Module 3: Introduction to Bayesian estimation of a population mean () and standard deviation ()

## Learning Objectives

By the end of Module 3, you should be able to:

1. Estimate a mean and standard deviation using the R package brms
2. Interpret Bayesian parameter estimates
3. Interpret Bayesian standard deviations of parameter estimates
4. Interpret Bayesian credible intervals
5. Explain the difference in interpretations of a Bayesian credible interval versus a frequentist confidence interval
6. Specify prior distributions

## Getting Started

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

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

Once you’ve installed the packages, you need to call them 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 them.

library("brms")  
library("bayesplot")

## The Example

Imagine a nursing home has hired you to conduct some statistical analyses for them. One of the things the nursing home would like to know is the mean and standard deviation of cognitive ability for their patients. You will use Bayesian methods to conduct these analyses.

## The Data

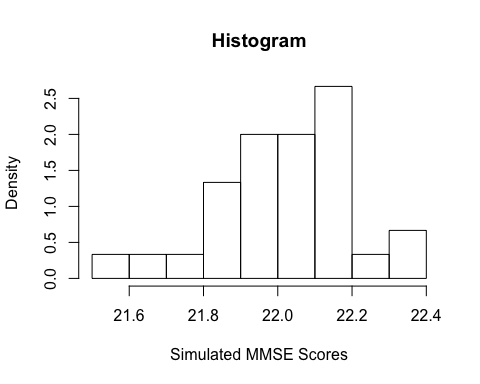
We will use the simulated MMSE scores from Module 1. For a refresher on the details of the simulation, see Module 1. To simulate the MMSE scores, copy and paste the following code into your R script and run it.

set.seed(1) # Setting the seed ensures that your random numbers match my random numbers   
 # As long as our seed values match, the chosen seed value is arbitrary  
MMSE <- rnorm(n = 30, mean = 22, sd = 0.2) # Simulate 30 observations from a normal distribution with a mean of 22 and standard deviation of 0.2  
simdata <- as.data.frame(MMSE)

## Visualizing the Simulated Scores

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

hist(simdata$MMSE, main = "Histogram", breaks = 10, prob = T, xlab = "Simulated MMSE Scores")



## 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) # Shapiro-Wilk test of normality

The results of the Shapiro-Wilk test 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

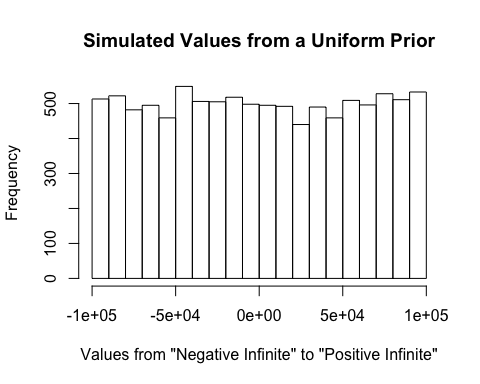
Now we present a detailed introduction to the Bayesian perspective so that you’ll be prepared to estimate and using **Bayesian estimation methods**.

In your introductory undergraduate psychology statistics course, you were probably taught statistics through the lens of the frequentist notion of probability. The frequentist perspective views the probability of an outcome as the value approached by the relative frequency of occurence of the outcome in a very long series of replications in a chance experiment. A classic example of frequentist probability is an infinite coin toss. **In this experiment, the probability of a particular outcome in the sample space (heads, tails) is defined as the value approached by the relative frequency of occurrence of the particular outcome in a very long series of coin tosses.** In contrast, the Bayesian perspective views probability as the subjective experience of uncertainty (De Finetti 1974; Van de Schoot et al. 2014). A classic example of Bayesian probability is placing a bet—for example, on a playoff game. **The better takes prior knowledge and personal judgement into account as they form their educated guess. The outcome of the game can then be used to update this knowledge, which can be incorporated by the next better.**

The Bayesian perspective conceptualizes parameters as **random variables with probability distributions** characterized by measures of central tendency and variability (Gelman et al. 2013). These probability distributions are called **prior distributions** and are specified a priori by the researcher according to their prior knowledge and their confidence in that knowledge. As an example of prior knowledge, consider our example of placing a bet on a playoff game: the prior distribution summarizes all of the better’s prior knowledge (team point differentials, three point percentages, points per possession, etc.) and personal judgement before the outcome of the game is known.

