Sample size determination for Bayesian testing of informative hypothesis in linear

regression models

Qianrao Fu

Department of Methodology and Statistics, Utrecht University

⁹   Abstract

¹⁰ It is a tradition that goes back to Jacob Cohen to calculate the sample size before

¹¹ collecting data. The most commonly asked question is: "How many subjects do we need to

¹² obtain a significant result if we use the p-value to evaluate the hypothesis if an effect size

¹³ exists?" In the Bayesian framework, we may want to know how many subjects are needed

¹⁴ to get convincing evidence if we use the Bayes factor to evaluate the hypothesis. This

¹⁵ paper proposes a solution to the above question by reaching two goals: firstly, the size of

¹⁶ the Bayes factor reaches a given threshold, and secondly the probability that the Bayes

¹⁷ factor exceeds the given threshold reaches a required value. Researchers can express their

¹⁸ expectations through the order or the sign hypothesis of the parameters in a linear

¹⁹ regression model. For example, the researchers may expect the regression coefficient to be

²⁰ $\beta_1 > \beta_2 > \beta_3$, which is an order constrained hypothesis; or the researchers may expect a

²¹ regression coefficient $\beta_1 > 0$, which is a sign hypothesis. The greatest advantage of using a

²² specific hypothesis is that the sample size required is reduced compared to an

²³ unconstrained hypothesis to achieve the same probability that the Bayes factor exceeds

²⁴ some threshold. This article provides sample size tables for the null hypothesis, order

²⁵ hypothesis, sign hypothesis, complement hypothesis, and unconstrained hypothesis. To

²⁶ enhance the applicability, an `R` package is developed via a Monte Carlo simulation, which

²⁷ can facilitate psychologists while planning the sample size even if they do not have any

²⁸ statistical programming background.

²⁹   *Keywords:* Bayes Factor, Informative Hypothesis, Multiple Linear Regression, Sample

³⁰ Size Determination, `SSDbain`

³¹   Sample size determination for Bayesian testing of informative hypothesis in linear

³²   regression models

Introduction

³⁴ Sample size determination is a crucial step in the design of a study. If the sample size is

³⁵ insufficient, then the study will not be able to draw valid conclusions. Conversely, if the

³⁶ sample size is much larger than required, the study will become expensive, time-consuming

³⁷ and ethically unacceptable. When the required sample size cannot be achieved to

³⁸ demonstrate convincing results, the researchers may consider not going ahead with this

³⁹ study to save money and efforts. In most universities, sample size determination and

⁴⁰ statistical power analysis are increasingly becoming a requirement for most research

⁴¹ proposals, applications for ethical clearance and journal articles. Based on Cohen's

⁴² research (Cohen, 1988, 1992), software for sample size calculation and power analysis, such

⁴³ as `G*Power` (Faul et al., 2009; Faul et al., 2007; Mayr et al., 2007), `nQuery Advisor`

⁴⁴ (Elashoff, 2007) and `PASS` (Hintze, 2011) has been developed. Through these software the

⁴⁵ researcher can obtain a sample size plan easily.

⁴⁶ The multiple linear model is one of the most often used models in the social and behavioral

⁴⁷ sciences. The multiple linear regression is widely used to evaluate how a response variable

⁴⁸ (Y) is related to a set of predictors ($X_1, X_2, \cdots, X_K$). Suppose that a group of researchers

⁴⁹ wants to investigate the relationship between the response variable *Income* and two

⁵⁰ predictor variables *Intelligence* (IQ) and *Socio-Economic Status* (SES) using multiple linear

⁵¹ regression. The regression coefficient corresponding to IQ is denoted as β_1, and the

⁵² regression coefficient corresponding to SES is denoted as β_2. The hypothesis of interest for

⁵³ the study is that the predictor of IQ has a stronger effect than SES on the response

⁵⁴ variable *Income.* This hypothesis can be expressed using notation as H_1: $\beta_1 > \beta_2$. To

⁵⁵ demonstrate this order relationship, the researchers want to detect a coefficient of

⁵⁶ determination, for instance, $R^2 = 0.13$, where R^2 is the proportion of the variance in the

dependent variable that is explained by the independent variable(s). Sample-size tables in the framework of null-hypothesis significance testing (NHST) based on the F-test show that in the case of two predictors, $R^2 = 0.13$, and a significance level of $\alpha = 0.05$, 67 subjects are necessary to obtain a power of 0.80 if the null hypothesis H_0: $\beta_1 = \beta_2 = 0$ is compared with the unconstrained hypothesis H_a: β_1, β_2. However, the expected ordering of the means ($\beta_1 > \beta_2$) is completely ignored in NHST.

NHST has been harshly criticized in numerous articles in recent years although it is the most commonly used method for statistical hypothesis testing. Among them, there are three crucial points:

1) The p-value derived from NHST is a measure of evidence against the null hypothesis H_0 (Hurlbert & Lombardi, 2009). What's more, it exaggerates the evidence against the null hypothesis H_0 (Berger, 1986; Berger & Sellke, 1987). That is, the p-value makes it relatively easy to obtain statistically significant findings. For example, if p-values of 0.05, 0.01, and 0.001 are considered, the posterior probabilities of the null, $P(H_0|x)$, for sample size $N = 50$ are 0.52, 0.22, and 0.034, respectively, which indicate these discrepancies between p-value and posterior probability are pronounced.

2) A significance level α of 0.05 typically reduces NHST to a binary decision rule. That is, the null hypothesis is rejected if the p-value is smaller than 0.05, and not rejected it if it is above 0.05 (Harlow et al., 2016; Nickerson, 2000; Wagenmakers, 2007). This leads to phenomena such as publication bias (Ioannidis, 2005; Simmons et al., 2011; Van Assen et al., 2014), and questionable research practices (Fanelli, 2009; Masicampo & Lalande, 2012; Wicherts et al., 2016), which both contributed to the replication crisis (Open Science Collaboration, 2015);

3) In practical applications, the null hypothesis is never exactly true. Therefore it is always rejected as the number of observations becomes large. (Cohen, 1994; Raftery, 1995; Royall, 1997).

An alternative that has gained notable attention over the past years is Bayesian hypothesis testing using the Bayes factor (Lee & Wagenmakers, 2014; Van de Schoot et al., 2017; Vandekerckhove et al., 2018; Wagenmakers et al., 2016). In contrast to NHST, the Bayes factor has the following advantages:

1) The Bayes factor cannot only provide evidence in favor of the alternative hypothesis but, in contrast to the p-value, it also provides evidence in favor of the null hypotheses;

2) As elaborated in Hoijtink, Mulder, et al. (2019), the Bayes factor is a continuous value that quantifies the degree of the evidence in favor of one hypothesis compared to another hypothesis instead of making a hard "accept/reject" decision about the null hypothesis. This helps to reduce the problem of replication crisis. This is because the evidence of supporting the null hypothesis can be obtained in the Bayesian framework. This makes it more likely to be published in scientific journals even when there are "non-significant" results as encountered in NHST;

3) The Bayes factor will approach 0 or ∞ when the sample size is very large (i.e., the Bayes factor for the null hypothesis H_0 goes to infinity if H_0 is true, and goes to 0 if the alternative hypothesis H_1 is true, as the sample size goes to infinity), that is, the property of consistency for Bayes factor as presented in Ly et al. (2016). This property guarantees that the Bayes factor will always support the true hypothesis when the sample size is large enough.

4) The Bayes factor can compare the null, unconstrained, complement, and informative hypotheses, where the informative hypothesis (Hoijtink, 2011) can express the researcher's expectations with regard to the sign or order of the regression coefficients of the predictors. For example, revisiting the example introduced earlier, the researcher may be interested in that *IQ* and *SES* both have a positive effect on *income*, that is, H_1: $\beta_1 > 0$, $\beta_2 > 0$.

As the Bayes factor becomes more and more often used (Van de Schoot et al., 2017;

Vandekerckhove et al., 2018; Wagenmakers et al., 2016), the Bayes factor calculation tools emerge. Currently, there are three `R` packages that can be used to compute the Bayes factor: for the evaluation of a null hypothesis versus the alternative hypothesis `BayesFactor` [1] (Morey et al., 2018); and, additionally, for the evaluation of informative hypothesis `bain` [2] (Gu et al., 2021) and BFpack [3] (Mulder et al., 2021). The first two packages are also available in JASP [4] (Love et al., 2019), which is an easy to use statistical software with intuitive interface.

In line with the popularity of Bayesian hypothesis testing, more attention to sample size determination should be paid in this framework. The purpose of this paper combined with Fu et al. (2021), and Fu et al. (unpublished) is to introduce a new `R` package `SSDbain` to help researchers who are not mathematicians and/or statisticians to obtain the minimum sample size required when the Bayes factor is used to evaluate informative hypotheses. The sample size is determined such that the probability that the Bayes factor is larger than a threshold denoted by BF_{thresh} is η under the competing hypotheses considered, where BF_{thresh} is a value that constitutes sufficient evidence for the researchers, and η is the probability to correctly find sufficient support for the true hypothesis. Throughout this paper, sample size determination (SSD) for the comparison of null, informative, and unconstrained hypotheses under a multiple linear regression in the Bayesian framework as implemented in the `R` package `bain` will be performed. This work will build on the sample size calculations for two-sample t-test discussed in Fu et al. (2021), one-way ANOVA discussed in Fu et al. (unpublished) and Bayes factor design analysis discussed in Schönbrodt and Wagenmakers (2018) and Stefan et al. (2019). Several tables based on $R^2 = 0.13$, which corresponds to Cohen's medium effect size $f^2 = 0.15$, are presented as an

---

[1] `https://richarddmorey.github.io/BayesFactor/`

[2] `https://informative-hypotheses.sites.uu.nl/software/bain/`

[3] `https://github.com/jomulder/BFpack`

[4] `https://jasp-stats.org/`

132 example to assist researchers in determining the minimum sample size required.

133 The outline of this paper is as follows. First, the multiple linear regression models that are

134 used in the article are introduced, the (informative) hypotheses that are evaluated are

135 described, and the Bayes factor as implemented in the R package `bain` is further elaborated

136 on. Subsequently, sample size determination will be introduced, followed by the

137 introduction of the function `SSDRegression` in the R package `SSDbain`, features of SSD will

138 be highlighted, and examples will be provided and discussed. The paper ends with a short

139 conclusion.

Multiple Linear Regression and (Informative) Hypotheses

141 In this paper, K regression coefficients, $\beta_1, \beta_2, \cdots, \beta_K$ are considered, where K is an integer

142 that is greater than or equal to 1. Let us consider the following linear regression model

143 where a dependent variable Y is regressed on K predictor variables $X_1, X_2, \cdots, X_K$, say,

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{i,k} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2), \tag{1}$$

144 where y_i for $i = 1, \cdots, N$ is the i-th observation of the dependent variable Y, N denotes

145 the size of the sample, $x_{i,k}$ denotes the i-th observation of the k-th predictor variable X_k,

146 where $k = 1, 2, \cdots, K$, β_0 is the intercept of the regression model, β_k is the regression

147 coefficient of the k-th predictor, and ϵ_i are independently and normally distributed errors

148 with variance σ^2.

149 In this paper, sample size will be determined for the comparison of null, informative,

150 complement, and unconstrained hypotheses. These hypotheses concern the regression

151 coefficients from the multiple linear regression model in Equation 1. The null and

152 unconstrained hypotheses are already well known in NHST. The informative hypothesis

153 will be introduced using the following example: a group of researchers wants to explore the

154 relationship between the response variable IQ and three predictor variables, namely social

skills, interest in artistic activities, and use of complicated language patterns. The multiple linear regression model is used to fit this relationship. The corresponding regression coefficients are denoted by β_1, β_2, and β_3, respectively. The following pairs of hypotheses may be compared: 1) A group of researchers is interested in whether at least one predictor has an effect on the dependent variable IQ, that is, H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ vs H_a: at least one predictor has an effect on IQ; 2) The same researchers may also have evidence that the first three predictor variables are expected to be positively associated with IQ, that is, H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ vs H_1: $\beta_1 > 0$, $\beta_2 > 0$, $\beta_3 > 0$; 3) In confirmatory studies, the interest is typically in testing specific hypotheses with order constraints on the relative importance of the predictors based on scientific expectations or psychological theories (Hoijtink, 2011, pp. 5–20). The researchers may expect that social skills is the strongest predictor, followed by interest in artistic activities, and then use of complicated language patterns. They may formulate the hypotheses as H_0: $\beta_1 = \beta_2 = \beta_3$ vs H_2: $\beta_1 > \beta_2 > \beta_3$; 4) The researchers may be interested in supporting H_1 or precluding H_1. That is, H_1: $\beta_1 > 0$, $\beta_2 > 0$, $\beta_3 > 0$ vs H_{1c}: not H_1, where the subscript c refers to the complement of H_1; 5) The researchers may want to know if H_2 is preferred over other hypotheses. That is, H_2: $\beta_1 > \beta_2 > \beta_3$ vs H_{2c}: not H_2, where the subscript c refers to the complement of H_2. In this paper the generic situations shown below will be studied,

Situation 1:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_K = 0$ vs H_a: at least one predictor has an effect on the dependent variable,

Situation 2:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_K = 0$ vs H_1: $\beta_1 > 0$, $\beta_2 > 0$, $\cdots$, $\beta_K > 0$,

where some or all the regression coefficients may be smaller than zero. That is, both "<" and ">" can exist in a hypothesis.

