

Testing Bayesian Informative Hypotheses in Five Steps With JASP and R



Sara Garofalo^{ORCID}, Gianluca Finotti, Matteo Orsoni^{ORCID},
 Sara Giovagnoli, and Mariagrazia Benassi

Department of Psychology, University of Bologna, Bologna, Italy

Advances in Methods and
 Practices in Psychological Science
 October-December 2024, Vol. 7, No. 4,
 pp. 1–23
 © The Author(s) 2024
 Article reuse guidelines:
 sagepub.com/journals-permissions
 DOI: 10.1177/25152459241260259
 www.psychologicalscience.org/AMPPS



Abstract

One of the most common applications of Bayes's theorem for inferential purposes consists of computing the ratio between the probability of the alternative and the null hypotheses via a Bayes factor, which allows quantifying the most likely explanation of the data between the two. However, the actual scientific questions that researchers are interested in are rarely well represented by the classically defined alternative and null hypotheses. Bayesian informative-hypothesis testing offers a valid and easy way to overcome such limitations via a model-selection procedure that allows comparing highly specific hypotheses formulated in terms of equality ($A = B$) or inequality ($A > B$) constraints among parameters. Although packages for testing informative hypotheses in the most used statistical software have been developed in recent years, they are still rarely used, possibly because their implementation and interpretation may not be straightforward. Starting from a brief theoretical overview of the Bayesian theorem and its applications in statistical inference (i.e., Bayes factor and Bayesian informative hypotheses), in this article, we provide two step-by-step tutorials illustrating how to test, interpret, and report in a scientific article Bayesian informative hypotheses using JASP and R/RStudio software for running a 2×2 analysis of variance and a multiple linear regression. The complete JASP files, R code, and data sets used in the article are freely available on the OSF page at <https://osf.io/dez9b/>.

Keywords

Bayes factor, Bayesian informative hypotheses, ANOVA, multiple linear regression

Received 6/2/24; Revision accepted 5/17/24

Imagine you booked a horseback riding experience for yourself and your loved ones. You gather all the equipment necessary to get you through this exciting day and drive your car toward the farm where an expert equestrian is waiting for you. Once you get there, you immediately start looking around for horses, but none seem to be at eye distance until you see something moving behind a bush. It is quite far away, so you start squinting in that direction and catch a glimpse of a donkey. Now, according to a famous joke (Senn, 2007), if you are a Bayesian-minded person, chances are you will conclude that you have seen a mule. Despite simplifying a lot, this funny story is useful to give an immediate picture of the core of Bayesian thinking. Indeed, from a Bayesian perspective, the probability of an event is given by both the evidence one is presented with (the donkey) and the expectation one has about the event before one observes

it (the horse). More precisely, the theorem proposed by the English mathematician and minister Thomas Bayes (Bayes, 1763) provides a formula to determine the conditional (or posterior) probability of an event, that is, the probability of an event given that another event has taken place. Despite appearing unintuitive, we apply this reasoning in many daily life circumstances. For instance, we have different expectations about getting rain if we are having a sunny or cloudy day. That is the case because people use the available information to create an expectation (Knill & Pouget, 2004).

Corresponding Author:

Sara Garofalo, Department of Psychology, University of Bologna, Bologna, Italy
 Email: sara.garofalo@unibo.it



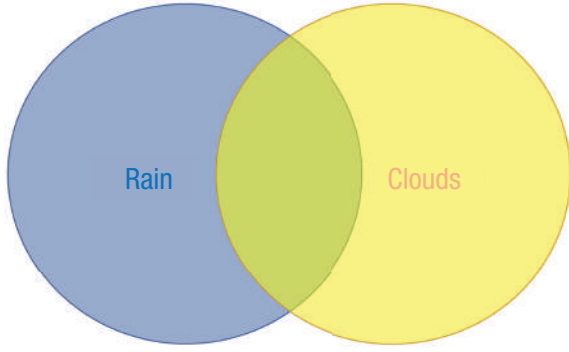


Fig. 1. Visualization of Bayes's theorem on Euler diagram. The blue (rain) and yellow (clouds) circles represent the individual probabilities of the two single events, and their intersection in green represents the conditional probability of the two events co-occurring.

Formally, the Bayes theorem is represented by the following formula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (1)$$

where A and B are two events, B has a probability different from 0, $P(A|B)$ is the conditional probability of A occurring given that B is true (posterior probability), $P(B|A)$ is the conditional probability of B occurring given that A is true (likelihood), $P(A)$ is the unconditional (i.e., not secondary to other events) probability of observing A (prior probability), and $P(B)$ is the unconditional (i.e., not secondary to other events) probability of B (marginal probability).

This formula can be read as follows: The probability of A given B equals the probability of B given A multiplied by the probability of A over the probability of B. This means that the conditional probability of the first event (A) when the second event (B) is true depends on the likelihood that the second event occurs when the first one occurs— $P(B|A)$ —and on the individual probabilities of the two single events— $P(A)$ and $P(B)$.

In more concrete terms, using Bayes's theorem, one could, for instance, calculate the probability of getting rain (A) when one sees clouds in the sky (B; Fig. 1). This could be achieved by taking into account the probability of seeing clouds when it rains (likelihood), the overall probability of getting rain (prior probability), and the overall probability of seeing clouds (marginal probability) in that season.

Applied to hypothesis testing, Bayes's theorem can be used to estimate the posterior probability of a hypothesis (H, or model) given the set of data (D, or evidence)

that one has collected, as shown in the following formula:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}, \quad (2)$$

where H indicates the hypothesis at hence and D the data collected. Equation 2 is identical to Equation 1 in all other respects.

In this scenario, the prior probability represents the expectation (or representation of uncertainty) regarding the truth status of the hypothesis.¹ The prior distribution is then combined with the information from the data (likelihood) to obtain the posterior distribution of the hypothesis, which will represent the probability of that hypothesis given the observed events (Wagenmakers, Marsman, et al., 2018). In other words, Bayes's theorem is used to go from the probability of the data given the hypothesis (e.g., how likely it is to observe these data if the hypothesis is true) to the probability of the model given the data (i.e., how likely the hypothesis is after having seen the data; Kruschke, 2014).

An increasingly popular application of Bayes's theorem for inferential purposes consists of providing a direct comparison between two different hypotheses by computing the ratio of the probability of one hypothesis over the other. This approach is known as the “Bayes factor” (BF), and it is typically used to compare the probability of the alternative hypothesis (H_1 , which assumes the presence of an effect or difference between parameters) with the probability of the null hypothesis (H_0 , which assumes the absence of an effect or difference between parameters; Kruschke & Liddell, 2018; Wagenmakers, Love, et al., 2018; Wagenmakers, Marsman, et al., 2018). The BF of the alternative hypothesis over the null hypothesis (BF_{10}) can be computed as follows:

$$BF_{10} = \frac{P(D|H_1)}{P(D|H_0)} = \frac{P(H_1|D)/P(H_0|D)}{P(H_1)/P(H_0)}. \quad (3)$$

Following the same formula, the ratio of the null over the alternative hypothesis (BF_{01}) can be obtained, too. For indications on how to interpret the BF, see Box 1.

Bayesian Informative Hypotheses

Despite presenting several advantages compared with other inferential approaches, the application of BFs as described above (BF_{10} or BF_{01}) still presents some limitations that can be overcome by testing informative hypotheses.

Box 1. Interpreting the Bayes Factor

Interpreting Bayes factors (BFs) involves assessing the strength of evidence (or predictive performance) of one hypothesis relative to another. For instance, $BF_{10} = 10$ indicates that H_1 is 10 times more likely than H_0 given this set of data. Likewise, $BF_{10} = 0$ indicates that both hypotheses are equally likely, whereas $BF_{10} < 0$ indicates that H_0 is more likely than H_1 . In the latter case, BF_{10} can also be reversed to BF_{01} to simplify the interpretation. Although some authors have suggested discrete criteria that characterize the strength of such evidence from weak ($1 > BF < 3$) to extreme ($BF > 100$; Andraszewicz et al., 2015; Jeffreys, 1961; van Doorn et al., 2021), we note that these numeric criteria are not universal and are subject to debate. One of the main advantages of the BF is, indeed, that it overcomes the dualistic reasoning typical of null hypothesis significance testing (accept/reject H_0) by quantifying and comparing the evidence in favor of one hypothesis over the other (van Doorn et al., 2021; Wagenmakers, Marsman, et al., 2018). For these reasons, the BF works at its best when interpreted on a continuous scale (Van Lissa et al., 2021). Even if $BF_{10} = 3$ is weaker evidence of H_1 prevailing over H_0 than $BF_{10} = 100$, it still indicates that H_1 is 3 times more likely than H_0 . Whether this evidence is enough to support a conclusion strongly depends on the research question and design. Researchers may adjust and comment on these thresholds based on the specific context of their study or field, the specific hypotheses being tested, and the practical significance of the findings.

Defining and testing meaningful research hypotheses

Especially in psychology and cognitive sciences, research hypotheses can be a lot more specific than the classically defined alternative ($A \neq B$) and null ($A = B$) hypotheses. Bayesian informative-hypothesis testing offers a valid and easy way to directly compare highly specific informative hypotheses (Béland et al., 2012; Garofalo et al., 2022; Gu et al., 2018; Hoijtink, 2012; Hoijtink et al., 2016), that is, hypotheses formulated to reflect research expectations in terms of equality (e.g., $A = B$) or inequality (e.g., $A > B$) constraints among the parameters (Gu et al., 2018; Hoijtink, 2012; Hoijtink, Mulder, et al., 2019).

For instance, although factorial designs are often used to test highly specific hypotheses in the form of

interaction effects, classical analysis of variance (ANOVA) is not necessarily the best way to go in such a case because the corresponding alternative and null hypothesis would not really reflect the precise research expectations (see Box 2; Garofalo et al., 2022). If, for example, one wants to test whether patients suffering from arachnophobia show higher anxiety levels when presented with spiders than cockroaches compared with a control group, this hypothesis can be represented as follows:

$$H_1 = (\mu_{\text{patients-spiders}} - \mu_{\text{patients-cockroaches}}) > (\mu_{\text{controls-spiders}} - \mu_{\text{controls-cockroaches}}).$$

Crucially, other hypotheses might be relevant to the researcher as well, such as the possibility that spiders

Box 2. Interaction Effect

The first hypothesis (H_1) reported as an example of informative hypothesis in the introduction and the second hypothesis (H_2) reported in Example A are, in fact, a mathematical formulation of an interaction effect. Especially in psychological and cognitive sciences, factorial experimental designs (e.g., a 2×2 analysis of variance) are frequently used to test very specific predictions (i.e., higher anxiety levels to spider vs. cockroaches in arachnophobia patients compared with a control group). However, using interaction effects—and follow-up post hoc analyses—should be considered as an exploratory approach, to be used when no specific hypotheses are present and the researcher aims to contrast all subgroups to investigate all possible relationships among them (Garofalo et al., 2022). When a specific research hypothesis is present, model-selection procedures such as Bayesian informative hypotheses can more powerfully test precise expectations (Degni et al., 2022; Garofalo et al., 2022) by enabling the definition of all relevant hypotheses in terms of inequality constraints among parameters and the comparison of their associated probability within a Bayesian inferential framework (Gu et al., 2018; Hoijtink, 2012; Hoijtink, Mulder, et al., 2019). For more details on the use and misuse of interaction effects and how to use Bayesian informative hypotheses in this context, see Garofalo et al. (2022).

Box 3. Definitions

Hu: Unconstrained hypothesis, a model that contains all possible sets of relationships between the parameters. Thus, it also contains the informative hypotheses defined by the researcher.

Hc: Complement hypothesis, a model that contains all possible sets of relationships between the parameters except the one represented by the hypothesis being tested. For example, the complement hypothesis of H_1 is not $\neg H_1$. Note that if a hypothesis is specified using equality constraints ($=$) alone or in addition to inequality constraints ($>$ or $<$), then Hc corresponds to Hu.

BFu: Bayes factor (BF) of any hypothesis (i) against its unconstrained hypothesis (u). It is given by the ratio between the two respective marginal likelihoods. This can also be seen as the ratio between the fit and the complexity of the hypothesis (see definitions below). In formula, $BF_{iu} = P(H_i | D) / P(H_u | D) = f_i / c_i$.

