



UvA-DARE (Digital Academic Repository)

A default Bayesian hypothesis test for correlations and partial correlations

Wetzels, R.; Wagenmakers, E.J.

Published in:
Psychonomic Bulletin & Review

DOI:
[10.3758/s13423-012-0295-x](https://doi.org/10.3758/s13423-012-0295-x)

[Link to publication](#)

Citation for published version (APA):

Wetzels, R., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 19(6), 1057-1064. <https://doi.org/10.3758/s13423-012-0295-x>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

A default Bayesian hypothesis test for correlations and partial correlations

Ruud Wetzels · Eric-Jan Wagenmakers

Published online: 14 July 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract We propose a default Bayesian hypothesis test for the presence of a correlation or a partial correlation. The test is a direct application of Bayesian techniques for variable selection in regression models. The test is easy to apply and yields practical advantages that the standard frequentist tests lack; in particular, the Bayesian test can quantify evidence in favor of the null hypothesis and allows researchers to monitor the test results as the data come in. We illustrate the use of the Bayesian correlation test with three examples from the psychological literature. Computer code and example data are provided in the journal archives.

Keywords Bayesian inference · Correlation · Statistical evidence

Introduction

A correlation coefficient indicates how strongly two variables are related. The concept is basic, and it comes as no surprise that the correlation coefficient ranks among the most popular statistical tools in any subfield of psychological science. The first correlation coefficient was developed by Francis Galton in 1888 (Stigler, 1989); further work by Francis Edgeworth and Karl Pearson resulted in the correlation measure that is used most frequently today, the *Pearson product-moment correlation coefficient*, or r (Pearson, 1920). The coefficient r is a measure of the linear relation between two variables, where $r = -1$ indicates a perfectly negative linear relation, $r = 1$ indicates a perfectly positive relation, and $r = 0$ indicates the absence of any linear relation.

In this article, we focus on the two-sided hypothesis test for the Pearson correlation coefficient. The standard (i.e., classical, orthodox, or frequentist) test produces a p value for drawing conclusions; the common rule is that when $p < .05$, one can reject the null hypothesis that no relation is present. Unfortunately, frequentist p value tests have a number of drawbacks (e.g., Edwards, Lindman, & Savage, 1963; Wagenmakers, 2007). For instance, p values do not allow researchers to quantify evidence in favor of the null hypothesis (Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wetzels et al., 2011). In addition, p values depend on the sampling plan, and hence, its users may not stop data collection when an interim result is compelling, nor may they continue data collection when the fixed sample size result is ambiguous (Edwards et al., 1963). These drawbacks are not merely theoretical but have real consequences for the way in which psychologists carry out their experiments and draw conclusions from their data.

An alternative to frequentist tests is provided by Bayesian inference and, in particular, the so-called Bayes factor (Jeffreys, 1961; Kass & Raftery, 1995). The Bayes factor computes the probability of the observed data under the null hypothesis vis-a-vis the alternative hypothesis. In contrast to the frequentist p value, the Bayes factor allows researchers to quantify evidence in favor of the null hypothesis. Moreover, with the Bayes factor, “it is entirely appropriate to collect data until a point has been proven or disproven, or until the data collector runs out of time, money, or patience” (Edwards et al., p. 193). Thus, the Bayes factor altogether eliminates the optional stopping phenomenon, where researchers can bias their results by collecting data until $p < .05$ (e.g., Simmons, Nelson, & Simonsohn, 2011). Researchers are allowed to monitor the Bayes factor as the data come in and stop whenever they feel that the evidence is compelling.

In the field of psychology, interest in hypothesis testing using the Bayes factor has greatly increased over the last years. For instance, a method for variable selection in regression models (Liang, Paulo, Molina, Clyde, & Berger, 2008) is

R. Wetzels (✉) · E.-J. Wagenmakers
Department of Psychology, University of Amsterdam,
Weesperplein 4,
1018 XA, Amsterdam, The Netherlands
e-mail: wetzels.ruud@gmail.com

used to develop a Bayesian ANOVA (Wetzels, Grasman & Wagenmakers, *in press*) and a Bayesian t test (Rouder et al., 2009; Wetzels, Raaijmakers, Jakab, & Wagenmakers, 2009); Masson has shown how statistical output from SPSS can be translated to Bayes factors using the BIC approximation (Masson, 2011); Hoijtink, Klugkist, and colleagues have promoted Bayes factors for order-restricted inference (e.g., Hoijtink, Klugkist, & Boelen, 2008).

