

A simple method for Bayesian model evaluation with ANOVA designs

Thomas J. Faulkenberry

Tarleton State University

Psychological Sciences Day 2019

Consider the test scores from students in three different treatment conditions:

- Treatment 1 - read and reread
- Treatment 2 - read, then answer prepared questions
- Treatment 3 - read, then create and answer questions

Treatment 1	Treatment 2	Treatment 3
2	5	8
3	9	6
8	10	12
6	13	11
5	8	11
6	9	12
$M = 5$	$M = 9$	$M = 10$

Typical question – are there differences among these condition means?

Standard approach - analysis of variance (ANOVA)

- model $Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$
- assume “null hypothesis” $\mathcal{H}_0 : \alpha_j = 0$
- compute probability of observing data Y_{ij} under \mathcal{H}_0
- if data is *rare* under \mathcal{H}_0 , reject \mathcal{H}_0

ANOVA computations

variance source	SS	df	MS	F
between treatments				
residual				
total				

ANOVA computations

variance source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
between treatments				
residual				
total	172			

$$\begin{aligned}SS_{\text{total}} &= \sum Y^2 - \frac{(\sum Y)^2}{N} \\&= 1324 - \frac{144^2}{18} \\&= 172\end{aligned}$$

ANOVA computations

variance source	SS	df	MS	F
between treatments	84			
residual				
total	172			

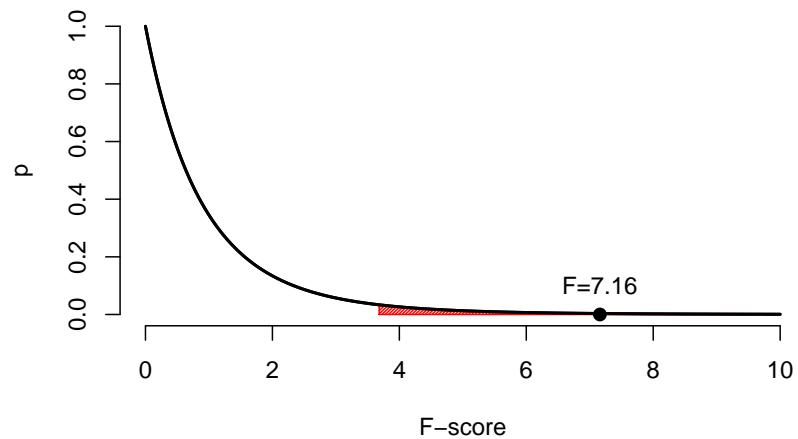
$$\begin{aligned}SS_{\text{bet tmts}} &= n \sum_{j=1}^3 (\bar{Y}_j - \bar{Y})^2 \\&= 6 \left[(5 - 8)^2 + (9 - 8)^2 + (10 - 8)^2 \right] \\&= 84\end{aligned}$$

ANOVA computations

variance source	SS	df	MS	F
between treatments	84	2	42	7.16
residual	88	15	5.87	
total	172	17		

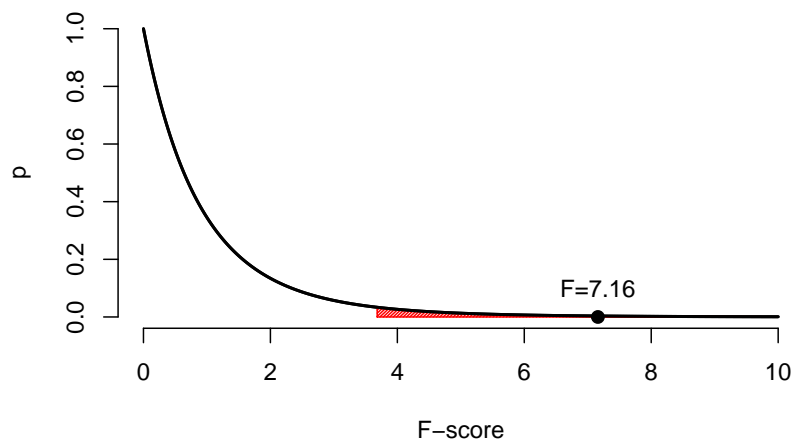
ANOVA computations

source	SS	df	MS	F
between treatments	84	2	42	7.16
within treatments	88	15	5.87	
total	172	17		



ANOVA computations

source	SS	df	MS	F
between treatments	84	2	42	7.16
within treatments	88	15	5.87	
total	172	17		



Since our data Y_{ij} is rare under \mathcal{H}_0 ($p = 0.007$), we reject \mathcal{H}_0 as an implausible model restriction.