BFc: BF of each hypothesis versus its complement hypothesis (Hc), which is a model that contains all sets of restrictions between the parameters except the one represented by the current hypothesis.

Fit: It indicates the extent to which the data are in agreement with the restrictions specified in the hypothesis. It can range between 0 and 1; higher values indicate a higher fit and vice versa.

Com: Complexity indicates how specific the hypothesis is so that when two hypotheses fit the data equally well, the less complex hypothesis is preferred (Occam's razor principle; Hoijtink, 2012; Hoijtink, Mulder, et al., 2019). It can range between 0 and 1; higher values indicate a higher complexity and vice versa.

PMP: The posterior model probability associated with each hypothesis is, by default, calculated in three ways, that is, relative to all other hypotheses tested (**PMPa**), to all other hypotheses tested plus the unconstrained hypothesis Hu (**PMPb**), and to all other hypotheses tested plus the complement hypothesis Hc (**PMPc**).

Bayes factor matrix: The BF matrix consists of a table showing, in each cell, the BF deriving from the ratio between all possible pairs of hypotheses tested. For instance, the first cell in the first row shows BF_{11} , that is, H_1 over H_1 ; the second cell in the first row shows BF_{12} , that is, H_1 over H_2 ; and so on.

merely elicit higher anxiety levels than cockroaches regardless of being a patient or a control:

$$H_2 = (\mu_{\text{patients-spiders}}, \mu_{\text{controls-spiders}}) > (\mu_{\text{patients-cockroaches}}, \mu_{\text{controls-cockroaches}})$$

or that patients present a systematically higher response than controls:

$$H_3 = \mu_{\text{patients-spiders}} > \mu_{\text{patients-cockroaches}} > \mu_{\text{controls-spiders}} > \mu_{\text{controls-cockroaches}}$$

Informative hypotheses could go far beyond this, for instance, including how much bigger this difference should be to be considered clinically relevant (see Characteristics of the Informative Hypotheses section).

Bayesian informative-hypothesis testing allows one to compare and contrast a set of predefined hypotheses—similar to the one described above—via a model-selection procedure in which each hypothesis (or model) represents a possible explanation of the phenomenon.

For each model, the associated posterior model probability (PMP) is calculated via the BFs of all hypotheses tested versus the so-called unconstrained hypothesis (Hu; see Box 3). Specifically, considering the three

hypotheses reported above, the formula for the PMP for H_1 would be

$$PMP_{H1} = \frac{BF_{1u}}{(BF_{1u} + BF_{2u} + BF_{3u})}. \quad (4)$$

PMPs are calculated as the ratio between the BF of the hypothesis at hence versus the Hu (BF_{1u}) over the sum of all BF_{iu} , where i indicates all informative hypotheses tested. Because it is a relative index, PMPs are expressed as a value ranging between 0 and 1 (the sum of all posterior model probabilities adds up to 1), which can be interpreted as the relative amount of support for each hypothesis given the data at hand and the set of competing hypotheses included. The model with the highest PMP reflects the hypothesis with the highest relative probability (Béland et al., 2012; Hoijtink, 2012; Hoijtink, Gu, & Mulder, 2019; Hoijtink, Mulder, et al., 2019; Kluytmans et al., 2012). All PMPs are based on equal prior probabilities because it is assumed that each hypothesis is equally likely.

Having a fail-safe test

Another advantage of Bayesian informative-hypothesis testing consists in the possibility to check whether there

are other—perhaps even more—plausible explanations for data compared with the ones hypothesized. Although the BF between any two specific hypotheses (e.g., BF_{10} or BF_{12}) merely indicates that one is more likely than the other, it says very little about how well the data are predicted by the prevailing hypothesis (van Doorn et al., 2021). In other words, it provides no information about the possibility that other sets of relationships among the parameters may be a better fit for the data.

To prevent this loss of information, PMPs are by default calculated not only for the informative hypothesis defined by the researcher but also for two additional models that can inform the inferential decision, that is, the unconstrained hypothesis (H_u) and the complement hypothesis (H_c ; see Box 3). Both H_c and H_u can be intended as a fail-safe test (Béland et al., 2012; Hoijtink, Mulder, et al., 2019; Van Lissa et al., 2021) because they will prevail over the hypotheses specified by the researcher if a better explanation is possible that has not been currently considered (see Box 4).

Characteristics of the informative hypotheses

We offer a few remarks that may be useful to understand how to define informative hypotheses. First, it is critical to understand that the formulation of a model representing the null hypothesis (e.g., $\mu_{\text{patients-spiders}} = \mu_{\text{patients-cockroaches}} = \mu_{\text{controls-spiders}} = \mu_{\text{controls-cockroaches}}$) is not mandatory and, as for any other model, should be included only if meaningful from a scientific point of view (Béland et al., 2012; Hoijtink, 2012; Hoijtink, Mulder, et al., 2019). Indeed, including a large number of hypotheses increases the risk that researchers will end up selecting the hypothesis that best fits the sampled data rather than one that best describes the population (for more information, see Box 4; Hoijtink, Mulder, et al., 2019).

Moreover, although in the following examples we define fairly simple models, more complex hypotheses can be tested. For instance, one may be interested in defining a successful reduction in anxiety levels only if the difference between two means is higher than 2 *SD* from the mean:

$$H_4: (\mu_{\text{patients-spiders}} - \mu_{\text{controls-spiders}}) > \mu + 2 \text{ SD.}$$

Another possibility is to use the ampersand (“&”) to combine two predictions in one hypothesis. For example, one may want to ensure that there is a positive reduction in anxiety levels (> 0), rather than just a mere difference, in patients presented with cockroaches

versus spiders in addition to the fact that such reduction is stronger in patients than controls:

$$\begin{aligned} H_5 = & (\mu_{\text{patients-spiders}} - \mu_{\text{patients-cockroaches}}) \\ & > 0 \ \& \ (\mu_{\text{patients-spiders}} - \mu_{\text{patients-cockroaches}}) \\ & > (\mu_{\text{controls-spiders}} - \mu_{\text{controls-cockroaches}}). \end{aligned}$$

On a final note, an important requirement is that all hypotheses included have to be compatible, possible, and nonredundant. In brief, this mainly means that the resulting set of equations must have a possible solution (e.g., one cannot expect $\mu_A > \mu_C$ and $\mu_A < \mu_C$) and that the hypotheses have to be specified with the smallest possible number of constraints (e.g., in H_2 , there is no need to specify that $\mu_C - \mu_D > 0$ because $[\mu_A - \mu_B] = [\mu_C - \mu_D]$). For a deeper understanding of this concept, see Gu et al. (2018) and Mulder et al. (2010). For the purpose of the present tutorial, it will suffice to say that if the hypotheses specified are not compatible, an error warning will be displayed in both JASP and R/RStudio so that an incorrect use of the analysis is prevented.

A model-comparison approach

A BF is no other than a model-selection criterion within a Bayesian inferential framework (Hoijtink, 2012). Like many other model-comparison approaches (e.g., those based on Akaike information criterion or Bayesian information criterion), it presents several advantages compared with classical hypothesis testing, including those described in the following.

Direct comparison of multiple models. Multiple models are evaluated simultaneously, and the model with the best fit is chosen. This approach allows researchers to directly compare the goodness of fit of multiple models and choose the one that provides the best explanation for the data, which can improve the accuracy and interpretability of the results. In contrast, hypothesis testing typically considers only one null hypothesis at a time.

Increased generalizability. By comparing multiple models, researchers can determine which models are more likely to generalize well to new data, which is important for making accurate predictions and drawing reliable conclusions.

Identify important variables. By comparing models that include different subsets of variables, researchers can determine which variables are most important in explaining the data and how they interact with one another.

Box 4. Interpreting the Results of Informative Hypotheses

There are at least three ways in which the posterior model probabilities (PMPs) and Bayes factors (BFs) provided in the results can inform inferential conclusions (Hojtink, Mulder, et al., 2019; Van Lissa et al., 2021). Put another way, there are three main questions that a researcher can answer with Bayesian informative hypotheses (see Box 3 for definitions):

1. Which of a set of hypotheses is the best?

The first question can be answered by looking at the relative proportion of PMP associated with all other hypotheses. If the PMP associated with H_u or H_c is higher than that associated with the informative hypotheses defined by the researcher, this indicates that the hypotheses tested do not constitute the best representation of the data. More precisely, because H_u also contains the hypothesis tested, H_c is the most informative for this purpose; hence, to answer this first question, we want to look at PMPc.

If the greatest proportion of PMPc is associated with one of the researcher-defined informative hypotheses, this can be identified as the hypothesis most supported by the data relative to all other hypotheses. Whereas if the portion of PMPc associated with H_c is higher than that associated with the researcher-defined informative hypotheses, the most likely conclusion is that the proposed hypotheses do not constitute an adequate description of the data, which could find a better fit in other models not currently considered.

Nevertheless, not all is lost if that is the case: (a) If H_c prevails, this information can be useful to further explore the data set and define new hypotheses that—and this is crucial to avoid hypothesizing after results are known (HARKing; Kerr, 1998)—shall be tested in future experiments; and (b) a comparison between the informative hypotheses defined by the researcher is still possible because although better descriptions of the data may exist, the researcher may still be particularly interested in a comparison between those specific hypotheses to check whether one may prevail. PMPa may be useful for this purpose (see Question 3 in this Box).

2. How much more likely a given hypothesis is relative to other possible explanations?

Bayes factors (BFs) can be used to quantify the strength of support for each hypothesis (or perhaps, just the best one) relative to the unconstrained (BF_u) and the complement (BF_c) hypotheses. Because H_u —and thus BF_u—also contains the hypothesis tested, BF_c is the most informative for inferential purposes. BF_c indicates how much more likely a given hypothesis is compared with any other set of relationships or restrictions among the parameters (i.e., any other possible model). For instance, $BF_{1c} = 30.5$ indicates that H_1 is 30.5 times than H_c , that is, 30.5 times more likely than other possible hypotheses. On the other hand, $BF_{1c} = 0.5$ indicates that H_c is more likely than H_1 , thus suggesting that other relationships among the parameters may be a better fit for the current data. See Box 1 for more on the interpretation of the BFs.

3. How much more likely a given hypothesis is relative to another specific hypothesis?

BFs can also be used to compute a direct comparison between each pair of informative hypotheses tested by the researcher (thus, not considering H_u and H_c). This information is usually reported in the so-called BF matrix, which is a table reporting the BFs between each pair of hypotheses (e.g., BF_{12} , BF_{13} , BF_{23}), calculated as the ratio between the two corresponding BF_us (e.g., $BF_{12} = BF_{1u} / BF_{2u}$). Such a comparison is possible because all BFs have the same denominator. This can be useful to quantify how much more likely a given hypothesis is compared with the other tested hypotheses. For instance, $BF_{12} = 5$ indicates that H_1 is 5 times more likely than H_2 .

Increased transparency. Model comparison provides a clear and objective way to evaluate the performance of different models, which can increase the transparency of research and help to prevent the use of inappropriate or biased models.

Flexibility. Model comparison allows for the use of a wide range of modeling techniques and assumptions, which can be tailored to specific research questions and data types.

Overcome dualistic thinking. Hypothesis testing typically provides a binary answer (reject or fail to reject the null hypothesis), whereas model comparison can provide more nuanced results. For example, model comparison can provide information on the relative importance of different variables or the degree to which a model explains the data.

Note that both model comparison and classical hypothesis testing can be useful in different situations. Researchers should choose the approach that best suits their research question and data.

Model assumptions

An important question concerns the model assumptions (e.g., normality, sphericity, independence) that the data must fulfill to ensure the correct use and interpretation of the results. Although there is still an ongoing debate (Hojtink, Mulder, et al., 2019; Van Rossum et al., 2013), current evidence indicates that BFs are robust if model violations are not too extreme. Nevertheless, they are still sensitive to their effect. The reader is advised to test in advance whether model assumptions (for ANOVA or regression in the context of this tutorial, but the same applies to any other statical model) are fulfilled. For more details and practical guidance on how to deal with outliers and violations of model assumptions, see Hoijtink, Mulder, et al. (2019).

