Bayesian data analysis for Sport Science

Supplementary materials:
https://github.com/JorgeDelro/Intro_Bayesian
*For correspondence:
jorgedelrosario.fernandez@uca.es

Jorge R Fernandez Santos[1,3*], Jesus G Ponce-Gonzalez[2,3], Jose Castro-Piñero[1,3], Jose L. Gonzalez-Montesinos[4]
[1]GALENO Research group and Department of Physical Education, Faculty of Education Sciences, University of Cádiz, Spain
[2]MOVE-IT Research group and Department of Physical Education, Faculty of Education Sciences, University of Cádiz, Spain
[3]Biomedical Research and Innovation Institute of Cádiz (INiBICA) Research Unit, Cádiz, Spain
[4]Department of Physical Education, Faculty of Education Sciences, University of Cádiz, Spain

Please cite as: Fernandez Santos, J.R., Ponce-Gonzalez, J.G., Castro-Piñero J. & Gonzalez-Montesinos, J.L. (2022). Bayesian Data analysis for Sport Science. *SportRχiv*.

ABSTRACT

Bayesian data analysis (BDA) is method of statistical inference that make use of the probability for quantify uncertainty in inferences based on statistical data analysis. Along the manuscript different concepts are addressed to those sport scientists that want to start to use BDA: the Bayes 'rule, hierarchical modeling, Markov Chain Monte Carlo techniques, predictive modeling, hypothesis testing and Bayesian workflow. Finally, an applied example is performed to help to apply of the previous concepts form a practical point of view and to report results.

All authors have read and approved this version of the manuscript. This article was last modified on 28th Jan 2022.

INTRODUCTION

Bayesian data analysis (BDA) is already a well-established method of statistical inference in many different disciplines like psychology, ecology, economy or health science [1–4]. Briefly, BDA make use of the probability for quantify uncertainty in inferences based on statistical data analysis [5]. This approach has some advantages over the traditional methods (also known as frequentist statistics) like the incorporation of prior knowledge to the statistical model via prior distribution, the result obtained is only based on the specific data under consideration, regardless of model complexity the final estimation is always a posterior probability distribution not depend on the stopping or testing intentions of the analyst and the straightforward interpretation of results [6–8].

Although the use of BDA in sport analytics has increased substantially in the last years [9], it has been argued recently that the current statistical practices in sport science are based on the null hypothesis significant testing under the frequentist approach and that this approach is flawless so sport scientists should shift towards alternative statistical methods like Bayesian statistics [10]. Nevertheless, the major drawback is that most current sport scientists are not trained in BDA.

Precisely, one of these studies aimed to introduce to sport scientists the Bayesian inference as a method of estimation in small sample size and small effects studies [11]. However, an introduction to BDA should have focused more in all the steps of a proper Bayesian workflow: First, the construction of informative priors when dealing with small sample size; Second, mathematical formulation of the model connecting it with the computer code; Third, model checking (i.e. prior and posterior predictive checking); and fourth, readers would have more benefits if every step of the analysis is presented with the computer code that executed it.

Many different software packages have been developed in recent years for Bayesian modeling [12]. Of all of them, the R package *brms* gather some characteristics that make it an ideal starting point to learn BDA: is user-friendly; models are specified using lme4-like formula syntax; It can be used to fit from single-level linear regression to multivariate or non-linear multilevel models; It uses the probabilistic programming language Stan to fit the models; and it has a large and growing user community [13,14].

Therefore, the main aim of this paper is to provide a practical introduction for those sport scientists who want to start to apply BDA and especially to those who usually have to deal with small sample size. This paper will be structured in two main sections: 1) a brief introduction to BDA fundamentals and 2) a detailed description of the Bayesian workflow using a working example. Throughout this paper it is only assumed that the reader is familiar with the R programming language for data analysis.

Fundamentals of BDA

Bayes 'rule and probability distributions

BDA consists in formulate a full probability model to make conclusions about a vector of parameters (θ) conditionally on the outcome variable (y), predictors (x) and the information we know a priori about these parameters ($p\theta$). These conclusions are expressed in terms of probability statements and they are calculated via the Bayes' rule. For regression models the Bayes' rule is expressed as:

$$p(\theta|y, x) \propto p(y|\theta, x)\, p(\theta|x),$$

where $p(y|\theta, x)$ is the likelihood function as it described the generative process of the outcome y given the parameters θ and the predictors x. Usually, researchers choose one of the member of the exponential family to describe the likelihood of the outcome. Sport scientists may be familiar with some of the members of this family like the normal distribution to describe a continuous outcome in linear regression, the binomial distribution for a binary outcome or the Poisson distribution for count outcome in generalized linear regression.

$p(\theta|x)$ represents the prior distribution and it contains all the information we have about the parameters θ from previous studies. Generally, three different classes of prior distributions can be distinguished related the amount of (un)certainty they incorporate to the model. *Non-informative* priors (also known as *vague* prior) have been used commonly on parameters where the researcher has no knowledge about its possible values. This class of prior can be found easily in the literature when performing BDA with software like BUGS or JAGS [15]. However, they should be replaced by a more informative prior to improves inferences due to theorical and computational reasons. *Weakly informative* priors encoded information to restrict the plausible range of values of a specific parameter but still leave a wide range of values to be cover [16]. This class of prior distribution has been recently proposed as default prior when there is no information about a parameter of the model. Lastly, a prior is considered to be *informative* when a researcher includes all the available information in a prior distribution restricting considerably the parameter space. Prior distributions play a key role in BDA when dealing with small sample size due to we can increase the precision of the estimated model parameters by excluding values that are not plausible through the use of informative priors [17].

$p(\theta|y, x)$ is the posterior probability distribution and it contains all the information about the model parameters after multiplying the likelihood function and the prior distribution. It is important to note that the posterior distribution is a compromise between the data we have at hand and the prior information. This means that with a small sample size the result obtained in the posterior distribution will be mainly determined by the prior distribution.