Sometimes researchers lack information; other times they have access to information gathered over years of research and replication. According to Kaplan and Depaoli (2013), both kinds of information are important to incorporate into Bayesian analyses. A lack of information is often incorporated into an analysis through specification of a uniform prior that ranges from negative to positive infinite. Uniform priors indicate that every value between negative and positive infinite is equally likely to occur. That is, they don’t provide information as to which value is most/least likely. To visualize scores sampled from a uniform prior, copy and paste the following code into your R script and run it.

set.seed(1)  
uniform\_p <- runif(n = 10000, min = -100000, max = 100000) # Simulate 10,000 observations from a uniform distribution with -100,000 as proxy for negative infinite and 100,000 as proxy for positive infinity  
hist(uniform\_p, main = "Simulated Values from a Uniform Prior", xlab = 'Values from "Negative Infinite" to "Positive Infinite"') # Visualize the simulated values



**To obtain a Bayesian parameter estimate, the parameter’s prior distribution is combined with the likelihood function (information from the data) via Bayes’ theorem to obtain the posterior distribution of the parameter upon which Bayesian statistical inference is based. A point estimate of a Bayesian parameter is the value that summarizes the center of its posterior distribution (usually the mean).**

With this in mind, let’s obtain Bayesian estimates of the population mean () and standard deviation () of our simulated MMSE scores using the brm() function from the R package brms. The R package brms is a package for fitting Bayesian models (Bürkner 2017). It has an intuitive formula syntax, which should be familiar to you if you have any experience with the R package lme4.

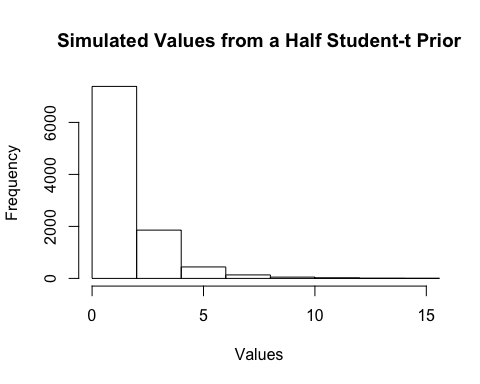
Output from the linear model we specify below includes:

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

**Before getting started, we need to specify a prior distribution for each parameter in the model ( and ).** We will specify the uniform prior discussed above for . The same uniform prior would be inappropriate for because standard deviations are never negative and do not have upper bounds. Instead, let’s 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 brms’s default prior for ). The lower bound at zero ensures that the distribution is strictly positive.

Student-*t* priors are *t* distributed. When you learned to conduct frequentist *t* tests in your introductory undergraduate psychology statistics course, you learned all about the *t* distribution. (Reminder: it’s similar to the standard normal distribution, but has an additional parameter and thicker tails.) To visualize scores sampled from this prior, copy and paste the following code into your R script and run it.

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 half student-t distributed  
hist(halft\_p, main = "Simulated Values from a Half Student-t Prior", xlab = "Values", xlim = c(0, 15)) # Visualize the simulated values

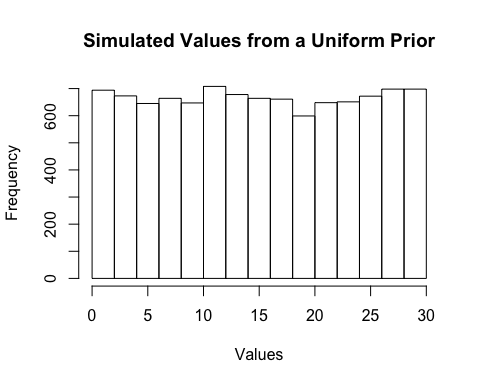


Now we can specify the model with the priors for and described above. (By the way, brms calls and “Intercept” and “sigma”, respectively.) Copy and paste the following code into your R script and run it.

prior1 <- c(prior\_string("uniform(-100000, 100000)", class = "Intercept"), # Specify the uniform prior for the population mean  
 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 a lower bound argument  
mod3 <- brm(formula = MMSE ~ 1, data = simdata, seed = 1, prior = prior1) # Fit a linear model to the simulated MMSE scores using the brm() function from the R package brms  
posterior\_summary(mod3) # Examine the model output

As discussed earlier, a lack of information is often incorporated into an analysis through specification of a uniform prior that ranges from negative to positive infinite. However, our current model (mod3) fails to incorporate the knowledge we have that MMSE scores range from 0-30, which implies that the mean of MMSE scores must be bounded by 0 and 30. We can incorporate this knowledge by specifying a uniform prior for that ranges from 0 to 30. That is, every value between 0 and 30 is equally likely to occur. To visualize scores sampled from this prior, copy and paste the following code into your R script and run it.