Supported Statistical Models and Other Tutorials

At the time of writing, the *bain* (short for Bayesian informative hypothesis) package in R can be applied to a wide range of statistical models, including Welch's *t* test (paired samples and one sample), ANOVA (within-subjects and between-subjects designs), analysis of covariance (ANCOVA), logistic regression, linear regression, structural equation modeling, and confirmatory factor analysis (Gu et al., 2014, 2019; Hoijtink, Gu, & Mulder, 2019; Hoijtink, Mulder, et al., 2019; Van Lissa et al., 2021). The *bain* package in JASP can be currently applied to a smaller number of models: Welch's *t* test (paired samples and one sample), ANOVA (between-subjects design only), ANCOVA, linear regression, and structural equation modeling. Note that both packages are still being developed, and more applications may be available in the near future.

Currently, there are two tutorial articles on Bayesian informative hypotheses. Hoijtink, Mulder, and colleagues (2019) covered the topic of one-way ANOVA, and Van Lissa and colleagues (2021) focused on structural equation modeling; both provided examples in R programming language. In the present article, we aim to extend previous tutorials in several ways by (a) covering new models that are frequently used in psychological and cognitive sciences, such as factorial ANOVA and multiple linear regression; (b) providing a gradual introduction from a more general way of Bayesian thinking to the use of BFs, reaching the definition and implementation of Bayesian informative-hypothesis testing; (c) showing how to test Bayesian informative-hypothesis testing in both JASP (JASP Team, 2021) and R/RStudio (Gu et al., 2021), illustrating the same five steps in both software; (d) starting from scratch, thus allowing a wider readership to familiarize with these analyses, even with no prior knowledge of either software or Bayesian testing;

(e) providing practical guidance not only on how to run the analysis but also on how to interpret and report the results in a research article, with a dedicated section for each example; and (f) guiding toward a visual exploration of the data and/or get novel insights.

Downloading and Installing JASP and R/RStudio

The very first step consists of installing the chosen software, in this case, either JASP (JASP Team, 2021) or R (R Core Team, 2023) and R studio (RStudio Team, 2023). Please ensure to download and install the latest version available of the software that you intend to use before proceeding.

JASP is a free software for statistical computing and graphics that offers a very intuitive graphical user interface. Compared with R, it is easier to use because it does not require programming skills; however, it offers a more limited set of analyses. To familiarize yourself with JASP's environment, there are many online tutorials available on the software website. The JASP installer can be downloaded at <https://jasp-stats.org/download/>.

R is a free open-source programming language for statistical computing and graphics. Installing R is necessary to enable your computer to process the R programming language in R and RStudio. RStudio is an integrated development environment that offers a more convenient graphical user interface and additional functionalities compared with R. We recommend working in RStudio rather than R, but this is not mandatory. In this article, we assume basic knowledge of R programming language. We suggest familiarizing with its functioning with one of the many tutorials available online or using the *swirl* package (<https://swirlstats.com/>). The R installer can be downloaded at <https://cran.r-project.org/bin/>. The RStudio Desktop installer can be downloaded at <https://rstudio.com/products/rstudio/download/>.

Open Materials

The complete R code, JASP files, and data sets used in the article are freely available, following FAIR (findable, accessible, interoperable, and reusable) data principles (Wilkinson et al., 2016), at the following OSF page: <https://osf.io/dez9b/>.

Example A: 2 × 2 ANOVA

Let us assume we want to test the efficacy of a new anxiety treatment on different levels of symptoms. To this aim, we recruit 200 volunteers suffering from anxiety-related disorders equally divided into two groups, one presenting high levels of symptoms and one presenting

low levels of symptoms. We then randomly assign half of each group to the experimental group (receiving a real drug treatment) and the other half to the control group (receiving a placebo treatment). We measure the anxiety level before and after the treatment, and we compute an index of treatment efficacy by subtracting the score recorded before the treatment from the score recorded after the treatment (pre–post) so that a positive score would indicate an improvement (i.e., a reduction) in anxiety, a negative score would indicate an increase in anxiety, and 0 indicates no difference between before and after the treatment.

We thus have a 2×2 factorial design with treatment (drug/placebo) and symptoms (high/low) as between-subjects independent variables and the treatment efficacy index as the dependent variable.

The ANOVA data set

The data set is composed of five columns: (a) “ID” (containing participants’ identification number), (b) “treatment” (drug/placebo), (c) “symptoms” (high/low), and (d) “score” (containing a continuous variable reporting the treatment efficacy score). Note that because the *bain* package requires all the levels of a factorial design to be stored in one single variable, an additional column named (e) “groups” containing all four groups (i.e., drug high, drug low, placebo high, and placebo low) is needed.

The data were simulated by sampling 50 cases from a normal distribution using different means and standard deviations ($M \pm SD$) for each group, that is, 7 ± 1.5 for the drug group with high symptoms, 6 ± 1.5 for the drug group with low symptoms, 0 ± 1.5 for the placebo group with high symptoms, and 6 ± 1.5 for the placebo group with low symptoms. This data set can be reproduced using the “creating the datasets.R” file available on the OSF page (<https://osf.io/dez9b/>).

The hypotheses

Let us assume we want to compare and contrast a set of three hypotheses. The first hypothesis poses that the two groups receiving a real treatment (drug) show a positive increase (> 0) in the treatment efficacy index and (&) that the strength of such treatment, intended as the difference between the two groups (drug/placebo), is similar regardless of symptoms level (high/low). This hypothesis can formally be represented as follows:

$$H_1: (\mu_{\text{Drug.High}}, \mu_{\text{Drug.Low}}) > 0 \ \& \ (\mu_{\text{Drug.High}} - \mu_{\text{Placebo.High}}) \\ = (\mu_{\text{Drug.Low}} - \mu_{\text{Placebo.Low}}).$$

The second hypothesis poses again that the two groups receiving a real treatment (drug) show a positive

increase (> 0) in the treatment efficacy index and (&) that the strength of such treatment, intended as the difference between the two (drug/placebo), is higher for the group with high symptoms compared with the group with low symptoms:

$$H_2: (\mu_{\text{Drug.High}}, \mu_{\text{Drug.Low}}) > 0 \ \& \ (\mu_{\text{Drug.High}} - \mu_{\text{Placebo.High}}) \\ > (\mu_{\text{Drug.Low}} - \mu_{\text{Placebo.Low}}).$$

The third hypothesis poses that there are comparable scores regardless of treatment and symptom level:

$$H_3: \mu_{\text{Drug.High}} = \mu_{\text{Placebo.High}} = \mu_{\text{Drug.Low}} = \mu_{\text{Placebo.Low}}.$$

For more information on how to define informative hypotheses, see the Characteristics of the Informative Hypotheses section.

Step-by-step tutorial in JASP

The analysis presented in this section is performed using the *bain* module on JASP (JASP Team, 2021). The file “bain_ANOVAinJASP.jasp” available on the OSF page (<https://osf.io/dez9b/>) contains all the steps reported below and can be directly opened in JASP.

Preliminary steps. To run Bayesian informative hypothesis (*bain*) in JASP, you first need to add the *bain* module to the software. To do that, open JASP, click on the “+” icon at the top right (do not confuse it with the “+” sign to “add a computed column”), and select the “BaIn” module. This will now add a “BaIn” icon to the top menu (Fig. 2). After loading the data set (Step 1a), you can click on the icon to see the list of all the *bain* analyses currently available (Fig. 3).

It is important to know that many analyses require generating a series of random numbers for computational purposes. The *bain* package, in particular, uses sampling to compute BFs and PMPs. To get reproducible results, it is possible to set a specific seed number from which the series of pseudorandom numbers is generated. Knowing the seed and the generator makes it always possible to reproduce the same output. Otherwise, you can change it to any number you like. In the “Additional Options” section, you will see that by default, the seed is set to “100.” A good practice to ensure the stability of your results is to repeat the analysis with different seeds and check whether the results are coherent. For this reason, we use different seeds in the JASP and R examples and ensure that although slightly different, the same trend should emerge from both analyses. To match the results obtained with both software, set both seeds to the same value.

Step 1a: load the data set. The first step is to load a file containing the ANOVA data set described, available on the



Fig. 2. Adding the “BaIn” module to JASP.

OSF page (<https://osf.io/dez9b/>). To do that, go to the main-menu icon on the top left, select “Open → Computer,” browse to the folder on your computer in which you downloaded the file, and then select and load the “dataset_anova.txt” file. You will now visualize a spreadsheet containing the entire data set (Fig. 3).

Step 2a: fit the model. By selecting “BaIn → ANOVA” in the top menu (Fig. 3), you will be presented with a new graphical user interface that can be used to select the variables of interest (Fig. 4). Set the variable “score” as the dependent variable and the variable “group” as the fixed factor.

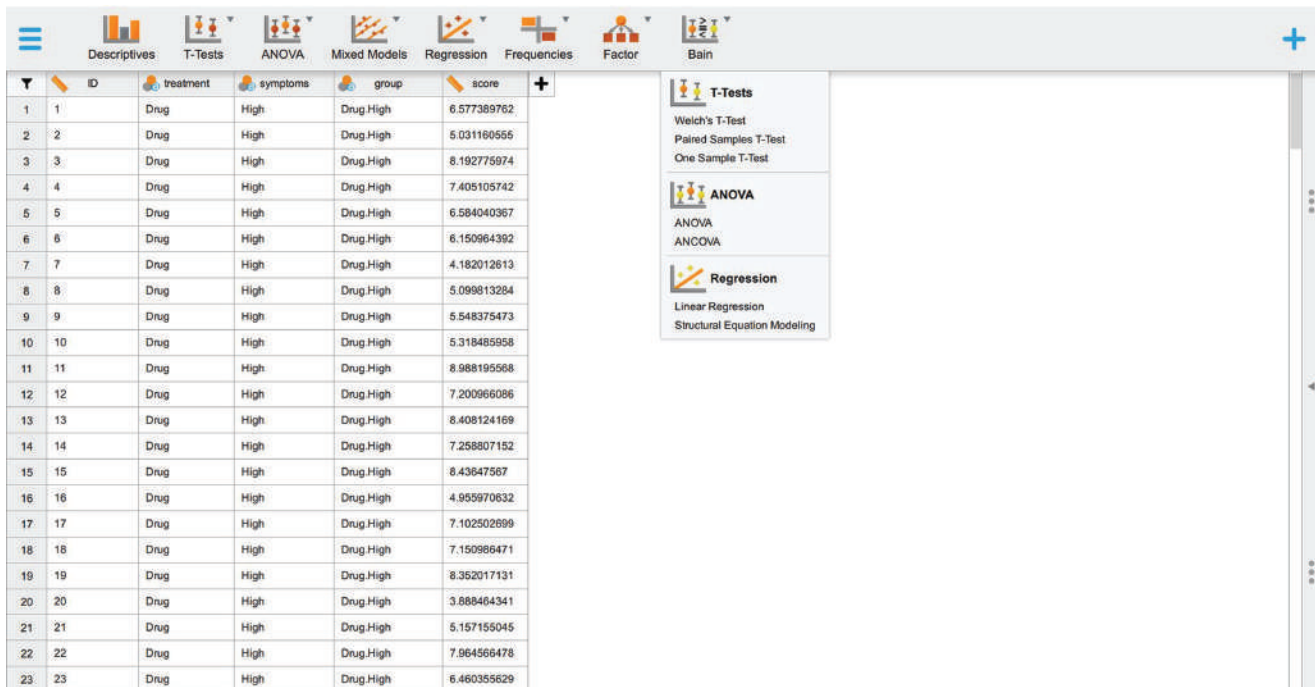


Fig. 3. Loading the data set and selecting the analysis.