One of the key aspects of BDA is to think that we are creating a generative model from which we can simulate new data. This generative model can let us to make predictive inference about data that we haven´t observed yet. The probability distribution obtained simulating new observations of the outcome from our model using only the prior distribution on the parameters is called the prior predictive distribution. Once the observed data have been included in the

model, we can simulate again new data from the posterior predictive distribution. Both predictive distributions are fundamental in model checking.

Bayesian hierarchical modeling

BDA involve the formulation of a full probability model starting from the likelihood function of the data to the prior distribution of the parameters. This mathematical formulation of the model where the values of some parameters depend on the values of other parameters is known as hierarchical modeling and represent the parametization of the model. Consider the following example with one outcome (y) and two predictor variables (X_1 and X_2):

$$
\begin{aligned}
y_i &\sim \text{Normal}(\upsilon_i, \sigma) && \text{[likelihood]} \\
\upsilon_i &= \alpha + \beta_1 X_1 + \beta_2 X_2 && \text{[linear model]} \\
\alpha &\sim \text{Normal}(\upsilon_\alpha, \sigma_\alpha) && [\alpha \text{ prior}] \\
\beta_1 &\sim \text{Normal}(\upsilon_{\beta 1}, \sigma_{\beta 1}) && [\beta_1 \text{ prior}] \\
\beta_2 &\sim \text{Normal}(\upsilon_{\beta 2}, \sigma_{\beta 2}) && [\beta_2 \text{ prior}] \\
\sigma &\sim \text{Uniform}(L_\sigma, U_\sigma) && [\sigma \text{ prior}]
\end{aligned}
$$

This formulation is the classical linear model where every observation of the outcome variable y is assumed to be distributed according to a Gaussian probability distribution with mean υ and standard deviation σ. Additionally, the mean υ is assumed to be equals a linear combination of the parameters α (i.e., the intercept), the effects of X_1 (β_1) and X_2 (β_2). The novel part is that prior probability distribution has been set on the model parameters α, β_1, β_2 and σ. In fact, these priori distributions also have parameters (also known as **hyperparameters**) that are also estimated from the data.

Markov Chain Monte Carlo methods

Different numerical techniques could be used to compute the posterior distribution of the model parameters like grid or quadratic approximation [18]. However, the most popular method to fit complex multiparameter models is **Markov Chain Monte Carlo** (MCMC). This method is the combination of two different techniques, Markov Chains and Monte Carlo simulation [19]. The former is a stochastic process (i.e., set of random quantities) where the probability of change to a new state at time $t + 1$ is dependent only of the current state of the process at time t and conditionally independent of the previous values. The latter is a powerful computational method used to generate independent random samples from a sampling distribution, this empirical samples could be used to summarize the distribution without using analytical calculations. Therefore, a MCMC is a process where random samples are drawn sequentially from the approximate posterior distribution of each model parameter simultaneously. At each step of the sequence, the algorithm corrects the draws using the

Markov property of the chain to better approximate the posterior distribution. The key point is that if we run the chain long enough it will converge to a stationary posterior distribution [5]. Metropolis and Gibbs sampling are probably the most widely known methods implemented both in BUGS and JAGS [15]. Recently, a probabilistic programming language called Stan have been developed [20]. This software make use of the No-U-Turn sampler, a variant of Hamiltonian Monte Carlo to compute the posterior distribution [21]. Hamiltonian Monte Carlo sampling have been showed to outperforms Metropolis and Gibbs sampling for complex multiparameter models [22].

MCMC methods are implemented by default in the Bayesian software so researchers do not have to worry about manually code it. However, it is essential to assess the representativeness of the posterior distribution and that the estimates of central tendency and limits are accurate and stable using numerical and graphical convergence diagnostics [23]. The potential scale reduction factor ($\hat{R}$) and the effective sample size (ESS) are probably the numerical converge diagnostics most used in the Bayesian software. $\hat{R}$ is a measure of how much variance there is between the chains relative to how much variance there is within chains and its value is 1.0 the chains are fully converged or greater if they are not converged to a common distribution. ESS is a measure of how much independent information there is in autocorrelated chains. Recently, an improved version of these numerical diagnostics has been developed and implemented in the probabilistic programming language Stan [24]. Stan output reports for every parameter estimated the maximum of rank normalized split-$\hat{R}$ and rank normalized folded-split-$\hat{R}$ which work for thick tailed distributions and is sensitive also to differences in scale. Moreover, the bulk effective sample size (bulk-ESS) and tail effective sample size (tail-ESS) are reported. The former informs about the sampling efficiency in the bulk of the distribution (related to efficiency of mean and median estimates) whereas the latter is a measure for sampling efficiency in the tails of the distribution (related to efficiency of variance and tail quantile estimate). It is recommended from a practical point of view to run at least four chains by default to estimate the posterior distribution of model parameters using MCMC and use 1.01 (or lower) and 400 (or greater) as thresholds for $\hat{R}$ and ESS respectively to trust in the posterior distribution estimated.

Model comparison and predictive accuracy

Once the model is fitted sport researchers assess how well the model fit to the sample. Probably, the most common measure used is R^2 described as "variance explained" or "goodness-of-fit". This measure has the problem that it increases when more predictors are added to the model even when the variables you add are random numbers [18]. Moreover, while models with many parameters fit the data better, they tend to *overfit* more than simple models. Overfitting occurs when the model learns too much from the sample which leads to poor out-of-sample predictions. In contrasts, when a model has too few parameters, they are inaccurate both within and out-of-sample producing a statistical error called underfitting. To deal with the

overfitting/underfitting dichotomy we can use two different approaches: cross-validation and information criteria.

The first approach consists basically on leave out a small part of our sample to test the model´s predictive accuracy. Therefore, the sample is divided into chunks (i.e., folds) which the statistical model is asked to predict one by one using the remaining chunks of the sample. Then, an average score of the out-of-sample accuracy if obtained. For Bayesian models we are going to use a special cross-validation method called the Pareto smoothed importance sampling cross-validation (PSIS-LOO) to estimate the model´s out of sample accuracy [25]. Leaving aside the mathematical aspects, this method computes the expected log pointwise predictive density which it is a useful measure to compare models and the Pareto k diagnostics which informs us about the reliability of the estimate by pointing to influential observations. Specifically, those data points associated with a k value higher than 0.7 are supposed to have a negative on PSIS-LOO score. A difference less than 4 in the expected log pointwise predictive density is considered small from a practical point of view when comparing models with a number of observations larger than 100.

