statistical essentials for modeling

Ben Bolker

Philosophy

What do we mean by statistical inference?

answering **scientific questions**

- clear, well-posed questions (theory) >
  - good experimental design >
    * statistical practice

  ... all are necessary, all connected!

- statistics is for:
  - quantifying best guesses (point estimates)
  - quantifying uncertainty (confidence intervals)
  - statements about *clarity* (statistical significance testing)

reproducibility crisis

many scientific results are unreproducible

- lack of openness (data/methods)
- questionable research practices (QRPs)
- p-hacking; snooping (Simmons, Nelson, and Simonsohn 2011); "Texas sharpshooter fallacy"
- "garden of forking paths" (Gelman)

  analytic decisions must be made independently of the data
 pre-registration (formal or informal);
  at least recognize the line between confirmatory and exploratory analyses

scientific hell

power analysis

- experimental design: **before** you observe/experiment
- think about *biological* effect sizes: what is the *smallest* effect that would be biologically interesting?
- need to specify effects and variances (standard deviations)
- simple designs (t-test, ANOVA, etc.)
  - base R: `apropos("power")`

The Nine Circles of Scientific Hell
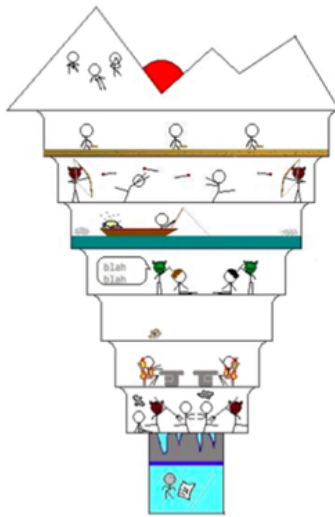
First Circle: Limbo

Second Circle: Overselling

Third Circle: Post-Hoc Storytelling

Fourth Circle: P-Value Fishing

Fifth Circle: Creative Use of Outliers

Sixth Circle: Plagiarism

Seventh Circle: Non-Publication of Data

Eighth Circle: Partial Publication of Data

Ninth Circle: Inventing Data

- – Russ Lenth home page
- most power analyses are crude/order-of-magnitude
- simulation-based power analysis (Bolker (2008) ch. 5)

goals of analysis (Harrell 2001)

- exploration
- prediction
- inference

exploration

- looking for patterns *only*
- no p-values at all
- confidence intervals (perhaps),
  but taken with an inferential grain of salt

prediction

- want quantitative answers about specific cases
- consider algorithmic approaches (esp. for big data)
- penalized approaches:
  automatically reduce model complexity
- confidence intervals are hard

inference

most typical scientific goal
  qualitative statements about clarity and importance of effects:

- effects that are distinguishable from null hypothesis of noise

- test among discrete hypotheses

  quantitative statements:

- relative strength/magnitude of effects
- importance (e.g. fraction variance explained)

what do p-values really mean?

- something about "strength of evidence"
- **not** "evidence for no effect" or "no difference"
- null hypotheses in ecology are never (?) true

- "the difference between significant and non-significant is not significant" (Gelman and Stern 2006)
- try talking about statistical **clarity** instead