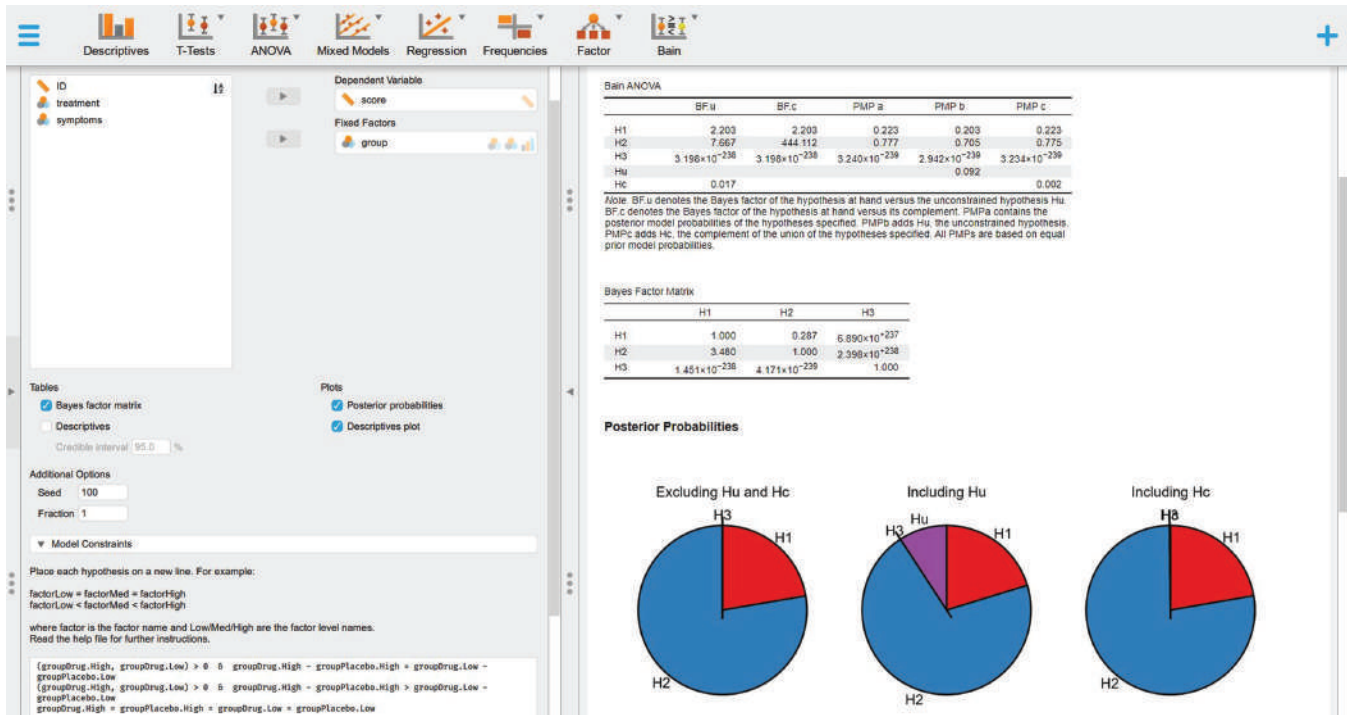


Fig. 4. Selecting the variables, defining the informative hypothesis, and visualizing the results.

Step 3a: define and test informative hypotheses with bain. We can now set the hypotheses defined above and compare them. To do that, you can type or copy and paste the text reported below in the “Model Constraints” box that you can find on the bottom of the left panel (Fig. 4):

```

(groupDrug.High, groupDrug.Low) > 0 &
groupDrug.High - groupPlacebo.High =
groupDrug.Low - groupPlacebo.Low
(groupDrug.High, groupDrug.Low) > 0 &
groupDrug.High - groupPlacebo.High >
groupDrug.Low - groupPlacebo.Low
groupDrug.High = groupPlacebo.High =
groupDrug.Low = groupPlacebo.Low
    
```

Then press Ctrl+Enter (if you use a Windows PC) or Cmd+Return (if you use a Mac). JASP will start testing the informative hypothesis specified, and the results will be presented on the right panel, where a table called “Bain ANOVA” will appear (also reported in Table 1), containing all the information outlined in Box 3.

For more complete results, check the boxes on the left panel (Fig. 4) to display the “Bayes factor matrix” table (Table 2) and the “Posterior probabilities” plot (Fig. 5), which provide a pie chart of the proportion of the three PMPs reported in Table 1.

Step 4a: interpret and report the results. In the following, we show how to interpret and report these results based on the indications outlined in Box 4.

Table 1. Bain Analysis of Variance

	BF _u	BF _c	PMP _a	PMP _b	PMP _c
H ₁	2.21	2.21	0.221	0.201	0.221
H ₂	7.778	438.875	0.779	0.708	0.777
H ₃	3.20 × 10 ⁻²³⁸	3.20 × 10 ⁻²³⁸	3.20 × 10 ⁻²³⁹	2.91 × 10 ⁻²³⁹	3.20 × 10 ⁻²³⁹
H _u				0.091	
H _c	0.018				0.002

Note: BF_u = Bayes factor versus H_u; BF_c = Bayes factor versus H_c; PMP_a = posterior model probability excluding H_u and H_c; PMP_b = posterior model probability including H_u; PMP_c = posterior model probability including H_c. Because the values reported here are particularly high (e+) or particularly low (e-), scientific notation is used for convenience.

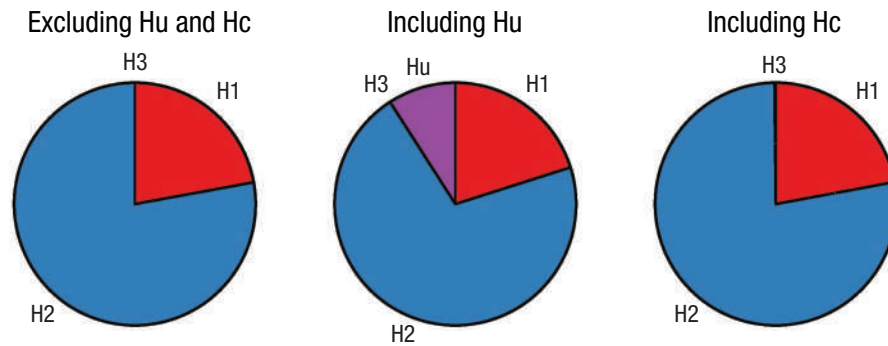


Fig. 5. Posterior model probabilities (PMPs). Posterior model probabilities associated with the set of hypotheses tested (PMPa), also including Hu (PMPb) or Hc (PMPc).

All results obtained point in the same direction, indicating that H_2 is the hypothesis (or model) most supported by the data. H_2 is indeed associated with a largely higher relative PMP (Fig. 5) not only when considering the set of informative hypotheses tested (77.9%; Table 1: PMPa) but also when including Hc (77.7%; Table 1: PMPc). This last result, in particular, indicates that other models would not best represent the data (see Box 4, Question 1: Which of a set of hypotheses is the best?). In addition, H_2 resulted 438.857 times more likely than Hc (Table 1: BFc), thus showing strong support for this hypothesis compared with other possible models (see Box 4, Question 2: How much more likely a given hypothesis is relative to other possible explanations?). Finally, H_2 resulted 3.519 times more likely than H_1 (Table 2: BF_{21}) and $2.432 \times 10^{+238}$ times more likely than H_3 (Table 2: BF_{23}), thus showing stronger support for H_2 relative to the other hypotheses tested (see Box 4, Question 3: How much more likely a given hypothesis is relative to another specific hypothesis?). Overall, these results indicate strong support for a higher treatment efficacy in the group presenting high levels of symptoms compared with the one presenting low levels of symptoms.

Step 5a: visual exploration of the data. Visually informative plots, such as estimation plots (Ho et al., 2019), raincloud plots (Allen et al., 2021), or plots representing model-estimated marginal means with confidence intervals (Garofalo et al., 2022), offer an immediate representation of the results based on a measurement scale that makes sense for the research question (Cumming, 2007, 2009; Cumming & Finch, 2005; Degni et al., 2022; Garofalo et al., 2022; Loftus & Masson, 1994; Masson, 2007). Thus, they can support data interpretation and possibly suggest new insights if our hypotheses result in a poor explanation of the data (see Box 4). Based on the software adopted, there are different options available to the user.

In JASP, by clicking on the relative boxes available on the left panel, it is possible to obtain a “Descriptives” table and a “Descriptive plots,” respectively, containing descriptive statistics in text or visual format.

The results thus generated (Fig. 6) confirm that although the strongest difference is predictably between the two drugs and the two placebo groups, there is a higher difference between the drug and placebo treatments in the high-symptoms group compared with the low-symptoms group.

Step-by-step tutorial in R

There are two ways to proceed with this example: (a) Open a new R script and use the following lines of code to play along with the tutorial or (b) open and use the file “BaIn_ANOVAinR.R” available on the OSF page (<https://osf.io/dez9b/>), which contains all the steps and code reported below.

Preliminary steps.

Clear workspace. Before starting any new analysis, it is generally considered good practice to clear all the variables, data sets, functions, and so on possibly loaded in the R environment. To do that, you can use the `rm()` function as follows:

```
rm(list=setdiff(ls(), ". . ."))
```

Table 2. Bayes Factor Matrix

	H_1	H_2	H_3
H_1	1	0.284	6.91×10^{-237}
H_2	3.519	1	2.43×10^{-238}
H_3	1.45×10^{-238}	4.11×10^{-239}	1

Note: Because the values reported here are particularly high (e+) or particularly low (e-), scientific notation is used for convenience.

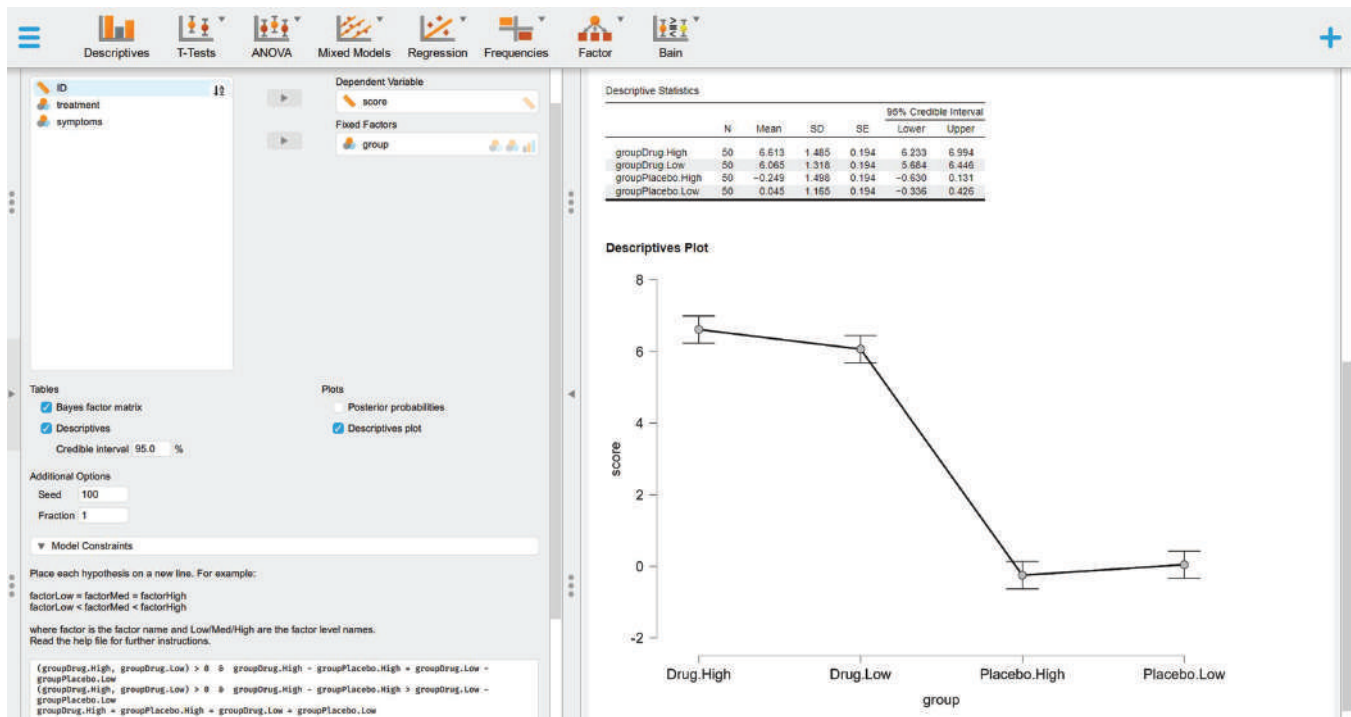


Fig. 6. Visual exploration of the data in JASP.