Perhaps the greatest impediment to the large-scale adoption of the Bayes factor is the lack of easy-to-use tests for statistical models that psychologists use in practice. For example, the test for the presence of a correlation (and partial correlation) is one of the most popular workhorses in experimental psychology, yet many psychologists will struggle to find a Bayes factor equivalent. In this article, we remove this hurdle by providing an easy-to-use Bayes factor alternative to the Pearson correlation test.

In this article, we first discuss the standard, frequentist tests for the presence of correlation and partial correlation. Next, we explain Bayesian model selection in general and then focus on a Bayesian test for correlation and partial correlation that is considered default. By *default* (or *objective*, or *uninformative*), we mean that the test is suitable for situations in which the researcher is unable or unwilling to use substantive information about the problem at hand. Key concepts and computations are illustrated with three examples of recent psychological experiments.

Frequentist test for the presence of correlation

We discuss the frequentist correlation test in the context of a study where participants were involved in an intensive meditation training program (MacLean et al., 2010). The aim of this program was to investigate whether there is an effect of meditation on visual acuity. To assess visual acuity, participants were asked to judge repeatedly whether a vertical line was long or short. Perceptual threshold was defined as the difference in visual angle between the short and the long lines that allowed the participant to classify the lines correctly 75 % of the time. The main result of the experiment was that the intensive meditation program decreased participants' perceptual threshold.

In addition to this main result, MacLean et al. (2010) explored whether the improved visual acuity is retained 5 months after termination of the meditation program and, more specifically, whether at follow-up the participants who had meditated the most also had the lowest threshold. The follow-up involved 54 participants, whose data are replotted in Fig. 1. On the basis of these data, MacLean et al. concluded that “this result indicates a correlation between the long-term stability of training-induced discrimination improvement and the maintenance of regular, but less intensive, meditation practice.”

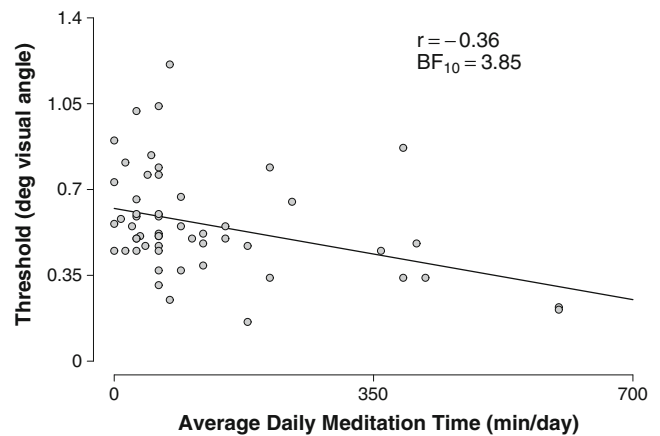


Fig. 1 Relationship between average daily meditation time and discrimination threshold. A negative correlation suggests that time spent in meditation improves visual perception (i.e., lowers the threshold). Data are replotted from MacLean et al. (2010)

To calculate the correlation between threshold and meditation time, we first define the following variables. For person i , mean daily meditation time is denoted x_i , and threshold is denoted y_i . For meditation time and threshold, the sample variances are $s_X^2 = 20,916.68$ and $s_Y^2 = 0.05$, and the sample means are $\bar{x} = 121$ and $\bar{y} = 0.56$, respectively. Then, the sample correlation coefficient of X and Y is calculated as follows:

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_X s_Y} = \frac{-589}{1629} = -.36, \quad (1)$$

where n is the number of participants ($n = 54$).

In order to test whether we can reject the null hypothesis that the correlation coefficient is zero, $\rho_{XY} = 0$, we calculate the t statistic (using $r_{XY} = -.36$ and $n = 54$):

$$t = r_{XY} \sqrt{\frac{(n-2)}{(1-r_{XY}^2)}} = -2.80, \quad (2)$$

which follows the Student t distribution with $n - 2$ degrees of freedom. This t statistic corresponds to a p value of 0.01. Therefore, with a significance level of $\alpha = 0.05$, researchers may feel that they can confidently reject the null hypothesis of no correlation.

Frequentist test for the presence of partial correlation

Partial correlation is the correlation between two variables, say X and Y , after the confounding effect of a third variable Z has been removed. Variable Z is known as the control variable. In psychological research, there are many situations in which one might want to partial out the effects of a control variable.

