Multiple Perspectives on Inference for Two Simple Statistical Scenarios

WRCI 2019

Noah N. N. van Dongen, Johnny B. van Doorn, Quentin F. Gronau, Don van Ravenzwaaij, Rink Hoekstra, Matthias N. Haucke, Daniel Lakens, Christian Hennig, Richard D. Morey, Saskia Homer, Andrew Gelman, Jan Sprenger & Eric-Jan Wagenmakers

This work was supported in part by a Vici grant from the Netherlands Organisation of Scientific Research awarded to EJW (016.Vici.170.083) and a Vidi grant awarded to DL (452.17.013). NvD's and JS's work was supported by ERC Starting Investigator Grant No. 640638.

Multiple Perspectives on Inference for Two Simple Statistical Scenarios

WRCI 2019

Noah N. N. van Dongen, Johnny B. van Doorn, Quentin F. Gronau, Don van Ravenzwaaij, Rink Hoekstra, Matthias N. Haucke, Daniel Lakens, Christian Hennig, Richard D. Morey, Saskia Homer, Andrew Gelman, Jan Sprenger & Eric-Jan Wagenmakers

This work was supported in part by a Vici grant from the Netherlands Organisation of Scientific Research awarded to EJW (016.Vici.170.083) and a Vidi grant awarded to DL (452.17.013). NvD's and JS's work was supported by ERC Starting Investigator Grant No. 640638.

Outline

Rationale

Design

Results of the analyses

Results of the discussion

Conclusion

Q & A

All the material is available at: https://osf.io/hykmz/

Rationale

The high ground of scientific objectivity has been seized by the frequentists.

Efron (1986, p.4)

Every statistician would be a Bayesian if he took the trouble to read the literature thoroughly and was honest enough to admit that he might have been wrong.

Lindley (1986, p.7)

Rationale

- Ongoing debates between statistical paradigms on merits and deficiencies of their and the other's approaches to hypothesis testing and estimation.
- These statistical paradigms are philosophically incompatible.
- These debates date back to the birth of modern statistics.
- These debates show no sign of stalling or abating.

HOWEVER…

Does it matter in practice which statistical approach you use?

Rationale

Your [Ronald Fisher's] *letter confirms my* [Harold Jeffreys'] *previous impression that it would only be once in a blue moon that we would disagree about the inference to be drawn in any particular case, and that in the exceptional cases we would both be a bit doubtful.*

Bennett (1990, p.162)

Rationale

Research Questions:

1. Do research scenarios give rise to different statistical approaches within and between statistical paradigms?
2. Do these approaches lead to similar or dissimilar conclusions in written reports and round-table discussion?

Design

Two scenarios

1. Comparison of two proportions: risk of a drug tested between drug and control group
2. Correlation: association between two properties.

Two data sets

1. Birth defects between women that took Cetirizine during pregnancy (n=181) and controls (n=1685) from Weber-Schoendorfer and Schaefer (2008)
2. Correlation between perceived stress and amygdalar activity (n=13) from Tawakol et al. (2017).

Four teams

1. Daniel Lakens and Christian Hennig
2. Richard Morey and Sakia Homer
3. Quentin Gronau, Johnny van Doorn, and Eric-Jan Wagenmakers
4. Andrew Gelman

Output

Written research reports and round-table discussion

Contingency table data

	Birth Defects		
Cetirizine exposure	**No**	**Yes**	**Total**
No	1588	97	1685
Yes	167	14	181
Total	1755	111	1866

Contingency table analysis results

Lakens and Hennig

- Frequentist test of equivalence
- 10% equivalence region
- Data deemed inconclusive

Gronau, van Doorn, and Wagenmakers

- Default Bayes factor BF_{10} =1.6
- Evidence “not worth more than a bare mention”
- Data deemed inconclusive

