Module 4: Introduction to Bayesian hypothesis testing for two independent samples

## Learning Objectives

By the end of Module 4, you should be able to:

1. Conduct a Bayesian *t* test using the R package brms
2. Understand and interpret Bayes factors

## Getting Started

If you’ve never installed the R package brms, you will need to install it by opening a new R script in RStudio, copying and pasting the following code, and running it:

install.packages(“brms”, dependencies = TRUE)

Once you’ve installed the package, you need to call it from the library by opening a new R Script, copying and pasting the following code into your R script and running it before you’ll be able to use it.

library("brms")

## The Example

Imagine a nursing home has hired you to conduct some statistical analyses. In your literature review, you came across a few articles documenting the increased risk of cognitive impairment for individuals with low education (see Stern et al. 1994; Letenneur et al. 1999; Le Carret et al. 2003). **Given this information, the nursing home would like to explore the relationship between the education levels and cognitive ability of their patients.**

In Module 2, the nursing home wanted to know **whether or not** there was sufficient evidence to suggest a significant difference between the mean of cognitive ability for patients with low levels of education and the mean of cognitive ability for patients with high levels of education. Now, the nursing home would like to know the **degree to which** the data favor one hypothesis over the other. We will use a statistical tool available in the Bayesian framework, called the Bayes factor, to obtain this knowledge.

## The Data

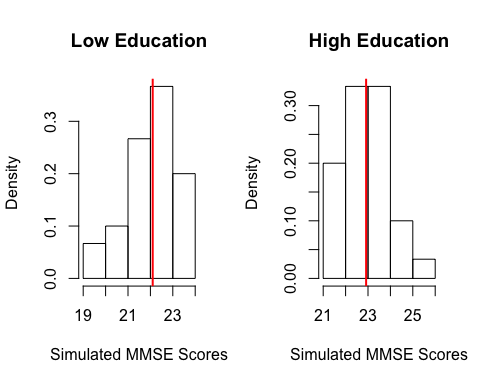
We will use the simulated data from Module 2. For a refresher on the details of these data, see Module 2. To simulate the data, copy and paste the following code into your R script and run it.

set.seed(1)  
nrep <- 30 # Tell the simulation that there will be 30 observations per group  
ngroup <- 2 # Tell the simulation that there will be two groups (low and high education)  
b0 <- 22 # Tell the simulation that the low education group's MMSE scores will be simulated from a distribution with a mean of 22  
b1 <- 0.75 # Tell the simulation that the high education group's MMSE scores will be simulated from a distribution with a mean of 22.5 (b0 + b1)  
educ <- rep(c("0", "1"), each = nrep) # Simulate the predictor variable  
err <- rnorm(n = ngroup\*nrep, mean = 0, sd = 1.25) # Simulate 60 residuals from a distribution with a mean of 0 and standard deviation of 1.25  
MMSE <- b0 + b1\*(educ == "1") + err # Simulate the outcome variable  
simdata <- data.frame(MMSE, educ) # Create a dataframe containing the simulated variables MMSE and educ

## Visualizing the Simulated Scores

Copy and paste the following code into your R script and run it to produce a histogram of the simulated MMSE scores for each group.

par(mfrow = c(1, 2)) # Create a plotting window that will display the histograms side-by-side  
hist(simdata$MMSE[educ=="0"], main = 'Low Education', breaks = 5, prob = T, xlab = "Simulated MMSE Scores")   
abline(v = mean(simdata$MMSE[educ=="0"]), col = "red", lwd=2) # Overlay a red line at the sample mean for group "0"  
hist(simdata$MMSE[educ=="1"], main = 'High Education', breaks = 5, prob = T, xlab = "Simulated MMSE Scores")   
abline(v = mean(simdata$MMSE[educ=="1"]), col = "red", lwd=2) # Overlay a red line at the sample mean for group "1"



## Testing the Assumption of Normality

In the following analysis, we will fit a statistical model that assumes the outcome variable is normally distributed in the population. Although the histograms above suggest that the data could have been sampled from a normal population distribution, let’s use the Shapiro-Wilk test of normality to evaluate this. Copy and paste the following code into your R script and run it.

shapiro.test(simdata$MMSE[educ=="0"]) # Shapiro-Wilk test of normality for group "0"  
shapiro.test(simdata$MMSE[educ=="1"]) # Shapiro-Wilk test of normality for group "1"

The results of the Shapiro-Wilk tests are insignificant (*p* > .05), so we fail to reject the null hypothesis. In other words, there is not enough evidence to suggest that the simulated MMSE scores are non-normally distributed.