What does this small p -value tell us?

Unfortunately, not much!

Unfortunately, not much!

It is a common misconception that the p -value indexes the probability of the null hypothesis \mathcal{H}_0 being true (i.e., $p(\mathcal{H}_0 \mid \text{data})$).

Unfortunately, not much!

It is a common misconception that the p -value indexes the probability of the null hypothesis \mathcal{H}_0 being true (i.e., $p(\mathcal{H}_0 \mid \text{data})$).

- but $p = p(\text{data} \mid \mathcal{H}_0)$
- in general, this is not equal to $p(\mathcal{H}_0 \mid \text{data})$
- to get $p(\mathcal{H}_0 \mid \text{data})$, we need Bayes' Theorem

Bayes' Theorem

$$\underbrace{p(\mathcal{H} \mid \text{data})}_{\text{Posterior beliefs about hypothesis}} = \underbrace{p(\mathcal{H})}_{\text{Prior beliefs about hypothesis}} \times \underbrace{\frac{p(\text{data} \mid \mathcal{H})}{p(\text{data})}}_{\text{predictive updating factor}}$$

Bayes' Theorem

$$\underbrace{p(\mathcal{H} \mid \text{data})}_{\text{Posterior beliefs about hypothesis}} = \underbrace{p(\mathcal{H})}_{\text{Prior beliefs about hypothesis}} \times \underbrace{\frac{p(\text{data} \mid \mathcal{H})}{p(\text{data})}}_{\text{predictive updating factor}}$$

If we want to compare \mathcal{H}_1 to \mathcal{H}_0 , we can take ratios:

$$\underbrace{\frac{p(\mathcal{H}_1 \mid \text{data})}{p(\mathcal{H}_0 \mid \text{data})}}_{\text{posterior odds}} = \underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\text{prior odds}} \times \underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\text{predictive updating factor}}$$

The predictive updating factor

$$B_{10} = \frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}$$

tells us how much better \mathcal{H}_1 predicts our observed data than \mathcal{H}_0 .

This ratio is called the **Bayes factor**

The predictive updating factor

$$B_{10} = \frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}$$

tells us how much better \mathcal{H}_1 predicts our observed data than \mathcal{H}_0 .

This ratio is called the **Bayes factor**

Computing Bayes factors is HARD¹

¹to see why (and how), see Faulkenberry, T. J. (2019). A tutorial on generalizing the default Bayesian t-test via posterior sampling and encompassing priors. *Communications for Statistical Applications and Methods*, 26(2), 1-22.

With some assumptions, we can compute Bayes factors for ANOVA designs using a method due originally to Kass and Raftery (1995) (but also see Masson, 2011, and Faulkenberry, 2018)

Basic idea:

1. set up two models: \mathcal{H}_0 and \mathcal{H}_1
2. compute BIC (Bayesian information criterion) for each model:

$$BIC = N \ln(SS_{\text{residual}}) + k \ln(N)$$

where

- N =total number of independent observations
- k =number of parameters in the model
- SS_{residual} = variance NOT explained by the model

3. compute Bayes factor as $e^{\frac{\Delta BIC}{2}}$

Example

source	SS	df	MS	F
bet tmts	84	2	42	7.16
residual	88	15	5.87	
total	172	17		

We'll set up our two models:

Null model: $\mathcal{H}_0 : \mu_1 = \mu_2 = \mu_3$

- this model has $k = 1$ parameter (the data is explained by a SINGLE mean)
- $SS_{\text{residual}} = 172$ (the model has only one mean, so **all** variance is left unexplained)

Example

source	SS	df	MS	F
bet tmts	84	2	42	7.16
residual	88	15	5.87	
total	172	17		

Null model: $\mathcal{H}_0 : \mu_1 = \mu_2 = \mu_3$

$$\begin{aligned} BIC_0 &= N \ln(SS_{\text{residual}}) + k \ln(N) \\ &= 18 \ln(172) + 1 \cdot \ln(18) \\ &= 95.55 \end{aligned}$$

Example

source	SS	df	MS	F
bet tmts	84	2	42	7.16
residual	88	15	5.87	
total	172	17		

Alternative model: $\mathcal{H}_1 : \mu_1 \neq \mu_2 \neq \mu_3$

- this model has $k = 3$ parameters (the data is explained by THREE means)
- $SS_{\text{residual}} = 88$ (the model accounts for variance between treatments with the three means, so SS_{residual} is left unexplained)