Consider a recent experiment on the role of implicit prediction in visual search by Lleras, Porporino, Burack, and Enns (2011). Implicit prediction was studied using an

interrupted search task featuring three groups of children and one group of adults (i.e., mean ages of 7, 9, 11, and 19 years). In the search task, participants had to identify a target among a set of distractors (i.e., one “T” among 15 “L” shapes). Crucially, brief looks at the search display (100–500 ms) were interrupted by longer “waits” in which the participant was shown a blank screen (1,000–3,500 ms). The focus of this study was on *rapid resumption*, the phenomenon that, in contrast to the first look at the stimulus (where only 2 % of the correct responses are faster than 500 ms), subsequent looks often show 30 % – 50 % correct responses faster than 500 ms.

On the basis of $n = 40$ observations, Lleras et al., (2011) calculated the correlation between mean successful search time (X) and the proportion of rapid resumption responses (Y): $r_{XY} = .51$, a highly significant correlation ($p < .01$). However, Lleras et al. also observed that this correlation does not take the participants’ age into account. The correlation between search time (X) and age (Z) is relatively high (i.e., $r_{XZ} = -.78$), and so is the correlation between rapid resumption (Y) and age (i.e., $r_{YZ} = -.66$). Hence, the authors computed a partial correlation to exclude the possibility that age Z caused the correlation between search time X and rapid resumption Y . This is accomplished by the following formula:

$$r_{XY|Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{[(1 - r_{XZ}^2)(1 - r_{YZ}^2)]^{1/2}}$$

$$= \frac{.51 - (-.78)(-.66)}{[(1 - (-.78)^2)(1 - (-.66)^2)]^{1/2}} = -.01. \quad (3)$$

This result shows that by controlling for the variable age, the correlation between search time and rapid resumption is virtually eliminated. The correlation, r_{xy} , is .51, but the *partial* correlation, $r_{XY|Z}$, is -.01. The p value for the partial correlation can be calculated by computing the t statistic (using $r_{XY|Z} = -.01$ and $n=40$):

$$t = r_{XY|Z} \sqrt{\frac{(n - 3)}{(1 - r_{XY|Z}^2)}} = -0.06, \quad (4)$$

which follows the Student t distribution with $n - 3$ degrees of freedom. This t statistic corresponds to a p value of .95. Hence, Lleras et al. did not reject the null hypothesis of no correlation between search time and rapid resumption.

Note that this nonsignificant result leaves the null hypothesis in a state of suspended disbelief. It is not statistically correct to conclude from a nonsignificant result that the data support the null hypothesis; after all, the same nonsignificant result could have been due to the fact that the data were relatively noisy. This is one of the prominent p value problems that does not occur in the alternative framework of Bayesian inference, which enables researchers to directly gather evidence in favor of the null.

Bayesian hypothesis testing

In Bayesian model selection or hypothesis testing, the competing statistical hypotheses are assigned prior probabilities. Suppose that we have two competing hypotheses: the null hypothesis, H_0 , and the alternative hypothesis, H_1 . These hypotheses are assigned prior probabilities of $p(H_0)$ and $p(H_1)$. Then, after observing the data Y , Bayes’ theorem is applied to obtain the posterior probability of both hypotheses. The posterior probability of the alternative hypothesis, $p(H_1|Y)$, is calculated as follows:

$$p(H_1|Y) = \frac{p(Y|H_1)p(H_1)}{p(Y|H_1)p(H_1) + p(Y|H_0)p(H_0)}, \quad (5)$$

where $p(Y|H_1)$ denotes the marginal likelihood of the data under the alternative hypothesis (and equivalently for the null hypothesis). The marginal likelihood of the alternative hypothesis is calculated by integrating the likelihood with respect to the prior:

$$p(Y|H_1) = \int_{\Theta} p(Y|\theta, H_1)p(\theta|H_1)d\theta. \quad (6)$$

Because the posterior model probabilities are sensitive to the prior probabilities of both hypotheses, $p(H_0)$ and $p(H_1)$, it is common practice to quantify the evidence by the ratio of the marginal likelihoods, also known as the *Bayes factor* (Jeffreys, 1961):

$$\frac{p(H_1|Y)}{p(H_0|Y)} = \frac{p(Y|H_1)}{p(Y|H_0)} \times \frac{p(H_1)}{p(H_0)} = BF_{10} \times \frac{p(H_1)}{p(H_0)}. \quad (7)$$

The Bayes factor, BF_{10} , is a weighted average likelihood ratio that indicates the relative plausibility of the data under the two competing hypotheses. Another way to conceptualize the Bayes factor is as the change from prior odds $p(H_1)/p(H_0)$ to posterior odds $p(H_1|Y)/p(H_0|Y)$ brought about by the data (cf. Eq. 7). This change is often interpreted as the *weight of evidence* (Good, 1983), and as such, it represents “the standard Bayesian solution to the hypothesis testing and model selection problems” (Lewis & Raftery, 1997, p. 648).