Situation 3:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_K$ vs H_2: $\beta_{1^*} > \beta_{2^*} > \cdots > \beta_{K^*}$,

where 1^*, 2^*, $\cdots$, K^* are a re-ordering of the numbers 1, 2, $\cdots$, K,

Situation 4:

H_1: $\beta_1 > 0$, $\beta_2 > 0$, $\cdots$, $\beta_K > 0$ vs H_{1c}: not H_1,

It should be noted that in this situation only ">" or "<" is allowed. The complexity of the complement hypothesis if both ">" and "<" exist in one hypothesis prevent me from discussing it in this paper.

and, Situation 5:

H_2: $\beta_{1^*} > \beta_{2^*} > \cdots > \beta_{K^*}$ vs H_{2c}: not H_2.

The standardized regression coefficients are used in Situation 3 and Situation 5 to ensure the regression coefficients are comparable. The reason is that the regression coefficients β_1, $\cdots$, β_K may all be in different units of measurement and direct comparison is illogical. In the next section it will be elaborated how the Bayes factor implemented in the R package bain can be used to evaluate these pairs of hypotheses.

Bayes Factor

In order to evaluate the competing hypotheses introduced in the previous section, the Bayes factor will be used to quantify the relative evidence provided by the data. The Bayes factor was proposed in pioneering work by Jeffreys (1961), and it was further discussed in Edwards et al. (1963), Kass and Raftery (1995), and Myung and Pitt (1997). The Bayes factor for informative hypotheses was elaborated in the tutorial by Hoijtink, Mulder, et al. (2019), which contains all the references to the statistical background of these Bayes factors. The Bayes factor can be explained as follows, in the case of H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ versus H_1: $\beta_1 > \beta_2 > \beta_3 > \beta_4$, a Bayes factor $\text{BF}_{01} = 19$, for

204   instance, means there is 19 times more support for the model specifying the equality on the

205   regression coefficients than for the model specifying order constraints. The Bayes factor for

206   the null hypotheses H_0, signed hypothesis H_1, or order hypothesis H_2 versus the

207   unconstrained hypothesis H_a (the hypothesis without restrictions on the regression

208   coefficients) is defined by the marginal likelihood ratio (Jeffreys, 1961; Kass & Raftery,

209   1995)

$$\text{BF}_{za} = \frac{m_z(\boldsymbol{Y}, \boldsymbol{X})}{m_a(\boldsymbol{Y}, \boldsymbol{X})} = \frac{\iint f(\boldsymbol{Y}, \boldsymbol{X} \mid \boldsymbol{\beta}, \sigma^2)\pi_z(\boldsymbol{\beta}, \sigma^2)d\boldsymbol{\beta}d\sigma^2}{\iint f(\boldsymbol{Y}, \boldsymbol{X} \mid \boldsymbol{\beta}, \sigma^2)\pi_a(\boldsymbol{\beta}, \sigma^2)d\boldsymbol{\beta}d\sigma^2}, \tag{2}$$

210   where z=0, 1, 2, c, where $\pi_z(\boldsymbol{\beta}, \sigma^2) = \pi_z(\boldsymbol{\beta})\frac{1}{\sigma^2}$ and $\pi_a(\boldsymbol{\beta}, \sigma^2) = \pi_a(\boldsymbol{\beta})\frac{1}{\sigma^2}$, where $\pi_z(\boldsymbol{\beta})$ and

211   $\pi_a(\boldsymbol{\beta})$ denote the prior distribution under H_z and H_a, respectively, and $f(\boldsymbol{Y}, \boldsymbol{X} \mid \boldsymbol{\beta}, \sigma^2)$ is

212   the density of the data based on the model in Equation 1.

213   According to Klugkist et al. (2005), Equation 2 can be simplified to

$$\text{BF}_{za} = \frac{f_z}{c_z}, \tag{3}$$

214   where f_z is the fit of hypothesis H_z and c_z is its complexity.

215   The complexity c_z can be expressed as

$$c_z = \int_{\beta \in H_z} \pi_a(\boldsymbol{\beta})d\boldsymbol{\beta}. \tag{4}$$

216   It is the proportion of the prior distribution in agreement with H_z if z=1, 2, and is reduced

217   to the density $\pi_a(\boldsymbol{\beta} = \boldsymbol{0})$ if z=0, where $\boldsymbol{\beta} = (\beta_1 - \beta_2, \beta_2 - \beta_3, \cdots, \beta_{k-1} - \beta_k)$ in case of

218   $\beta_1 = \beta_2 = \cdots = \beta_K$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_K)$ in case of $\beta_1 = 0$ & $\beta_2 = 0 \cdots$ & $\beta_K = 0$.

219   The complexity stands for how specific the hypothesis H_z is if $z =$1, 2. The more specific

220   the inequality constrained hypothesis, the lower the complexity, but the complexity for the

221   null hypothesis H_0 and the inequality constrained hypotheses H_1 or H_2 cannot be

222 compared because the first is a density and the latter a probability.

223 The fit f_z can be expressed as

$$f_z = \int_{\beta \in H_z} \pi_a(\boldsymbol{\beta} \mid \boldsymbol{X}) d\boldsymbol{\beta}. \tag{5}$$

224 It is the proportion of the posterior distribution in agreement with H_z if $z=1$, 2, and is

225 reduced to the density $\pi_a(\boldsymbol{\beta} = \boldsymbol{0} \mid \boldsymbol{X})$ if $z=0$. The fit stands for how much the data

226 supports H_z relative to H_a if $z = 1$, 2. The more the support from the data, the larger the

227 fit.

228 Based on BF_{za}, the Bayes factor BF_{zc} for $z = 1$, 2 that expresses the support in the data

229 for H_z relative to its complement hypothesis H_c, can be derived as:

$$\mathrm{BF}_{zc} = \frac{\mathrm{BF}_{za}}{\mathrm{BF}_{ca}} = \frac{f_z}{c_z} \Big/ \frac{1 - f_z}{1 - c_z}, \tag{6}$$

230 where $1 - f_z$ denotes the fit of hypothesis H_c, and $1 - c_z$ denotes the complexity of

231 hypothesis H_c. The Bayes factor BF_{02}, that expresses the support in the data for H_0

232 against the competing hypothesis H_2, is represented by:

$$\mathrm{BF}_{02} = \frac{\mathrm{BF}_{0a}}{\mathrm{BF}_{2a}} = \frac{f_0}{c_0} \Big/ \frac{f_2}{c_2}, \tag{7}$$

233 and the Bayes factor BF_{01} that expresses the support in the data for H_0 against the

234 competing hypothesis H_1 is represented by:

$$\mathrm{BF}_{01} = \frac{\mathrm{BF}_{0a}}{\mathrm{BF}_{1a}} = \frac{f_0}{c_0} \Big/ \frac{f_1}{c_1}. \tag{8}$$

235 In this paper, the calculation of the Bayes factor as implemented in the R package `bain`

236 (Gu et al., 2018; Hoijtink, Gu, et al., 2019) will be used. In `bain`, the posterior distribution

237 of the regression coefficients is approximated by a normal distribution based on large

238 sample theory (Gelman et al., 2013, pp. 101)

$$\pi_a(\boldsymbol{\beta} \mid \boldsymbol{X}) \approx N(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}_\beta), \tag{9}$$

239 where $\hat{\boldsymbol{\beta}}$ is the MLE (maximum likelihood estimation) of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\Sigma}}_\beta$ is its covariance matrix

240 if Situations 1, 2, and 4 are considered, and $\hat{\boldsymbol{\beta}}$ is the MLE of the standardized parameters

241 and their covariance matrix if Situations 3 and 5 are considered. In `bain`, the adjusted

242 fractional normal prior distribution (Gu et al., 2018) is used.

$$\pi_a(\boldsymbol{\beta}) = N(\boldsymbol{0}, \hat{\boldsymbol{\Sigma}}_\beta/b). \tag{10}$$

243 The variance of this prior for each of the regression coefficient is calculated using a fraction

244 b of the information in the data (De Santis & Spezzaferri, 2001; Mulder, 2014; O'Hagan,

245 1995). The mean of the prior for each of the regression coefficient is chosen as zero, which

246 is located on the boundary of the constrained region of the competing hypotheses to make

247 sure the Bayes factor is consistent when equality constrained hypotheses are evaluated

248 (Mulder, 2014).

249 According to Berger and Pericchi (1996), the default value $b = J/N$ is used to specify the

250 variance of the prior distribution, where J is the minimal training sample size (a small part

251 of the observed data). In `bain`, J is replaced by the number of independent constraints.

252 This can be illustrated using an example. If H_0: $\beta_1 = \beta_2 = \beta_3$ versus H_1: $\beta_1 > \beta_2 > \beta_3$, the

253 number of independent constraints is $J = 2$, that is, there are two contrasts $\beta_1 - \beta_2$ and

254 $\beta_2 - \beta_3$ to be evaluated in the hypotheses.

255 In order to explore the influence of the variance of the prior distribution on the resulting of

256 the Bayes factor if hypothesis H_0 is included in the competing hypotheses, a sensitivity

257 analysis should be conducted. A sensitivity analysis can help fully understand the Bayesian

258 results combined with the prior and properly interpret the impact of the prior. In the

259 following sections, different choices of b value ($b = J/N$, $b = 2J/N$, and $b = 3J/N$) are

260 used. If only the inequality constraints are included in the competing hypotheses, the prior

261 has no influence on the value of Bayes factor, see Mulder (2014) for an explanation.

262 Therefore, the sensitivity analysis is relevant for Situations 1-3, but it does not affect the

263 value of Bayes factor when different bs are used for Situations 4-5.

The Criterion for Sample Size Determination

265 In the traditional a priori power analysis, the purpose of sample size determination is to

266 control the Type I error rate and the Type II error rate. The sample size can be calculated

267 if the significance level α, the desired statistical power $1 - \beta$, and the to-be-detected

268 population effect size f^2 in a multiple linear regression are given (Cohen, 1988, 1992). In

269 Bayesian hypothesis testing, instead of controlling the Type I error rate and Type II error

270 rate, the sample size is calculated to guarantee that the Bayes factor exceeds a user

271 specified threshold with a specific probability for the true hypothesis. In the following

272 paragraphs, it will be explained how the sample size is determined when the Bayes factor is

273 used to evaluate (informative) hypotheses under a multiple linear regression model.

274 The criterion that is proposed has also been used for the two-sample t-test (Fu et al., 2021)

275 and one-way ANOVA (Fu et al., unpublished). To help the readers understand how to

276 determine the sample size, we use Figure 1 to illustrate. The process of determining the

277 sample size can be divided into the following five steps:

278 1. Sample size determination always starts with the specification of two competing

279   hypotheses. The competing hypotheses used in Figure 1 are H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

280   and H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$.

281 2. One needs to specify plausible values of the parameters in order to perform a sample

282   size determination. For each population whether H_0 is true or H_1 is true, the

283   parameter values are unknown. In the next section it will be elaborated how these

284   parameter values can be chosen. For now, in Figure 1, when H_0 is true,

285     $\beta_1 = \beta_2 = \beta_3 = 0$ is used, and when H_1 is true, $\beta_1 = \beta_2 = \beta_3 = 0.208$ is used, which

286   corresponds to $R^2 = 0.13$.

287   3. From each population, data sets with a certain sample size are repeatedly sampled.

288   How the sample size is chosen will be elaborated in the next section.

289   4. For the population under H_0, the Bayes factor BF_{01} for each data set is computed

290   and for the population under H_1, the Bayes factor BF_{10} for each data set is also

291   computed. In Figure 1, in panel (a) on the left: the distribution of BF_{01} is shown,

292   and in panel (b) on the right: the distribution of BF_{10} is shown.

293   5. The required sample size should be large enough such that the Bayes factor is larger

294   than a user selected threshold BF_{thresh} with a specific probability η, where

295     $\text{BF}_{thresh} = 3$ is marked with a dashed line, and $\eta = 0.8$ is marked with the shaded

296   area in Figure 1. That is, $P(\text{BF}_{01} > \text{BF}_{thresh}|H_0) \geq \eta$ and $P(\text{BF}_{10} > \text{BF}_{thresh}|H_1) \geq \eta$

297   have to be satisfied. As shown in Figure 1, to satisfy the condition of BF_{01} under H_0:

298     $\beta_1 = \beta_2 = \beta_3 = 0$ that $P(\text{BF}_{01} > 3|H_0) \geq 0.80$, the sample size required is $N = 23$,

299   While the sample size should be more than 100 to satisfy the condition that

300     $P(\text{BF}_{10} > 3|H_1) \geq 0.80$ when H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$ with $R^2 = 0.13$. To

301   satisfy both conditions, at least a sample of $N = 100$ should be collected.
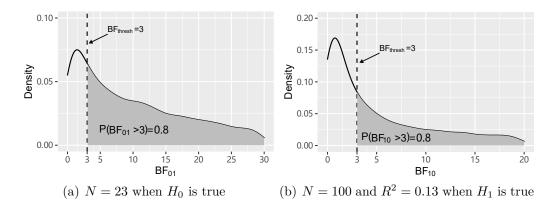
(a) $N = 23$ when H_0 is true   (b) $N = 100$ and $R^2 = 0.13$ when H_1 is true

Figure 1. The sampling distribution of BF_{01} under H_0 and BF_{10} under H_1. The vertical dashed line represents $BF_{thresh} = 3$, and the shaded area denotes the probability that the Bayes factor exceeds 3 if the target $\eta = 0.8$. The N in the label is the sample size needed to achieve the requirements.