p-value example

```
##    p value
## A  0.0011 **
## B  0.1913
## C  0.0011 **
## D  0.1913
## ---
## Signif. codes:
##    0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
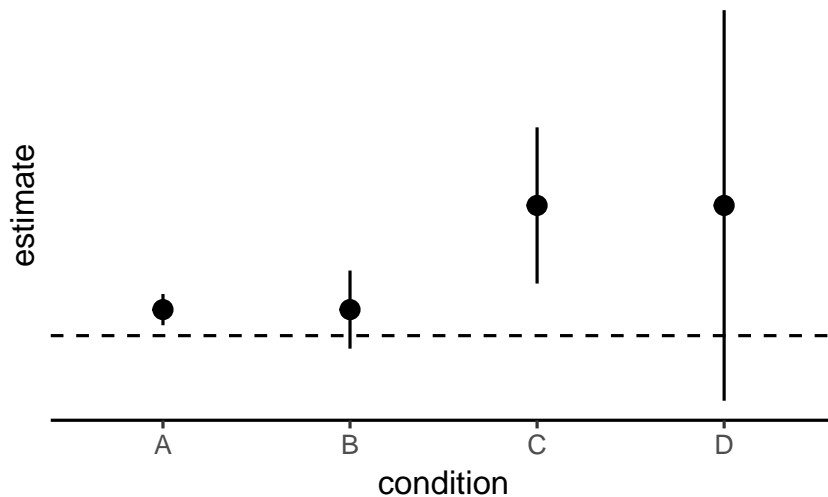

Real example (Dushoff et al. 2006)

From a study of influenza mortality, estimating fraction of mortality attributable to influenza A, influenza B, or weather alone . . .
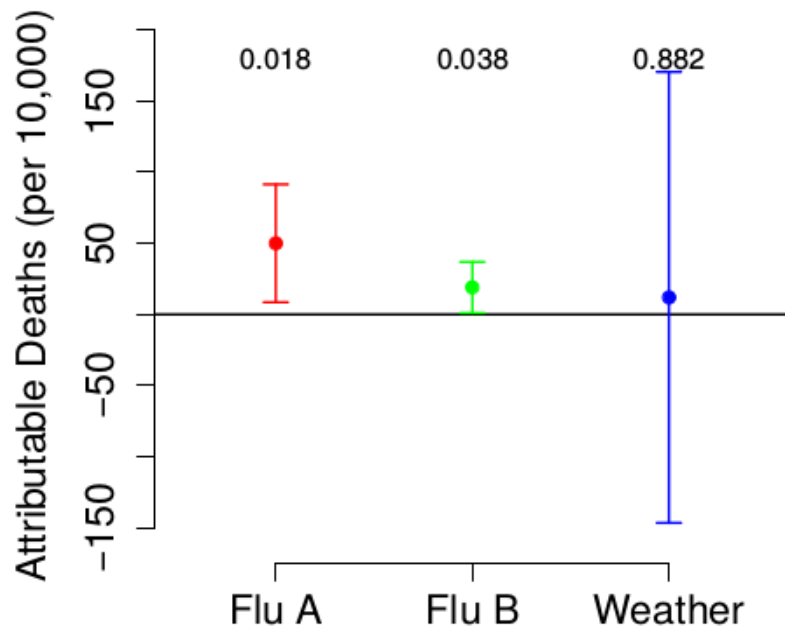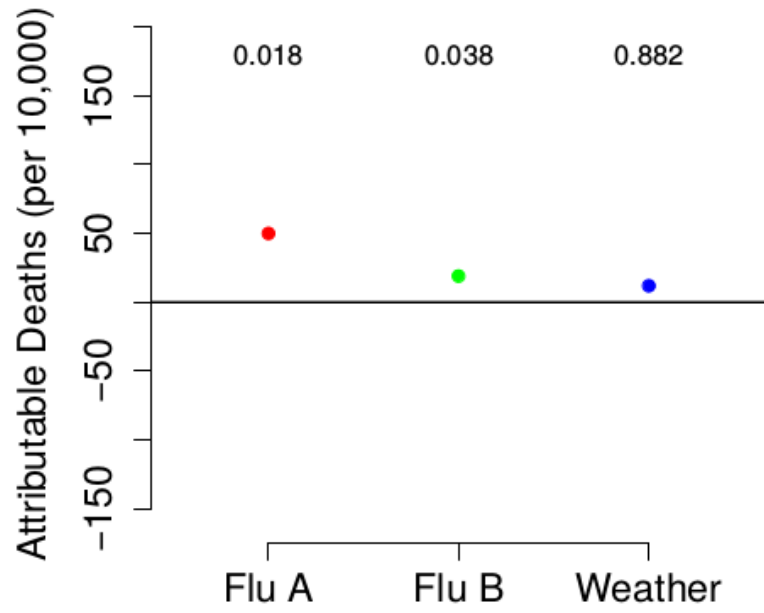 Why does weather not seem to have an effect???

The explanation

realism in data analysis

how much data do you need for a given model?

- link to video
- rule of thumb: 10-20 per data point
- rules for continuous, count, binomial data
- counting data points/"degrees of freedom" for clustered data?

dimension reduction

- must be *a priori*
- discard interactions
- simplify questions
- collapse variables, e.g. by PCA

overview of inference

modes of inference (Bolker (2008) chapter 6)

- Wald vs. likelihood ratio vs. Bayesian
- information-theoretic (AIC etc.) methods
- single vs. multiple parameters:
  e.g. Z vs χ^2
- finite-size vs asymptotic:
  e.g. Z vs. t
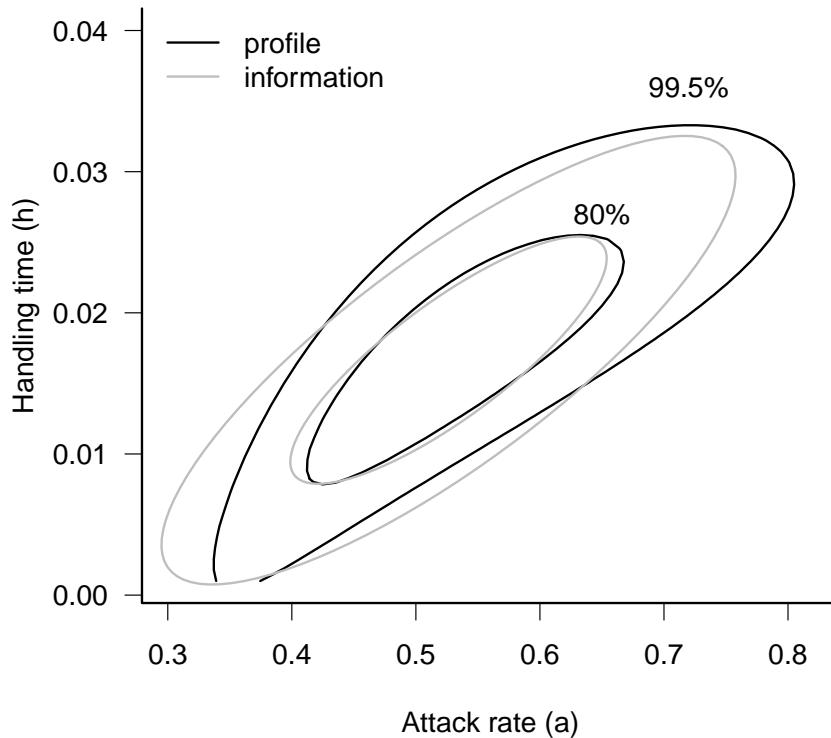
Bayesian stats 101

- frequentist inference: based on likelihood function + sampling properties
- $\mathcal{L} = \text{Prob}(\theta|x)$