When the Bayes factor has a value greater than 1, this indicates that the data are more likely to have occurred under the alternative hypothesis H_1 than under the null hypothesis H_0 , and vice versa when the Bayes factor is below 1. For example, when $BF_{10} = 4$, this indicates that the data are four times as likely to have occurred under the alternative hypothesis H_1 than under the null hypothesis H_0 .

Jeffreys (1961) proposed a set of verbal labels to categorize different Bayes factors according to their evidential impact. This set of labels, presented in Table 1, facilitates scientific communication but should be considered only an approximate descriptive articulation of different standards of evidence (Kass & Raftery, 1995).

Default prior distributions for the linear model

In order to calculate the Bayes factor, one needs to specify prior distributions for the parameters in H_0 and H_1 (cf. Eq. 6). A long line of research in Bayesian statistics has focused on finding appropriate default prior distributions—that is, prior distributions that reflect little information and have desirable characteristics. Much of this statistical development has taken place in the framework of linear regression. In order to capitalize on this work, we later restate the correlation test and the partial correlation test as linear regression:

$$Y = \alpha + \beta X + \varepsilon, \tag{8}$$

where X is the vector of predictor variables, which are assumed to be measured as deviations from their corresponding sample means.

For linear regression, one of the most popular priors is known as Zellner’s g -prior (Zellner, 1986). This prior corresponds to a normal distribution on the regression coefficients β , Jeffreys’s prior on the error precision ϕ (Jeffreys, 1961), and a uniform prior on the intercept α :

$$p(\beta|\phi, g, X) = N\left(0, \frac{g}{\phi} (X^T X)^{-1}\right), p(\phi, \alpha) \propto \frac{1}{\phi}. \tag{9}$$

Note that the information in the data about β can be conceptualized as $\phi^{-1}(X^T X)^{-1}$ (Kass & Wasserman, 1995). Hence, g is a scaling factor controlling the information that we give the prior on β , relative to the information in the sample. For example, when $g = 1$, the prior carries the same weight as the observed data; when $g = 10$, the prior carries one tenth as much weight as the observed data.

Obviously, the choice of g is crucial to the analysis, and much research has gone into choosing an appropriate g . This

Table 1 Evidence categories for the Bayes factor BF_{10} (Jeffreys, 1961). We replaced the label “not worth more than a bare mention” with “anecdotal”

Bayes factor BF_{10}		Interpretation
>	100	Decisive evidence for H_1
30	–	Very Strong evidence for H_1
10	–	Strong evidence for H_1
3	–	Substantial evidence for H_1
1	–	Anecdotal evidence for H_1
	1	No evidence
1/3	–	Anecdotal evidence for H_0
1/10	–	Substantial evidence for H_0
1/30	–	Strong evidence for H_0
1/100	–	Very Strong evidence for H_0
<	1/100	Decisive evidence for H_0

is a difficult problem: A default prior should not be very informative, but a prior that is too vague can lead to unwanted behavior. Various choices of g have been proposed; a popular setting is $g = n$, the *unit information prior* (n equals the sample size; Kass & Wasserman, 1995), but others have argued for $g = k^2$ (k equals the number of parameters; Foster & George, 1994) or $g = \max\{n, k^2\}$ (Fernandez, Ley, & Steel, 2001). However, the choice for a single g remains difficult.

The impact of the choice of g can be clarified using an example taken from Kanai et al. (in press) that concerned the correlation between the number of Facebook friends and the normalized gray matter density at the peak coordinate of the right entorhinal cortex. Figure 2 shows the data; people with more Facebook friends have higher gray matter density, $r = .48, p < .002$. The effect that a specific choice of g has on the Bayes factor for this data set is shown in Fig. 3. This figure demonstrates that when g is increased, the support for the null hypothesis can be made arbitrarily large. This is due to the fact that if g increases, the vagueness of M_1 does too. This phenomenon is known as the Jeffreys–Lindley–Bartlett paradox (Bartlett, 1957; Jeffreys, 1961; Lindley, 1980; but see Vanpaemel, 2010). One of the primary desiderata for a default Bayesian hypothesis test is to avoid this paradox.