## The Analysis

We will conduct a **two-tailed two independent samples Bayesian *t* test** to compare the population means of the low and high education groups ( and , respectively). Because the *t* test is a special case of the linear model, we specify a linear model with a binary predictor like in Module 2.

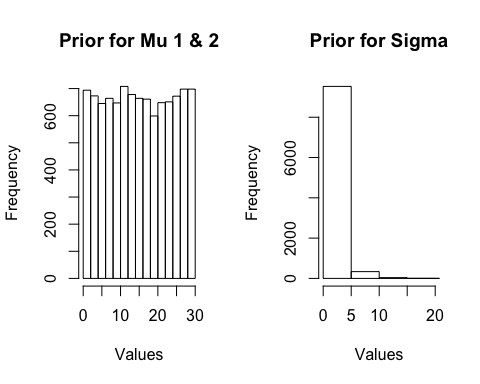
Since this is a Bayesian analysis, we will use the brm() function from the R package brms to fit the linear model to the simulated MMSE scores. Output from this linear model will include:

1. an estimate of the population mean ,
2. the standard deviation of the estimate of ,
3. an estimate of the population mean ,
4. the standard deviation of the estimate of ,
5. an estimate of the population standard deviation (referred to as the residual standard error), and
6. the standard deviation of the estimate of .

The parameters estimated by this model are , , and . (In contrast, the parameters estimated by the model specified in Module 2 were , , and , where .) As part of the model, we need to specify **a prior distribution for each model parameter**. For and , we will specify uniform priors ranging from 0-30 because we know that MMSE scores cannot fall below 0 or above 30. For , we will specify a half student-*t* prior with a lower bound at 0, 3 degrees of freedom, a mean of 0, and a standard deviation of 2.5 (this is the default prior for in brms).

To visualize scores sampled from these priors, copy and paste the following code into your R script and run it.

par(mfrow = c(1, 2))  
  
# Prior for the population means  
set.seed(1)  
bound\_p <- runif(n = 10000, min = 0, max = 30) # Simulate 10,000 observations from a uniform distribution from 0 to 30  
hist(bound\_p, main = "Prior for Mu 1 and 2", xlab = "Values") # Visualize the simulated values  
  
# Prior for sigma  
set.seed(1)  
tdist <- rt(n = 20000, df = 3)\*sqrt(((2.5^2)\*(3-2))/3) # Simulate 20,000 observations from a student-t distribution with 3 degrees of freedom and scale the observations so that they have a standard deviation of 2.5  
sd(tdist) # Verify that the standard deviation of the simulated scores is roughly equal to 2.5  
halft\_p <- tdist[tdist >= 0] # Remove any simulated values less than 0 so that the values are strictly positive   
hist(halft\_p, main = "Prior for Sigma", xlab = "Values", xlim = c(0, 20), breaks = 12) # Visualize the simulated values



To conduct our Bayesian analysis, implementing the priors described above, copy and paste the following code into your R script and run it.

prior3 <- c(prior\_string("uniform(0, 30)", class = "b", lb = 0, ub = 30), # Specify the uniform prior from 0 to 30 for the population means  
 prior("student\_t(3, 0, 2.5)", class = "sigma")) # Specify the half student-t distribution with 3 degrees of freedom, a mean of 0, and a standard deviation of 2.5 for the population standard deviation  
 # brms automatically restricts class = "sigma" parameters (standard deviation parameters) to be non-negative, so we don't need to specify the argument lb = 0  
set.seed(1) # Ensure reproducible random sampling  
mod5 <- brm(MMSE ~ 0 + educ,   
 prior = prior3,  
 sample\_prior = TRUE,  
 data = simdata) # Fit a linear model to the simulated MMSE scores using the brm() function from the R package brms  
posterior\_summary(mod5) # Examine the model output

## Interpreting the Model Results

Remember, rather than attempting to obtain the mean of the actual posterior distribution, brms uses a variety of sampling algorithms to draw random samples of parameter values from each parameter’s posterior distribution. In this way, the respective means of the sampled values can easily be calculated to produce the Bayesian parameter estimates.

The line of code

fixef(mod5)[,1:2]

outputs the means (in the first column) and standard deviations (in the second column) of the values sampled from the respective posterior distributions of and , where and are the means of the simulated MMSE scores for groups “0” and “1”, respectively. The line of code

VarCorr(mod5)$residual$sd[,1:2]

outputs the mean (in the first column) and standard deviation (in the second column) of the values sampled from the posterior distribution of , where represents the spread of MMSE scores about their mean. Like in Module 2, this model output only contains a single estimate of (as opposed to containing estimates of both and ).

