Bayesian analysis of Linear Regression models: A workshop using JASP

Tom Faulkenberry

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

Ulster University Psychology Seminar
September 29, 2021

Outline:

- intro to Bayes
- 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.

Define hypotheses about (population) correlation ρ

$\mathcal{H}_0 : \rho = 0$, $\mathcal{H}_1 : \rho \neq 0$

↓

Collect data

frequentist ↙ ↘ **Bayesian**

Frequentist:

Compute:

$p(\text{data} \mid \mathcal{H}_0)$

"p-value"

↓

Interpretation:

If p is small, data is <u>rare</u> under $\mathcal{H}_0$, so we reject $\mathcal{H}_0$ in favor of $\mathcal{H}_1$.

Bayesian:

Compute:

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

"Bayes factor"

↓

Interpretation:

if $BF_{01} > 1$, data more likely under $\mathcal{H}_0$.

if $BF_{01} < 1$, data more likely under $\mathcal{H}_1$.

p-value = $p(\text{data} \mid \mathcal{H}_0)$

1) only considers fit of $\mathcal{H}_0$ as a potential model for data

2) ignores fit of $\mathcal{H}_1$

Thus, "support" for $\mathcal{H}_1$ is only indirect

Bayes factor = $\dfrac{p(\text{data} \mid \mathcal{H}_0)}{p(\text{data} \mid \mathcal{H}_1)}$

1) considers relative adequacy of both models as predictors of data.

2) can directly index support for either $\mathcal{H}_0$ or $\mathcal{H}_1$.

Ex: $BF_{01} = 8 \longrightarrow$ "The observed data are 8 times more likely under $\mathcal{H}_0$ than $\mathcal{H}_1$."

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 $\mathcal{H}$:

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

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

for <u>two</u> models:

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

$$\text{posterior odds} = \text{prior odds} \times \text{Bayes factor}$$

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:

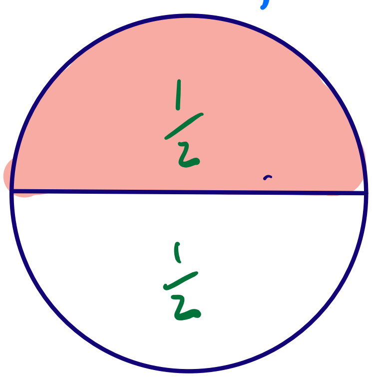
$$\boxed{p(\mathcal{H}_0 \mid \text{data}) = \frac{BF_{01} \cdot p(\mathcal{H}_0)}{BF_{01} \cdot p(\mathcal{H}_0) + p(\mathcal{H}_1)}}$$

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

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

Example:

$p(\mathcal{H}_0)$



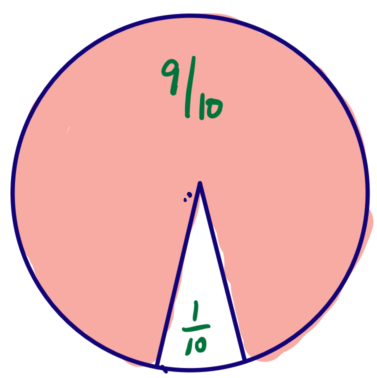
$p(\mathcal{H}_1)$

observe data

$\xrightarrow{\hspace{2cm}}$

$BF_{01} = 9$

$p(\mathcal{H}_0 \mid data)$



$p(\mathcal{H}_1 \mid data)$

Prior odds = 1:1

Posterior odds = 9:1

$$\boxed{\begin{aligned} {}^{*}\, p(\mathcal{H}_0 \mid data) &= \frac{BF_{01}}{BF_{01}+1} \\ &= \frac{9}{9+1} \\ &= 0.9 \end{aligned}}$$

② <u>priors on parameters</u> within a given model.