In a different but related approach, Zellner and Siow (1980) extended the work of Jeffreys (1961) and proposed assigning the regression coefficients a multivariate Cauchy prior, with a precision based on the concept of unit information (Liang et al., 2008). However, the marginal likelihood for this model specification is not analytically tractable, and therefore, this approach did not gain much popularity (but note that these priors are well-studied nonetheless; see Bayarri & Garcia-

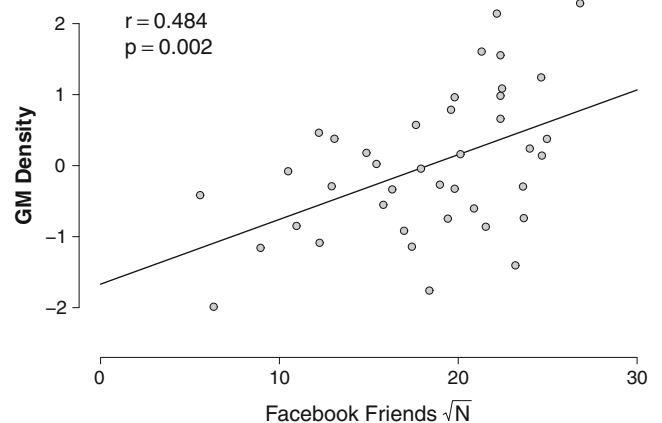


Fig. 2 Relation between the number of Facebook friends and the normalized gray matter (GM) density at the peak coordinate of the right entorhinal cortex. A positive correlation indicates that people with many Facebook friends have denser gray matter in the right entorhinal cortex. Data are replotted from Kanai, Bahrami, Roylance, and Rees (in press)

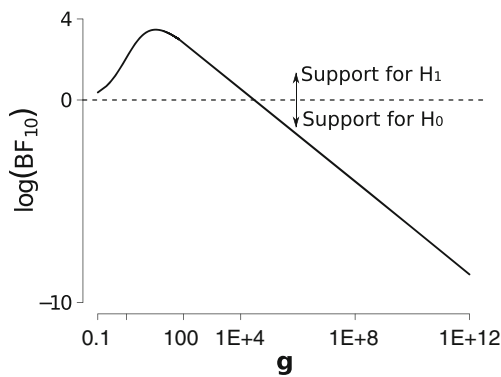


Fig. 3 Illustration of the Jeffreys–Lindley–Bartlett paradox when the Zellner g prior is applied to the data from Kanai, Bahrami, Roylance, and Rees (in press). By increasing g , the Bayes factor can be made arbitrarily close to 0, signifying close to infinite support for the null model

Donato, 2007; Berger, Ghosh, & Mukhopadhyay, 2003; Berger & Pericchi, 2001).

Recently, however, Liang et al. (2008) represented this Jeffreys–Zellner–Siow (JZS) prior as a mixture of g -priors—that is, an inverse-gamma $(1/2, n/2)$ prior on g and Jeffreys’s prior on the precision ϕ :

$$\begin{aligned}
 p(\beta|\phi, g, X) &= \int N\left(0, \frac{g}{\phi} (X^T X)^{-1}\right) p(g) dg \\
 p(\phi) &\propto \frac{1}{\phi} \\
 p(g) &= \frac{(n/2)^{1/2}}{\Gamma(1/2)} g^{-3/2} e^{-n/(2g)}.
 \end{aligned}
 \tag{10}$$

From this (mathematically equivalent) perspective, the problem of selecting a single g has been mitigated by assigning g a prior. The formulation above combines the computational advantages of the g -prior with the theoretical advantages of the Cauchy prior (see Liang et al., 2008, for details). Moreover, the mixture representation also facilitates the calculation of the Bayes factor, leaving only one integral that has to be estimated numerically. Note that the same prior setup underlies the JZS ANOVA (Wetzels et al., in press) and the JZS t test (Rouder et al., 2009; Wetzels et al., 2009). In the following, we will use this setup for our correlation and partial correlation test.

The JZS Bayes factor for correlation and partial correlation

In order to calculate the Bayes factor for the JZS (partial) correlation test, we conceptualize these Bayesian tests as a comparison between two regression models, such that the test becomes equivalent to a variable selection test for linear regression (i.e., a test of whether or not the regression coefficient β should be included in the model). This conceptualization allows us to exploit the JZS prior distribution. Computer code for calculating the JZS Bayes factors is presented in the Appendix.

The JZS Bayes factor for correlation

Suppose that we have observed data from two variables, X and Y , and are interested in their correlation. Consider the regression from Eq. 8, where α is the intercept, β is the regression coefficient, and ϵ is the error term, normally distributed with precision ϕ .