Example

source	SS	df	MS	F
bet tmts	84	2	42	7.16
residual	88	15	5.87	
total	172	17		

Alternative model: $\mathcal{H}_1 : \mu_1 \neq \mu_2 \neq \mu_3$

$$\begin{aligned} BIC_1 &= N \ln(SS_{\text{residual}}) + k \ln(N) \\ &= 18 \ln(88) + 3 \cdot \ln(88) \\ &= 89.26 \end{aligned}$$

Thus,

$$\begin{aligned} B_{10} &= e^{\frac{\Delta BIC}{2}} \\ &= e^{\frac{95.55 - 89.26}{2}} \\ &= 22.87 \end{aligned}$$

Thus,

$$\begin{aligned} B_{10} &= e^{\frac{\Delta BIC}{2}} \\ &= e^{\frac{95.55 - 89.26}{2}} \\ &= 22.87 \end{aligned}$$

This means that the data are approximately 23 times more likely under \mathcal{H}_1 than \mathcal{H}_0

What about $p(\mathcal{H}_1 \mid \text{data})$?

What about $p(\mathcal{H}_1 \mid \text{data})$?

It is easy to show

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

What about $p(\mathcal{H}_1 \mid \text{data})$?

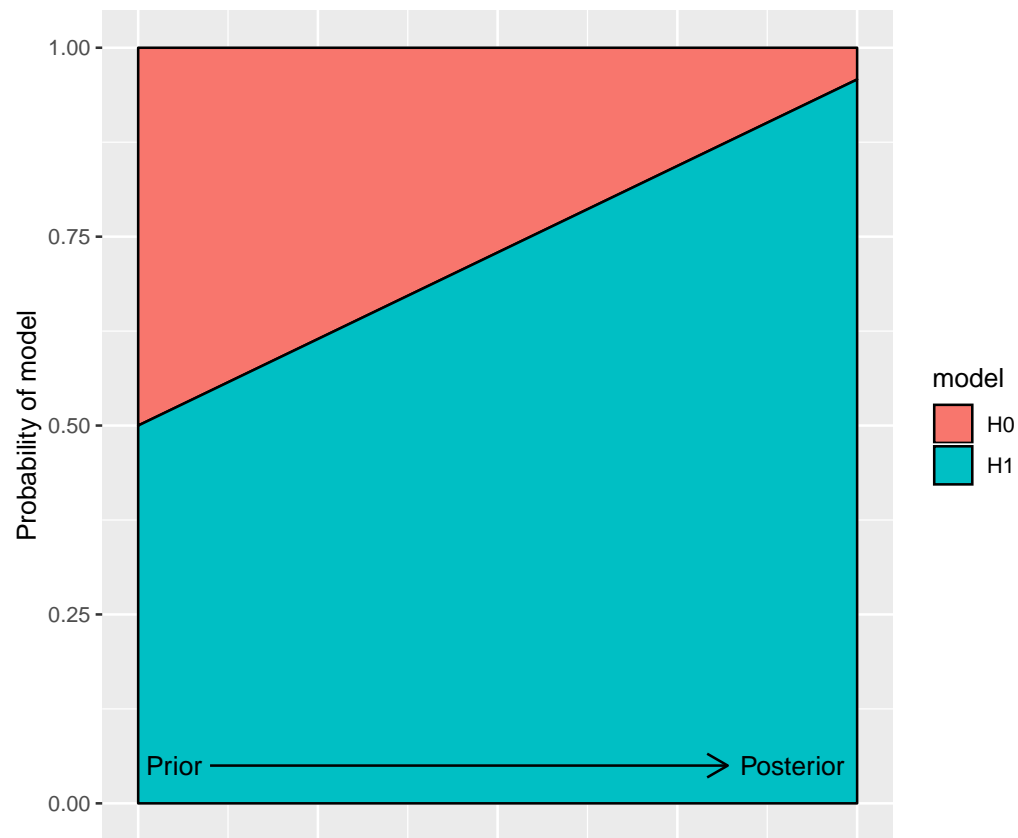
It is easy to show

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

Thus, we have

$$\begin{aligned} p(\mathcal{H}_1 \mid \text{data}) &= \frac{22.87}{1 + 22.87} \\ &= 0.958 \end{aligned}$$

Visualizing “flow of model belief”:



Thank you!

- Thanks to Tarleton Office of Research and Innovation for funding!
- slides available at github.com/tomfaulkenberry/talks
- more details in Faulkenberry, T. J. (2019). Computing Bayes factors to measure evidence from experiments: An extension of the BIC approximation. *Biometrical Letters*, 55(1), 31-43.
- Twitter: @tomfaulkenberry
- Email: faulkenberry@tarleton.edu