The second approach is to compute an information criterion which report an estimate of the relative out-of-sample divergence. The widely applicable information criterion (WAIC) is an information criterion that is invariant to parametrization and it uses entire posterior distribution [26]. WAIC can be used together with PSIS-LOO for model comparison.

Hypothesis testing

Sport scientists maybe know the model of the section 2.2 as ANCOVA where the interest resides in estimate mean difference among groups by using some kind of planned contrasts or post-hoc analysis while adjusting the model with a continuous variable. In this way, the decision of whether or not there is a statistically significant difference among training groups is based on the computation of a p-value and if it is less than an established threshold (traditionally if $p<0.05$). However, several publications have alarmed about the misuse and misinterpretation of p-values as an index of significance [27–30]. As an alternative, Bayesian inference offers two different approaches to analyze the presence or absence of an effect: estimation approach based on the description of the posterior distribution and Bayes factor (BF) approach [31].

The first approach uses an interval called a credible interval (CrI) which define a percentage of values that are found in the central portion of the posterior distribution. Although its aim is similar, CrI should not be confused with confidence intervals since its computation and meaning are different. A special CrI is the highest-density interval (HDI) which summarizes the uncertainty of the parameter estimated in such way that any parameter value inside a 95%HDI are the 95% most credible values. Then, it is calculated what percentage of the HDI falls inside a region of practical equivalence (ROPE) that represent a range of parameter values that equivalent to the null value for practical purposes [32]. Thus, if for example the 95%HDI falls completely inside the ROPE means that the most credible values of the parameter a practically

equivalent to the null value. Obviously, the ROPE has had to be established by the researcher based on previous studies.

The second approach is based on the comparison of two probability distributions: one (prior distribution) where all the probability is allocated over the null value (or ROPE), and one (posterior distribution) where the probability mass has shifted away from the null value once the observed data have been taken into account. Therefore, a BF indicates the degree to which the posterior distribution has move further away or closer to the null value and it is represented as BF.

It is important to note some advantages of using these approaches to perform post-hoc analysis like there is no need to correct for multiple tests due to type I error rate inflation due to BDA does not rely on sampling distributions; Conversely to frequentists confidence intervals, the interpretation of a HDI is intuitive; and BF assess both evidence in favor or against an effect (in contrasts to p-values) [33].

Bayesian workflow

It has been highlighted recently the need to establish a Bayesian workflow due to different reasons as computation of complex statistical models, model extensions and comparing multiples models fit to the same data [34]. This workflow currently includes the steps of model building, checking and inference. Several authors have already proposed steps for BDA specifically and regression-type analysis in general [35–39]. Here, we are going to summarize the key steps of BDA:

Gather prior information

BDA starts even before analyzing the database. Researchers can use the results reported in previous studies to get an idea of the possible values that parameters of interest may have. These values can be incorporated to our analysis via prior distributions and thus exclude values that are not possible to reach. Therefore, to include informative prior in our model is going to provide us the possibility of increase the precision of the result even if the sample size is small. If previous information is not available maybe researchers are able to specify the limit of the parameter space. Practical guidelines to construct informative priors are (ref): 1) research in high quality scientific literature and ask experts on the subject matter; 2) to use a good method to gather information systematically; 3) to specify where you got the information and 4) always visualize the prior distribution.

Exploratory data analysis

To summarize the data collected by using graphs like histograms or boxplots and summary statistics of central and dispersion tendency is an essential step to investigate about the distribution of the variables and the relationship established among them. To explore the nature of the data will allow researchers to define a proper likelihood function for the outcome

(e.g., observations of the outcome variable are generated following a normal distribution or Student-t distribution).

Define the statistical model: prior distributions and likelihood function

Once the prior distributions on model parameters and the likelihood function have been defined, researchers must represent using mathematical notation the dependencies among the variables in the model as we have seen in section 2.2. It is important to ensure that the results are replicable by specifying the outcome, the predictor variables, the parameters and hyperparameters and their probability distributions along with a text clarifying the mathematical description of the model.

Prior predictive checking

We must ensure that the statistical model that we just defined is consistent with domain knowledge. Thus, one solution can be to simulate data from the model using only the information incorporated via prior distribution and then check if the data generated fall in the range we expected or there are a high number of extreme observations. This method is essential to evaluate the validity of the model denied in the previous step.

Parameter estimation and model checking

The posterior distribution of the parameters can be obtained by fitting the model using a statistical software. Then, we must check if the results obtained are reliable using graphical and numerical MCMC diagnostics (i.e., $\widehat{\boldsymbol{R}}$, ESS and traceplots; section 2.3) and performing posterior predictive checking.

Posterior predictive checking

We want to highlight this method since it is a very simple and powerful tool to evaluate the model fitted. The idea is the same seen for prior predictive checking but now the data have been added to the model: if the model is good then the new data simulated from the model must resemble the data observed. This method makes use of the data twice: once to fit the model and another one to perform the checking so it is recommended to choose statistics that are orthogonal to the model parameters [40].

Model selection

Several models can be fitted to answer to the same research question. Therefore, we need to know which model is the best and should be used to perform inference. Recall from section 2.4 that we must worry with the overfitting-underfitting dichotomy and we can use PSIS-LOO and WAIC to deal with it.

Further analysis and report results

When the final model is fitted and all the checks have been successfully passed, sport researchers are very often interested in testing hypothesis to report "a significant statistical result". These hypothesis in the context of regression analysis are related to the effect of the parameters estimated or to perform additional pairwise-comparisons between the levels of a categorical variable. Final results reported must include a point estimate (i.e., mean or median) and a credible interval that summarize the posterior samples obtained.