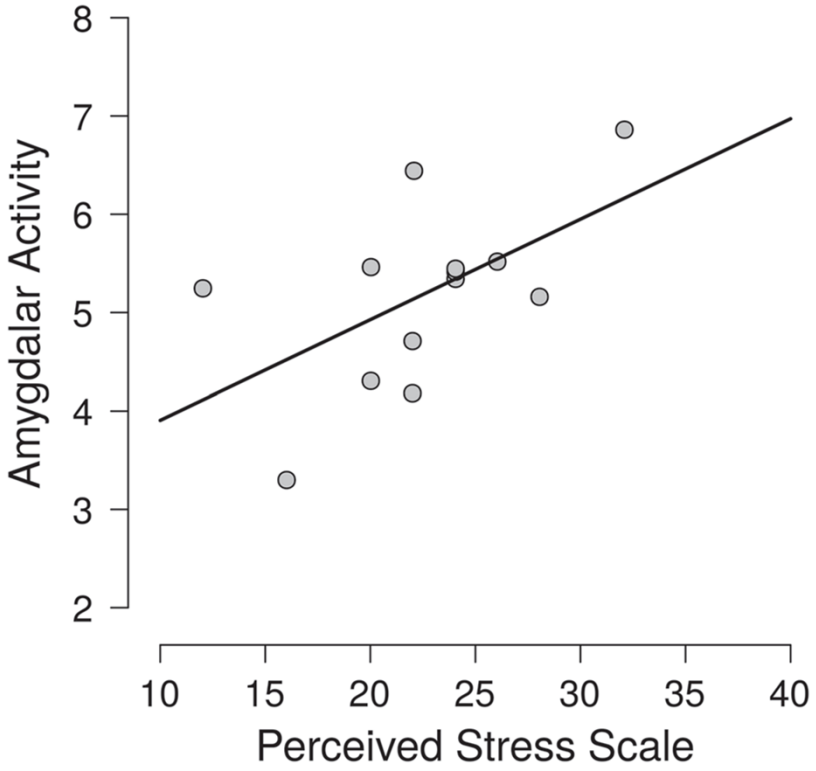
Morey and Homer

- Frequentist logistic model, p=.287
- Observational, so possible confounds
- Data deemed inconclusive

Gelman

- Simple comparison and logistic regression
- Bayesian analysis needs good prior
- A key question is who takes the drug in the population
- Data likely to be inconclusive

Correlation data



Perceived Stress Scale	Amygdalar Activity
12.0103	5.2418
32.035	6.8601
22.0296	6.4402
20.0079	5.462
24.0155	5.4439
24.0155	5.3349
24.0155	5.4216
26.0082	5.5176
28.012	5.1615
21.9872	4.7114
21.9872	4.1844
20.0138	4.3079
16.0088	3.3015

Correlation analysis results

Lakens and Hennig

- Frequentist correlation, p = .047
- Concerns about multiple comparisons
- Data deemed inconclusive

Gronau, van Doorn, and Wagenmakers

- Default Bayes factor BF10 = 2
- Small sample size
- Data deemed inconclusive

Morey and Homer

- Frequentist correlation, confidence intervals
- Small n, means assumptions unverifiable
- Is the effect specific for amygdala?
- Data deemed inconclusive

Gelman

- Linear regression with logarithms of the variables.
- Bayesian analysis needs good prior
- Problems with generalizing to population
- Data likely to be inconclusive

Discussion Results

Opening statement:

In statistics, it doesn't matter what approach is used. As long as you do conduct your analysis with care, you will invariably arrive at the same qualitative conclusion.

Discussion Results

Two points of agreement:

1. All teams agreed that much of problems in research start before statistics get involved.
2. They were all sceptical of the adequacy of the models they had used.