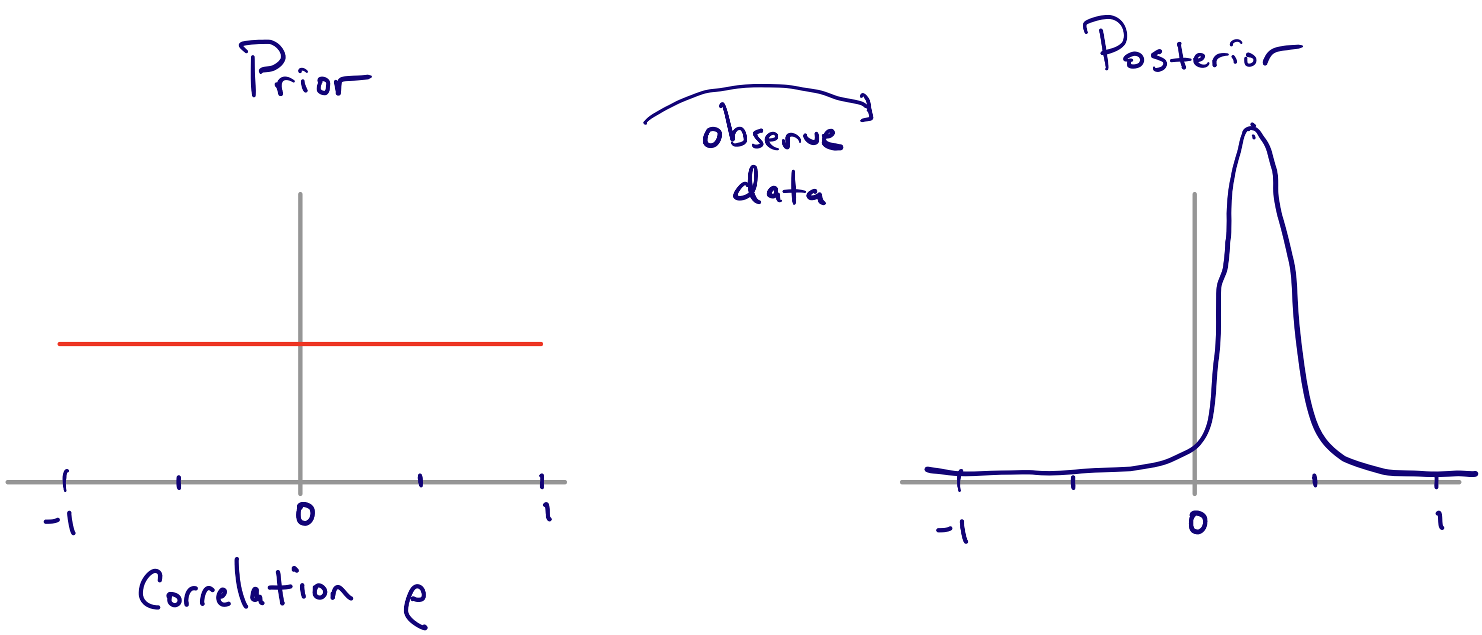
- model definitions : $\mathcal{H}_0 : \rho = 0$

  $\mathcal{H}_1 : \rho \neq 0$ ← 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 $(-1, 1)$



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,

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

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

- "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 $p(\mathcal{H}_1 \mid \text{data}) = 0.93$."

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

Example: does synchronous attendance matter in hybrid courses?

- see my JASP blog post
- for 33 students in my Fall 2020 statistics course I recorded:
  * final course grade (max 100)
  * mode of attendance (0 = asychronous, 1 = synchronous)
  * average standardized viewing time for recorded lectures (in minutes - max 75)

Data available at https://osf.io/yf2sb

Let's perform a Bayesian linear regression

* Dependent variable = "grade"
* Covariates = "avgView"
  "sync"

Note: under "Advanced options", choose Uniform model prior

Output from JASP :

1. Model comparison : gives prior / posterior model probabilities, sorted from best fit to worst fit.

Model Comparison – grade

Models	P(M)	P(M|data)	BF_M	BF_{10}	R^2
avgView	0.250	0.746	8.822	1.000	0.338
sync + avgView	0.250	0.220	0.847	0.295	0.338
sync	0.250	0.023	0.069	0.030	0.137
Null model	0.250	0.011	0.033	0.015	0.000

Interpretation: the model containing only average viewing time is the most probable after observing data

Note:

(1) all models are set to be equally likely, a priori.

(2) BF_M : change in model odds after observing data.

Ex: for "avgView" model

$$\text{Prior odds} = \frac{0.25}{0.25 + 0.25 + 0.25} = 0.333$$

$$\text{Posterior odds} = \frac{0.746}{0.220 + 0.023 + 0.011} = 2.937$$

$$\text{so } BF_M = \frac{2.937}{0.333} = 8.82$$

(3) BF_{10} = relative predictive adequacy against best model

* including "sync" in the model gives $BF_{10} = 0.295$

→ the data are only 0.295 times as likely if we include the effect of attendance mode.

OR: the data are $1/0.295 = 3.39$ times more likely if we exclude the effect of attendance mode.