Applied Bayesian workflow example

In this section we are going to consider the case study of Mergensen et al. [11] who in turn used a study of Humberstone-Gough et al. [41] to illustrate the workflow analysis. Briefly, they compared the effects of three different training regimens "Live High Training Low" altitude training (LHTL, n = 7), "Intermittent Hypoxic Exposure" (IHE, n = 7), and "Placebo" (n = 7) on different variables using a pre-post design. For the sake of simplicity, the difference in the concentration of hemoglobin mass (Hbmass, units of grams) is going to be the outcome of our example while the percentage change in weekly training load (ChangeWtr, %) and training group membership (Group, three levels: LHTL, IHE and Placebo) are the predictor variables. Our interest as researchers lies in analyzing differences among the training groups. A box plot of the Hbmass by group show us that the outcome follows a Gaussian distribution in each group, the presence of an outlier in the IHE group and a possible effect of group membership (figure 1).

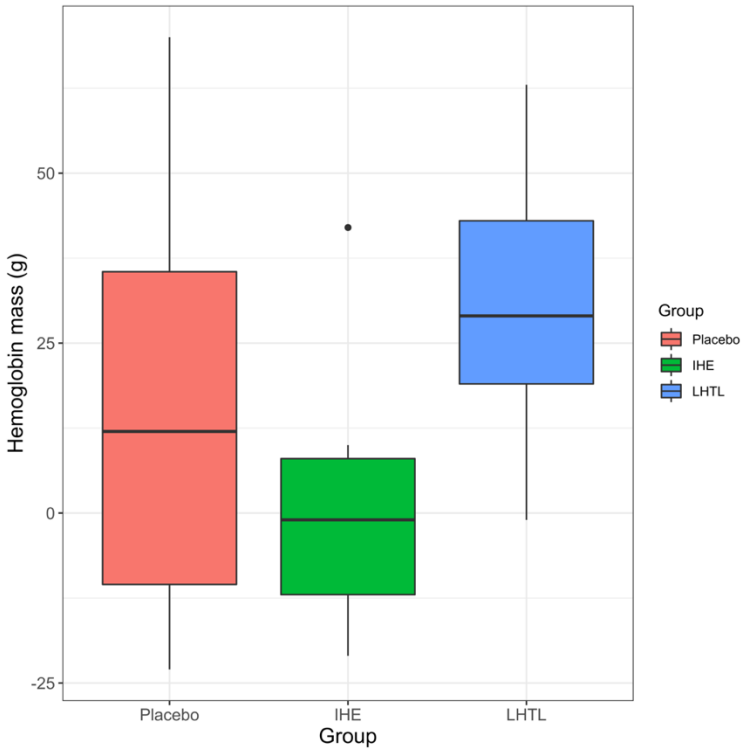

Figure 1. Box-plot of the distribution of hemoglobin mass across training groups.

Note that there are only 7 participants in each group so prior information about the parameters can help us to get a reliable estimate. In this case, an informative prior about the effect of LHTL was placed based on a meta-analysis about training regimens on Hbmass [42].

Model definition

We assumed that the Hbmass is distributed according a Gaussian distribution while the intercept (α) and standard deviation (σ) follow a Student´s T distribution. Therefore, the statistical model can be described as follow:

$$Hbmass_i \sim \text{Normal}(u_i, \sigma) \quad \text{[likelihood]}$$
$$u_i = \alpha + \beta_1 \, ChangeWtr + \beta_2 \, Group \quad \text{[linear model]}$$
$$\alpha \sim \text{StudentT}(12, 7, 3) \quad [\alpha \text{ prior}]$$
$$\beta_1 \sim \text{Normal}(0, 2) \quad [\beta_1 \text{ prior}]$$
$$\beta_{2IHE} \sim \text{Normal}(0, 2) \quad [\beta_{IHE} \text{ prior}]$$
$$\beta_{2LHTL} \sim \text{Normal}(22.6, 1) \quad [\beta_{LHTL} \text{ prior}]$$
$$\sigma \sim \text{StudentT}(0, 15, 3) \quad [\sigma \text{ prior}]$$

Recall that the variable Group is categorical so the effect of this variable (β_2) is interpreted as deviation from the reference group (the control group in our model). It is a good practice to plot prior distribution to check the range of plausible values for each parameter (Figure 2).

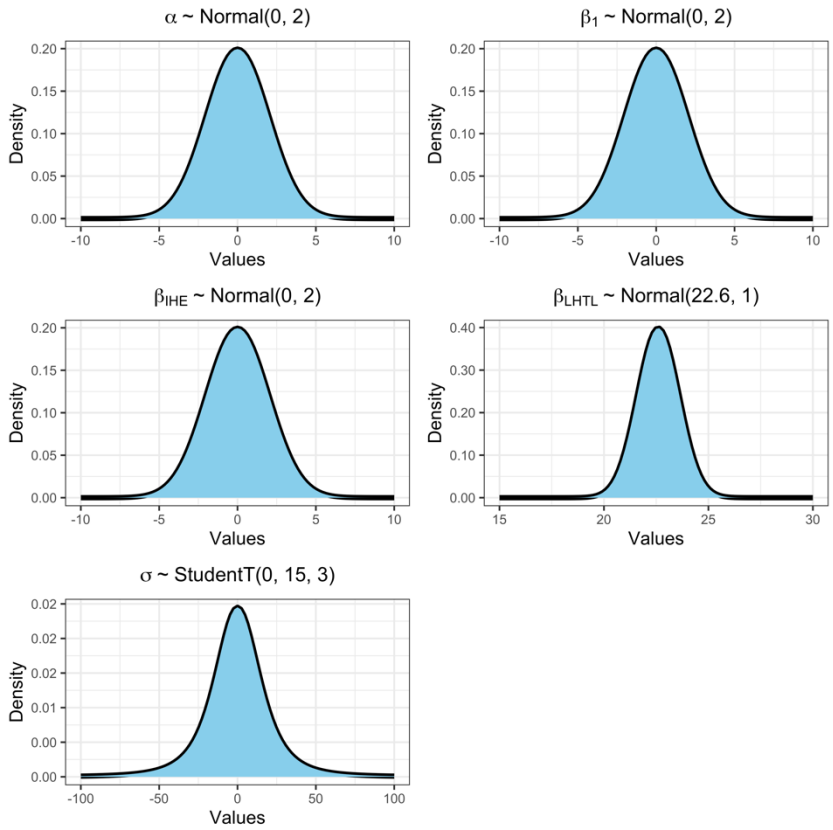

Figure 2. Prior distributions on model parameters.

