up: "But then how do you know that the effect is real?!"
- Reply: "You can never know that, even with statistical significance."

WHAT IF JOURNALS REQUIRE NHST?

- That would be madness!
- I've never got the impression that NHST is a requirement. Rather, some form of statistical inference is.

WHAT ABOUT BAYES FACTORS/BAYESIAN P-VALUES/ETC...

- They can be as bad as NHST p-values, if are not used correctly.
- Bayes Factors require a lot of EXTRA TIME. And they add little information (unless you are particularly interested in comparing two or more contrasting hypotheses, especially if they are all not null).