Bayesian Model Selection - a Deeper Dive into Bayesian Correlation and Linear Regression

Tom Faulkenberry

Tarleton State University
Stephenville, TX, USA
https://tomfaulkenberry.github.io
Twitter: @tomfaulkenberry

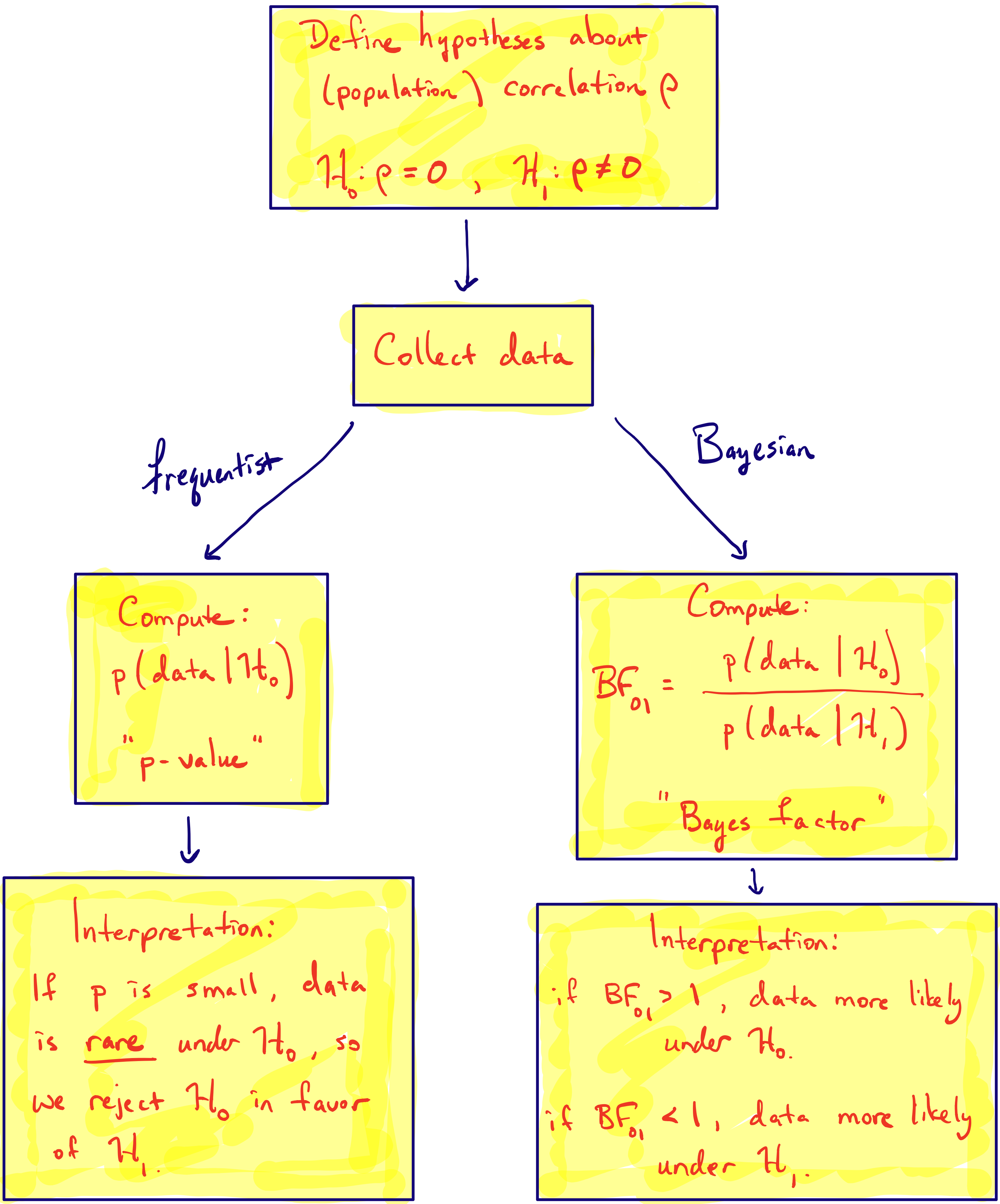
Ulster University Psychology Seminar
June 4, 2020

Outline:

- review of differences between p-values & Bayes factors
- priors on models vs. priors on parameters
- correlation example using JASP, w/ reporting template.
- linear regression example, w/ introduction to Bayesian model averaging

* These slides can be downloaded from https://tomfaulkenberry.github.io/talks.html

Suppose we are interested in the relationship between maths anxiety and performance on a standardized assessment.



p-value = $p(\text{data} \mid H_0)$

1) only considers fit of H_0 as a potential model for data
2) ignores fit of H_1

Thus, "support" for H_1 is only indirect

Bayes factor = $\dfrac{p(\text{data} \mid H_0)}{p(\text{data} \mid H_1)}$

1) considers relative adequacy of both models as predictors of data.
2) can directly index support for either H_0 or H_1.