Prior predictive checking

Once the model is defined, the prior predictive distribution can be computed to check whether the model and prior distributions are consistent with domain expertise removing extreme but not impossible parameters values [38]. Hence, adding information via prior distribution allows the Bayesian computation and interpretation of the parameters estimated. Prior predictive distribution can be computed via brms by using the function **`brm`** and setting the argument **`sample_prior = "only"`**. The function **`brm`** can be considered the main function of the package since is the one used to fit the models. Consider special attention to the argument related to the MCMC, **`warmup`** to set the number of iterations used by the MCMC algorithm to figure out how to explore the posterior distribution efficiently; **`chains`** to specify the number of Markov chains and **`iter`** to set the number of iterations per chain. In our example, we create an object called **`bmod1Prior`** which will store all the information about the model. Additional arguments like **`data`** to select a data frame that contains all the variables in the model; **`family`** to set the likelihood function of the outcome (see section 2.1) and **`prior`** to use the prior distribution on parameters previously defined are necessary. Note that our model assume that the outcome follows a Gaussian distribution with an identity link function (**`family =`**

`gaussian(link = "identity")`). The link function is used to establish a relationship between predictor variables and the mean of the outcome distribution and in the case of linear regression models the link function used is the identity.

Prior predictive distribution is showed in figure 3. In this figure y represent the distribution of Hbmass and y_{rep} the distribution of simulated sets using only information from prior distributions. Note that most of the distribution area is over the value 0 and values ± 100 g for Hbmass are very unlikely.

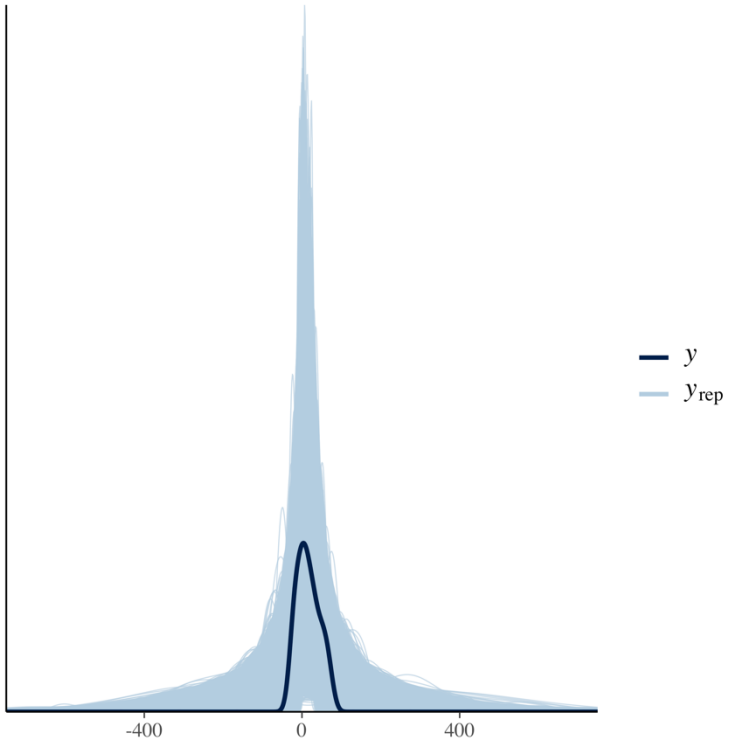

Figure 3. Prior predictive distribution.

Model fitting

Next, we fit the model by updating the object bmod1Prior by changing the argument sample_prior = "no". Once brms fits the model we should check that the parameters have been estimated correctly (see section 2.2). The function summary(bmod1) shows us the parameter values estimated and additional information about the reliability of these results (table 1). Here, all the parameters have a $\hat{R}$ (Rhat in the table) of 1 and both ESS > 400 so we can trust that these results have been obtained with accuracy.

Table 1. Parameter estimation results.

Population-Level Effects:							
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	0.02	1.97	-3.77	3.87	1.00	5084	3642
ChangeWtr	0.20	0.10	-0.01	0.40	1.00	4156	2962
GroupIHE	-0.20	1.97	-3.94	3.65	1.00	5406	3246
GroupLHTL	22.66	1.01	20.68	24.62	1.00	4647	2702
Family Specific Parameters:							
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	25.705	4.3.92	18.86	34.04	1.00	4364	2967

l-95% CI indicates lower-bound of 95% credible interval; u-95% CI, upper-bound of 95% credible interval; Rhat, potential scale reduction factor; Bulk_ESS bulk effective sample size; Tail_ESS tail effective sample size.

Additional checks can be done by checking the traceplot of each Markov chain used in the MCMC estimation (figure 4). This traceplot should be concentrated around the estimated value for each parameter.