Install and/or load the necessary packages. Once R is installed, it comes with several built-in functions, such as `sum()`, which returns the sum of the values inserted between the brackets, or `sqrt()`, which returns the square root of all the values present in its arguments. To know what a function is used for, you can type “?” before the function name (i.e., `?sum`) in the R console to open the help. Sets of functions that work together are usually grouped and contained in so-called packages (or libraries). For instance, `sum()` and `sqrt()` are part of the *base* package, which is automatically loaded when you use R. Packages for more specific analysis or plots have to be directly installed and loaded. For this tutorial, you need to install and load the *bain* package (Hojtink, Mulder, et al., 2019), which contains all the functions needed to run Bayesian informative-hypotheses analyses. To do so, you can use the `install.packages()` function by running in R or RStudio the following line:

```
install.packages("bain")
```

This line of code is needed only if you have not already installed this package. Once a package is installed on your computer, you can use the `library()` function to load it and access its functions. This loading needs to be done every time you start a new R session.

```
library(bain)
```

These two steps are required each time you need to use functions that are not part of base R but need to be loaded from an external package. In this tutorial, a few more packages will be used to generate plots or manipulate the data. These packages are not strictly required to run Bayesian informative hypotheses but can be very useful for several purposes. Each time, we suggest one of these packages, we explain their purpose and how to use them. However, the same results could be obtained with different packages.

Step 1a: load the data set. There are several options to load the file containing your data. The easiest one is to use the RStudio graphical user interface by clicking on “File → Import dataset” and choosing your file type (e.g., text, excel, SPSS; see Fig. 7). For this example, the data set is saved in a .txt file named “dataset_anova.txt.” To open it, go to “File → Import dataset → From Text (base)” and browse to the folder on your computer that contains the “dataset_anova.txt” file (downloaded from OSF page <https://osf.io/dez9b/>) and then click on “Import” to load the file. If you accept the default values, your data set will now appear in the R Environment with the same name as the original file name.

If you prefer to use R code for loading a file, there are dedicated packages containing functions that can be used based on the file extension (e.g., a .txt file can be loaded with the `read.table()` function). An overview of

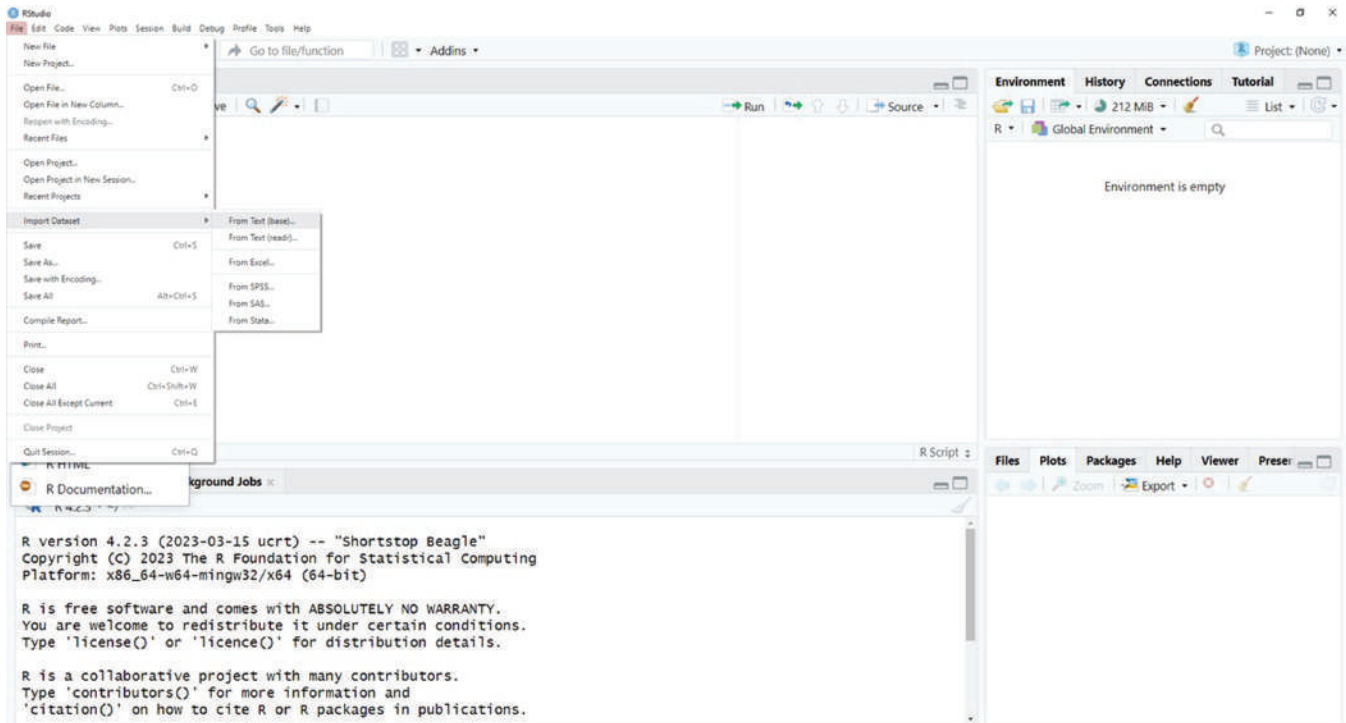


Fig. 7. Loading the data set from the graphical user interface of RStudio.

such functions would go beyond the scope of this tutorial, but in the R script available online, you will find a few commented lines that can be used for this purpose.

Once the file has been loaded, an object named “dataset_anova” should appear in your R environment. By clicking on this object, the entire data set will appear (Fig. 8) containing the variables described above in the ANOVA Data Set section. The `str()` function can be used to familiarize with your data set. This function shows the structure of your data set and has the additional benefit of showing the data type for each variable:

```
str(dataset_anova)
```

Step 2a: fit the model. The following code can be used to fit a linear model, with the function `lm()`, which returns estimated means for each level contained in the column “group.” The outcome of the fitted model will be stored in an object called “fit.” Note that the `-1` is used to estimate the means for each level of the variables. The subsequent line uses the `coef()` function to store the means estimated by the model for each group and session in an object called “estimated”:

```
fit = lm(score ~ group - 1, data =
  dataset_anova)
```

Step 3a: define and test informative hypotheses with bain. We can now set the hypotheses defined above and compare them. Before doing that, you can use the `coef()` function to take a look at the estimated means and their names because these are the variable names to use to define the informative hypotheses:

```
coef(fit)
```

It is also important to know that many analyses require generating a series of random numbers for computational purposes. The *bain* package, in particular, uses sampling to compute BFs and PMPs. To get reproducible results, it is possible to use the `set.seed()` function, which sets a specific seed number from which the series of pseudorandom numbers is generated. Knowing the seed and the generator makes it always possible to reproduce the same output. It can be set to any number you like. A good practice to ensure the stability of your results is to repeat the analysis with different seeds and check whether the results are coherent. For this reason, we may use different seeds in the JASP and R examples and ensure that although slightly different, the same trend should emerge from both analyses. The following code can be used to set the seed to 123:

```
set.seed(123)
```

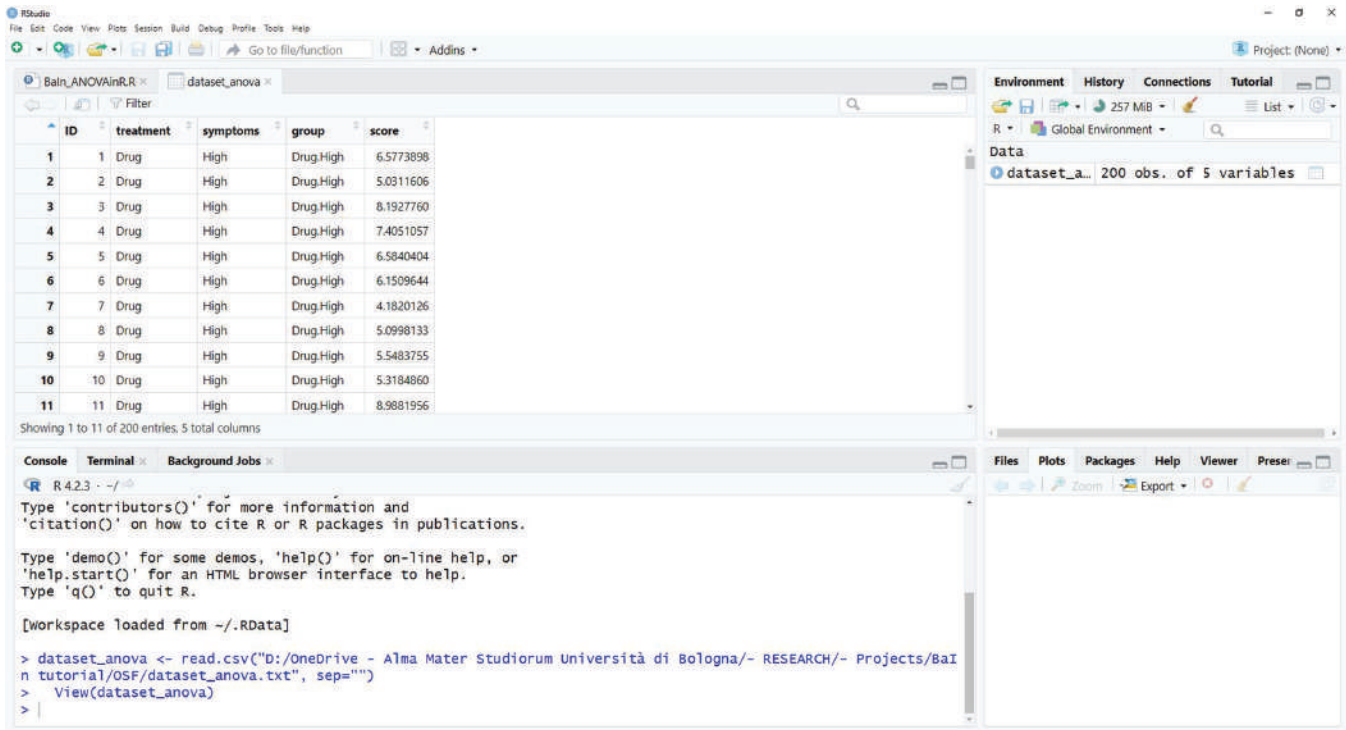


Fig. 8. Data set loaded in RStudio.

The following code can be used to test the informative hypotheses described above. The output is stored in an object called “results”:

```

results <- bain(x = fit,
               hypothesis =
               "(groupDrug.High, groupDrug.Low) > 0 &
                groupDrug.High - groupPlacebo.High =
                groupDrug.Low - groupPlacebo.Low;
                groupDrug.High, groupDrug.Low) > 0 &
                groupDrug.High - groupPlacebo.High
                > groupDrug.Low - groupPlacebo.Low;
                groupPlacebo.High = groupDrug.High =
                groupPlacebo.Low = groupDrug.Low")
  
```

In this code, the first argument “x” is a vector containing the estimated means for each group and condition, and the second argument “hypothesis” contains the hypotheses.

There are a few options available to display the results. The function print() returns in the console the main results or bain ANOVA table (Table 3):

```
print(results)
```

This table is identical to Table 1 reported above in Step 4a except for the presence of fit (Fit) and complexity (Com) scores. For more information and definitions, see Box 3.

A pie chart displaying the three PMPs reported in Table 1 can be created using the pie() function or similar

Table 3. Bain Analysis of Variance

	Fit	Com	BFu	BFc	PMPa	PMPb	PMPc
H ₁	0.097	0.044	2.212	2.212	0.225	0.204	0.224
H ₂	0.986	0.129	7.637	472.992	0.775	0.704	0.774
H ₃	0.000	0.012	0.000	0.000	0.000	0.000	0.000
H _u						0.092	
H _c	0.014	0.871	0.016				0.002

Note: Hu = unconstrained hypothesis; Hc = complement hypothesis; fit = model fit; com = model complexity; BFu = Bayes factor versus Hu; BFc = Bayes factor versus Hc; PMPa = posterior model probability excluding Hu and Hc; PMPb = posterior model probability including Hu; PMPc = posterior model probability including Hc.

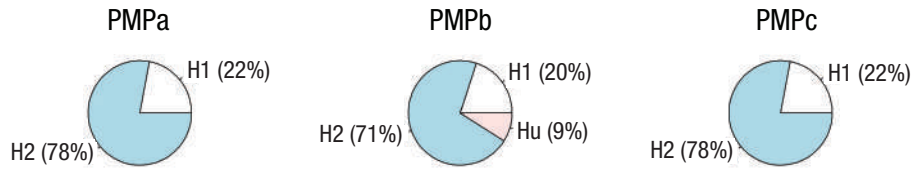


Fig. 9. Posterior model probabilities (PMPs). PMPs associated with the set of hypotheses tested (PMPa), also including Hu (PMPb) or Hc (PMPc). The code used to generate this figure is available on the R script associated with this example.

ones (Fig. 9). A detailed explanation of how to generate plots in R falls beyond the scope of the present article; however, the code used to generate Figure 9 is available on the R script associated with this example.