Ex: $BF_{01} = 8$ → "The observed data are 8 times more likely under H_0 than H_1."

Jeffreys (1961):

BF	Evidence*
1 – 3	anecdotal
3 – 10	moderate
10 – 30	strong
30 – 100	very strong
> 100	extreme

* these are only guidelines!

How does Bayes work?

for single model $\mathcal{H}$:

$$p(\mathcal{H} \mid data) = p(\mathcal{H}) \times \frac{p(data \mid \mathcal{H})}{p(data)}$$

posterior belief in $\mathcal{H}$ = prior belief in $\mathcal{H}$ × updating factor

for two models:

$$\frac{p(\mathcal{H}_0 \mid data)}{p(\mathcal{H}_1 \mid data)} = \frac{p(\mathcal{H}_0)}{p(\mathcal{H}_1)} \times \frac{p(data \mid \mathcal{H}_0)}{p(data \mid \mathcal{H}_1)}$$

posterior odds = prior odds × Bayes factor

What do we mean by prior?

Two types of "priors":

1) priors on models

2) priors on parameters within a given model

① Priors on models — before observing data, what is relative likelihood of competing models?

- common default: $p(\mathcal{H}_0) = p(\mathcal{H}_1) = 1/2$

  ↳ i.e., "1-1 prior odds"

- these prior model probabilities must add to 1

  ↳ $p(\mathcal{H}_0) + p(\mathcal{H}_1) = \frac{1}{2} + \frac{1}{2} = 1$

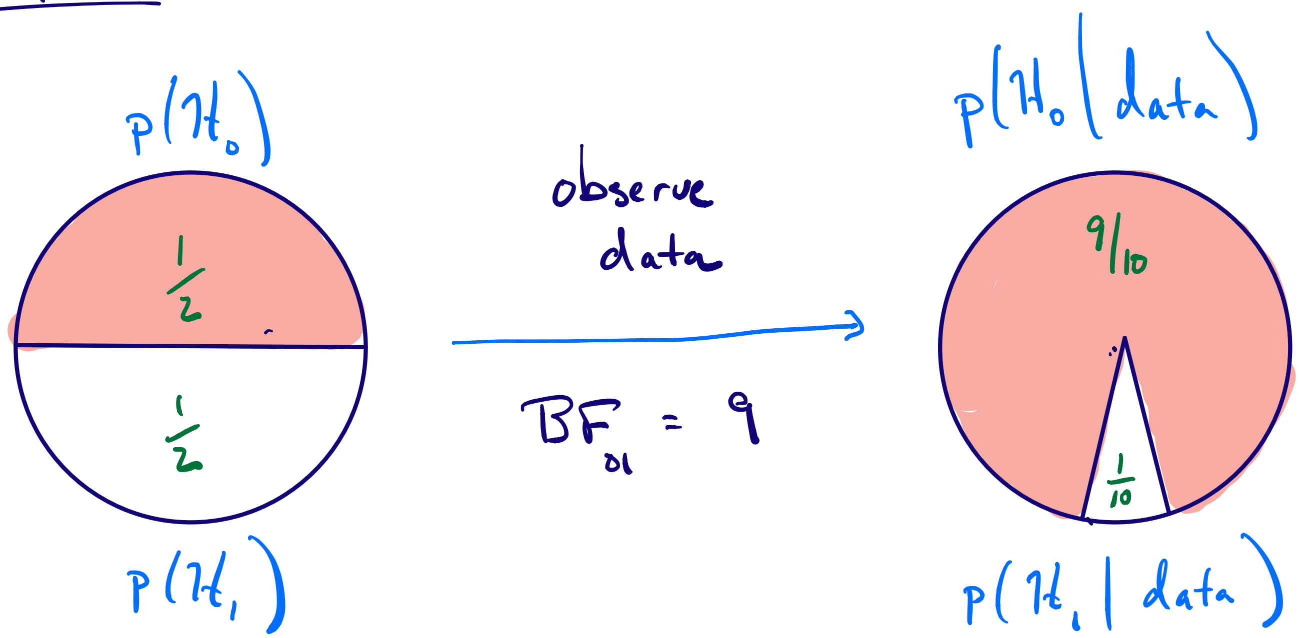
- prior model probabilities are updated after observing data:

$$p(\mathcal{H}_0 \mid \text{data}) = \frac{BF_{01} \cdot p(\mathcal{H}_0)}{BF_{01} \cdot p(\mathcal{H}_0) + p(\mathcal{H}_1)}$$

* Note: if $p(\mathcal{H}_0) = p(\mathcal{H}_1) = \frac{1}{2}$,

$$p(\mathcal{H}_0 \mid \text{data}) = \frac{BF_{01}}{BF_{01} + 1}$$

Example:



Prior odds = 1:1

Posterior odds = 9:1

$$* \; p(\mathcal{H}_0 \mid \text{data}) = \frac{BF_{01}}{BF_{01} + 1}$$

$$= \frac{9}{9+1}$$

$$= 0.9$$

(2) priors on parameters within a given model.

- model definitions : $\mathcal{H}_0 : \rho = 0$

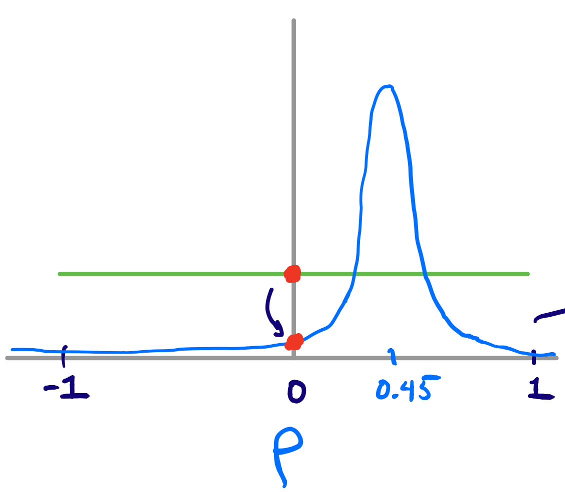
  $\mathcal{H}_1 : \rho \neq 0$ ← what exactly do we mean here?

- we quantify our uncertainty about the correlation ρ under $\mathcal{H}_1$ by placing a distribution on ρ

- suppose we have no idea what to expect. Here, we might believe any value of ρ is equally likely to occur.

  ↳ we say ρ is uniformly distributed on $(-1, 1)$

Prior



observe data →

Posterior

ratio of $\rho = 0$ in prior & posterior $= BF_{01}$

– "Savage-Dickey Density Ratio"

Let's continue our working example. Suppose we tested $N = 65$ participants and observed a correlation of $r = 0.37$.

- use JASP "Summary Statistics" module

Elements to report:

1. report results of hypothesis test
   - define $\mathcal{H}_0$, $\mathcal{H}_1$ and specify prior under $\mathcal{H}_1$.

     "Under the null hypothesis we expect a correlation of 0 between maths anxiety and performance. Thus, we define $\mathcal{H}_0: \rho = 0$. The alternative hypothesis is two-sided, $\mathcal{H}_1: \rho \neq 0$, and we assigned a uniform prior probability to all values of ρ between -1 and $+1$."

   - report and interpret Bayes factor

     "We found a Bayes factor of $BF_{10} = 13.93$, which means that the observed data are approximately 14 times more likely under $\mathcal{H}_1$ than $\mathcal{H}_0$. This result indicates strong evidence in favor of $\mathcal{H}_1$."

- (optional) calculate and report posterior model probability for preferred model.

  - from earlier,

$$p(\mathcal{H}_1 \mid data) = \frac{BF_{10}}{BF_{10} + 1}$$

$$= \frac{13.93}{13.93 + 1} = 0.93.$$

- "Assuming prior odds of 1-1 for $\mathcal{H}_1$ and $\mathcal{H}_0$, our observed data updated these odds to 13.93-to-1 in favor of $\mathcal{H}_1$. This is equivalent to a posterior model probability of $p(\mathcal{H}_1 \mid data) = 0.93$."