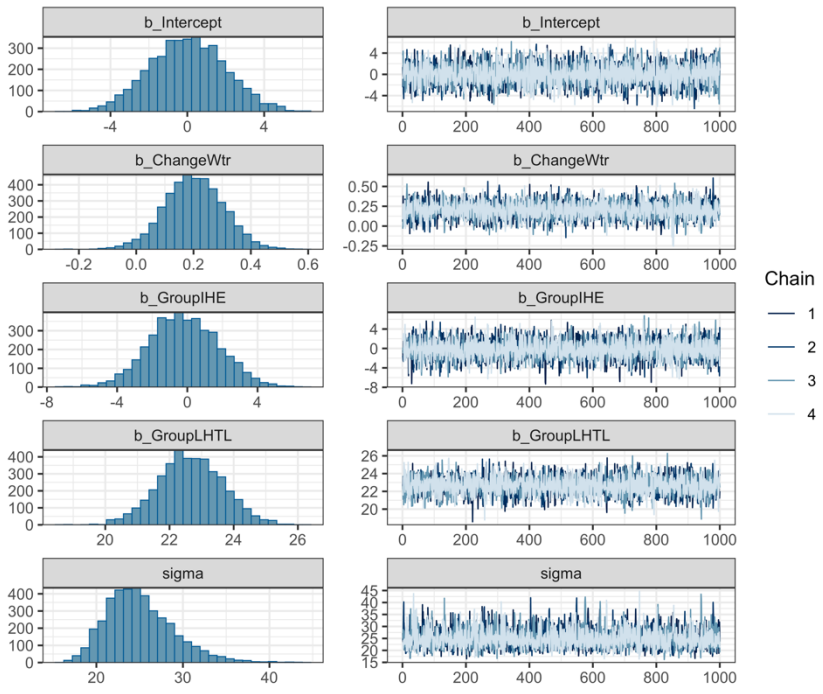

Figure 4. Posterior parameter distribution on the left column. Traceplots for each parameter estimated on the right column.

Posterior predictive checking

We are going to simulate data sets (y_{rep}) to compare with the distribution of the observed data (y) like in section 3.2 but in this case, it is used the posterior distribution. This method is used to asses model adequacy. Figure 5 shows the posterior predictive distribution of our model. Look like the fit is reasonable but there is a high variation that it is not capture by model´s prediction.

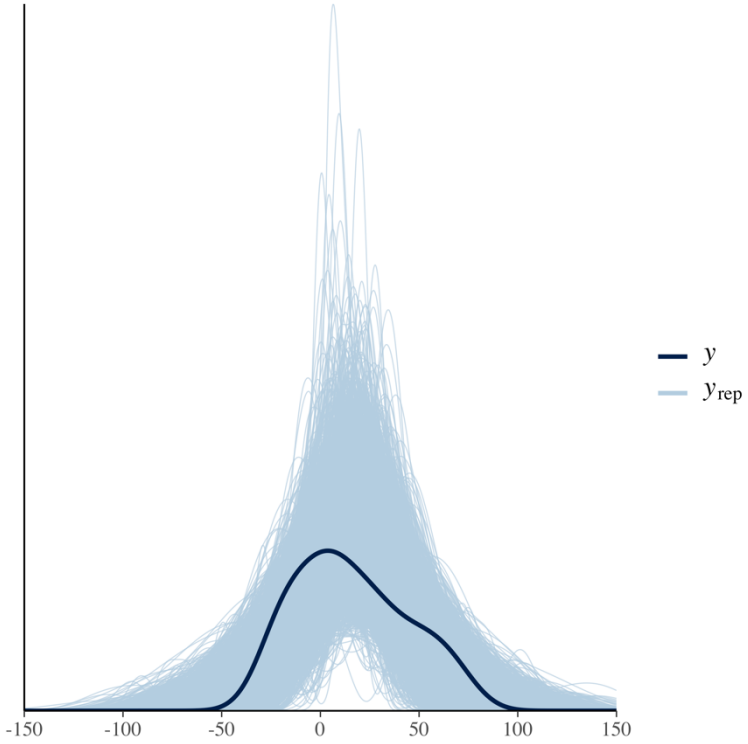

Figure 5. Posterior predictive distribution.

Model updating and selection

Figure 1 showed us the presence of an outlier in IHE group so perhaps we could improve the model if we use a likelihood function that allows the presence of extreme values. This kind of method is commonly known as robust regression and makes use of the Student-t distribution. Like the Gaussian distribution, the Student-t distribution is defined by the mean μ and the scale σ parameters, but it has also the shape parameter υ that controls the thick of the tails of the distribution. This Bayesian robust regression have been described previously and can be used

in practice by sport researchers. Robust regression can be easily performed using brms by changing the argument family (family = student(link = identity)).

Once the model is fitted, we can compare the predictive accuracy of both model (section 2.4). First, the PSIS-LOO is estimated for each model via loo function setting the argument save_psis = T and then function loo_compare is used. This function computes pairwise comparisons between the model with the largest expected predictive density (first row, better accuracy). In our case, the difference can be considered insignificant due to the small numbers computed (table 2). Interestingly, the Gaussian model has better accuracy so we are going to use that model to perform contrasts.

60. Table 2. Loo_compare function results comparing the predictive accuracy of bmod1 and bmod2

	elpd_diff	se_diff
bmod1	0.0	0.0
bmod2	-0.5	0.3

Elpd_diff means difference in the expected log predictive density; se_diff, standard error of the difference in the expected log predictive density.

Post-hoc hypothesis contrasts

Recall that our interest resides in compute contrasts among the levels off the group variable. The function **`hypothesis`** allows to perform multiple non-linear hypothesis test for model parameters. The specific contrasts (i.e. pairwise differences) should be encoded as a character string by using the name of the model parameters. In our example, to test the differences between placebo vs IHE, placebo vs LHTL and IHE vs LHTL respectively we should use the following string: **`c("Intercept = Intercept + GroupIHE", "Intercept = Intercept + GroupLHTL", "Intercept + GroupIHE + Intercept + GroupLHTL")`** in the argument called **`hypothesis`**.

Table 3. Pairwise comparisons among the training groups.

Hypothesis	Estimate	Est.	CI.	CI.Upper	Evid.	Post.	Star
1	0.20	1.97	-3.65	3.94	0.97	0.49	
2	-22.66	1.01	-24.62	-20.68	0.00	0.00	*
3	-22.85	2.24	-27.26	-18.52	0.00	0.00	*

Hypothesis 1 = (Intercept) – (Intercept + GroupIHE) = 0; Hypothesis 2 = (Intercept) – (Intercept + GroupLHTL) = 0; Hypothesis 3 = (Intercept + IHE) – (Intercept + GroupLHTL) = 0. Est. Error means standard error; CI. Lower, lower-bound of the credible interval; CI. Upper, upper-bound of the credible interval; Evid. Ratio, evidence ratio; Post.Prob, posterior probability.