2. Inclusion Bayes factors: we can do Bayesian model averaging and compute Bayes factors for each predictor

Posterior Summary

Posterior Summaries of Coefficients

								95% Credible Interval	
Coefficient	P(incl)	P(excl)	P(incl|data)	P(excl|data)	$BF_{inclusion}$	Mean	SD	Lower	Upper
Intercept	1.000	0.000	1.000	0.000	1.000	78.000	2.621	72.245	83.267
sync	0.500	0.500	0.243	0.757	0.321	0.337	3.988	−8.542	12.160
avgView	0.500	0.500	0.966	0.034	28.817	0.394	0.138	0.000	0.616

$$\text{Inclusion Bayes factor} = \frac{\text{posterior odds of including predictor}}{\text{prior odds of including predictor}}$$

Ex: BF_{incl} for "avgView"

$$\text{Prior odds} = \frac{0.250 + 0.250}{0.250 + 0.250} = 1$$

$$\text{Posterior odds} = \frac{0.746 + 0.220}{0.023 + 0.011} = \frac{0.966}{0.034} = 28.41$$

Interpretation: "the observed data are 28.41 times more likely under a model which contains average viewing time as a predictor"

3. Estimates: whereas the first half of table gives us model comparison, the second half gives us (model averaged) estimates of the effects

Ex: coefficient of avgView = 0.394

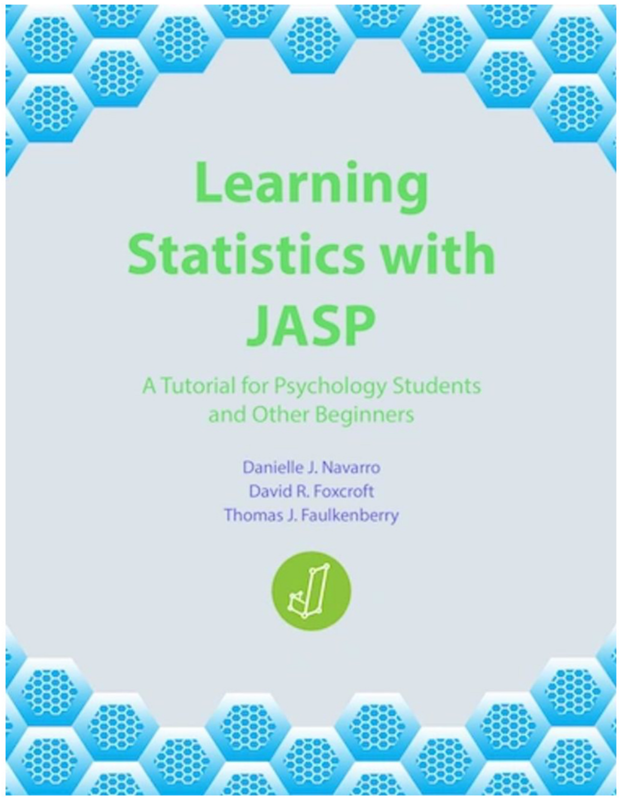
* can interpret this as: "each additional minute of recorded lecture viewing time increased the predicted grade by 0.394 points, 95% credible interval = (0.133, 0.695)"

Note: this estimate is model-averaged, meaning that it also takes into account the small probability (0.0335) that avgView is not a predictor of grades.

* see "Marginal posterior distributions" plot in JASP

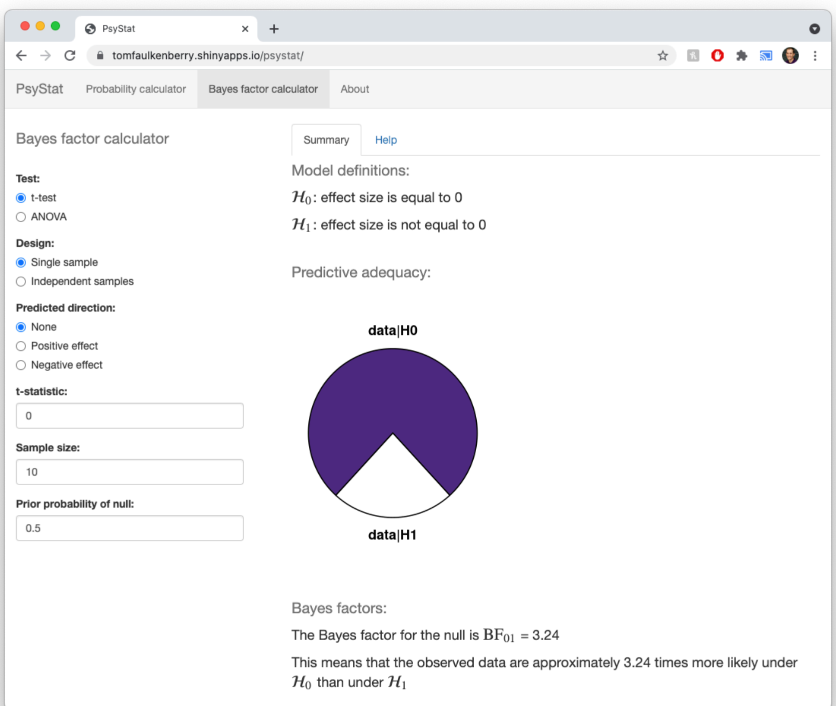
Before you go — a couple of shameless plugs!

1) our FREE statistics textbook!



2) a free online Bayes factor calculator!

https://tomfaulkenberry.shinyapps.io/psystat



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