2. report results of parameter estimation
   - only if $\mathcal{H}_1$ is the preferred model!
   - specify parameter of interest and remind reader of prior under $\mathcal{H}_1$
     - "of interest is the posterior distribution for ρ, the population-level correlation between maths anxiety and performance. Under $\mathcal{H}_1$, ρ was assigned a uniform prior over the interval from -1 to $+1$."

- report the 95% credible interval.
  - "The posterior distribution for ρ had a median of 0.356, with a central 95% credible interval that ranges from 0.134 to 0.554."

---

see van Doorn et al. (2019). The JASP guidelines for conducting and reporting a Bayesian analysis. https://psyarxiv.org/yqxfr.

Bayesian Linear Regression

- basic ideas remain the same, but:
  - multiple competing models, depending on # predictors
  - uncertainty across models and within models.
    - ↳ use Bayesian model averaging

Lets use JASP to walk through a Bayesian linear regression.

- details found in

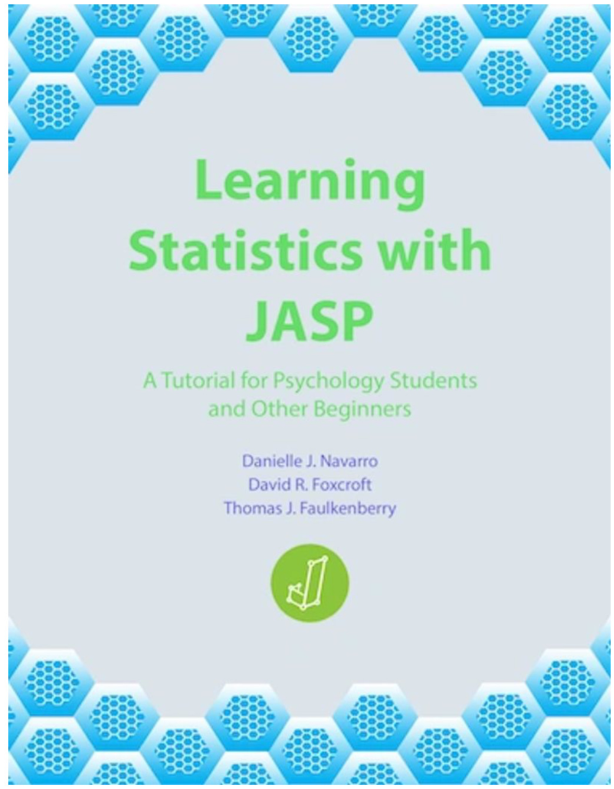
  Faulkenberry, Ly, & Wagenmakers (2020). Bayesian inference in numerical cognition: A tutorial using JASP. https://psyarxiv.org/vg9pw

* dataset: https://git.io/JeXui
  - ↳ "regression example (csv)"

Before you go — a couple of shameless plugs!

1) our FREE statistics textbook!

www.learnstatswithjasp.com



2) a free online Bayes factor calculator!

https://tomfaulkenberry.shinyapps.io/anovaBFcalc

5:50 AM Thu Jun 4 99%
tomfaulkenberry.shinyapps.io

BF calculator for single-factor ANOVA summaries

F-statistic: 5.3

df1: 3

df2: 20

Design:
- (•) Between-subjects
- ( ) Repeated-measures

Prior probability of null: 0.5

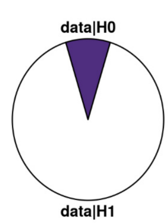
Designed by Tom Faulkenberry based on methods described here and here

For source code, vist my Github page

Model definitions:

$\mathcal{H}_0$: all condition means are equal

$\mathcal{H}_1$: not all condition means are equal



Bayes factors:

The Bayes factor for the null is 0.11
The Bayes factor for the alternative is 9.52
The observed data is approximately 9.52 times more likely under the alternative than the null

Posterior probabilities:

The posterior probability for the null is 0.0951
The posterior probability for the alternative is 0.9049

* for more details, see Faulkenberry (2019), Advances in Methods & Practices in Psychological Science

Take home points:

- Bayes is easy, especially with the right software.
- Bayes answers the questions you thought you were asking
- testing or estimation? No need to choose — Bayes gives you both!

More questions — contact me!

Email: faulkenberry@tarleton.edu

Twitter: @tomfaulkenberry

Web: https://tomfaulkenberry.github.io