# Tutorial: Bayesian data analysis

**John K. Kruschke (johnkruschke@gmail.com)**
Department of Psychological and Brain Sciences, Indiana University
1101 E. 10th St., Bloomington, IN 47405 USA
**Tutorial web site: `http://tiny.cc/BayesAtCogSci2015`**

**Keywords:** Bayesian; data analysis; regression; hierarchical model; statistics; methods; Markov chain Monte Carlo; R programming language; *p* value; confidence interval; puppies

BAYESIAN DATA ANALYSIS is superseding traditional methods in sciences from anthropology to zoology. Bayesian methods solve many problems inherent in *p* values and confidence intervals. More importantly, Bayesian methods are more richly and intuitively informative. Bayesian analysis applies flexibly and seamlessly to simple situations or complex hierarchical models and realistic data structures, including small samples, large samples, unbalanced designs, missing data, censored data, outliers, etc. Bayesian analysis software is flexible and can be used for a wide variety of data-analytic and psychometric models.

This full-day tutorial presents a ground-level, hands-on introduction to doing Bayesian data analysis. The presenter is an award-winning teacher who has honed new materials from many previous courses. He has written an acclaimed textbook on the topic, now greatly expanded in its second edition. The tutorial materials include free software and numerous programs that can be used for real data analysis.

## Objectives

Attendees will learn:

- the rich and intuitive information provided by Bayesian analysis and how it differs from traditional (frequentist) methods.
- the concepts and hands-on use of modern algorithms ("Markov chain Monte Carlo") that achieve Bayesian analysis for realistic applications.
- how to use the free software, called R and JAGS, for Bayesian analysis, along with many programs created by the instructor that are readily useable and adaptable for your research.
- many useful applications, including comparison of two groups, regression models, and hierarchical models.

## Prerequisites

No specific mathematical expertise is presumed. In particular, no matrix algebra and no calculus is used in the tutorial. Some previous familiarity with statistical methods such as a *t*-test or linear regression can be helpful, as would be some previous experience with computer programming, but these are not crucial.

## Audience

The intended audience is graduate students, faculty, and other researchers, from all disciplines, who want a ground-floor introduction to doing Bayesian data analysis.

## Content and Schedule

**9:00–10:20.** The day begins with a genuine beginner's introduction to foundational concepts. An introductory chapter that covers this material is available online at the tutorial's web site (shown under the title / byline of this document). The session continues with a complete example of Bayesian comparison of two groups. A video summarizing this material is also available at the tutorial's web site.

10:20–10:40. Break

**10:40–12:00.** The second morning session covers the ideas behind the essential algorithms that make modern Bayesian analysis possible: Markov chain Monte Carlo (MCMC). While it is important to understand the ideas of MCMC, fortunately we don't have to deal with MCMC directly because the programming language JAGS makes it easy to implement a huge variety of models. By lunch time you will have a chance to make JAGS do Bayesian analyses for you.

12:00–1:30. Lunch (on your own).

**1:30–2:50.** The first afternoon session considers frequently used models, including various regression models. You will see how easy it is to implement hierarchical models in JAGS.

2:50–3:10. Break

**3:10–4:30.** The second afternoon explores how null values are assessed in Bayesian and frequentist analyses. After briefly reviewing the perils of *p* values and the con game of confidence intervals, two Bayesian approaches to null value assessment will be explored.

**4:30–5:00.** The tutorial concludes with an open question-and-answer period.

## Presenter

The presenter is eight-time winner of Teaching Excellence Recognition Awards from Indiana University. He has given numerous well-received workshops on Bayesian data analysis, and is the author of the acclaimed book, *Doing Bayesian Data Analysis* (Second Edition, Kruschke, 2015), along with numerous articles on Bayesian data analysis (e.g., Kruschke, 2010, 2013). The presenter is Professor of Psychological and Brain Sciences, and Adjunct Professor of Statistics, at Indiana University in Bloomington. He was awarded the Troland Research Award from the National Academy of Sciences, and the Remak Distinguished Scholar Award from Indiana University. He has been on the editorial boards of various scientific journals, including *Psychological Review*, the *Journal of Experimental Psychology: General*, and the *Journal of Mathematical Psychology*. (On the other hand, he put pictures of puppies on the cover of the book he wrote.)

## Computer software and background material

Attendees are strongly encouraged to bring a notebook computer to the tutorial. Install software on your notebook computer *before arriving*. See instructions at the tutorial's web site. Be sure to install R, RStudio, JAGS, and the programs from the book. All software is free.