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 = "Simulated Values from a Uniform Prior", xlab = "Values") # Visualize the simulated values



To estimate the linear model with the updated prior for , copy and paste the following code into your R script and run it.

prior2 <- c(prior\_string("uniform(0, 30)", class = "Intercept"), # Specify the uniform prior from 0 to 30 for the population mean  
 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  
mod4 <- brm(formula = MMSE ~ 1, data = simdata, seed = 1, prior = prior2) # Fit a linear model to the simulated MMSE scores using the brm() function from the R package brms  
posterior\_summary(mod4) # Examine the model output  
fixef(mod4) # The "Estimate" column is an estimate of the population mean of MMSE scores  
 # The "Est.Error" column is the standard deviation of the estimate of the population mean  
 # The "Q2.5" and "Q97.5" columns are the lower and upper bounds (respectively) of the Bayesian credible interval for the population mean  
VarCorr(mod4) # The "Estimate" column is an estimate of the population standard deviation of MMSE scores  
 # The "Est.Error" column is the standard deviation of the estimate of the population standard deviation  
 # The "Q2.5" and "Q97.5" columns are the lower and upper bounds (respectively) of the Bayesian credible interval for the population standard deviation

## Interpreting the Model Results

Before interpreting the model results, it’s worth re-stating that **Bayesian parameter estimates are obtained by combining each parameter’s prior distribution with the likelihood function (information from the data) via Bayes’ theorem to obtain the respective posterior distributions of the parameters.** Earlier, we stated that a point estimate of a Bayesian parameter is the value that summarizes the center of its posterior distribution (usually the mean). However, the math necessary to obtain the mean of a posterior distribution is often impossible for complicated models. 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(mod4)

outputs the mean and standard deviation of the values sampled from the posterior distribution of . The “Estimate” column contains the mean of the sampled values of , and the “Est.Error” column contains the standard deviation of the sampled values of . The line of code

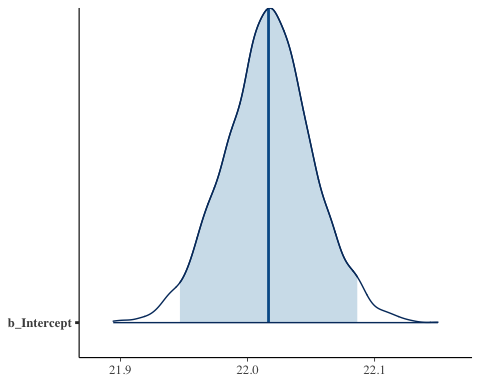
VarCorr(mod4)

outputs the mean and standard deviation of the values sampled from the posterior distribution of . The “Estimate” column contains the mean of the sampled values of , and the “Est.Error” column contains the standard deviation of the sampled values of .

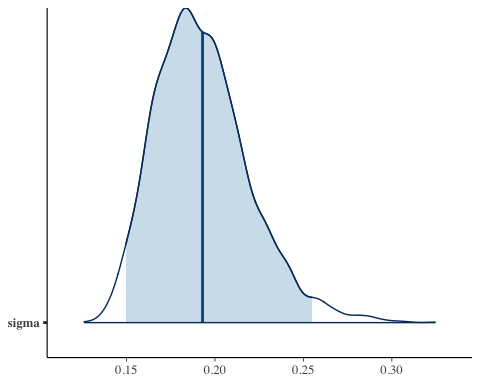
## Visualizing the Posterior Distributions of and

It can be useful to create a visual display of the values sampled from each parameter’s posterior distribution. Copy and paste the following code into your R script and run it to produce histograms of the values sampled from the posterior distributions of and .

posterior <- as.array(mod4)  
mcmc\_areas(  
 posterior,  
 pars = c("b\_Intercept"), # Tell the function to plot the posterior distribution of the mean  
 prob = 0.95, # 95% credible intervals  
 point\_est = "mean" # Specify which measure of central tendency you want to represent the point estimate  
)



mcmc\_areas(  
 posterior,  
 pars = c("sigma"), # Tell the function to plot the posterior distribution of the standard deviation  
 prob = 0.95, # 95% credible intervals  
 point\_est = "mean" # Specify which measure of central tendency you want to represent the point estimate  
)



The vertical blue line at the center of each posterior distribution represents the mean of the sampled values. The shaded light blue area of each posterior distribution represents the credible interval. The Bayesian credible interval is similar to the frequentist confidence interval but has a different interpretation.

## Interpreting Credible Intervals for the Parameters

**The Bayesian credible interval expresses uncertainty in regards to the value of a parameter estimate** (Kruschke and Liddell 2018). The larger the spread of the posterior distribution of a parameter, the wider the credible interval, the less certain we are (and vice versa).