In Module 2, the main piece of model output that needed to be evaluated to address the research question was whether or not the estimate of (the estimated difference between and ) was statistically significant, as determined by the frequentist *p*-value. The *p*-value from that analysis was an estimate of the proportion of test statistics in the sampling distribution of the test statistic that were equal to or more extreme than the test statistic obtained, assuming a true null hypothesis. **Importantly, the frequentist *p*-value cannot evaluate evidence in favor of the null hypothesis, only against it (Rouder et al. 2009).**

**In the Bayesian framework, it is possible to quantify the amount of evidence from the data in support of one hypothesis over another (Wetzels et al. 2011). One Bayesian measure used to quantify this evidence is the Bayes factor, which is an odds ratio of the probability of the data under one hypothesis relative to the probability of the data under another.**

We will explain the Bayes factor and its role in Bayesian hypothesis testing in the next section.

## Bayesian Hypothesis Testing

In the context of our current analysis, the null and alternative hypotheses are:

and

.

Before observing the data , and are assigned prior probabilities and , where represents a distribution where the null value (zero) is the only probable value, and represents a distribution where values other than zero are probable (Kruschke and Liddell 2018). The ratio is referred to as the prior odds. After observing the data , the prior odds are updated to posterior odds , where represents the change from prior odds to posterior odds. **This change from prior to posterior odds is referred to as the Bayes factor (Wetzels et al. 2011; Wagenmakers et al. 2010).**

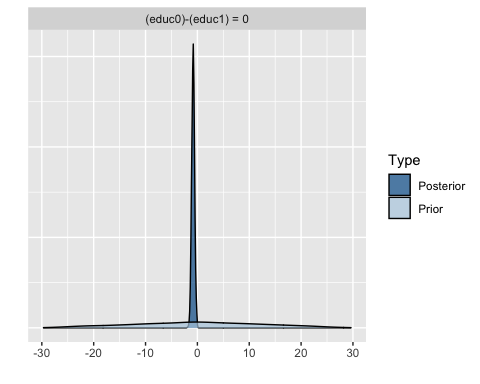
We will use the hypothesis() function from the R package brms to calculate the Bayes factor for our *t* test. Copy and paste the following code into your R script and run it to obtain the Bayes factor .

set.seed(1)  
h1 <- hypothesis(mod5, "educ0 = educ1") # Test the point hypothesis that the population means are equal  
BF01 <- h1$hypothesis$Evid.Ratio # Obtain the Bayes factor

The above code yields , which, according to Wagenmakers et al. (2010), means that the data are times more likely under than .

Now, let’s create a visual to help solidify our intuitive understanding of the Bayes factor. Copy and paste the following code into your R script and run it.

plot(h1)



The above code produced histograms for the posterior and prior distributions for . The Bayes factor is a comparison of the heights of the posterior and prior distributions at the value specified by the null hypothesis ().

According to Jeffreys (1961), Bayes factors from 1 to .33 indicate anecdotal evidence in favor of , Bayes factors from .33 to .10 indicate substantial evidence in favor of , Bayes factors from .10 to .03 indicate strong evidence in favor of , Bayes factors from .03 to .01 indicate very strong evidence in favor of , and Bayes factors <.01 indicate decisive evidence in favor of . Our Bayes factor of indicates only anecdotal evidence in favor of .

In summary, the *p*-value is a tool within the frequentist framework that can be used to gather evidence against the null hypothesis (but not in favor of it), whereas the Bayes factor is a tool within the Bayesian framework that can be used to gather evidence either in favor of the null or the alternative hypotheses (Wetzels et al. 2011). **In other words, while the *p*-value only allows one to evaluate whether or not there is enough evidence against the null hypothesis to reject it, the Bayes factor allows one to make a statement about which hypothesis the data are more likely under, and moreover, how much more likely the data are under one hypothesis than the other.** It is important that we understand the differences in interpretations between the two tools so that we arrive at proper conclusions.

## References

Gigerenzer, G., Krauss, S., and Vitouch, O. (2004), “The null ritual”, *The Sage handbook of quantitative methodology for the social sciences*, 391-408.

Jeffreys, H. (1961), *Theory of probability*, 3rd Edition, Oxford University Press.

Kruschke, J. K., and Liddell, T. M. (2018), “Bayesian data analysis for newcomers”, *Psychonomic bulletin and review*, 25(1), 155-177.

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., and Iverson, G. (2009), “Bayesian *t* tests for accepting and rejecting the null hypothesis”, *Psychonomic Bulletin and Review*, 16, 225–237.

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010), “Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method”, *Cognitive psychology*, 60(3), 158-189.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., and Wagenmakers, E. J. (2011), “Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests”, *Perspectives on Psychological Science*, 6(3), 291-298.