To further support model selection, PMPs can also be compared via a BF matrix (Box 3) in which all hypotheses are compared with each other (Table 4). The following code can be used to extract this information from the “results” object created in the previous steps:

```
results$BFmatrix
```

Step 4a: interpret and report the results. Please refer to Step 4a of the JASP example above for a complete guide on how to interpret and report these results.

Step 5a: visual exploration of the data. Please refer to Step 4a of the JASP example reported above for more on the importance of visual exploration of the data and how to interpret the results obtained.

In R/RStudio, in particular, several packages can be used for this purpose. Figure 10 shows a raincloud plot generated with the ggplot() and geom_rain() functions from the ggplot2 and ggrain packages (Allen et al., 2021). A detailed explanation of how to generate plots in R falls beyond the scope of the present article; however, the code used to generate Figure 10 is available on the R script associated with this example.

The following line of code can be used to print a table with descriptive statistics:

```
summary(results, ci = 0.95)
```

Example B: Multiple Linear Regression

Let us assume we have already established that the previously tested anxiety treatment works (drug > placebo) and now wish to investigate if factors other than symptom severity can affect its effectiveness. To this aim, we want to evaluate on a continuous scale the impact of symptom severity along with drug dosage and age (independent variables). To this aim, we recruit 100 volunteers currently undergoing such anxiety treatment and

obtain the same treatment-efficacy index used before (dependent variable). Because all our variables are continuous, multiple linear regression is a viable approach to test the influence that each independent variable and their combination can exert on the dependent variable.

The regression data set

The data set is composed of five columns: “ID” (containing participants’ identification number), “treatment.effect” (containing the treatment-efficacy score), “age” (containing participants’ age), “dosage” (containing participants’ drug dosage), and “symptoms” (containing participants’ symptoms). All variables are on a continuous scale.

The data were simulated by sampling 100 cases from a normal distribution using different means and standard deviations ($M \pm SD$) for each variable: 0 ± 10 for the treatment effect, 40 ± 15 for age, 10 ± 5 for dosage, and 50 ± 5 for symptoms. All variables were subsequently standardized as z scores. This data set can be reproduced using the “creating the datasets.R” file available on the OSF page (<https://osf.io/dez9b/>).

The hypotheses

The first hypothesis poses that dosage has a higher impact than symptoms on treatment effect, which has a higher impact than age on treatment effect:

$$H_1: \text{dosage} > \text{symptoms} > \text{age}.$$

Table 4. Bayes Factor Matrix

	H ₁	H ₂	H ₃
H ₁	1	0.275	6.87E+237
H ₂	3.639	1	2.50E+238
H ₃	0.000	0.000	1

Note: Because the values reported here are particularly high (e+) or particularly low (e−), scientific notation is used for convenience.

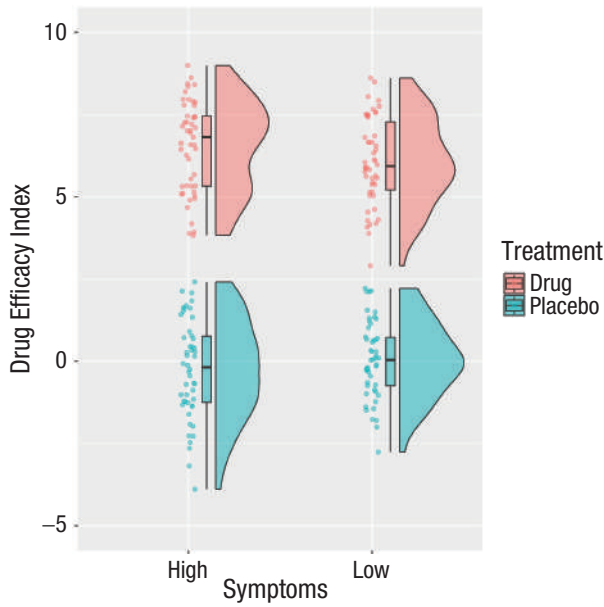


Fig. 10. Visualizing the data. Raincloud plots of the data set representing the groups divided based on treatment (drug/placebo) and symptoms (high/low). The code used to generate this figure is available on the R script associated with this example.

The second hypothesis poses that symptoms have a higher impact than dosage on treatment effect, which has a higher impact than age on treatment effect:

$$H_2: \text{symptoms} > \text{dosage} > \text{age}.$$

The third hypothesis poses that dosage and symptoms have a comparably higher impact than age on treatment effect:

$$H_3: (\text{dosage, symptoms}) > \text{age}.$$

For more on how to define informative hypotheses, see the Characteristics of the Informative Hypotheses section.

Step-by-step tutorial in JASP

The analysis presented in this section is performed using the *bain* module on JASP (JASP Team, 2021). The file “bain_REGRESSIONinJASP.jasps” available on the OSF page (<https://osf.io/dez9b/>) contains all the steps reported below and can be directly opened in JASP.

Preliminary steps. Follow the Preliminary Steps section of the JASP tutorial for Example A.

Step 1b: load the data set. Follow Step 1a of the JASP tutorial to load the file “dataset_regression.txt.” You will now visualize a spreadsheet containing the entire data set (Fig. 11) described above in The Regression Data Set section.

Step 2b: fit the model. By selecting “Linear Regression” in the *bain* menu (Fig. 11), you will be presented with a

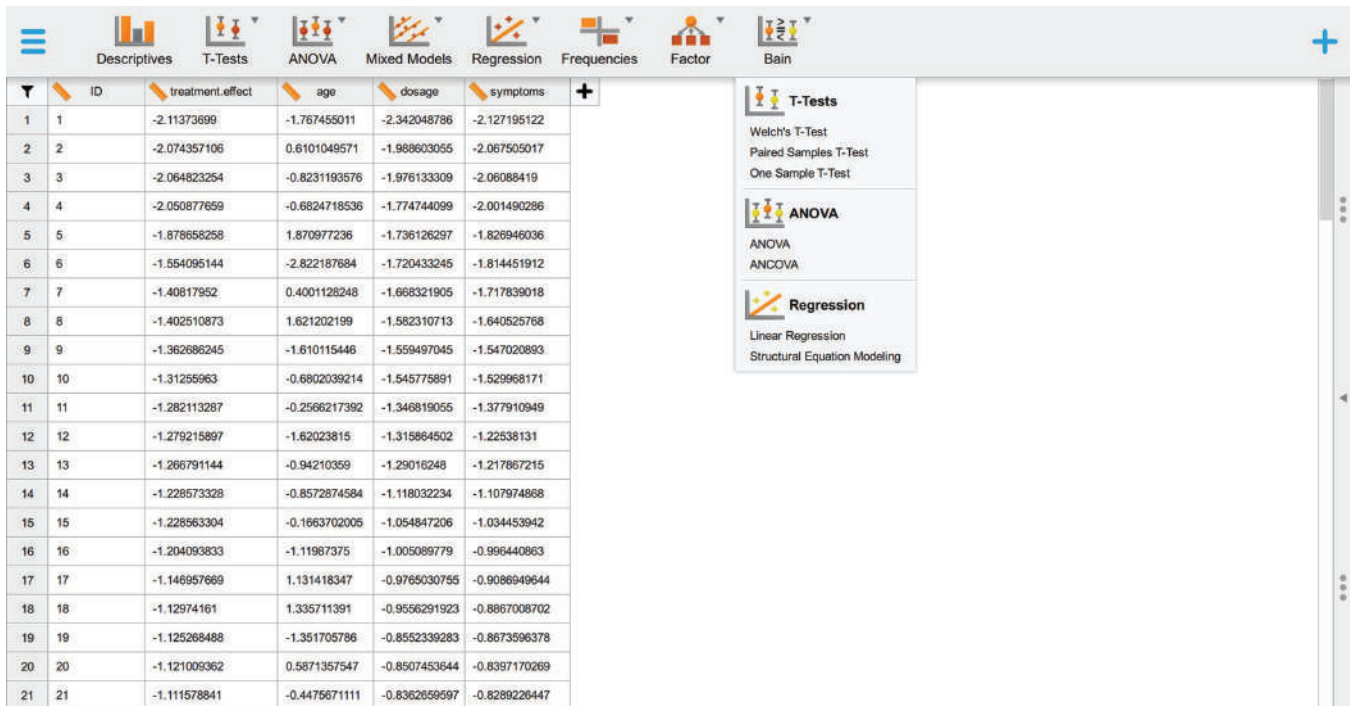


Fig. 11. Loading the data set and choosing the analysis in JASP.

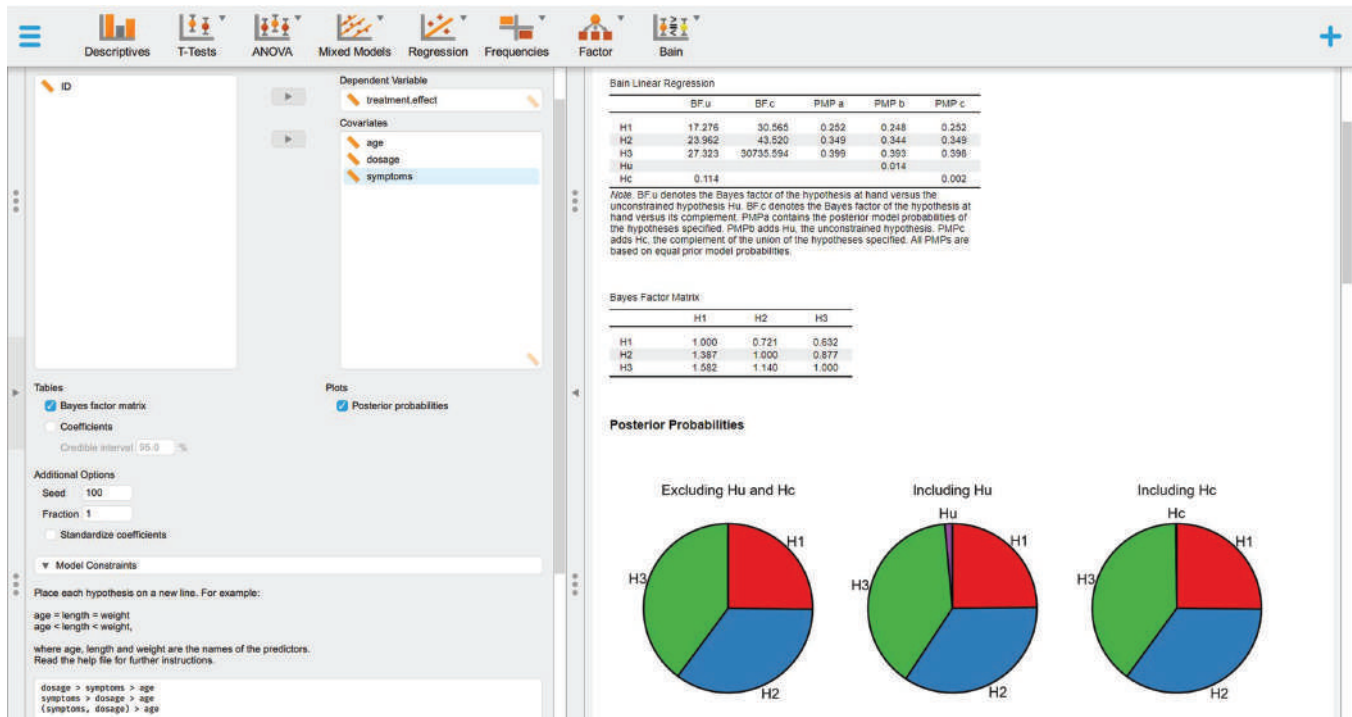


Fig. 12. Definition of informative hypothesis and results.

new graphical user interface that can be used to select the variables of interest (Fig. 12). Set the variable “treatment. effect” as the dependent variable and the variables “age,” “dosage,” and “symptoms” as covariates.

Step 3b: define and test informative hypotheses with bain. Follow Step 3a of the JASP tutorial to set the hypotheses reported below, visualize the bain linear regression results (Table 5), display the BF matrix (Table 6), and plot the PMPs (Fig. 13):

```
dosage > symptoms > age
symptoms > dosage > age
(symptoms, dosage) > age
```

Step 4b: interpret and report the results. In the following, we show how to interpret and report these results based on the indications outlined in Box 4.