- Bayesian inference: based on likelihood + prior

  – Posterior$(\theta) \propto \mathcal{L}(\theta|x)$Prior$(\theta)$

- priors are important
- Bayesian *credible intervals* based on highest posterior density or quantiles

  results **more explicitly** based on model + prior choices

Bayesian stats 102

- *Markov chain Monte Carlo*: computational methods for sampling from the posterior
- once we have a a sample we can compute mean, confidence intervals . . .
- e.g. `MCMCglmm, brms, rethinking`

References

Bolker, Benjamin M. 2008. *Ecological Models and Data in R*. Princeton, NJ: Princeton University Press.

Dushoff, Jonathan, Joshua B. Plotkin, Cecile Viboud, David J. D. Earn, and Lone Simonsen. 2006. "Mortality Due to Influenza in the United States—An Annualized Regression Approach Using Multiple-Cause Mortality Data." *American Journal of Epidemiology* 163 (2): 181–87. doi:10.1093/aje/kwj024.

Gelman, Andrew, and Hal Stern. 2006. "The Difference Between 'Significant' and 'Not Significant' Is Not Itself Statistically Significant." *The American Statistician* 60 (4): 328–31. doi:10.1198/000313006X152649.

Harrell, Frank. 2001. *Regression Modeling Strategies*. Springer.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22 (11): 1359–66. doi:10.1177/0956797611417632.