Researchers require the relative support in the data under one hypothesis compared to the other hypothesis, or vice versa, to be at least BF_{thresh}. The η is used to control the error rates when either of the competing hypothesis is true. If $\eta = 0.8$, this means that the error rate is not more than $1 - \eta = 0.2$ if either of the competing hypotheses is true. The remaining issue is that there is still a lack of a standard to choose the sizes of BF_{thresh} and η. What constitutes sufficient evidence, and what is the appropriate probability to convincingly support the true hypothesis? The selection of these values depends on the area of research and whether primary or secondary outcome measures are investigated. The researchers can consult the professionals in the field of behavior and social science, for example, to fill in a questionnaire, ask them about what constitutes sufficient evidence for various scenarios in their respective fields. The responses from the professionals can be modelled with the wisdom-of-the-crowd paradigm (Lee et al., 2012; Surowiecki, 2004), which states that an aggregate of the judgment and estimates of many people is more accurate than the judgement of one person. Based on the risk level of the research, the appropriate value of the BF_{thresh} also varies. For example, to verify the effectiveness of vaccines against the Covid-19, a large value of the BF_{thresh} is recommended, such as 10. Contrary to that, a small value, such as 3, is preferred for the investigation of the height of

319   elementary school students in different regions. The η is introduced to limit the error rates.

320   For example, when η is equal to 0.8, the Type I error and Type II error rates would be no

321   more than 20%. If the consequences of missing an effect may be significant, for example in

322   a toxicity test, one may need a relatively high η, e.g. 0.90. In a survey, one would be

323   interested only in large effects, and errors in detecting effects may not have such serious

324   consequences. In this case, $\eta = 0.80$ may be sufficient.

The Basic Algorithm Used for Sample Size Determination

326   In the Bayesian framework, most of the research questions and data issues are sufficiently

327   complicated such that the problems cannot be solved analytically. In this paper, when the

328   researchers use the Bayes factor to evaluate hypotheses, Monte Carlo methods will be used

329   to determine the sample size. Figure 2 displays the process of the simulation-based

330   algorithm. The corresponding steps in Figure 2 are discussed below.

331   1. Before proceeding with the sample size determination, the following ingredients need

332   to be specified:

333   (1) The competing hypotheses are specified using the regression coefficients. The

334   options for the competing hypotheses are the null, the complement, the

335   unconstrained, the order and the sign hypotheses.

336   (2) The regression coefficients for each population in which the hypothesis is true can

337   be calculated if the fixed coefficient of determination R^2 ($R^2 > 0$) and the ratio

338   among the regression coefficients $\beta_1, \beta_2, \cdots, \beta_K$ are given. Appendix B shows

339   how to achieve the regression coefficients if these two factors are known. If

340            $R^2 = 0$, the regression coefficients for the population are all equal to 0.

341   (3) The correlation matrix among the predictor variables

$$
\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1K} \\ \rho_{21} & 1 & \cdots & \rho_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{K1} & \cdots & \rho_{K(K-1)} & 1 \end{bmatrix}.
$$

This ingredient is used to generate correlated data. This is because in practice, correlated predictor variables are the rule rather than the exception.

(4) The desired threshold for the Bayes factor BF_{thresh}.

(5) The probability η that the Bayes factor is larger than BF_{thresh} under each of the two competing hypotheses.

2. Randomly draw T data sets with sample of size $N = 10$ if the hypothesis H_s is true and if the competing hypothesis H_v is true. The data sets are denoted as

$(\boldsymbol{x}_i^{(t)}, y_{si}^{(t)}) = (x_{i1}^{(t)}, \cdots, x_{iK}^{(t)}, \beta_0 + \sum_{k=1}^{K} \beta_k^s x_{ik} + \epsilon_i)$, and

$(\boldsymbol{x}_i^{(t)}, y_{vi}^{(t)}) = (x_{i1}^{(t)}, \cdots, x_{iK}^{(t)}, \beta_0 + \sum_{k=1}^{K} \beta_k^v x_{ik} + \epsilon_i)$, respectively, where s=0, 1, 2, v=1, 2, a, $1c$, or $2c$, and β_k^s is the regression coefficient when the hypothesis H_s is true, and β_k^v is the regression coefficients when the competing hypothesis is true, where $t = 1, \cdots, T$. As is elaborated in Appendix B, the intercept β_0 is set to zero, $\boldsymbol{x}_i^{(t)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ and the regression coefficients are chosen such that $y_{si}^{(t)} \sim \mathcal{N}(0,1)$, and $y_{vi}^{(t)} \sim \mathcal{N}(0,1)$, that is, in the population the regression coefficients are standardized.

3. Compute the values of the Bayes factor for each simulated data set. If H_s is true the Bayes factor is denoted by $\mathrm{BF}_{sv}^{(t)}$ $(t = 1, 2, \cdots, T)$ and if H_v is true, the Bayes factor is denoted by $\mathrm{BF}_{vs}^{(t)}$.

4. Calculate the proportion of the Bayes factor that is larger than BF_{thresh}, that is, $P(\mathrm{BF}_{sv}^{(t)} > \mathrm{BF}_{thresh}|H_s)$ denoted by p_s and the probability $P(\mathrm{BF}_{vs}^{(t)} > \mathrm{BF}_{thresh}|H_v)$ denoted by p_v.

5. If both p_s and p_v are larger than η, the algorithm stops and the sample sizes

364   computed are provided. Otherwise, the sample size N is progressively increased by

365   one, return to Step 2, and repeat Steps 3-5 until both p_s and p_v are larger than η.

366 The computing effort of the basic algorithm can be extremely high when the required

367 sample size is very large, since the number of the iterations is $N - 10 + 1$. In addition, to

368 execute the sensitivity analyses, the process from Step 1 to Step 5 has to be performed with

369 three different fraction values, namely $b = \frac{J}{N}$, $b = \frac{2J}{N}$ and $b = \frac{3J}{N}$ (see the section "Bayes

370 Factor"). Therefore, the computation effort is tripled. To reduce the computation effort, an

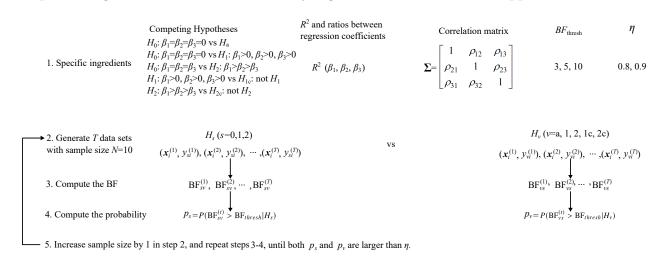371 improved algorithm based on a dichotomy algorithm is introduced in Appendix A.

Figure 2. The process for sample size determination with $K = 3$.

SSDRegression: A Function for Sample Size Determination for Multiple Linear

Regression

374 Sample size determination using the Bayes factor for evaluating null, informative,

375 complement, and unconstrained hypotheses within the multiple linear regression models

376 was implemented in a function `SSDRegression` in the `R` package `SSDbain` to facilitate

377 general utilization of the methodology. It has to be emphasized that the user can refer to

378 the help file for further elaboration of the function, and the function has been tested in the

379 test-that file. The code is available on GitHub [5]. This package already includes three

_____

[5] https://github.com/Qianrao-Fu/SSDbain

functions called "SSDttest", "SSDANOVA", "SSDANOVA_robust", which have been

introduced in Fu et al. (2021) and Fu et al. (unpublished). As a part of the `R` package

`SSDbain`, the function "SSDRegression" will now be introduced. This section describes the

specific input and the return results for function "SSDRegression". After installing the `R`

package `SSDbain` (which automatically installs `bain` if not already installed on your

computer), the following call is used to calculate the sample size required.

```
library(SSDbain)
Res<-SSDRegression(Hyp1="beta1=beta2=beta3=0",Hyp2="Ha",k=2,rho=matrix(c
    (1,0.2,0.2,0.2,1,0.2,0.2,0.2,1),nrow=3),R_square1=0,R_square2=0.13,T_
    sim=10000,BFthresh=3,eta=0.8,seed=10,standardize=FALSE,ratio=c(1,1))
```

The following arguments appear in this call:

1. `Hyp1` and `Hyp2`, strings that specify one pair of hypotheses of interest. For example, if

   H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ versus H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$,

   `Hyp1='beta1=beta2=beta3=0'`, `Hyp2='beta1>0 & beta2>0 & beta3>0'`. Attention

   should be paid to the following situations. If the unconstrained hypothesis is

   involved, `Hyp2='Ha'`; if the complement hypothesis is engaged, `Hyp2='Hc'`.

2. `K`, a positive integer that specifies the number of predictors. For example, if the

   model is $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$, `K=3`.

3. `rho`, a matrix that specifies the correlation between the predictors, which is a

   symmetric matrix with ones on the diagonal and values of ρ elsewhere. Here ρ is the

   correlation between two different predictors.

4. `R_square1` and `R_square2`, parameters used to specify the coefficients of

   determination R_1^2 and R_2^2 under `Hyp1` and `Hyp2`, respectively. Typically, the expected

   `R_square1` and `R_square2` are identified from 1) a pilot study; 2) a similar study; 3)

   a field defined meaningful effect; or 4) an educated guess based on informal

   observations and knowledge of the field.

5. `BFthresh`, a numeric value not less than 1 that specifies the required size of the Bayes factor for the true hypothesis. For example, if H_0 is compared with H_2, $\mathrm{BF}_{02} \geq \mathrm{BF}_{thresh}$ or $\mathrm{BF}_{20} \geq \mathrm{BF}_{thresh}$ should be reached. The default setting is `BFthresh=3`. In the next section it will be elaborated why this is the default value.

6. `eta`, a numeric value that specifies the probability that the Bayes factor is larger than `BFthresh` if either of the competing hypotheses is true. For example if H_0 versus H_2 and H_2 is true, $P(\mathrm{BF}_{20} > BF_{thresh}|H_2) \geq \eta$. The default setting is `eta=0.80`. In the next section it will be elaborated why this is the default value.

7. `T_sim`, a positive integer that specifies the number of data sets sampled from the populations corresponding to the two hypotheses of interest. A larger number of samples returns a more accurate sample size estimate but takes a longer time to run. Users are advised to start with a smaller number of samples (e.g., `T_sim=1000`) to get a rough estimate of the required sample size before confirming it with the default setting `T_sim=10000`.

8. `seed`, a positive integer that specifies the seed of `R`'s random number generator. The sample size required may be different with different seed values, but the number of simulated data sets `T_sim` can be large enough to ensure the stability of the results. It should be noted that at least `T_sim=10000` is required to guarantee the stability of the results. The default setting is `seed=10`. In Table 4 it will be illustrated that using `T_sim=10000` renders stable results.

9. `standardize`, a logical value that specifies whether hypotheses regarding standardized or unstandardized regression coefficients are evaluated. With `standardize = TRUE` hypotheses with respect to standardized regression coefficients are evaluated. With `standardize = FALSE` hypotheses with respect to unstandardized regression coefficients are evaluated. In the `R` package `bain`, the Bayes factors are calculated differently when standardized and unstandardized

432 coefficients are evaluated. The function `seBeta` in the R package `fungible` is used to

433 estimate the standardized regression coefficients and their corresponding covariance

434 matrix, and the function `lm` in the R package `stats` is used to estimate the

435 unstandardized regression coefficients and their corresponding covariance matrix. It

436 should be highlighted that if the ordered hypothesis for regression coefficients is

437 included, `standardize = TRUE` should be used. This is because, as explained earlier,

438 standardized coefficients are comparable, but unstandardized coefficients are not. For

439 other cases, `standardize = FALSE`.

440 10. `ratio`, an optional vector that specifies the ratio among the regression coefficients for

441 the population if H_1 is one of the interested hypothesises. For Situation 1, the ratio

442 of the regression coefficients 1: 1: $\cdots$: 1 is used for H_a. For Situation 2, the ratio of

443 the regression coefficients 1: 1: $\cdots$: 1 is used for H_1. For Situation 3, the ratio of the

444 regression coefficients 1: 1: $\cdots$: 1 is used for H_0 and the regression coefficients are

445 computed such that $R^2 = 0.13$, and the ratio of the regression coefficients Kd:

446 $(K-1)d$: $\cdots$: $3d$: $2d$: d is used for H_2, where d can be calculated by Equation 13. If

447 the order of regression coefficients in the hypothesis H_2 changes, the corresponding

448 ratio will follow the variation of the regression coefficients in the hypothesis. For

449 Situation 4, the ratio is consistent with H_1 in Situation 2, and the ratio is reordered

450 using the representative hypothesis (see Appendix B) for H_{1c}. For Situation 5, the

451 ratio is consistent with H_2 in Situation 3, and the ratio is reordered using the

452 representative hypothesis (see Appendix B) for H_{2c}, where K is the number of

453 predictors involved in the hypotheses. The elaborations of how this leads to

454 regression coefficients for the considered hypothesis can be found in Appendix B.

