Lecture 2 – Bayesian Hypothesis Testing.

Suppose we are interested in assessing effectiveness of some mathematics instruction program.
After implementing the program, we measure mathematical ability using the Scale for Advancing Mathematical Ability (SAMA), a national assessment with a known mean score of 50.

Our sample (who had the training) had a mean of 54.4 and a standard deviation of 10.

Did the training work?

Basic statistical framework:

(1) define two competing models ("hypotheses") that could possibly generate our observed data.

- one where the training worked (H_1)
- one where it didn't (H_0)

(2) assess the fit of these models against the observed data.

How to proceed?

Define hypotheses about (population) effect size δ "delta"

$\mathcal{H}_0 : \delta = 0$, $\mathcal{H}_1 : \delta > 0$

Collect data

Compute:

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

"p-value"

Interpretation:

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

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-values versus Bayes Factors

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: $BF_{10} = 9 \rightarrow$ "The observed data are 9 times more likely under $\mathcal{H}_1$ than $\mathcal{H}_0$".

$BF_{01} = 9 \rightarrow$ "The observed data are 9 times more likely under $\mathcal{H}_0$ than $\mathcal{H}_1$".

Interpreting Bayes factors:

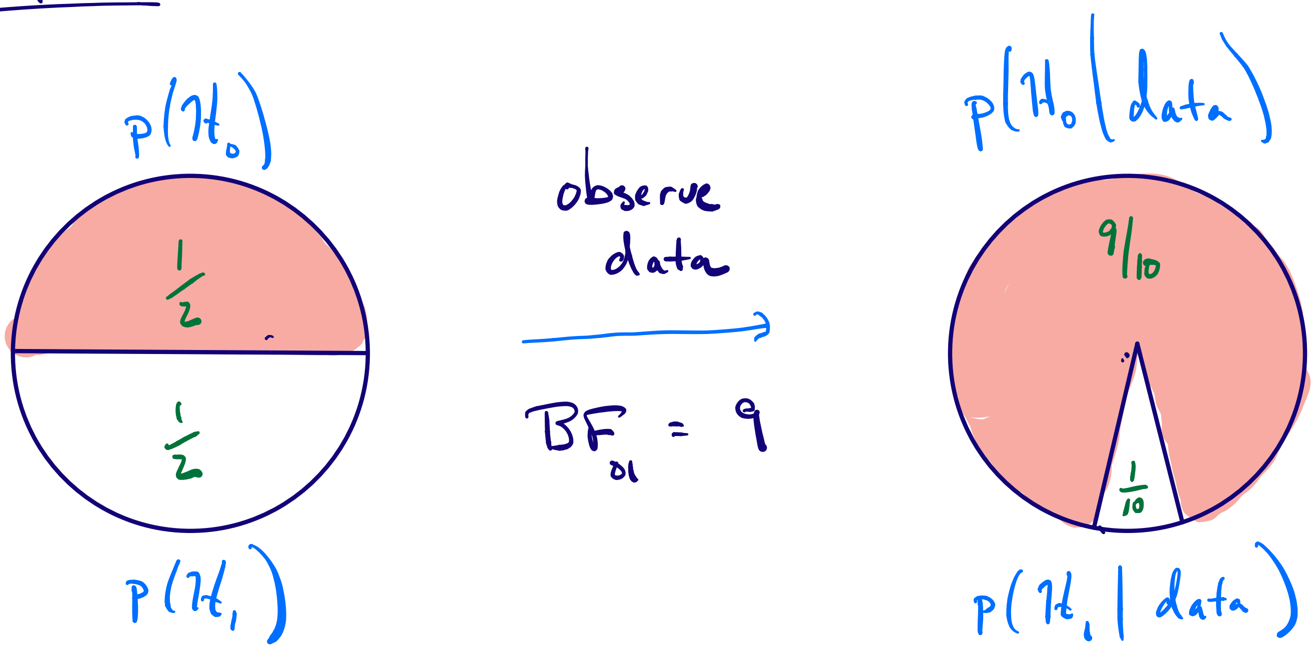
(1) relative predictive adequacy of models

* $BF_{01} = 9 \longrightarrow$ "The observed data are 9 times more likely under $\mathcal{H}_0$ than $\mathcal{H}_1$".

(2) updating factor

* $BF_{01} = 9 \longrightarrow$ "After observing data, my prior odds for $\mathcal{H}_0$ over $\mathcal{H}_1$ have been increased by a factor of 9"

Example:



Prior odds = 1:1

Posterior odds = 9:1

Let's continue our working example. Recall that we tested $N = 65$ participants and observed a sample mean of 54.4 with $\hat{\sigma} = 10$.

Step 1 – convert our observed data to test statistic

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{N}} = \frac{54.4 - 50}{10/\sqrt{65}} = \frac{4.4}{1.24}$$

$$= 3.55$$

Step 2 – convert t-score to Bayes factor

* https://tomfaulkenberry.shinyapps.io/bayesFactorCalc

* or, use "Summary Stats Module" in JASP.

Elements to report:

- define H_0, H_1

"Under the null hypothesis we expect an effect size of 0. Thus, we define $H_0: \delta = 0$. Under the alternative, we expect a positive effect. That is, $H_1: \delta > 0$"

- report and interpret Bayes factor

  "We found a Bayes factor of $BF_{10} = 69.4$, which means that the observed data are approximately 70 times more likely under $\mathcal{H}_1$ than $\mathcal{H}_0$."

- calculate and report posterior model probability for preferred model.

- "Assuming prior odds of 1-1 for $\mathcal{H}_1$ and $\mathcal{H}_0$, our observed data updated these odds to 69.4-to-1 in favor of $\mathcal{H}_1$. This is equivalent to a posterior model probability of $p(\mathcal{H}_1 \mid \text{data}) = 0.986$"

What about "statistical significance"?

- we usually don't use this phrase in Bayesian context.
- instead, we speak of <u>model evidence</u> from data

Jeffreys (1961):

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

* these are only guidelines!

"In summary, these data provide very strong evidence in favor of $\mathcal{H}_1$; that is, the training program had a positive effect on maths performance."