This function computes a Bayes factor between the hypothesis and its alternative and is expressed as BF_{01}. This result is showed in the column called "evidence ratio" and more specifically it refers to the evidence of H_0 (i.e., null hypothesis = no significant difference) over H_1 (i.e., alternative hypothesis = significant difference). We prefer to show from practical point of view BF_{10} that means the evidence H_1 over H_0. To calculate it divide 1 by the result of the evidence ratio so for hypothesis 1, 2 and 3 the BF_{10} is 1,03, >100 and >100 respectively. This evidence can be classified as anecdotical for hypothesis 1 and extreme for hypothesis 2 and 3 [1]. Therefore, it can be reported that there is a significant effect of LHTL training regimen over placebo (Difference (95% Credible interval; BF_{10}) = 22.66 g (20.68, 24.62; BF_{10} = >100) and IHE training regimen (22.85 g (18.52, 27.26; BF_{10} = >100).

CONCLUSION

BDA offers a very interesting alternative for sport scientists who want to overcome the limitation of traditional statistics, especially those who need to analyze databases with low sample size. Obviously, there are lot of concepts and methods that have not been treated in the text. However, through this manuscript the basic concepts, benefits, workflow and a practical example are presented as a starting point for those who are interested in learn how to perform Bayesian inference.

Acknowledgments

Different R packages have been used together with *brms* to perform the analysis of this manuscript: *tidyverse* and *ggpubr* for data manipulation and plotting and *bayestestR*, *bayesplot* and *loo* for further analysis [43–47].

Contributions

JRFS and JLGM wrote the first draft of the manuscript. JRFS wrote all the code and perform all the analyses. JGPG and JCP read, revised and completed the final version of the manuscript.

Data and Supplementary Material Accessibility

The data, reproducible code and figures of this manuscript can be found in https://github.com/JorgeDelro/Intro_Bayesian

REFERENCES

1. Lee, M.D.; Wagenmakers, E.-J. *Bayesian Cognitive Modeling: A Practical Course*; Cambridge University Press: Cambridge, 2014; ISBN 978-1-107-01845-7.

2. Kery, M. *Introduction to WinBUGS for Ecologists: Bayesian Approach to Regression, ANOVA, Mixed Models and Related Analyses*; Academic Press, Inc., 2010; ISBN 0-12-378605-3.
3. Greenberg, E. *Introduction to Bayesian Econometrics*; 2nd ed.; Cambridge University Press: Cambridge, 2012; ISBN 978-1-107-01531-9.
4. Lesaffre, E.; Lawson, A.B. *Bayesian Biostatistics*; John Wiley & Sons, 2012;
5. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*; CRC press, 2013;
6. Wagenmakers, E.-J.; Marsman, M.; Jamil, T.; Ly, A.; Verhagen, J.; Love, J.; Selker, R.; Gronau, Q.F.; Šmíra, M.; Epskamp, S.; et al. Bayesian Inference for Psychology. Part I: Theoretical Advantages and Practical Ramifications. *Psychon. Bull. Rev.* 2018, *25*, 35–57, doi:10.3758/s13423-017-1343-3.
7. Kruschke, J.K.; Liddell, T.M. Bayesian Data Analysis for Newcomers. *Psychon. Bull. Rev.* 2018, *25*, 155–177, doi:10.3758/s13423-017-1272-1.
8. Dienes, Z.; Mclatchie, N. Four Reasons to Prefer Bayesian Analyses over Significance Testing. *Psychon. Bull. Rev.* 2018, *25*, 207–218, doi:10.3758/s13423-017-1266-z.
9. Santos-Fernandez, E.; Wu, P.; Mengersen, K.L. Bayesian Statistics Meets Sports: A Comprehensive Review. *J. Quant. Anal. Sports* 2019, *15*, 289–312, doi:10.1515/jqas-2018-0106.
10. Bernards, J.R.; Sato, K.; Haff, G.G.; Bazyler, C.D. Current Research and Statistical Practices in Sport Science and a Need for Change. *Sports Basel Switz.* 2017, *5*, doi:10.3390/sports5040087.
11. Mengersen, K.L.; Drovandi, C.C.; Robert, C.P.; Pyne, D.B.; Gore, C.J. Bayesian Estimation of Small Effects in Exercise and Sports Science. *PloS One* 2016, *11*, e0147311, doi:10.1371/journal.pone.0147311.
12. Mai, Y.; Zhang, Z. Software Packages for Bayesian Multilevel Modeling. *Struct. Equ. Model. Multidiscip. J.* 2018, *25*, 650–658, doi:10.1080/10705511.2018.1431545.
13. Bürkner, P.-C. Brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw. Vol 1 Issue 1* 2017.
14. Bürkner, P.-C. Advanced Bayesian Multilevel Modeling with the R Package Brms. *R J.* 2018, *10*, 395–411, doi:10.32614/RJ-2018-017.
15. Lunn, D.; Jackson, C.; Best, N.; Thomas, A.; Spiegelhalter, D. *The BUGS Book: A Practical Introduction to Bayesian Analysis*; CRC press, 2012;
16. Gelman, A.; Simpson, D.; Betancourt, M. The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy* 2017, *19*, doi:10.3390/e19100555.
17. Zondervan-Zwijnenburg, M.; Peeters, M.; Depaoli, S.; Van de Schoot, R. Where Do Priors Come From? Applying Guidelines to Construct Informative Priors in Small Sample Research. *Res. Hum. Dev.* 2017, *14*, 305–320, doi:10.1080/15427609.2017.1370966.
18. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*; CRC press, 2020;
19. Gill, J. *Bayesian Methods: A Social and Behavioral Sciences Approach*; CRC press, 2014; Vol. 20;.

20. Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. Stan: A Probabilistic Programming Language. *J. Stat. Softw. Vol 1 Issue 1* 2017.
21. Hoffman, M.D.; Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* 2014, *15*, 1593–1623.
22. Monnahan, C.C.; Thorson, J.T.; Branch, T.A. Faster Estimation of Bayesian Models in Ecology Using Hamiltonian Monte Carlo. *Methods Ecol. Evol.* 2017, *8*, 339–348, doi:10.1111/2041-210X.12681.
23. Kruschke, J. Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2014.
24. Vehtari, A.; Gelman, A.; Simpson, D.; Carpenter, B.; Burkner, P.-C. Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC. *Bayesian Anal* 2021, doi:10.1214/20-BA1221.
25. Vehtari, A.; Gelman, A.; Gabry, J. Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC. *Stat. Comput.* 2017, *27*, 1413–1432, doi:10.1007/s11222-016-9696-4.
26. Watanabe, S. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *J Mach Learn Res* 2010, *11*, 3571–3594.
27. Wasserstein, R.L.; Lazar, N.A. The ASA Statement on P-Values: Context, Process, and Purpose. *Am. Stat.* 2016, *70*, 129–133, doi:10.1080/00031305.2016.1154108.
28. Greenland, S.; Senn, S.J.; Rothman, K.J.; Carlin, J.B.; Poole, C.; Goodman, S.N.; Altman, D.G. Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations. *Eur. J. Epidemiol.* 2016, *31*, 337–350, doi:10.1007/s10654-016-0149-3.
29. Amrhein, V.; Greenland, S.; McShane, B. *Scientists Rise up against Statistical Significance*; Nature Publishing Group, 2019;
30. Wasserstein, R.L.; Schirm, A.L.; Lazar, N.A. Moving to a World Beyond "p < 0.05." *Am. Stat.* 2019, *73*, 1–19, doi:10.1080/00031305.2019.1583913.
31. Rouder, J.N.; Haaf, J.M.; Vandekerckhove, J. Bayesian Inference for Psychology, Part IV: Parameter Estimation and Bayes Factors. *Psychon. Bull. Rev.* 2018, *25*, 102–113, doi:10.3758/s13423-017-1420-7.
32. Kruschke, J.K. Rejecting or Accepting Parameter Values in Bayesian Estimation. *Adv. Methods Pract. Psychol. Sci.* 2018, *1*, 270–280, doi:10.1177/2515245918771304.
33. Kruschke, J.K.; Liddell, T.M. The Bayesian New Statistics: Hypothesis Testing, Estimation, Meta-Analysis, and Power Analysis from a Bayesian Perspective. *Psychon. Bull. Rev.* 2018, *25*, 178–206, doi:10.3758/s13423-016-1221-4.
34. Gelman, A.; Vehtari, A.; Simpson, D.; Margossian, C.C.; Carpenter, B.; Yao, Y.; Kennedy, L.; Gabry, J.; Bürkner, P.-C.; Modrák, M. *Bayesian Workflow*; 2020;
35. Muth, C.; Oravecz, Z.; Gabry, J. User-Friendly Bayesian Regression Modeling: A Tutorial with Rstanarm and Shinystan. *Quant. Methods Psychol.* 2018, *14*, 99–119, doi:10.20982/tqmp.14.2.p099.

36. Vasishth, S.; Nicenboim, B.; Beckman, M.E.; Li, F.; Kong, E.J. Bayesian Data Analysis in the Phonetic Sciences: A Tutorial Introduction. *J. Phon.* 2018, *71*, 147–161, doi:10.1016/j.wocn.2018.07.008.
37. Bautista, J.R.; Pavlakis, A.; Rajagopal, A. Bayesian Analysis of Randomized Controlled Trials. *Int. J. Eat. Disord.* 2018, *51*, 637–646, doi:10.1002/eat.22928.
38. Schad, D.J.; Betancourt, M.; Vasishth, S. *Toward a Principled Bayesian Workflow in Cognitive Science*; 2020;
39. Zuur, A.F.; Ieno, E.N. A Protocol for Conducting and Presenting Results of Regression-Type Analyses. *Methods Ecol. Evol.* 2016, *7*, 636–645, doi:10.1111/2041-210X.12577.
40. Gabry, J.; Simpson, D.; Vehtari, A.; Betancourt, M.; Gelman, A. Visualization in Bayesian Workflow. *J. R. Stat. Soc. Ser. A Stat. Soc.* 2019, *182*, 389–402, doi:10.1111/rssa.12378.
41. Humberstone-Gough, C.E.; Saunders, P.U.; Bonetti, D.L.; Stephens, S.; Bullock, N.; Anson, J.M.; Gore, C.J. Comparison of Live High: Train Low Altitude and Intermittent Hypoxic Exposure. *J. Sports Sci. Med.* 2013, *12*, 394.
42. Gore, C.J.; Sharpe, K.; Garvican-Lewis, L.A.; Saunders, P.U.; Humberstone, C.E.; Robertson, E.Y.; Wachsmuth, N.B.; Clark, S.A.; McLean, B.D.; Friedmann-Bette, B.; et al. Altitude Training and Haemoglobin Mass from the Optimised Carbon Monoxide Rebreathing Method Determined by a Meta-Analysis. *Br. J. Sports Med.* 2013, *47*, i31, doi:10.1136/bjsports-2013-092840.
43. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* 2019, *4*, 1686, doi:10.21105/joss.01686.
44. Kassambara, A. *Ggpubr: "ggplot2" Based Publication Ready Plots*; 2020;
45. Makowski, D.; Ben-Shachar, M.S.; Lüdecke, D. BayestestR: Describing Effects and Their Uncertainty, Existence and Significance within the Bayesian Framework. *J. Open Source Softw.* 2019, *4*, 1541, doi:10.21105/joss.01541.
46. Gabry J, Mahr T *Bayesplot: Plotting for Bayesian Models*;
47. Vethari, A.; Gabry, J.; Magnusson, M.; Yao, Y.; Bürkner, P.; Paananen, T.; Gelman, A. *Loo: Efficient Leave-One-out Cross-Validation and WAIC for Bayesian Models.*; 2020;