455 After running the function, the main output resulting from analyses are the sample size

456 required and the corresponding probability that the Bayes factor is larger than BF_{thresh}

457 when either of the competing hypotheses is true. As an example, if the following call to

SSDRegression is executed,

```
library(SSDbain)
SSDRegression(Hyp1='beta1=beta2=beta3=0',Hyp2='Ha',k=3,rho=matrix(c
    (1,0.2,0.2,0.2,1,0.2,0.2,0.2,1),nrow=3)
,R_square1=0,R_square2=0.13,T_sim=10000,BFthresh=3,eta=0.8,seed=10,
    standardize=FALSE,ratio=c(1,1,1)),
```

the results are obtained using fractions $b = J/N$, $b = 2J/N$ and $b = 3J/N$ (with the aim to address the sensitivity to the specification of the prior distribution):

```
    using N=146 and fraction b=0.0205
    P(BF0a>3|H0)=0.973
    P(BFa0>3|Ha)=0.803

    using N=120 and fraction b=0.0500
    P(BF0a>3|H0)=0.918
    P(BFa0>3|Ha)=0.804

    using N=105 and fraction b=0.0857
    P(BF0a>3|H0)=0.840
    P(BFa0>3|Ha)=0.806
```

According to the results, the sample size required is 146 if the minimum fraction $b = J/N$ is used. Moreover, the results of sensitivity analysis were summarized to develop a deeper understanding of the impact of the prior distributions in an applied Bayesian research. In this paper, the sensitivity analysis would entail adjusting the variance of the prior distribution to see how much impact the variance of the prior distribution makes on the final sample size. For example, if $b = 2J/N$ is used, the required sample size is 120, and it is 105 if $b = 3J/N$ is used. Therefore, the probabilities $P(\text{BF}_{0a} > 3|H_0)$ and $P(\text{BF}_{a0} > 3|H_a)$ are becoming closer with the increase of the fraction b. If the researchers want to obtain a conservative result (e.g., a convincing evidence should be required before another hypothesis is preferred over the null hypothesis), researchers can collect data with

487 sample size 146; and collect data with sample size 105 if they want to obtain a similar

488 probability for $P(\text{BF}_{0a} > 3|H_0)$ and $P(\text{BF}_{a0} > 3|H_a)$.

Sample Size Tables for Multiple Linear Regression

490 To investigate the sample size and highlight the properties of sample size determination for

491 multiple linear regression, a total of seven tables were made. Tables 1-6 containing two

492 predictors, three predictors, and four predictors, where the predictors are uncorrelated

493 ($\rho = 0$), weakly correlated ($\rho = 0.2$), and strongly correlated ($\rho = 0.5$) are shown. Tables

494 1-3 show the regression coefficients for the populations under the hypotheses H_0, H_1, H_2,

495 H_a, H_{1c}, and H_{2c} for the standard situation, which will be introduced in the next

496 paragraph. These regression coefficients are obtained via the approach elaborated in

497 Appendix B. Tables 4-6 demonstrate the required sample size and the corresponding

498 probability that the Bayes factor is larger than $\text{BF}_{thresh} = 3$, which can be used if users

499 agree with the "standard". Of course, users can differ in opinion, use other values and

500 compute the sample size using the R package SSDbain. Finally, the sample size comparison

501 under frequentist and Bayesian frameworks are shown in Table 7.

502 At the present, standard choices for Bayesian sample size determination do not exist and

503 will be proposed in this paper (see Tables 1-3). The standard situation is defined as

504 $R^2 = 0.13$, $\text{BF}_{thresh} = 3$, $\eta = 0.8$, and the ratios of the regression coefficients for different

505 hypotheses are described as follows. For Situation 1, the ratio of the regression coefficients

506 1: 1: $\cdots$: 1 is used for H_a. For situation 2, the ratio of the regression coefficients 1: 1: $\cdots$:

507 1 is used for H_1. For Situation 3, the ratio of the regression coefficients Kd: $(K-1)d$: $\cdots$:

508 $3d$, $2d$: d is used for H_2, where d can be calculated by Equation 13. If the order of

509 regression coefficients in the hypothesis H_2 changes, the corresponding ratio will follow the

510 variation of the regression coefficients in the hypothesis. The ratio of the regression

511 coefficients 1: 1: $\cdots$: 1 is used for H_0 and the coefficients are computed such that

512 $R^2 = 0.13$. For Situation 4, the ratio is consistent with H_1 in Situation 2, and the ratio is

reordered using the representative hypothesis (see Appendix B) for H_{1c}. For Situation 5,

the ratio is consistent with H_2 in Situation 3, and the ratio is reordered using the

representative hypothesis (see Appendix B) for H_{2c}. As for the definition of standard

situation, the reasons are as follows.

1. The coefficient of determination $R^2 = 0.13$ is selected, which corresponds to Cohen's

    medium effect size $f^2 = 0.15$. As phrased by Cohen (1988) the medium effect size is

    conceived as a size large enough to be visible to the naked eye. Meta-analyses showed

    that the average published effect size is around the medium effect size (Bakker et al.,

    2012), which therefore coincides with the needs of most psychologists.

2. The threshold of Bayes factor BF_{thresh} is 3 in the standard situation because the

    Bayes factor of 3 often matches the amount of evidence with $p-$value<0.05 (Jeffreys,

    1961; Wetzels et al., 2011). Besides, Dienes (2014) argued that the corresponding

    Bayes factor is about 3 when a result is just significant. Furthermore, the Bayes

    factor of 3 deserving attention is a consensus in the scientific community, which

    represents a just convincing evidence boundary (Dienes & Mclatchie, 2018; Jeffreys,

    1961; Kass & Raftery, 1995).

3. The value of $\eta = 0.8$ means that the probability that the Bayes factor exceeds the

    $\text{BF}_{thresh} = 3$ is at least 0.8 no matter which hypothesis in one pair of hypotheses is

    true. The value of 0.80 is used because it is a commonly accepted value for sufficient

    power in the classical framework. A pair of hypotheses is considered because the

    Bayes factors is symmetric in the sense that it allows accumulating evidence for

    either of these two hypothesis. This is in contrast with the p-values in null hypothesis

    significance testing where the type I error rate is 0.05, while the type II error rate is

    0.2.

4. According to the guidance in Vanbrabant et al. (2015), the differences between the

    regression coefficients should be equally spaced, and the common difference is denoted

539 by d. Therefore, if H_2 is considered, the ratio of Kd: $\cdots$: $3d$: $2d$: d, where d can be

540 calculated by Equation 13, or a reordering of the arithmetic sequence $Kd, \cdots, 3d, 2d,$

541 d for the variations of H_2 is chosen. If H_1 and H_a are considered, the ratio of 1: 1:

542 $\cdots$: 1 is chosen, where 1 may be replaced by -1 if the smaller than symbol $<$ is used

543 in H_1. If H_0 from Situation 3 is considered, the ratio of the regression coefficients is

544 1: 1: $\cdots$: 1. This ratio makes the absolute value of regression coefficients equal.

545 The results in Tables 4-6 are obtained with `set.seed=10`. To illustrate the stability of the

546 results with `T=10000`, Table 4 also presents (within parenthesis) the obtained sample sizes

547 using `set.seed=1234`. As can be seen, the results of sample size determination with

548 $T = 10000$ is not sensitive to the choice of the seed. Based on the results presented in these

549 tables, several features of SSD can be highlighted. 1) The required sample size for H_1

550 versus H_0 is smaller than for H_a versus H_0. For example, in Table 4, when $\rho = 0$, the

551 sample size is 121 if H_0 is compared with H_a, while a sample size of 90 is needed if H_1 is

552 used instead of H_a. The reason is that H_1 is more specific than H_a, which is illustrated in

553 Figure 3. The shaded area on the left of Figure 3 is the parameter space of H_a (no

554 constraints are imposed on the standardized regression coefficients β_1 and β_2), and the

555 shaded area on the right of Figure 3 is the parameter space of H_1. The parameter space for

556 H_1 is contained in H_a. Therefore, it is easier to distinguish H_0 from H_1 than from H_a.

557 Hence, a higher probability that the Bayes factor exceeds a determined threshold is

558 obtained and consequently a smaller sample-size is needed. 2) As the fraction used in the

559 prior distribution increases from $b = J/N$ to $b = 2J/N$, then to $b = 3J/N$, the required

560 sample size is reduced if H_0 is one of the hypotheses under consideration. For example, in

561 Table 5, when H_0 is compared with H_2 for $\rho = 0$, the required sample size is 776 for

562 $b = J/N$, 675 for $b = 2J/N$, and 624 for $b = 3J/N$. This can be explained as follows.

563 Firstly, the sample size is $N = \max\{N_1, N_2\}$, where N_1 is the sample size when H_0 is true,

564 and N_2 is the sample size when the competing hypothesis is true. Secondly, from Tables

565 4-6, the probability p_0 is much larger than 0.8, indicating that it is easier for the hypothesis

566 H_0 than for its competing hypothesis to reach the threshold of 3, thus the required sample

567 size is N_2 obtained if the competing hypothesis is true. Thirdly, the complexity c_0 becomes

568 larger as the fraction b increases. The reason is that a larger b implies a prior with a

569 smaller variance as shown in Equation 10 such that the prior density evaluated at

570 $\beta_1 = \beta_2 = \cdots = \beta_K$ or $\beta_1 = \beta_2 = \cdots = \beta_K = 0$ in Equation 4 is larger. Therefore, the

571 Bayes factors BF_{0a}, BF_{01}, and BF_{02} decrease as b increases. Taking the inverse yields the

572 opposite, the Bayes factor BF_{a0}, BF_{10}, and BF_{20} increase as b increases. Thus, the sample

573 size N_2 decreases with b. The advice about how to choose b is described in the final

574 paragraph of section "SSDRegression: A Function for Sample Size Determination for

575 Multiple Linear Regression". 3) When H_0 is compared with the order hypothesis H_2, the

576 required sample size is much larger than that in other cases in the same table. For

577 example, in Table 4, when H_0 versus H_2 for $\rho = 0.5$, the required sample size is 2721 for

578 $b = J/N$, 2527 for $b = 2J/N$, and 2426 for $b = 3J/N$. This occurs because the regression

579 coefficients are relatively close to each other as presented in Tables 1-3. However, the

580 sample size can be adjusted through enlarging the common difference from d to a multiple

581 of d. For example, the ratio with three predictors can be chosen as ($7d$: $4d$: d) instead of

582 ($3d$: $2d$: d) in Table 5, which will reduce the sample size because it is easier to distinguish

583 regression coefficients that have more unequal sizes. 4) In general, as the number of

584 predictors increases, the sample size required increases. For example, if H_0 is compared

585 with H_a, $\rho = 0$, and fraction $b = J/N$, from Tables 4-6, we observe that the sample size

586 increases from 121 to 147, then to 170 when the number of predictors increases from 2 to 3,

587 then to 4. This is consistent with the property for the classical sample size in Table 7.

588 From Table 7, we can see that the sample size increases from 68 to 77, then to 85 for 2, 3,

589 and 4 predictors, respectively. This is because when more predictors are added a higher

590 level of evidence is required to state that any of them are significant (classical) or

591 substantial (Bayesian).

(a) H_0: $\beta_1 = \beta_2 = 0$ vs. H_a

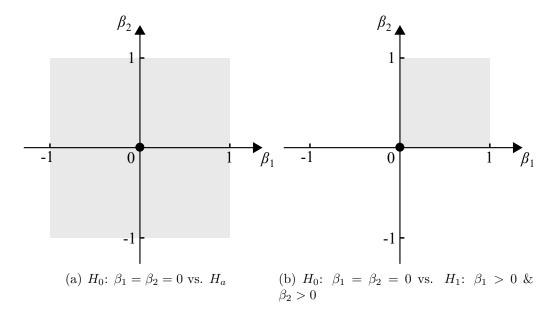(b) H_0: $\beta_1 = \beta_2 = 0$ vs. H_1: $\beta_1 > 0$ & $\beta_2 > 0$

Figure 3. The grey area in Figure (a) is the admissible parameter space for H_a, the grey area in Figure (b) is the admissible parameter space for H_1, and the black bold dot is the admissible parameter space for H_0.

As shown in Table 7, the comparison of sample size for the standard situation in the Bayesian framework and standard a priori power analysis are illustrated. The focused Situation is H_0 vs H_a because this situation exists in both classical and Bayesian hypothesis testing. Situation H_0 vs H_1 in Table 7 is used for supplementary illustration. From Table 7, compared with the classical sample size, a larger sample size is required for the SSD based on the Bayes factor. For instance, if H_0: $\beta_1 = \beta_2 = 0$ versus H_a, the sample size is 68 in the classical framework, while the sample size is 121 if the fraction $b = 2/N$ is used for Bayes factor. When $b = 2/N$ is used, the Bayes factor is very conservative (see, Hoijtink, Mulder, et al., 2019). In other words, the sample size has to be large enough to provide a convincing evidence to support the non-null hypothesis. Furthermore, for a less conservative value such as $b = 2J/N$ and $b = 3J/N$ the required sample size decreases. According to the first item in Table 7, the sample size is 104 for $b = 2J/N$, and 95 for $b = 3J/N$, which approaches the classical sample size of 68 although it remains slightly larger. Next, when H_0 is compared with the sign hypothesis H_1, the required sample sizes

606 (depending on b) may even be smaller than the ones in the classical framework. For

607 example, when the number of predictors $K = 3$, the required sample size is 71 for

608 $b = 2J/N$ and 66 for $b = 3J/N$, which are smaller than the classical sample size of 77.

609 The sample size resulting from Bayesian may be larger than the sample size resulting from

610 power analysis, but the Bayes factor provides more information than the p-value. 1) The

611 Bayes factor can quantify the degree of evidence supporting one hypothesis over another

612 hypothesis. For example, in Tables 4-6, the sample size is calculated such that the degree

613 of evidence of supporting the true hypothesis is 3 times larger than the competing

614 hypothesis. 2) Researchers can obtain more specific knowledge if they use an informative

615 hypothesis instead of the traditional alternative hypothesis. For example, if H_0 is

616 compared with H_2, researchers do not only know if the three coefficients are equal or not,

617 but also know the order of the coefficients. 3) The required sample size can be adjusted

618 through sensitivity analysis by modifying the scaling parameter of the prior distribution. If

619 researchers favor the null hypothesis, a larger scaling parameter that corresponds to a

620 smaller fraction b would be chosen because it makes it easier to reject the alternative in

621 favor of the null hypothesis; and if they want to obtain relatively symmetrical evidence for

622 supporting one pair of competing hypothesis, the fraction b can be chosen such that both

623 probabilities become as equal as possible. This can be illustrated using the comparison of

624 H_0 with H_1 in Table 5. If $b = J/N$ is chosen, the probability that the Bayes factor is larger

625 than 3 is 0.964 when H_0 is true, but the probability that the Bayes factor is larger than 3 is

626 0.802 when H_1 is true. If $b = 3J/N$ is chosen, the probability that the Bayes factor is larger

627 than 3 is 0.811 when H_0 is true, and the probability that the Bayes factor is larger than 3

628 is 0.833, which are relatively close.