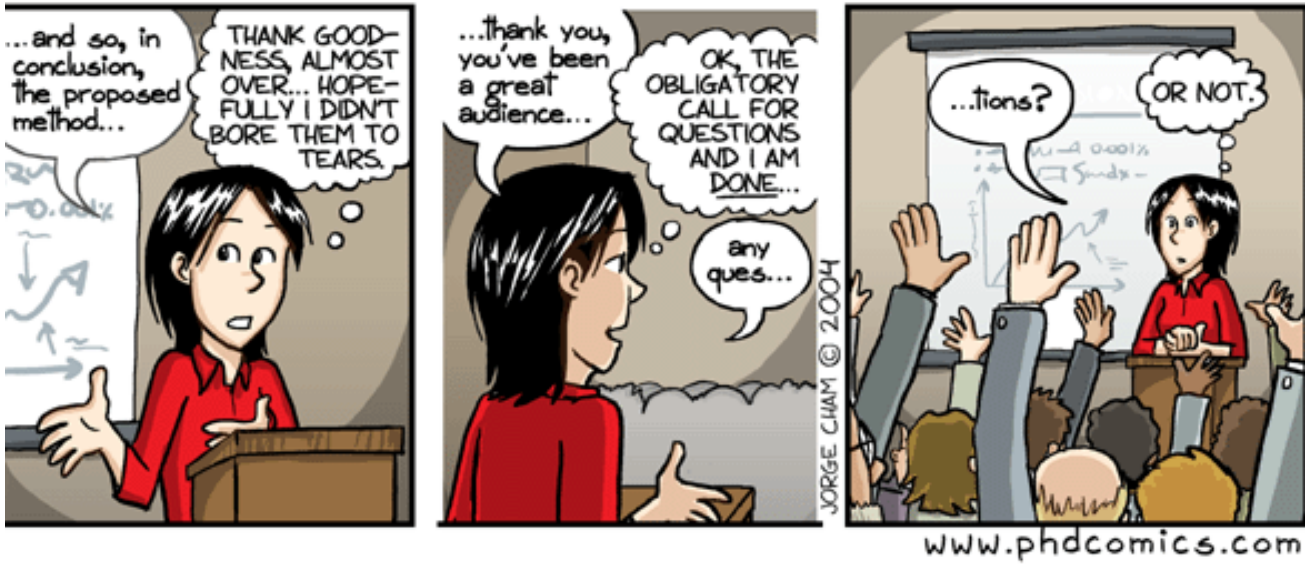
Two points of disagreement:

1. Whether or not one could learn anything without invoking Bayes rule, overtly or covertly.
2. When and how strictly to adhere to the modelling assumptions.

Conclusion(s)

- All teams employed widely different approaches.
- There is tentative support for the Fisher–Jeffreys conjecture:
  - regardless of the statistical framework in which they operate, careful analysts will often come to similar conclusions.
- Statistical analysis, even for the simplest of scenarios, requires more than a mechanical application of a set of rules
  - a careful analysis is a process that involves both skepticism and creativity.
- Incorporation of independent re-analysis into the peer-review process might be possible and fruitful.
- Investigation of how we learn from statistical results might be useful.
- Future researchers might benefit from allocation of attention and resources to the improvement of methodological education of research aspects that precede statistical analysis.

Questions?



All the material is available at: https://osf.io/hykmz/

References

Bennett, J. H. (1990). *Statistical Inference and Analysis: Selected Correspondence of R. A. Fisher*. Oxford, Clarendon Press.

Efron, B. (1986). Why Isn't Everyone a Bayesian? *The American Statistician*, 40, 1–5.

Lindley, D. V. (1986). Comment on 'Why Isn't Everyone a Bayesian?' By Bradley Efron. *The American Statistician*, 40, 6–7.

Tawakol, A., Ishai, A., Takx, R. A. P., Figueroa, A. L., Ali, A., Kaiser, Y., Truong, Q. A., Solomon, C. J. E., Calcagno, C., Mani, V., Tang, C. Y., Mulder, W. J. M., Murrough, J. W., Hoffmann, U., Nahrendorf, M., Shin, L. M., Fayad, Z. A., and Pitman, R. K. (2017). Relation Between Resting Amygdalar Activity and Cardiovascular Events: A Longitudinal and Cohort Study. *The Lancet*, 389, 834–845

Weber-Schoendorfer, C. and Schaefer, C. (2008). The Safety of Cetirizine During Pregnancy: A Prospective Observational Cohort Study. *Reproductive Toxicology*, 26, 19–23.