Although all results obtained point in the same direction, indicating that H₃ is the hypothesis most supported by the data, the strength of the evidence is not clear-cut in this case, thus suggesting careful consideration. H₃ is indeed associated with the highest relative PMP (Fig. 13) both when considering only the set of informative hypotheses tested (39.9%; Table 5: PMPa) and when including H_c (39.8%; Table 5: PMPc); however, such proportion is not strikingly different from that associated with H₂ (34.9%). Despite this, the three hypotheses do

seem to be good models for the data (H_c = 0.02%; see Box 4, Question 1: Which of a set of hypotheses is the best?). In addition, although all hypotheses resulted to be more likely than the complement hypothesis (Table 5: BF_c), H₃ reported a much stronger support in this regard, being 30,735.594 times more likely than H_c (see Box 4, Question 2: How much more likely a given hypothesis is relative to other possible explanations?). Finally, H₃ proved to be 1.582 times more likely than H₁ (Table 6: BF₃₁) and 1.14 times more likely than H₂ (Table 6: BF₃₂), thus showing weak but still present support in its favor (see Box 4, Question 3: How much more likely a given hypothesis is relative to another specific hypothesis?). Overall, these results indicate that although both symptoms and dosage undoubtedly appear to have a

Table 5. Bain Linear Regression

	BFu	BFc	PMPa	PMPb	PMPc
H ₁	17.276	30.565	0.252	0.248	0.252
H ₂	23.962	43.52	0.349	0.344	0.349
H ₃	27.323	30,735.594	0.399	0.393	0.398
Hu				0.014	
Hc	0.114				0.002

Note: BFu = Bayes factor versus Hu; BFc = Bayes factor versus Hc; PMPa = posterior model probability excluding Hu and Hc; PMPb = posterior model probability including Hu; PMPc = posterior model probability including Hc.

Table 6. Bayes Factor Matrix

	H ₁	H ₂	H ₃
H ₁	1	0.721	0.632
H ₂	1.387	1	0.877
H ₃	1.582	1.14	1

stronger impact on treatment efficacy than age, whether the two have a significantly different impact is yet to be clarified.

Step 5b: visual exploration of the data. Unfortunately, at the moment of writing, the *bain* module does not allow generating a plot of the model data directly. As an alternative, the standard module for running linear regression can be used for this aim. To do that, select the “Regression” module and then “Linear Regression” (Fig. 14). You will be now presented with a new graphical user interface that can be used to select the variables of interest (Fig. 14). Set the variable “treatment.effect” as the dependent variable and the variables “age,” “dosage,” and “symptoms” as covariates. Then, go to the “Plots” section on the bottom left panel and check the “Partial plots” box.

Partial regression plots can be used to display the relationship between the dependent variable and one independent variable at a time while controlling for the presence of other independent variables in the model (Fox & Weisberg, 2019). By visual inspection only (Fig. 14), it is possible to appreciate a few crucial things: (a) A linear estimation of the relationship between the variables appears appropriate; (b) dosage and symptoms appear to have the strongest relationship with the treatment effect, whereas no relationship with age emerges; and (c) there is an outlier score that needs to be taken into account for a correct interpretation of the results (Altman & Krzywinski, 2016; Greco et al., 2019).

Step-by-step tutorial in R

There are two ways to proceed with this example: (a) Open a new R script and use the following lines of code to play along with the tutorial or (b) open and use the file “BaIn_REGRESSIONinR.R” available on the OSF page (<https://osf.io/dez9b/>), which contains all the steps and code reported below.

Preliminary steps. Follow the Preliminary Steps section of the R tutorial for Example A.

Step 1b: load the data set. Follow Step 1a of the R tutorial to load the file “dataset_regression.txt” and visualize the data set described above in The Regression Data Set section.

Step 2b: fit the model. The following code can be used to fit a linear model with the function `lm()`, which returns estimated coefficients for each independent variable:

```
fit <- lm(treatment.effect ~ age +
  dosage + symptoms, dataset_
  regression)
```

Step 3b: define and test informative hypotheses with bain. Follow Step 3a of the R tutorial to test the hypotheses reported below, print the main results of the *bain* linear regression (Table 7), display the BF matrix (Table 8), and plot the PMPs (Fig. 15):

```
set.seed(123)
results = bain(x = fit,
  hypothesis =
  "dosage > symptoms > age;
  symptoms > dosage > age;
  (symptoms, dosage) > age",
  standardize = TRUE)
```

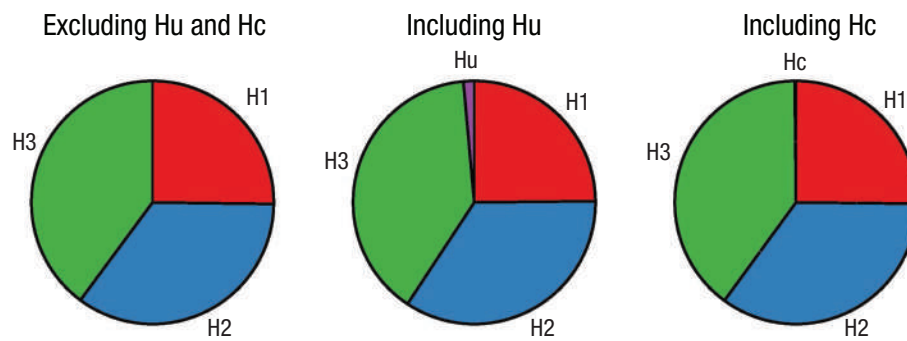


Fig. 13. Posterior model probabilities (PMPs). PMPs associated with the set of hypotheses tested (PMPa), also including Hu (PMPb) or Hc (PMPc).

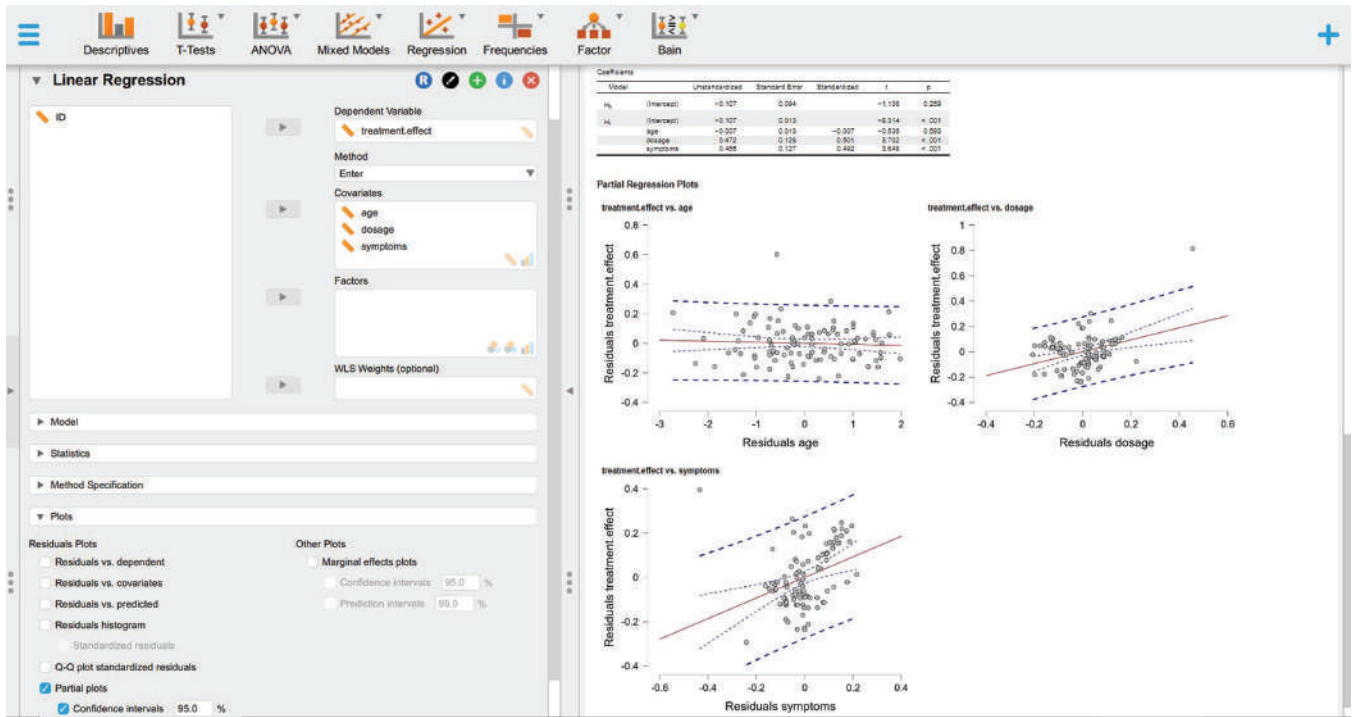


Fig. 14. Using the “Regression” module to obtain partial regression plots.

Step 4b: interpret and report the results. Please refer to Step 4b of the JASP example for a complete guide on how to interpret and report these results.

Step 5b: visual exploration of the data. Please refer to Step 4b of the JASP example reported above for more on the importance of visual exploration of the data and how to interpret the results obtained. In R/RStudio, in particular, the function `avPlots()` from the package `car` can be used to generate partial regression plots (Fig. 16). A

detailed explanation of how to generate plots in R falls beyond the scope of the present article; however, the code used to generate Figure 16 is available on the R script associated with this example.

Conclusion

Especially in—but not limited to—psychological and cognitive sciences, experimental hypotheses do not fit the conventional definitions of null ($A = B$) and

Table 7. *Bain* Regression

	<i>Fit</i>	<i>Com</i>	<i>BFu</i>	<i>BFc</i>	<i>PMPa</i>	<i>PMPb</i>	<i>PMPc</i>
H_1	0.418	0.013	31.418	53.267	0.364	0.36	0.364
H_2	0.452	0.019	24.113	43.194	0.28	0.276	0.279
H_3	1	0.033	30.684	1.18e ⁺⁰⁹	0.356	0.352	0.355
H_u						0.011	
H_c	0.173	0.965	0.179				0.002

Note: H_u = unconstrained hypothesis; H_c = complement hypothesis; *fit* = model fit; *com* = model complexity; *BFu* = Bayes factor versus H_u ; *BFc* = Bayes factor versus H_c . *PMPa* = posterior model probability excluding H_u and H_c ; *PMPb* = posterior model probability including H_u ; *PMPc* = posterior model probability including H_c . Because the values reported here are particularly high (e+) or particularly low (e-), scientific notation is used for convenience.

Table 8. Bayes Factor Matrix

	H ₁	H ₂	H ₃
H ₁	1	1.303	1.024
H ₂	0.767	1	0.786
H ₃	0.977	1.272	1

alternative hypotheses ($A \neq B$). Most times, actual research expectations involve very specific predictions

informative-hypothesis testing offers a powerful and easy way to directly compare predefined hypotheses through a Bayesian model-selection procedure in which each hypothesis represents a potential explanation or expectation for a phenomenon. Following the five steps outlined in the present tutorial, this can be easily achieved using two of the most commonly adopted software for statistical analysis: JASP and R/RStudio. In the present tutorial article, we encourage the adoption of such an inferential approach whenever specific research

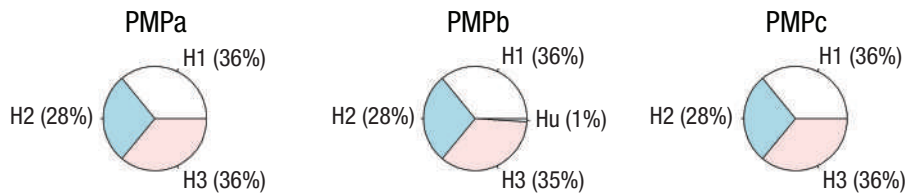


Fig. 15. Posterior model probabilities (PMPs). PMPs associated with the three hypotheses (PMPa), also including Hu (PMPb) or Hc (PMPc). The code used to generate this figure with the pie() function is available on the R script associated with this example.

that could be defined in terms of equality ($=$) and inequality ($>$ or $<$) constraints among the parameters. In other words, these are informative hypotheses that could be better tested via a model-comparison approach than via classical hypothesis testing (Gu et al., 2018; Hoijtink, 2012; Hoijtink, Mulder, et al., 2019; Kerr, 1998). Bayesian

hypotheses are present (Garofalo et al., 2022) to overcome old—and not necessarily more accurate—standards in statistical practice and embrace a richer and more meaningful interpretation of research results (Calin-Jageman & Cumming, 2019; Garofalo et al., 2022; Hoijtink et al., 2016).