629 **Illustrative example 1**

630 A study has been designed by a psychologist to explore the relationship between *the SAT*

631 *score of students* and *achievement levels* (β_1), *cultural factors* (β_2), *socioeconomic status*

632 (β_3), and *psychological factors* (β_4). To determine the sample size it is assumed that the

633 correlation between each pair of predictors equals 0.3. The psychologist plans to compare

634 hypothesis H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ with H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$ & $\beta_4 > 0$. Based

635 on experience, the psychologist expects that R^2 will be about 0.09. The sample size should

636 be large enough such that $\text{BF}_{thresh} = 3$ and $\eta = 0.8$ are satisfied. The required sample size

637 can be determined using the following call to `SSDRegression`.

```
638  Res<-SSDRegression(Hyp1='beta1=beta2=beta3=beta4=0',Hyp2='beta1>0&beta2>0&
639      beta3>0&beta4>0',k=4,rho=matrix(c
640      (1,0.3,0.3,0.3,0.3,1,0.3,0.3,0.3,0.3,1,0.3,0.3,0.3,0.3,1),nrow=k),R_
641      square1=0,R_square2=0.13,T_sim=10000,BFthresh=3,eta=0.8,seed=10,
642      standardize=FALSE,ratio=c(1,1,1,1))
```

643 The results are as follows.

```
644     using N=151 and fraction b=0.0265
645     P(BF01>3|H0)=0.990
646     P(BF10>3|H1)=0.803
647
648     using N=107 and fraction b=0.0748
649     P(BF01>3|H0)=0.948
650     P(BF10>3|H1)=0.805
651
652     using N=78 and fraction b=0.1538
653     P(BF01>3|H0)=0.862
654     P(BF10>3|H1)=0.802
```

655 Based on the above results, if the researchers use the minimum fraction $b = J/N$ of the

656 data for the prior distribution, the required sample size is 151. If the fraction of the data is

657 increased to *two times* of the minimum fraction $2J/N$, the required sample size is reduced

658 to 107. If the fraction of the data is increased to *three times* of the minimum fraction

659 $3J/N$, the required sample size is reduced to 78. If the resources are sufficient, and the

660 psychologist wants to get a larger probability that the Bayes factor supports the null

hypothesis if it is true, he or she should use a smaller b. If the resources are insufficient, or the psychologist wants to get two relatively close probability values regardless of whether the null or non-null hypothesis is true, a larger b is recommended. Specifically, from the result above, we can see that when $b = 0.0265$ is used, the probability that the Bayes factor supports the null hypothesis H_0 is 0.990, which is larger than the probability of 0.803 that the Bayes factor supports H_1. If $b = 0.1538$ is used, the probability that the Bayes factor supports H_0 is 0.862, and the probability that the Bayes factor supports H_1 is 0.802. These two probabilities are much closer.

Illustrative example 2 Considering the example from Vanbrabant et al. (2015): a group of psychologists wants to investigate the relation among IQ and social skills (β_1), interest in artistic activities (β_2), and use of complicated language patterns (β_3). The hypotheses of interest are H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$ versus H_c: not H_1. The psychologists expect a medium effect size $f^2 = 0.15$, which corresponding to a coefficient of determination $R^2 = 0.13$. The correlation matrix $\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{bmatrix}$. After looking up the corresponding results in Table 5, the following information can be obtained

```
1  using N=40 and fraction b=0.0750
2  P(BF1c>3|H1)=0.809
3  P(BFc1>3|Hc)=0.827
```

Based on the above results, the required sample size is 40. As elaborated earlier in this paper, for hypotheses $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$ and H_c, the Bayes factor is not sensitive to the prior distribution regardless which fraction b is used.

Conclusion

This paper proposed a sample size determination method to evaluate the classical null, unconstrained, and informative hypotheses (and their complement) in the context of the

685 multiple linear regression model. The presented sample size tables will benefit researchers

686 as they will be able to look up the necessary sample size if they aim to use the standard

687 Bayesian situation, that is, $\text{BF}_{thresh} = 3$, $\eta = 0.8$, $R^2 = 0.13$, and the ratios of the

688 regression coefficients for different hypotheses are described as follows. For Situation 1, the

689 ratio of the regression coefficients 1: 1: $\cdots$: 1 is used for H_a. For situation 2, the ratio of

690 the regression coefficients 1: 1: $\cdots$: 1 is used for H_1. For Situation 3, the ratio of the

691 regression coefficients Kd: $(K-1)d$: $\cdots$: $3d$: $2d$: d is used for H_2, where d can be

692 calculated by Equation 13. If H_0 from Situation 3 is considered, the ratio of the regression

693 coefficients is 1: 1: $\cdots$: 1. If the order of regression coefficients in the hypothesis H_2

694 changes, the corresponding ratio will follow the variation of the regression coefficients in

695 the hypothesis. For Situation 4, the ratio is consistent with H_1 in Situation 2, and the ratio

696 is reordered using the representative hypothesis (see Appendix B) for H_{1c}. For Situation 5,

697 the ratio is consistent with H_2 in Situation 3, and the ratio is reordered using the

698 representative hypothesis (see Appendix B) for H_{2c}, where K is the number of predictors

699 involved in the hypotheses. If researchers aim to use other situations than the standard

700 ones covered in the tables, the function `SSDRegression`, which is part of the `R` package,

701 `SSDbain` can be used to help researchers to calculate the sample size. Compared with the

702 unconstrained hypothesis, the introduction of informative hypotheses results in a

703 substantial gain in the probability that the Bayes factor exceeds the threshold and thus

704 reduces the required sample size.

705 This paper makes an important contribution to sample size determination for informative

706 hypotheses using Bayes factor within multiple linear regression models. However, it has

707 some limitations. First, sample size determination is available if the assumptions of the

708 regression model used to simulate the data also apply to real data. Secondly, since missing

709 data may occur in real data, and the researchers will have to guess which proportion of

710 their data is missing, and adjust the required sample size accordingly. Thirdly, commonly

711 used models at the present such as t-test, one-way ANOVA and multiple linear regression

712 have been considered. However, SSD extensions to more complex models, like structural

713 equation modeling, and general multivariate models still have to be developed and will be

714 added to the package of `SSDbain` in the future.

715 Despite these limitations, the `R` package `SSDbain` should be a welcome addition to the

716 applied researcher's toolbox, and can help the researcher identify the required sample sizes

717 while planning a research project.

⁷¹⁸ References

⁷¹⁹ Bakker, M., Van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called

⁷²⁰ psychological science. *Perspectives on Psychological Science*, *7*(6), 543–554.

⁷²¹ https://doi.org/10.1177/1745691612459060

⁷²² Berger, J. O. (1986). *Are p-values reasonable measures of accuracy*. In Pacific Statistical

⁷²³ Congress (I.S.Francis et al., eds.). North-Holland, Amsterdam.

⁷²⁴ Berger, J. O., & Pericchi, L. R. (1996). The intrinsic bayes factor for model selection and

⁷²⁵ prediction. *Journal of the American Statistical Association*, *91*(433), 109–122.

⁷²⁶ https://doi.org/http://dx.doi.org/10.1080/01621459.1996.10476668

⁷²⁷ Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of

⁷²⁸ p values and evidence. *Journal of the American statistical Association*, *82*(397),

⁷²⁹ 112–122. https://doi.org/https://doi.org/10.1080/01621459.1987.10478397

⁷³⁰ Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale,

⁷³¹ NJ: Erlbaum.

⁷³² Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

⁷³³ https://doi.org/10.1037/0033-2909.112.1.155

⁷³⁴ Cohen, J. (1994). The earth is round (p<. 05). *American Psychologist*, *49*(12), 997–1003.

⁷³⁵ https://doi.org/http://dx.doi.org/10.1037/0003-066X.49.12.997

⁷³⁶ De Santis, F., & Spezzaferri, F. (2001). Consistent fractional bayes factor for nested normal

⁷³⁷ linear models. *Journal of Statistical Planning and Inference*, *97*(2), 305–321.

⁷³⁸ https://doi.org/https://doi.org/10.1016/S0378-3758(00)00240-8

⁷³⁹ Dienes, Z. (2014). Using bayes to get the most out of non-significant results. *Frontiers in*

⁷⁴⁰ *Psychology*, *5*, 781. https://doi.org/https://doi.org/10.3389/fpsyg.2014.00781

⁷⁴¹ Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer bayesian analyses over

⁷⁴² significance testing. *Psychonomic Bulletin & Review*, *25*(1), 207–218.

⁷⁴³ https://doi.org/10.3758/s13423-017-1266-z

744 Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for

745   psychological research. *Psychological Review*, *70*(3), 193.

746   https://doi.org/https://doi.org/10.1037/h0044139

747 Elashoff, J. (2007). *nQuery version 7.0 advisor user's guide*. Los Angeles, CA, USA.

748 Fanelli, D. (2009). How many scientists fabricate and falsify research? a systematic review

749   and meta-analysis of survey data. *PloS One*, *4*(5), e5738.

750   https://doi.org/https://doi.org/10.1371/journal.pone.0005738

751 Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using

752   g* power 3.1: Tests for correlation and regression analyses. *Behavior Research*

753   *Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

754 Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible

755   statistical power analysis program for the social, behavioral, and biomedical sciences.

756   *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

757 Fu, Q., Hoijtink, H., & Moerbeek, M. (2021). Sample-size determination for the bayesian t

758   test and welch's test using the approximate adjusted fractional bayes factor.

759   *Behavior Research Methods*, *53*(1), 139–152.

760   https://doi.org/10.3758/s13428-020-01408-1

761 Fu, Q., Moerbeek, M., & Hoijtink, H. (unpublished). Sample size determination for

762   Bayesian anovas with informative hypotheses. Retrieved from

763   10.31234/osf.io/ymvb9.

764 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013).

765   *Bayesian data analysis* (3rd ed.). Chapman; Hall/CRC.

766 Gu, X., Hoijtink, H., Mulder, J., Van Lissa, C. J., Van Zundert, C., Jones, J., & Waller, N.

767   (2021). `bain`: Bayes factors for informative hypotheses [computer software manual].

768   R package version 0.2.6. Retrieved from

769   https://cran.r-project.org/web/packages/bain/index.html.

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

Harlow, L. L. E., Mulaik, S. A., & Steiger, J. H. (2016). *What if there were no significance
tests?* New York: Routledge.

Hintze, J. (2011). *Pass 11.* Kaysville, Utah, USA: NCSS, LLC.

Hoijtink, H. (2011). *Informative hypotheses: Theory and practice for behavioral and social
scientists.* Boca Raton: Chapman; Hall/CRC.

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

Howell, D. C. (2012). *Statistical methods for psychology* (8th ed.). Belmont, CA: Cengage
Learning.

Hurlbert, S. H., & Lombardi, C. M. (2009). Final collapse of the neyman-pearson decision
theoretic framework and rise of the neofisherian. *Annales Zoologici Fennici, 46*(5),
311–349. https://doi.org/10.5735/086.046.0501

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine,
2*(8), e124. https://doi.org/https://doi.org/10.1371/journal.pmed.0020124

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

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical
Association, 90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A bayesian approach. *Psychological Methods*, *10*(4), 477. https://doi.org/10.1037/1082-989X.10.4.477

Lee, M. D., Steyvers, M., De Young, M., & Miller, B. (2012). Inferring expertise in knowledge and prediction ranking tasks. *Topics in Cognitive Science*, *4*(1), 151–163. https://doi.org/10.1111/j.1756-8765.2011.01175.x

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course.* Cambridge university press.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., et al. (2019). Jasp: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*(1), 1–17. https://doi.org/10.18637/jss.v088.i02

Ly, A., Verhagen, J., & Wagenmakers, E.-J. (2016). Harold jeffreys's default bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32. https://doi.org/https://doi.org/10.1016/j.jmp.2015.06.004

Masicampo, E., & Lalande, D. R. (2012). A peculiar prevalence of p values just below. 05. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2271–2279. https://doi.org/10.1080/17470218.2012.711335

Mayr, S., Erdfelder, E., Buchner, A., & Faul, F. (2007). A short tutorial of gpower. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 51–59. https://doi.org/10.20982/tqmp.03.2.p051

Morey, R. D., Rouder, J. N., Jamil, T., & Urbanek, S. (2018). `BayesFactor`: Computation of bayes factors for common designs. R package version 0.9.12-4.2. Retrieved from https://cran.r-project.org/web/packages/BayesFactor/index.html.

Mulder, J. (2014). Prior adjusted default bayes factors for testing (in) equality constrained hypotheses. *Computational Statistics & Data Analysis*, *71*, 448–463. https://doi.org/https://doi.org/10.1016/j.csda.2013.07.017

Mulder, J., Van Lissa, C. J., Williams, D. R., Gu, X., Olsson-Collentine, A., Boeing-Messing, F., Fox, J.-P., et al. (2021). `BFpack`: Flexible bayes factor testing of scientific expectations [computer software manual]. R package version 0.3.2. Retrieved from https://cran.r-project.org/web/packages/BFpack/index.html.

