PSYC 5301 - Lecture 9

Thomas J. Faulkenberry, Ph.D.

April 23, 2019

Review from last time

Recall that by Bayes Theorem, we have

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

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**

We can compute Bayes factors for ANOVA models using the BIC:

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

where

- N=total number of independent observations
- k=number of parameters in the model
- SS_{residual} = variance NOT explained by the model
- Note: smaller BIC = better model fit

Steps:

- set up two models: $\mathcal{H}_0$ and $\mathcal{H}_1$
- compute BIC (Bayesian information criterion) for each model
- compute Bayes factor as $e^{\frac{\Delta BIC}{2}}$

Example

(this is from HW 8, #4)

Treatment 1	Treatment 2	Treatment 3
1	5	7
2	2	3
0	1	6
1	2	4

first, we model as ANOVA:

source	SS	df	MS	F
bet tmts	32.67	2	16.37	7.01
within tmts	21	9	2.33	
total	53.67	11		

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)
- $SS_{\text{residual}} = 53.67$ (the model has only one mean, so all variance is left unexplained)

$$
\begin{aligned}
BIC &= N\ln(SS_{\text{residual}}) + k\ln(N) \\
&= 12\ln(53.67) + 1\cdot\ln(12) \\
&= 50.28
\end{aligned}
$$

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)
- $SS_{\text{residual}} = 21$ (the model accounts for variance between treatments with the three means – SS_{within} is left unexplained)

$$
\begin{aligned}
BIC &= N\ln(SS_{\text{residual}}) + k\ln(N) \\
&= 12\ln(21) + 3 \cdot \ln(12) \\
&= 43.99
\end{aligned}
$$

Thus,

$$B_{10} = e^{\frac{\Delta BIC}{2}} = e^{\frac{50.28-43.99}{2}} = 23.22$$

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

Repeated measures designs

The same ideas will extend to work with repeated measures designs. The only difference is that we need to think carefully about:

- the number of *independent* observations
- residual SS

Example

Consider the following example from Exam 1, which asks about task performance as a function of computer desk layout:

Subject	Layout 1	Layout 2	Layout 3
#1	6	2	4
#2	8	6	7
#3	3	6	9
#4	3	2	4

Let's work through the ANOVA model, since it has been a while:

Step 1 - compute condition means AND subject means:

Subject	Layout 1	Layout 2	Layout 3	M
#1	6	2	4	4
#2	8	6	7	7
#3	3	6	9	6
#4	3	2	4	3
M	5	4	6	5

Remember that once we find SS, we remove subject variability and partition what's left:

$$
\begin{aligned}
SS_{\text{total}} &= \sum X^2 - \frac{(\sum X)^2}{N} \\
&= 360 - \frac{60^2}{12} \\
&= 60
\end{aligned}
$$

$$
\begin{aligned}
SS_{\text{bet subj}} &= n \sum_{i=1}^{4} (\overline{X}_{\text{subj } i} - \overline{X})^2 \\
&= 3\Big[(4-5)^2 + (7-5)^2 + (6-5)^2 + (3-5)^2\Big] \\
&= 30
\end{aligned}
$$

$$
\begin{aligned}
SS_{\text{bet tmts}} &= n\sum_{j=1}^{3}(\overline{X}_{\text{tmt }j} - \overline{X})^2 \\
&= 4\Big[(5-5)^2 + (4-5)^2 + (6-5)^2\Big] \\
&= 8
\end{aligned}
$$

Thus, our ANOVA table is as follows:

Source	SS	df	MS	F
bet tmts	8	2	4	1.09
residual	22	6	3.67	
subject	30	3	10	
total	60	11		

BIC computations

We'll set up our two models:

Null model: $\mathcal{H}_0 : \alpha_1 = \alpha_2 = \alpha_3$

- this model has $k = 1$ parameter (the data is explained by a SINGLE treatment effect)
- $SS_{\text{residual}} = 30$ (what is left after removing subject variance)
- $N = 8$ independent observations (for each of 4 subjects, there are $3 - 1 = 2$ independent observations)
- Note: general formula: $N = s(c - 1)$, where $s =$ number of subjects and $c =$ number of conditions

$$\begin{aligned} BIC &= N \ln(SS_{\text{residual}}) + k \ln(N) \\ &= 8 \ln(30) + 1 \cdot \ln(8) \\ &= 29.29 \end{aligned}$$

Alternative model: $\mathcal{H}_1 : \alpha_1 \neq \alpha_2 \neq \alpha_3$

- this model has $k = 3$ parameters (the data is explained by THREE treatment effects)
- $SS_{\text{residual}} = 22$

$$\begin{aligned} BIC &= N \ln(SS_{\text{residual}}) + k \ln(N) \\ &= 8 \ln(22) + 3 \cdot \ln(8) \\ &= 30.97 \end{aligned}$$

Thus,

$$B_{01} = e^{\frac{\Delta BIC}{2}} = e^{\frac{30.97-29.81}{2}} = 1.79$$

meaning that the data are approximately 2 times more likely under $\mathcal{H}_0$ than $\mathcal{H}_1$

Some lessons

The previous homework questions give us some lessons about p-values:

1. p-values are uniformly distributed under the null. The implication is that a single p-value gives us no information about the likelihood of any model
2. optional stopping inflates Type I error rate.
3. $p = p(\text{data} \mid \mathcal{H}_0)$. This is NOT equal to $p(\mathcal{H}_0 \mid \text{data})$

However, with some cleverness, we can actually calculate $p(\mathcal{H}_0 \mid \text{data})$. All we need is Bayes theorem:

Posterior model probabilities

Recall from Bayes theorem:

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

Let's assume $p(\mathcal{H}_0) = p(\mathcal{H}_1)$ (that is, $\mathcal{H}_0$ and $\mathcal{H}_1$ are equally likely, *a priori*).

Then the previous equation reduces to

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

Then we have:

$$\begin{aligned} p(\mathcal{H}_0 \mid \text{data}) &= B_{01} \cdot p(\mathcal{H}_1 \mid \text{data}) \\ &= B_{01}\Big[1 - p(\mathcal{H}_0 \mid \text{data})\Big] \\ &= B_{01} - B_{01} \cdot p(\mathcal{H}_0 \mid \text{data}) \end{aligned}$$

Let's solve this equation for $p(\mathcal{H}_0 \mid \text{data})$:

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

which implies by factoring:

$$p(\mathcal{H}_0 \mid \text{data}) \Big[1 + B_{01} \Big] = B_{01}$$

or equivalently

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

Note: by the same reasoning, we can prove

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

Let's compute these for the examples we've done tonight:

Example 1: $B_{10} = 23.22$

This example that $\mathcal{H}_1$ was a better fit. Thus,

$$\begin{aligned}
p(\mathcal{H}_1 \mid \text{data}) &= \frac{B_{10}}{1 + B_{10}} \\
&= \frac{23.22}{1 + 23.22} \\
&= 0.959
\end{aligned}$$

Example 2: $B_{01} = 1.79$

This example that $\mathcal{H}_0$ was a better fit. Thus,

$$\begin{aligned}
p(\mathcal{H}_0 \mid \text{data}) &= \frac{B_{01}}{1 + B_{01}} \\
&= \frac{1.79}{1 + 1.79} \\
&= 0.642
\end{aligned}$$