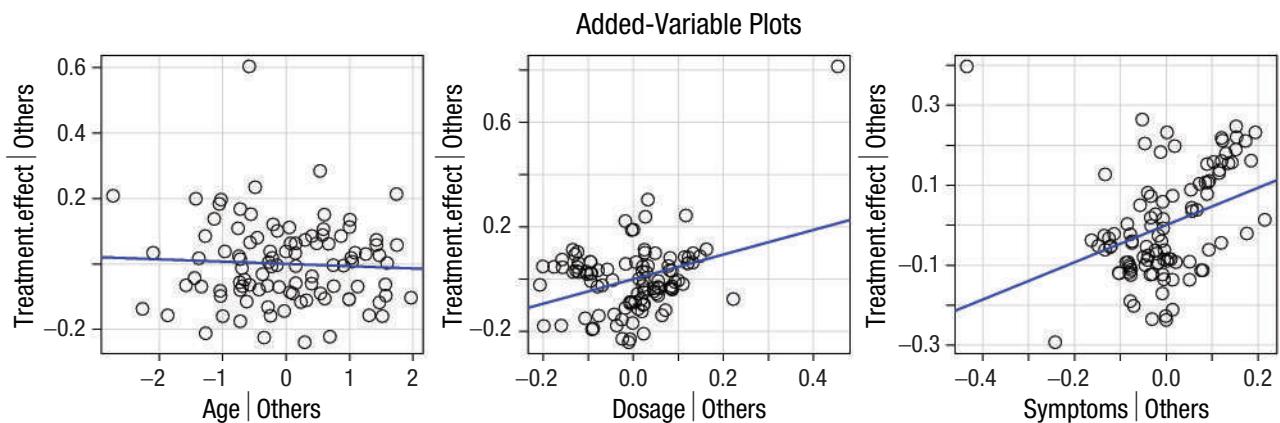


Fig. 16. Visualizing the data. Added-variable plots of the linear regression model. Each panel represents the dependent variable (treatment effect) on the y-axis and one of three independent variables (age, dosage, and symptoms, respectively) on the x-axis. The regression line in blue shows the association between the two variables displayed net of all other independent variables in the model.

Transparency

Action Editor: Katie Corker

Editor: Patricia J. Bauer

Author Contributions

Sara Garofalo: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Gianluca Finotti: Data curation; Software; Validation; Visualization; Writing – review & editing.

Matteo Orsoni: Project administration; Writing – review & editing.

Sara Giovagnoli: Project administration; Writing – review & editing.

Mariagrazia Benassi: Conceptualization; Funding acquisition; Methodology; Project administration; Supervision; Validation; Visualization; Writing – review & editing.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

The authors of this work are supported by the following grants: (a) European Union - NextGenerationEU - and the Italian Ministry of Research (MUR) National Recovery and Resilience Plan (NRRP-PNRR), in the framework of the project “Individual variations in resisting temptation: the balance between motivational and inhibitory control” (ES PNRR - M4C2 - I1.3 - P2022KAZ45_001 – CUP J53D23017240001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them. (b) Progetti di Rilevante Interesse Nazionale (PRIN) 2022 grant funded by the European Union - Next-GenerationEU and the Italian Ministry of Research (MUR) National Recovery and Resilience Plan (NRRP), project “Decision-making through the lens of the motor system: a neurocomputational perspective” (CUP J53D23019450006). (c) Progetti di Rilevante Interesse Nazionale (PRIN) 2020 project titled “Adaptive and multidimensional assessment of executive functions” (N. 20209WKCLL) funded by the Italian Ministry of Research (MUR). (d) #NEXTGENERATIONEU project titled “MNESYS” (N. PE0000006; DN. 1553 11.10.2022), funded by the Italian Ministry of University and the Research, National Recovery and Resilience Plan. (e) Human Brain Project titled “The Motor-way to Decision Making” (MoDeM) (CUP J32F20000870001), funded by the European Union’s Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (SGA3) under the FLAG-ERA JTC 2019 scheme in collaboration with the Italian Ministry of Research (MUR).

Open Practices

This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Sara Garofalo  <https://orcid.org/0000-0002-6023-2802>

Matteo Orsoni  <https://orcid.org/0000-0001-9675-9086>

Acknowledgments

We sincerely thank Herbert Hoijtink for his contribution to this work. His careful suggestions and feedback had a huge impact on the final version of the article, which immensely improved in terms of clarity and precision of the content thanks to his inputs.

Note

1. The topic of prior probabilities deserves a separate discussion that goes beyond the aim of the present article. To read more about this, please see Gelman and Yao (2021), Hoijtink et al. (2016), Kruschke (2014), Mikkola et al. (2021), van Wesel et al. (2011), and Zandonella Callegger et al. (2022).

References

- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, 4, Article 63. <https://doi.org/10.12688/wellcomeopenres.15191.2>
- Altman, N., & Krzywinski, M. (2016). Points of significance: Analyzing outliers: Influential or nuisance? *Nature Methods*, 13(4), 281–282. <https://doi.org/10.1038/nmeth.3812>
- Andrzejewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E. J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, 41(2), 521–543. <https://doi.org/10.1177/0149206314560412>
- Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, 53, 370–418. <https://doi.org/10.1098/rstl.1763.0053>
- Béland, S., Klugkist, I., Raïche, G., & Magis, D. (2012). A short introduction into Bayesian evaluation of informative hypotheses as an alternative to exploratory comparisons of multiple group means. *Tutorials in Quantitative Methods for Psychology*, 8(2), 122–126. <https://doi.org/10.20982/tqmp.08.2.p122>
- Calin-Jageman, R. J., & Cumming, G. (2019). The new statistics for better science: Ask how much, how uncertain, and what else is known. *American Statistician*, 73(Suppl. 1), 271–280. <https://doi.org/10.1080/00031305.2018.1518266>
- Cumming, G. (2007). Inference by eye: Pictures of confidence intervals and thinking about levels of confidence. *Teaching Statistics*, 29(3), 89–93. <https://doi.org/10.1111/j.1467-9639.2007.00267.x>
- Cumming, G. (2009). Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine*, 28(2), 205–220. <https://doi.org/10.1002/sim.3471>
- Cumming, G., & Finch, S. (2005). Inference by eye confidence intervals and how to read pictures of data. *American Psychologist*, 60(2), 170–180. <https://doi.org/10.1037/0003-066X.60.2.170>

- Degni, L. A. E., Dalbagnò, D., Starita, F., Benassi, M., di Pellegrino, G., & Garofalo, S. (2022). General Pavlovian-to-instrumental transfer in humans: Evidence from Bayesian inference. *Frontiers in Behavioral Neuroscience*, *16*, Article 945503. <https://doi.org/10.3389/fnbeh.2022.945503>
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.
- Garofalo, S., Giovagnoli, S., Orsoni, M., Starita, F., & Benassi, M. (2022). Interaction effect: Are you doing the right thing? *PLOS ONE*, *17*(7), Article e0271668. <https://doi.org/10.1371/journal.pone.0271668>
- Gelman, A., & Yao, Y. (2021). Holes in Bayesian statistics. *Journal of Physics G: Nuclear and Particle Physics*, *48*(1), 1–13. <https://doi.org/10.1088/1361-6471/abc3a5>
- Greco, L., Luta, G., Krzywinski, M., & Altman, N. (2019). Analyzing outliers: Robust methods to the rescue. *Nature Methods*, *16*(4), 275–276. <https://doi.org/10.1038/s41592-019-0369-z>
- Gu, X., Hoijtink, H., Mulder, J., & Rosseel, Y. (2019). Bain: A program for Bayesian testing of order constrained hypotheses in structural equation models. *Journal of Statistical Computation and Simulation*, *89*(8), 1526–1553. <https://doi.org/10.1080/00949655.2019.1590574>
- Gu, X., Hoijtink, H., Mulder, J., & van Lissa, C. (2021). *bain: Bayes factors for informative hypotheses* (Version 0.2.8) [R package]. <https://cran.r-project.org/package=bain>
- Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, *19*(4), 511–527. <https://doi.org/10.1037/met0000017>
- Gu, X., Mulder, J., & Hoijtink, H. (2018). Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. *British Journal of Mathematical and Statistical Psychology*, *71*(2), 229–261. <https://doi.org/10.1111/bmsp.12110>
- Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: Data analysis with estimation graphics. *Nature Methods*, *16*(7), 565–566. <https://doi.org/10.1038/s41592-019-0470-3>
- Hoijtink, H. (2012). *Informative hypotheses: Theory and practice for behavioral and social scientists*. Chapman & Hall/CRC. <https://doi.org/10.1201/B11158>
- Hoijtink, H., Gu, X., & Mulder, J. (2019). Bayesian evaluation of informative hypotheses for multiple populations. *British Journal of Mathematical and Statistical Psychology*, *72*(2), 219–243. <https://doi.org/10.1111/bmsp.12145>
- Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, *24*(5), 539–556. <https://doi.org/10.1037/met0000201>
- Hoijtink, H., van Kooten, P., & Hulsker, K. (2016). Why Bayesian psychologists should change the way they use the Bayes factor. *Multivariate Behavioral Research*, *51*(1), 2–10. <https://doi.org/10.1080/00273171.2014.969364>
- JASP Team. (2021). *JASP* (Version 0.16).
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4
- Kluytmans, A., van de Schoot, R., Mulder, J., & Hoijtink, H. (2012). Illustrating Bayesian evaluation of informative hypotheses for regression models. *Frontiers in Psychology*, *3*, Article 2. <https://doi.org/10.3389/fpsyg.2012.00002>
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. <https://doi.org/10.1016/j.tins.2004.10.007>
- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Elsevier.
- Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin and Review*, *25*(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. <https://doi.org/10.3758/BF03210951>
- Masson, M. E. J. (2007). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *58*(4), 289–289.
- Mikkola, P., Martin, O. A., Chandramouli, S., Hartmann, M., Pla, O. A., Thomas, O., Pesonen, H., Corander, J., Vehtari, A., Kaski, S., Bürkner, P.-C., & Klami, A. (2021). *Prior knowledge elicitation: The past, present, and future*. arXiv. <http://arxiv.org/abs/2112.01380>
- Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, *140*(4), 887–906. <https://doi.org/10.1016/j.jspi.2009.09.022>
- R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- RStudio Team (2023). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA. <http://www.rstudio.com/>
- Senn, S. (2007). *Statistical issues in drug development*. John Wiley & Sons. <https://doi.org/10.1002/9780470723586>
- van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin and Review*, *28*(3), 813–826. <https://doi.org/10.3758/s13423-020-01798-5>
- Van Lissa, C. J., Gu, X., Mulder, J., Rosseel, Y., Van Zundert, C., & Hoijtink, H. (2021). Teacher's corner: Evaluating informative hypotheses using the Bayes factor in structural equation models. *Structural Equation Modeling*, *28*(2), 292–301. <https://doi.org/10.1080/10705511.2020.1745644>
- Van Rossum, M., Van De Schoot, R., & Hoijtink, H. (2013). “Is the hypothesis correct” or “Is it not”: Bayesian evaluation of one informative hypothesis for ANOVA. *Methodology*, *9*(1), 13–22. <https://doi.org/10.1027/1614-2241/a000050>
- van Wesel, F., Hoijtink, H., & Klugkist, I. (2011). Choosing priors for constrained analysis of variance: Methods based

- on training data. *Scandinavian Journal of Statistics*, 38(4), 666–690. <https://doi.org/10.1111/j.1467-9469.2010.00719.x>
- Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E. J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin and Review*, 25(1), 58–76. <https://doi.org/10.3758/s13423-017-1323-7>
- Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin and Review*, 25(1), 35–57. <https://doi.org/10.3758/s13423-017-1343-3>
- Wilkinson, M. D., Dumontier, M., Aalbersberg Ij, J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018. <https://doi.org/10.1038/sdata.2016.18>
- Zandonella Callegher, C., Marci, T., De Carli, P., & Altoè, G. (2022). *Evaluating informative hypotheses with equality and inequality constraints: A tutorial using the Bayes factor via the encompassing prior approach*. PsyArXiv. <https://doi.org/10.31234/osf.io/6kc5u>