Myung, I. J., & Pitt, M. A. (1997). Applying occam's razor in modeling cognition: A bayesian approach. *Psychonomic Bulletin & Review*, *4*(1), 79–95. https://doi.org/10.3758/BF03210778

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241–301. https://doi.org/10.1037/1082-989X.5.2.241

O'Hagan, A. (1995). Fractional bayes factors for model comparison. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 99–118. https://doi.org/https://doi.org/10.1111/j.2517-6161.1995.tb02017.x

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 943–951. https://doi.org/10.1126/science.aac4716

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. https://doi.org/http://dx.doi.org/10.2307/271063

Royall, R. (1997). *Statistical evidence: A likelihood paradigm*. New York, NY: Chapman; Hall/CRC.

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as

significant. *Psychological science*, *22*(11), 1359–1366.

https://doi.org/https://doi.org/10.1177/0956797611417632

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E.-J. (2019). A tutorial on bayes factor design analysis using an informed prior. *Behavior Research Methods*, *51*(3), 1042–1058. https://doi.org/https://doi.org/10.3758/s13428-018-01189-8

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business*. New York, NY: Doubleday.

Van Assen, M. A., Van Aert, R. C., Nuijten, M. B., & Wicherts, J. M. (2014). Why publishing everything is more effective than selective publishing of statistically significant results. *PLoS one*, *9*(1), e84896.

https://doi.org/https://doi.org/10.1371/journal.pone.0084896

Van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*(2), 217–239. https://doi.org/10.1037/met0000100

Vanbrabant, L., Van De Schoot, R., & Rosseel, Y. (2015). Constrained statistical inference: Sample-size tables for anova and regression. *Frontiers in Psychology*, *5*, 1565.

https://doi.org/10.3389/fpsyg.2014.01565

Vandekerckhove, J., Rouder, J. N., & Kruschke, J. K. (2018). Editorial: Bayesian methods for advancing psychological science. *Psychonomic Bulletin & Review*, *25*(1), 1–4.

https://doi.org/10.3758/s13423-018-1443-8

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.

https://doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, *25*(3), 169–176.

https://doi.org/10.1177/0963721416643289

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J.

    (2011). Statistical evidence in experimental psychology: An empirical comparison

    using 855 t tests. *Perspectives on Psychological Science*, *6*(3), 291–298.

    https://doi.org/https://doi.org/10.1177/1745691611406923

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., &

    Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and

    reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in*

    Psychology, *7*, 1832. https://doi.org/10.3389/fpsyg.2016.01832

Appendix A: SSD based on dichotomy search algorithm

In this appendix, an improved algorithm that can effectively speed up the calculation process based on the dichotomy search is introduced. As described in the basic algorithm in the section "The Basic Algorithm Used for Sample Size Determination", a large amount of iterations are conducted for the calculation process from Steps 2-5, which brings great burdens on the computation and costs more time to reach the conditions in Step 5. The presented search-based process can sharply reduce the computing burden by reducing the number of iterations. The basic idea of the proposed algorithm is to gradually adjust the sample size using a dichotomy algorithm until $p_s \geq \eta$ and $p_v \geq \eta$ hold. The Steps 4-5 described in the basic algorithm in the section "The Basic Algorithm Used for Sample Size Determination" are replaced by the following steps.

Step 4: Define the lower and upper bound of sample size N for the dichotomy search method.

To use the dichotomy search method, the interval of the sample size needs to be determined. Let LB and UB denote the lower and upper bound of the optimal sample size N, where the LB cannot be smaller than 10. In order to narrow the distance between lower and upper bound, the following steps are conducted.

- Compute p_s and p_v using Steps 2-4 from the basic algorithm based on $N = 100$.

- (i) If $p_s \geq \eta$ and $p_v \geq \eta$, then set $N = \frac{N}{2}$, and repeat Steps 2-4 and (i) until $p_s < \eta$ or $p_v < \eta$. Then set $LB = N$, $UB = 2N$.

- (ii) If $p_s < \eta$ or $p_v < \eta$, then set $N = 2N$, and repeat Steps 2-4 and (ii) until $p_s \geq \eta$ and $p_v \geq \eta$. Then set $LB = \frac{N}{2}$, $UB = N$.

Step 5: Compute the optimal sample size

(I) Based on the LB and UB determined by above, let $N_{\text{mid}} = \frac{LB+UB}{2}$.

(II) Compute p_s and p_v with $N = N_{\text{mid}}$ using Steps 2-4.

(III) If $p_s \geq \eta$ and $p_v \geq \eta$, $UB = N_{\text{mid}}$; else, $LB = N_{\text{mid}}$.

(IV) Update the value of N_{mid} with $N_{\text{mid}} = \frac{LB+UB}{2}$.

(V) Return to Step II with $N = N_{\text{mid}}$ and repeat Steps II-V until $N_{\text{mid}} = LB + 1$ satisfied. Then $N = N_{\text{mid}}$.

Appendix B: How to calculate the regression coefficients based on the coefficient of determination and the ratio among the regression coefficients

In this appendix, the process of calculating the regression coefficients is described if the input ingredients coefficient of determination R^2 and the ratio among the regression coefficients are given instead of the regression coefficients are given directly. This matter has been mentioned in Section "The Basic Algorithm Used for Sample Size Determination". Based on the Model in Equation 1, the regression coefficients can be computed as follows.

1. Variance can be calculated on both sides of Equation 1:

$$VAR[y_i] = VAR[\sum_{k=1}^{K} \beta_k x_{ik}] + VAR(\epsilon_i). \tag{11}$$

918   2. Divide by VAR[y_i] on both sides of Equation 11,

$$1 = R^2 + \frac{VAR(\epsilon_i)}{VAR[y_i]}. \tag{12}$$

919   3. As presented in the section "The Basic Algorithm Used for Sample Size

920   Determination", VAR[x_{ik}]=1 and VAR[y_i]=1. Combining with Equation 10 and 11,

921   the formula for coefficient of determination R^2 can be rewritten as

$$R^2 = VAR[\sum_{k=1}^{K} \beta_k x_{ik}] = \sum_{k=1}^{K} \beta_k^2 + 2 \sum_{k<k\prime} \beta_k \beta_{k\prime} \rho_{kk\prime}, \tag{13}$$

922   where $\rho_{kk\prime}$ denotes the correlation between predictor variables x_{ik} and $x_{ik\prime}$, which is

923   the element in the correlation matrix $\boldsymbol{\Sigma}$.

924   The ratio $\beta_1 : \beta_2 : \cdots : \beta_K$ can be ascertained from pilot study, published results from a

925   similar study, and can be estimated based on the expert's advice and the prior knowledge

926   of the field. If hypotheses H_a and H_1 are considered, and the ratio

927     $\beta_1 : \beta_2 : \cdots : \beta_K = 1 : 1 : \cdots : 1$, then $\beta_1 = \beta_2 = \cdots = \beta_K$. For hypothesis

928     $H_0 : \beta_1 = \beta_2 = \cdots = \beta_K$ itself, the relation of the regression coefficients is also

929     $\beta_1 = \beta_2 = \cdots = \beta_K$. By substituting R^2, $\boldsymbol{\Sigma}$, $\beta_1 = \beta_2 = \cdots = \beta_K$ into Equation 13,

930     $\beta_1, \beta_2, \cdots, \beta_K$ can be derived. If hypotheses H_2 is considered, and the ratio

931     $\beta_1 : \beta_2 : \cdots : \beta_K = r_1 : r_2 : \cdots : r_K$, then $\beta_1 = r_1/r_K \beta_K$, $\beta_2 = r_2/r_K \beta_K$, $\cdots$,

932     $\beta_{K-1} = r_{K-1}/r_K \beta_K$. By substituting R^2, $\boldsymbol{\Sigma}$, $\beta_1 = r_1/r_K \beta_K$, $\beta_2 = r_2/r_K \beta_K$, $\cdots$,

933     $\beta_{K-1} = r_{K-1}/r_K \beta_K$ into Equation 13, β_K can be derived. Subsequently, $\beta_1, \beta_2, \cdots, \beta_{K-1}$

934   can be obtained. It should be noted that the default signs of the regression coefficients are

935   positive for all the hypotheses, unless they are designated to be negative in the hypotheses.

936   It will now be explained how the population regression coefficients for the complement of

937   H_1 will be determined. Changing the sign of the regression coefficients once is called a

938   violation. The number of violations is determined by the number of the signs of regression

coefficients changed. The complement hypothesis of H_1 can be divided into K categories

based on the number of violations. That is, if K predictors are considered, there will be K

categories, namely one deviation from H_1, two violations from H_1, $\cdots$, K violations from

H_1. To facilitate the reader's understanding of the proposed approach, two examples are

provided with two and five predictors. Firstly, this paper discusses the simplest situation

where only two predictors are needed. In this situation, the hypothesis H_1 can be

expressed as $\beta_1 > 0 \& \beta_2 > 0$. The complement of H_1 with two predictors includes the

following three cases:

$H_{c1} : \beta_1 < 0 \& \beta_2 > 0$.

$H_{c2} : \beta_1 > 0 \& \beta_2 < 0$.

$H_{c3} : \beta_1 < 0 \& \beta_2 < 0$.

In the above three cases, the hypotheses H_{c1} and H_{c2} have one thing in common. They

have one violation from H_1. The order could also be H_{c2}, H_{c1}, and H_{c3}, since the Bayes

factor $\text{BF}_{c1,1}$ for H_{c1} versus H_1 if data are simulated from populations H_{c1} and $\text{BF}_{c2,1}$ for

H_{c2} versus H_1 if data are simulated from populations H_{c2} are almost the same. That is,

there is no preference for one of these two hypotheses. Different from these two hypotheses,

the hypothesis H_{c3} has two violations. The more numbers of violations in the complement

of H_1, the easier it is to distinguish the complement hypothesis from the hypothesis H_1.

Therefore, the Bayes factor $BF_{c3,1}$ for H_{c3} versus H_1 if data are simulated from population

H_{c3} is larger than the $BF_{c1,1}$ and $BF_{c2,1}$ if data are simulated from populations H_{c1} and

H_{c2}, respectively. To determine the regression coefficients for the complement of H_1, a

representative hypothesis has to be selected. In this paper, the hypothesis corresponding to

the median of the number of hypotheses ordered using the number of violations is selected

as the representative hypothesis of the complement hypothesis of H_1, in this case, H_{c2}. The

measure of median is used because it refers to the most central value in the ascending

Bayes factors which are ordered using the number of violations.

Based on the ratio $\beta_1: \beta_2 = 1:1$ for H_1, the relation of the regression coefficients under the

966 complement hypothesis H_{1c} is $\beta_2 = -\beta_1$, where $\beta_1 > 0$. By substituting R^2, ρ, and

967 $\beta_2 = -\beta_1$ into Equation 13, the regression coefficients can be calculated.

968 To further clarify the proposed method, a more complex scenario with five predictors is

969 discussed. The hypothesis H_1 can be expressed as $\beta_1 > 0 \& \beta_2 > 0 \& \beta_3 > 0 \& \beta_4 > 0 \& \beta_5 > 0$.

970 To find the hypothesis representative of the complement of H_1, all the possible hypotheses

971 in the complement are ordered using the number of violations. First, only one violation is

972 considered. The number of hypotheses in this case will be $\binom{5}{1}$. In other words, five

973 hypotheses should be considered if one violation happens:

974 $H_{c1} : \beta_1 < 0 \& \beta_2 > 0 \& \beta_3 > 0 \& \beta_4 > 0 \& \beta_5 > 0$.

975 $H_{c2} : \beta_1 > 0 \& \beta_2 < 0 \& \beta_3 > 0 \& \beta_4 > 0 \& \beta_5 > 0$.

976 $H_{c3} : \beta_1 > 0 \& \beta_2 > 0 \& \beta_3 < 0 \& \beta_4 > 0 \& \beta_5 > 0$.

977 $H_{c4} : \beta_1 > 0 \& \beta_2 > 0 \& \beta_3 > 0 \& \beta_4 < 0 \& \beta_5 > 0$.

978 $H_{c5} : \beta_1 > 0 \& \beta_2 > 0 \& \beta_3 > 0 \& \beta_4 > 0 \& \beta_5 < 0$.

979 It should be noted that the order is arbitrary, and any permutation would also be

980 acceptable. The order is irrelevant because all hypothesis containing one violation will lead

981 to about the same Bayes factors $BF_{ci,1}$ for H_{ci} versus H_1 if data are simulated from

982 populations H_{ci} $(i = 1, \cdots, 5)$, respectively. Similarly, in the following, permutations are

983 also arbitrary for the same number of violations.

984 If there are two violations in a hypothesis, the total number of hypotheses will be $\binom{5}{2}$, that

985 is ten. We name these hypotheses as H_{c6}, H_{c7},...and H_{c15}, which are given as:

986 $H_{c6} : \beta_1 < 0 \& \beta_2 < 0 \& \beta_3 > 0 \& \beta_4 > 0 \& \beta_5 > 0$.

987 $H_{c7} : \beta_1 < 0 \& \beta_2 > 0 \& \beta_3 < 0 \& \beta_4 > 0 \& \beta_5 > 0$.

988 $H_{c8} : \beta_1 < 0 \& \beta_2 > 0 \& \beta_3 > 0 \& \beta_4 < 0 \& \beta_5 > 0$.

989 $H_{c9} : \beta_1 < 0 \& \beta_2 > 0 \& \beta_3 > 0 \& \beta_4 > 0 \& \beta_5 < 0$.

990 $H_{c10} : \beta_1 > 0 \& \beta_2 < 0 \& \beta_3 < 0 \& \beta_4 > 0 \& \beta_5 > 0$.

991 $H_{c11} : \beta_1 > 0 \& \beta_2 < 0 \& \beta_3 > 0 \& \beta_4 < 0 \& \beta_5 > 0$.