**The interpretation of a 95% Bayesian credible interval is that the 95% most probable values of a parameter lie within the interval, given the data (Kruschke and Liddell 2018).** This interpretation correctly applies the statement of probability to the parameter. Remember, in the frequentist perspective, statements of probability are applied to the sampling distribution, not the parameter. If you need a refresher on the interpretation of a frequentist confidence interval, see the section Calculating and Interpreting a Confidence Interval for from Module 1.

The posterior\_summary() function outputs the means and standard deviations of the values sampled from each parameter’s posterior distribution along with a credible interval for each parameter. Copy and paste the following code into your R script and run it.

posterior\_summary(mod4)

The “Q2.5” and “Q97.5” columns are the lower and upper bounds (respectively) of the Bayesian credible intervals for and .

## Recommended Introductory-Level Bayesian Overview Articles

De la Fuente, E. I., Cañadas, G. R., Guàrdia, J., and Lozano, L. M. (2009), “Hypothesis probability or statistical significance? Inference for the mean in a normal distribution”, *Methodology*, 5(1), 35-39.

Depaoli, S., and Boyajian, J. (2014), “Linear and nonlinear growth models: Describing a Bayesian perspective”, *Journal of consulting and clinical psychology*, 82(5), 784.

Dienes, Z. (2011), “Bayesian versus orthodox statistics: Which side are you on?”, *Perspectives on Psychological Science*, 6(3), 274-290.

Gallistel, C. R. (2009), “The Importance of proving the null”, *Psychological Review*, 116(2), 439-453.

Gelman, A., and Shalizi, C. R. (2013), “Philosophy and the practice of Bayesian statistics”, *British Journal of Mathematical and Statistical Psychology*, 66(1), 8-38.

Howard, G. S., Maxwell, S. E., and Fleming, K. J. (2000), “The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and bayesian analysis”, *Psychological Methods*, 5(3), 315-332.

Kennedy, J. E. (2014), “Bayesian and classical hypothesis testing: Practical differences for a controversial area of research”, *Journal of Parapsychology*, 78(2), 170-182.

Klugkist, I., Van Wesel, F., and Bullens, J. (2011), “Do we know what we test and do we test what we want to know?”, *International Journal of Behavioral Development*, 35(6), 550-560.

Kruschke, J. K. (2010), “Bayesian data analysis”, *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(5), 658-676.

Kruschke, J. K. (2013), “Bayesian estimation supersedes the *t* test”, *Journal of Experimental Psychology: General*, 142(2), 573-588.

Masson, M. E. J. (2011), “A tutorial on a practical Bayesian alternative to null-hypothesis significance testing”, *Behavior Research Methods*, 43(3), 679-690.

Rupp, A. A., Dey, D. K., and Zumbo, B. D. (2004), “To bayes or not to bayes, from whether to when: Applications of Bayesian methodology to modeling”, *Structural Equation Modeling*, 11(3), 424-451.

Van de Schoot, R., Hoijtink, H., and Romeijn, J. W. (2011), “Moving beyond traditional null hypothesis testing: Evaluating expectations directly”, *Frontiers in Psychology*, 2, 24.

Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., and van Aken, M. A. G. (2014), “A Gentle introduction to bayesian analysis: Applications to developmental research”, *Child Development*, 85(3), 842-860.

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

## References

Bürkner, P. (2017), “brms: An R Package for Bayesian Multilevel Models Using Stan”, *Journal of Statistical Software*, 80(1), 1–28.

De Finetti, B. (1974), *Theory of probability,* 1st Edition, New York, NY: Wiley.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975), “Mini-mental state: A practical method for grading the cognitive state of patients for the clinician”, *Journal of Psychiatric Research*, *12*(3), 189-198.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013), *Bayesian data analysis*, 3rd Edition, CRC press.

Kaplan, D., and Depaoli, S. (2013), *Bayesian statistical methods*, in T. D. Little (Ed.), Oxford handbook of quantitative methods (pp. 407–437), Oxford University Press.

Kochhann, R., Varela, J. S., de Macedo Lisboa, C. S., and Chaves, M. L. F. (2010), “The Mini Mental State Examination: review of cutoff points adjusted for schooling in a large Southern Brazilian sample”, *Dementia and Neuropsychologia*, 4(1), 35.

Siu, A. L. (1991), “Screening for dementia and investigating its causes”, *Annals of Internal Medicine*, 115(2), 122-132.

Van De Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., and Depaoli, S. (2017), “A systematic review of Bayesian articles in psychology: The last 25 years”, *Psychological Methods*, 22(2), 217.