The book, *Doing Bayesian Data Analysis, Second Edition*, is highly recommended as background and follow-up to the tutorial. Extensive information about the book, and a link to a publisher's discount, can be found at the tutorial's web site. Slides from the presentations at the tutorial will also be made available to attendees.

## What is Bayesian data analysis and why learn it?

Bayesian reasoning is simply the re-allocation of credibility across possibilities. For a given domain of data, we begin with a set of possible explanations and the prior credibility of each explanation. Then we observe some data, and re-allocate credibility toward the explanations that are more consistent with the data. This sort of re-allocation is intuitive in everyday reasoning, as when Sherlock Holmes argued that when you have eliminated the impossible, whatever remains must be the truth — as illustrated in the left side of Figure 1. The mathematically exact way to re-allocate credibility is done by Bayesian analysis. A realistic illustration appears in the right column of Figure 1.

Anything we want to know from the analysis is directly "read off" the posterior distribution; e.g., the most credible value and the exact uncertainty of the estimate. There is no need to derive $p$ values (and their associated confidence intervals) from auxiliary assumptions about sampling distributions and null hypotheses. The posterior distribution provides exactly the information that we intuitively already think that frequentist analysis provides but does not. Bayesian software applies seamlessly to simple and complex models.

The analysis methods are at the convergence of two historical trends in the practice of data analysis, shown in Figure 2. The trend from frequentist to Bayesian methods is shown across columns. A second trend, from a focus on null hypothesis testing to a focus on estimation with uncertainty, is shown across rows. Each cell of Figure 2 indicates combinations of method and focus. This tutorial aims at the convergence of the two trends in the lower-right cell.

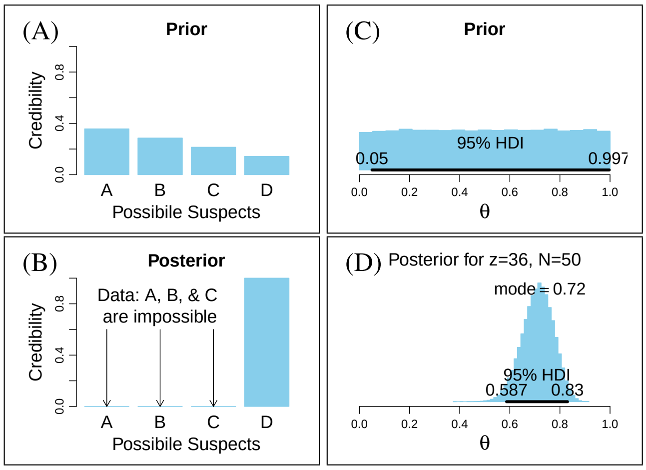


Figure 1: Bayesian analysis re-allocates credibility across possibilities. *Left column: Reasoning of Sherlock Holmes.* (A) Prior distribution of credibility (i.e., culpability) across four suspects for a crime. (B) After data indicate that suspects A, B, and C could not have committed the crime, the posterior distribution loads all credibility on suspect D. *Right column: Bayesian estimation of the cure probability, θ, of a drug.* (C) Prior distribution is broad, meaning a wide range of cure probabilities is possible. (D) After observing 36 cures in 50 patients, the posterior distribution is narrower, and precisely displays the most credible probability and the uncertainty of the estimate. (HDI = highest density interval.)

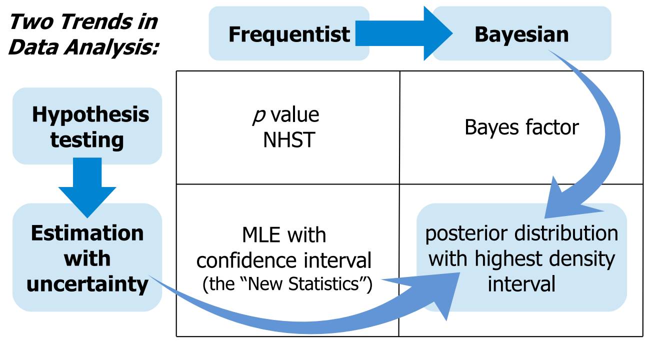


Figure 2: Two historical trends in data analysis converge on Bayesian estimation with uncertainty, as taught in this tutorial. (NHST = null hypothesis significance testing. MLE = maximum likelihood estimate.)

## References


Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*(7), 293–300. doi: 10.1016/j.tics.2010.05.001

Kruschke, J. K. (2013). Bayesian estimation supersedes the $t$ test. *Journal of Experimental Psychology: General*, *142*(2), 573–603. (DOI: 10.1037/a0029146)

Kruschke, J. K. (2015). *Doing Bayesian data analysis, Second Edition: A tutorial with R, JAGS, and Stan*. Burlington, MA: Academic Press / Elsevier.