$H_{c12} : \beta_1 > 0 \& \beta_2 < 0 \& \beta_3 > 0 \& \beta_4 > 0 \& \beta_5 < 0.$

$H_{c13} : \beta_1 > 0 \& \beta_2 > 0 \& \beta_3 < 0 \& \beta_4 < 0 \& \beta_5 > 0.$

$H_{c14} : \beta_1 > 0 \& \beta_2 > 0 \& \beta_3 < 0 \& \beta_4 > 0 \& \beta_5 < 0.$

$H_{c15} : \beta_1 > 0 \& \beta_2 > 0 \& \beta_3 > 0 \& \beta_4 < 0 \& \beta_5 < 0.$

Similarly, there are $\binom{5}{3}$ (ten) hypotheses with three violations, which are shown as:

$H_{c16} : \beta_1 < 0 \& \beta_2 < 0 \& \beta_3 < 0 \& \beta_4 > 0 \& \beta_5 > 0.$

$H_{c17} : \beta_1 < 0 \& \beta_2 < 0 \& \beta_3 > 0 \& \beta_4 < 0 \& \beta_5 > 0.$

$H_{c18} : \beta_1 < 0 \& \beta_2 < 0 \& \beta_3 > 0 \& \beta_4 > 0 \& \beta_5 < 0.$

$H_{c19} : \beta_1 < 0 \& \beta_2 > 0 \& \beta_3 < 0 \& \beta_4 < 0 \& \beta_5 > 0.$

$H_{c20} : \beta_1 < 0 \& \beta_2 > 0 \& \beta_3 > 0 \& \beta_4 > 0 \& \beta_5 < 0.$

$H_{c21} : \beta_1 > 0 \& \beta_2 < 0 \& \beta_3 < 0 \& \beta_4 < 0 \& \beta_5 > 0.$

$H_{c22} : \beta_1 > 0 \& \beta_2 < 0 \& \beta_3 < 0 \& \beta_4 > 0 \& \beta_5 < 0.$

$H_{c23} : \beta_1 > 0 \& \beta_2 < 0 \& \beta_3 > 0 \& \beta_4 < 0 \& \beta_5 < 0.$

$H_{c24} : \beta_1 > 0 \& \beta_2 < 0 \& \beta_3 > 0 \& \beta_4 < 0 \& \beta_5 < 0.$

$H_{c25} : \beta_1 > 0 \& \beta_2 > 0 \& \beta_3 < 0 \& \beta_4 < 0 \& \beta_5 < 0.$

There are $\binom{5}{4}$ (five) hypotheses with four violations, which are displayed as follows.

$H_{c26} : \beta_1 < 0 \& \beta_2 < 0 \& \beta_3 < 0 \& \beta_4 < 0 \& \beta_5 > 0.$

$H_{c27} : \beta_1 < 0 \& \beta_2 < 0 \& \beta_3 < 0 \& \beta_4 > 0 \& \beta_5 < 0.$

$H_{c28} : \beta_1 < 0 \& \beta_2 < 0 \& \beta_3 > 0 \& \beta_4 < 0 \& \beta_5 < 0.$

$H_{c29} : \beta_1 < 0 \& \beta_2 > 0 \& \beta_3 < 0 \& \beta_4 < 0 \& \beta_5 < 0.$

$H_{c30} : \beta_1 > 0 \& \beta_2 < 0 \& \beta_3 < 0 \& \beta_4 < 0 \& \beta_5 < 0.$

Finally, the hypotheses with five violations only have $\binom{5}{5}$ (one) case, which is given as

$H_{c31} : \beta_1 < 0 \& \beta_2 < 0 \& \beta_3 < 0 \& \beta_4 < 0 \& \beta_5 < 0.$

As mentioned in the last example, the Bayes factor $\text{BF}_{1c,1}$ for H_{1c} versus H_1 increases with the number of violations. There are totally $\binom{5}{1} + \binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5} = 2^5 - 1 = 31$ hypothesis for the complement of H_1. After all the hypotheses are presented, the

representative hypothesis H_{c16} is selected, which is the hypothesis corresponding the

median of the hypotheses ordered using the number of violations. Based on the ratio β_1:

β_2: β_3: β_4: β_5=1:1:1:1:1 for H_1, the relation of the regression coefficients under the

complement hypothesis H_{1c} is $\beta_1 = \beta_2 = \beta_3 = -\beta_4 = -\beta_5$, where β_4 and β_5 are larger than

zero. By substituting R^2, ρ, and $\beta_1 = \beta_2 = \beta_3 = -\beta_4 = -\beta_5$ into Equation 13, the

regression coefficients can be calculated.

In summary, we can conclude that the complement hypotheses of H_1 include

$\binom{k}{1} + \binom{k}{2} + ... + \binom{k}{k} = 2^k - 1$ cases. The hypothesis corresponding to the median number of

$\{1, 2, \cdots, 2^k - 1\}$ (i.e., the hypothesis $H_{2^{k-1}}$) is selected as representative of the

complement hypothesis of H_1. If the ratio $\beta_1 : \beta_2 : \cdots : \beta_K = 1 : 1 : \cdots : 1$, then

$\beta_1 = \beta_2 = \cdots \beta_{K-1} = \beta_K$. By substituting R^2, $\boldsymbol{\Sigma}$, $\beta_1 = \beta_2 = \cdots \beta_{K-1} = \beta_K$ into Equation

13, β_K can be derived. Subsequently, $\beta_1, \beta_2, \cdots, \beta_{K-1}$ can be obtained. The default signs

of the regression coefficients are positive for all the hypotheses at first. If the signs of the

regression coefficients are negative in the representative hypothesis, they will be finally

designated to be negative.

Another issue that needs to be addressed is how to calculate the regression coefficients for

the complement of H_2. Firstly, with two predictors H_2 is given as $\beta_1 > \beta_2$. There is only

one hypothesis for the complement of H_2, namely H_{c1}: $\beta_1 < \beta_2$. There is no doubt that H_{c1}

can be regarded as the representative hypothesis. Based on Table 1, if H_2 is true, the ratio

β_1: β_2=2:1 for H_2. The relation of the regression coefficients under the complement

hypothesis H_{2c} is β_1: β_2=1:2, where β_1 and β_2 are larger than zero. By substituting R^2, ρ,

and β_1: β_2=1:2 into Equation 13, the regression coefficients can be calculated.

In following examples, swapping the regression coefficients of adjacent positions once is

called a violation. The number of violations is determined by the number of the regression

coefficients of adjacent positions swapped. This process can be described as follows. For

the convenience of description, a specific example is used to illustrated. For example,

$H_2 : \beta_1 > \beta_2 > \beta_3$ is considered, β_1 and β_2 are at adjacent positions, and β_2 and β_3 are also

at adjacent positions. After swapping adjacent positions β_1 and β_2, a new order

$H_{c1} : \beta_2 > \beta_1 > \beta_3$ can be obtained. In the new order, β_1 and β_3 are at adjacent positions.

Swap them, and a new order can be obtained again, which is $H_{c2} : \beta_2 > \beta_3 > \beta_1$. By

swapping adjacent positions β_2 and β_3, the order $H_{c3} : \beta_3 > \beta_2 > \beta_1$ can be obtained.

During this process, the adjacent positions are swapped for three times. Therefore, there

are three violations from $H_2 : \beta_1 > \beta_2 > \beta_3$ to H_{c3}. The complement hypothesis of H_2 can

be divided into $\binom{K}{2}$ categories based on the number of violations. For three or more

predictors, there is more than one hypothesis for the complement of H_2. Therefore, all the

possible hypotheses should be considered and a representative one should be selected. For

three predictors, there are three categories for the complement hypothesis of H_2, namely

one violation from H_2, two violations from H_2, and three violations from H_2. There are

two hypotheses with one violation:

$H_{c1} : \beta_2 > \beta_1 > \beta_3$.

$H_{c2} : \beta_1 > \beta_3 > \beta_2$.

There are two hypotheses containing two violations:

$H_{c3} : \beta_2 > \beta_3 > \beta_1$.

$H_{c4} : \beta_3 > \beta_1 > \beta_2$.

There is only one hypothesis with three violations:

$H_{c5} : \beta_3 > \beta_2 > \beta_1$.

As the Bayes factor $\mathrm{BF}_{2c,2}$ for H_{2c} versus H_2 becomes larger with an increasing number of

violations, the hypothesis corresponding the median of the number of hypotheses ordered

using the number of violations ($H_{c3} : \beta_2 > \beta_3 > \beta_1$) is selected as the hypothesis

representing the complement of H_2. Based on Table 1, if H_2 is true, the ratio β_1: β_2:

β_3=3:2:1 for H_2. The relation of the regression coefficients under the complement

hypothesis H_c is β_1: β_2: β_3=1:3:2 (reordered using H_{c3}: $\beta_2 > \beta_3 > \beta_1$), where β_1, β_2, and

β_3 are larger than zero. By substituting R^2, ρ, and β_1: β_2: β_3=1:3:2 into Equation 13, the

1072 regression coefficients can be calculated.

1073 In order to better summarize the rule of the selection of the representative hypothesis, the

1074 situation with four predictors $H_2 : \beta_1 > \beta_2 > \beta_3 > \beta_4$ is further discussed. The hypotheses

1075 with one, two, three and four violations are:

1076 One violation:

1077 $H_{c1} : \beta_2 > \beta_1 > \beta_3 > \beta_4$.

1078 $H_{c2} : \beta_1 > \beta_3 > \beta_2 > \beta_4$.

1079 $H_{c3} : \beta_1 > \beta_2 > \beta_4 > \beta_3$.

1080 two violations:

1081 $H_{c4} : \beta_2 > \beta_3 > \beta_1 > \beta_4$.

1082 $H_{c5} : \beta_2 > \beta_1 > \beta_4 > \beta_3$.

1083 $H_{c6} : \beta_3 > \beta_1 > \beta_2 > \beta_4$.

1084 $H_{c7} : \beta_1 > \beta_3 > \beta_4 > \beta_2$.

1085 $H_{c8} : \beta_1 > \beta_4 > \beta_2 > \beta_3$.

1086 three violations:

1087 $H_{c9} : \beta_3 > \beta_2 > \beta_1 > \beta_4$.

1088 $H_{c10} : \beta_2 > \beta_3 > \beta_4 > \beta_1$.

1089 $H_{c11} : \beta_3 > \beta_1 > \beta_4 > \beta_2$.

1090 $H_{c12} : \beta_1 > \beta_4 > \beta_3 > \beta_2$.

1091 $H_{c13} : \beta_4 > \beta_1 > \beta_2 > \beta_3$.

1092 four violations:

1093 $H_{c14} : \beta_3 > \beta_2 > \beta_4 > \beta_1$.

1094 $H_{c15} : \beta_2 > \beta_4 > \beta_3 > \beta_1$.

1095 $H_{c16} : \beta_3 > \beta_4 > \beta_1 > \beta_2$.

1096 $H_{c17} : \beta_4 > \beta_1 > \beta_3 > \beta_2$.

1097 $H_{c18} : \beta_4 > \beta_2 > \beta_1 > \beta_3$.

1098 five violations:

1099 $H_{c19} : \beta_3 > \beta_4 > \beta_1 > \beta_2.$

1100 $H_{c20} : \beta_4 > \beta_2 > \beta_3 > \beta_1.$

1101 $H_{c21} : \beta_2 > \beta_4 > \beta_1 > \beta_3.$

1102 $H_{c22} : \beta_4 > \beta_3 > \beta_1 > \beta_2.$

1103   six violations:

1104 $H_{c23} : \beta_4 > \beta_3 > \beta_2 > \beta_1.$

1105 Overall, there are 23 hypotheses for the complement of H_2. Similarly, the hypothesis

1106 corresponding to the median of the number of hypotheses ordered using the number of

1107 violations ($H_{c12} : \beta_1 > \beta_4 > \beta_3 > \beta_2$) is recommended as the representative hypothesis of

1108 the complement hypothesis of H_2. Based on Table 1, if H_2 is true, the ratio β_1: β_2: β_3:

1109 β_4=4:3:2:1 for H_2. The relation of the regression coefficients under the complement

1110 hypothesis H_{2c} is β_1: β_2: β_3: β_4=4:1:2:3 (reordered using H_{c12}: $\beta_1 > \beta_4 > \beta_3 > \beta_2$), where

1111 β_1, β_2, β_3 and β_4 are larger than zero. By substituting R^2, ρ, and β_1: β_2: β_3: β_4=4:1:2:3

1112 into Equation 13, the regression coefficients can be calculated.

1113 By summarizing the current examples, the total number of hypotheses in the complement

1114 of H_2 is $K! - 1$. The hypothesis corresponding to the median of the number of hypotheses

1115 ordered using the number of violations (i.e., the hypothesis $H_{K!/2}$) can be selected as the

1116 representative hypothesis. If the ratio $\beta_1 : \beta_2 : \cdots : \beta_K = r_1 : r_2 : \cdots : r_K$ for H_2, the ratio

1117 of the complement hypothesis of H_2 would be obtained based on the order of hypothesis

1118 $H_{K!/2}$. By substituting R^2, $\boldsymbol{\Sigma}$, and the ratio of the complement hypothesis into Equation

1119 13, β_K can be derived. Subsequently, β_1, $\beta_2,\cdots$, β_{K-1} can be obtained.

1120 Some researchers may recommend placing the regression coefficients under the complement

1121 of H_1/H_2 on the boundary of H_1/H_2 (set all regression coefficients equal to 0). Although

1122 the boundary value does not belong to H_1 or H_2, it is the closest value to H_1 or H_2.

1123 However, the value of the Bayes factors $BF_{1c,1}$ for H_{1c} versus H_1 or $BF_{2c,2}$ for H_{2c} versus

1124 H_2 would be always around 1 no matter how large the sample size is. That is, neither

hypothesis is preferred over the other. Therefore, SSD cannot be performed if the regression coefficients for the complement are based on the boundary of the parameter space of H_1 or H_2.

Table 1

Chosen population values for the regression coefficients in the multivariate linear model with two predictors.

	$\rho = 0$		$\rho = 0.2$		$\rho = 0.5$	
	β_1	β_2	β_1	β_2	β_1	β_2
H_0: $\beta_1 = \beta_2 = 0$	0	0	0	0	0	0
H_0: $\beta_1 = \beta_2$	0.255	0.255	0.233	0.233	0.208	0.208
H_a	0.255	0.255	0.233	0.233	0.208	0.208
H_1: $\beta_1 > 0$ & $\beta_2 > 0$	0.255	0.255	0.233	0.233	0.208	0.208
H_2: $\beta_1 > \beta_2$	0.322	0.161	0.299	0.150	0.272	0.136
H_{1c}	-0.255	0.255	-0.233	0.233	-0.208	0.208
H_{2c}	0.161	0.322	0.150	0.299	0.136	0.272

Note: For H_1 and H_a, the ratio of the regression $\beta_1 : \beta_2 = 1 : 1$ is used. For hypothesis H_2, the ratio of the regression $\beta_1 : \beta_2 = 2 : 1$ is used.

Table 2
Chosen population values for the regression coefficients in the multivariate linear model with three predictors.

	$\rho = 0$			$\rho = 0.2$			$\rho = 0.5$		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
H_0: $\beta_1 = \beta_2 = \beta_3 = 0$	0	0	0	0	0	0	0	0	0
H_0: $\beta_1 = \beta_2 = \beta_3$	0.208	0.208	0.208	0.176	0.176	0.176	0.147	0.147	0.147
H_a	0.208	0.208	0.208	0.176	0.176	0.176	0.147	0.147	0.147
H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$	0.208	0.208	0.208	0.176	0.176	0.176	0.147	0.147	0.147
H_2: $\beta_1 > \beta_2 > \beta_3$	0.289	0.193	0.096	0.252	0.168	0.084	0.216	0.144	0.072
H_{1c}	-0.208	-0.208	0.208	-0.176	-0.176	0.176	-0.147	-0.147	0.147
H_{2c}	0.096	0.289	0.193	0.084	0.252	0.168	0.072	0.216	0.144

For H_1 and H_a, the ratio of the regression $\beta_1 : \beta_2 : \beta_3 = 1 : 1 : 1$ is used. For hypothesis H_2, the ratio of the regression $\beta_1 : \beta_2 : \beta_3 = 3 : 2 : 1$ is used.

Table 3
Chosen population values for the regression coefficients in the multivariate linear model with four predictors.

	$\rho = 0$				$\rho = 0.2$				$\rho = 0.5$			
	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_4
H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$	0	0	0	0	0	0	0	0	0	0	0	0
H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4$	0.180	0.180	0.180	0.180	0.142	0.142	0.142	0.142	0.114	0.114	0.114	0.114
H_a	0.180	0.180	0.180	0.180	0.142	0.142	0.142	0.142	0.114	0.114	0.114	0.114
H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$ & $\beta_4 > 0$	0.180	0.180	0.180	0.180	0.142	0.142	0.142	0.142	0.114	0.114	0.114	0.114
H_2: $\beta_1 > \beta_2 > \beta_3 > \beta_4$	0.263	0.197	0.132	0.066	0.217	0.163	0.109	0.054	0.179	0.134	0.089	0.045
H_{1c}	-0.180	-0.180	0.180	0.180	-0.142	-0.142	0.142	0.142	-0.114	-0.114	0.114	0.114
H_{2c}	0.132	0.263	0.066	0.197	0.109	0.217	0.054	0.163	0.089	0.179	0.045	0.134

For H_1 and H_a, the ratio of the regression $\beta_1 : \beta_2 : \beta_3 : \beta_4 = 1 : 1 : 1 : 1$ is used. For hypothesis H_2, the ratio of the regression $\beta_1 : \beta_2 : \beta_3 : \beta_4 = 4 : 3 : 2 : 1$ is used.

Table 4

The required sample size and the corresponding probability that the Bayes factor is larger than 3 when the $\mathrm{BF}_{thresh} = 3$, $\eta = 0.8$, $R^2 = 0.13$, ratio $\beta_1 : \beta_2 = 1 : 1$ for H_a and H_1, ratio $\beta_1 : \beta_2 = 2 : 1$ for H_2 and the number of predictors is 2.

		$\rho = 0$			$\rho = 0.2$			$\rho = 0.5$		
		$b = J/N$	$b = 2J/N$	$b = 3J/N$	$b = J/N$	$b = 2J/N$	$b = 3J/N$	$b = J/N$	$b = 2J/N$	$b = 3J/N$
H_0: $\beta_1 = \beta_2 = 0$	p_0	0.948	0.880	0.801	0.948	0.881	0.801	0.948	0.881	0.801
	N	121 (122)	104 (105)	95 (96)	121 (122)	105 (104)	93 (93)	121 (122)	105 (106)	93 (93)
H_a	p_a	0.804	0.800	0.804	0.804	0.805	0.802	0.804	0.805	0.802
H_0: $\beta_1 = \beta_2$	p_0	0.962	0.944	0.931	0.973	0.955	0.944	0.980	0.971	0.961
	N	887 (889)	816 (815)	775 (773)	1326 (1328)	1221 (1220)	1176 (1176)	2721 (2725)	2527 (2525)	2426 (2426)
H_2: $\beta_1 > \beta_2$	p_2	0.802	0.804	0.800	0.806	0.805	0.801	0.804	0.805	0.801
H_0: $\beta_1 = \beta_2 = 0$	p_0	0.939	0.862	0.805	0.940	0.862	0.804	0.938	0.860	0.803
	N	90 (92)	74 (72)	71 (70)	88 (88)	71 (73)	71 (70)	86 (88)	70 (71)	69 (69)
H_1: $\beta_1 > 0$ & $\beta_2 > 0$	p_1	0.802	0.811	0.837	0.808	0.805	0.843	0.804	0.805	0.842
H_1: $\beta_1 > 0$ & $\beta_2 > 0$	p_1		0.950			0.968			0.987	
	N		60 (60)			79 (77)			134 (133)	
H_{1c}: not H_1	p_{1c}		0.802			0.800			0.804	
H_2: $\beta_1 > \beta_2$	p_2		0.804			0.812			0.801	
	N		163 (162)			235 (233)			438 (435)	
H_{2c}: not H_2	p_{2c}		0.808			0.800			0.800	

The correlation between the predictors is denoted as ρ, where ρ is chosen as 0, 0.2 or 0.5, respectively in this table. The prior distributions with different b are used to perform a sensitivity analysis. It should be noted that b is not relevant for the evaluation of H_1 and H_2 versus their respective complements. The symbol p_i denotes the probability that the Bayes factor supports H_i when H_i is true, where $i = 0, 1, 2, a, 1c, 2c$. The sample size values in the parentheses is calculated with the set.seed 1234. Comparing the values without and within parenthesis shows that using T=10,000 renders stable sample size estimates.

Table 5

The required sample size and the corresponding probability that the Bayes factor is larger than 3 when the $\mathrm{BF}_{thresh} = 3$, $\eta = 0.8$, $R^2 = 0.13$, ratio $\beta_1 : \beta_2 : \beta_3 = 1 : 1 : 1$ for H_a and H_1, ratio $\beta_1 : \beta_2 : \beta_3 = 3 : 2 : 1$ for H_2 and the number of the predictors is 3.

		$\rho = 0$			$\rho = 0.2$			$\rho = 0.5$		
		$b = J/N$	$b = 2J/N$	$b = 3J/N$	$b = J/N$	$b = 2J/N$	$b = 3J/N$	$b = J/N$	$b = 2J/N$	$b = 3J/N$
H_0: $\beta_1 = \beta_2 = \beta_3 = 0$	p_0	0.974	0.909	0.831	0.973	0.918	0.840	0.973	0.918	0.840
	N	148	119	104	146	120	105	146	120	105
H_a	p_a	0.809	0.806	0.805	0.803	0.804	0.806	0.803	0.804	0.806
H_0: $\beta_1 = \beta_2 = \beta_3$	p_0	0.993	0.983	0.976	0.995	0.992	0.986	0.998	0.996	0.993
	N	776	675	624	1350	1210	1124	3301	2981	2794
H_2: $\beta_1 > \beta_2 > \beta_3$	p_2	0.803	0.806	0.803	0.802	0.804	0.801	0.802	0.804	0.800
H_0: $\beta_1 = \beta_2 = \beta_3 = 0$	p_0	0.964	0.879	0.811	0.964	0.887	0.809	0.968	0.899	0.806
	N	100	71	66	98	71	62	94	69	59
H_1: $\beta_1 > 0$ & $\beta_2 > 0$&$\beta_3 > 0$	p_1	0.802	0.802	0.833	0.807	0.802	0.833	0.802	0.801	0.825
H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$	p_1		0.801			0.809			0.809	
	N		35			40			47	
H_{1c}: not H_1	p_{1c}		0.833			0.827			0.814	
H_2: $\beta_1 > \beta_2 > \beta_3$	p_2		0.801			0.801			0.808	
	N		254			410			882	
H_{2c}: not H_2	p_{2c}		0.909			0.909			0.901	

The correlation between the predictors is denoted as ρ, where ρ is chosen as 0, 0.2 or 0.5, respectively in this table. The prior distributions with different b are used to perform a sensitivity analysis. It should be noted that b is not relevant for the evaluation of H_1 and H_2 versus their respective complements. The symbol p_i denotes the probability that the Bayes factor supports H_i when H_i is true, where $i = 0, 1, 2, a, 1c, 2c$.

Table 6

The required sample size and the corresponding probability that the Bayes factor is larger than 3 when the $\mathrm{BF}_{thresh} = 3$, $\eta = 0.8$, $R^2 = 0.13$, ratio $\beta_1 : \beta_2 : \beta_3 : \beta_4 = 1 : 1 : 1 : 1$ for H_a and H_1, ratio $\beta_1 : \beta_2 : \beta_3 : \beta_4 = 4 : 3 : 2 : 1$ for H_2 and the number of the predictors is 4.

		$\rho = 0$			$\rho = 0.2$			$\rho = 0.5$		
		$b = J/N$	$b = 2J/N$	$b = 3J/N$	$b = J/N$	$b = 2J/N$	$b = 3J/N$	$b = J/N$	$b = 2J/N$	$b = 3J/N$
H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$	p_0	0.986	0.934	0.840	0.984	0.935	0.840	0.984	0.935	0.840
	N	173	134	109	172	135	109	172	135	109
H_a	p_a	0.813	0.803	0.800	0.808	0.805	0.800	0.808	0.805	0.800
H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4$	p_0	0.998	0.993	0.989	0.999	0.998	0.996	0.999	0.999	0.998
	N	762	652	583	1580	1373	1257	4298	3901	3580
H_2: $\beta_1 > \beta_2 > \beta_3 > \beta_4$	p_2	0.800	0.803	0.805	0.802	0.800	0.800	0.801	0.807	0.810
H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$	p_0	0.982	0.896	0.800	0.980	0.905	0.805	0.981	0.906	0. 809
	N	109	72	65	103	70	61	102	67	56
H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$ & $\beta_4 > 0$	p_1	0.800	0.804	0.859	0.807	0.806	0.855	0.806	0.803	0.844
H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$ & $\beta_4 > 0$	p_1		0.892			0.948			0.986	
	N		56			83			167	
H_{1c}: not H_1	p_{1c}		0.805			0.802			0.803	
H_2: $\beta_1 > \beta_2 > \beta_3 > \beta_4$	p_2		0.803			0.802			0.802	
	N		272			488			1148	
H_{2c}: not H_2	p_{2c}		0.860			0.854			0.857	

The correlation between the predictors is denoted as ρ, where ρ is chosen as 0, 0.2 or 0.5, respectively in this table. The prior distributions with different b are used to perform a sensitivity analysis. It should be noted that b is not relevant for the evaluation of H_1 and H_2 versus their respective complements. The symbol p_i denotes the probability that the Bayes factor supports H_i when H_i is true, where $i = 0, 1, 2, a, 1c, 2c$.

Table 7

Comparison between classical sample size determined using power=0.8, $\alpha = 0.05$, $f^2 = 0.15$ and the Bayesian sample size determined using $\eta = 0.8$, $BF_{thresh} = 3$, $R^2 = 0.13$, the ratio between each pair of coefficients is 1:1.

$K = 2$	H_0: $\beta_1 = \beta_2 = 0$ vs H_a	classical	68		
		Bayesian	$b = J/N$	$b = 2J/N$	$b = 3J/N$
			121	104	95
	H_0: $\beta_1 = \beta_2 = 0$ vs H_1: $\beta_1 > 0$ & $\beta_2 > 0$	Bayesian	90	74	71
$K = 3$	H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ vs H_a	classical	77		
		Bayesian	$b = J/N$	$b = 2J/N$	$b = 3J/N$
			148	119	104
	H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ vs H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$	Bayesian	100	71	66
$K = 4$	H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs H_a	classical	85		
		Bayesian	$b = J/N$	$b = 2J/N$	$b = 3J/N$
			173	134	109
	H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs H_1: $\beta_1 > 0$ & $\beta_2 > 0$ & $\beta_3 > 0$ & $\beta_4 > 0$	Bayesian	109	72	65