Next, we are interested in how well this regression equation fits the data. The standard method for assessing this fit is by calculating the coefficient of determination R^2 :

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}}, \tag{11}$$

where SS_{err} denotes the residual sum of squares and SS_{tot} denotes the total sum of squares. Note that R^2 is the proportion of variance that is accounted for by the regression model. Specifically, R^2 is an indication of how much better the fit of model M_1 is, when compared with model M_0 :

$$\mathcal{M}_0 : Y = \alpha + \epsilon \tag{12}$$

$$\mathcal{M}_1 : Y = \alpha + \beta X + \epsilon.$$

When R^2 is low (i.e., near zero), the addition of the regression coefficient β to M_0 has caused only a small increase in explained variance. As R^2 increases, so does the importance of β . Because R^2 is the square of the sample correlation r , a test for whether or not the correlation equals zero is equivalent to a test for whether or not β equals zero. Hence, the correlation test can be recast as a comparison between two linear regression models, M_0 and M_1 (e.g., Draper & Smith, 1998).

The Bayes factor BF_{10} using the JZS prior setup can then be calculated as follows (see Liang et al., 2008):

$$\begin{aligned}
 BF_{10} &= \frac{p(Y|M_1)}{p(Y|M_0)} \\
 &= \frac{(n/2)^{1/2}}{\Gamma(1/2)} \times \int_0^\infty (1+g)^{(n-2)/2} \times [1 + (1-r^2)g]^{-(n-1)/2} g^{(-3/2)} e^{-n/(2g)} dg.
 \end{aligned}
 \tag{13}$$

Note that the only input to Eq. 13 is the usual sample correlation r and the number of observations n . The resulting Bayes factor BF_{10} quantifies the evidence in favor of the alternative hypothesis. Therefore, Bayes factors greater than 1 indicate evidence for the presence of a correlation, and Bayes factors smaller than 1 indicate evidence for the absence of a correlation.

The JZS Bayes factor for partial correlation

Again, we formalize the test as a model selection problem between two regression models. Assume that we have three variables, Y , X_1 , and X_2 , and we want to test whether the partial correlation between Y and X_2 is zero or not. Analogously to the correlation example, one is

interested in whether adding the variable X_2 increases R^2 when the variable X_1 is already included in the regression model. Hence, we compare the two models (e.g., Draper & Smith, 1998):

$$\mathcal{M}_0 : Y = \alpha + \beta_1 X_1 + \epsilon \tag{14}$$

$$\mathcal{M}_1 : Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

The Bayes factor BF_{10} using the JZS prior setup can then be calculated as follows:

$$BF_{10} = \frac{p(Y|M_1)}{p(Y|M_0)} = \frac{\int_0^\infty (1+g)^{(n-1-p_1)/2} \times [1 + (1-R_1^2)g]^{-(n-1)/2} g^{(-3/2)} e^{-n/(2g)} dg}{\int_0^\infty (1+g)^{(n-1-p_0)/2} \times [1 + (1-R_0^2)g]^{-(n-1)/2} g^{(-3/2)} e^{-n/(2g)} dg} \tag{15}$$

Input to Eq. 15 is the coefficient of determination for H_0 and for H_1 (i.e., R_0^2 and R_1^2), the number of regression coefficients H_0 and H_1 (i.e., p_0 and p_1), and the number of observations n . Note that the coefficient of determination for M_0 is found by squaring the sample correlation between the variable of interest and the controlling variable: $R_0^2 = r_{YX_1}^2$; the coefficient of determination for M_1 can be written as $R_1^2 = r_{YX_2|X_1}^2 (1 - r_{YX_1}^2) + r_{YX_1}^2$. As before, the resulting Bayes factor BF_{10} quantifies the evidence in favor of the alternative hypothesis.

Correlation example: The meditation data

In the meditation study, MacLean et al. (2010) tested the hypothesis of a relation between meditation time and visual acuity (see Fig. 1). The sample correlation between these two variables was found to be $r_{XY} = -.36$; the associated p value is .01, significant at the $\alpha = .05$ level.

We can now apply Eq. 13 to calculate the Bayes factor. Entering $r_{XY} = -.36$ and $n = 54$ in Eq. 13 yields a Bayes factor $BF_{10} = 3.86$, indicating that the data are 3.86 times more likely to have occurred under H_1 than under H_0 , a “substantial” Bayes factor according to the coarse category scheme proposed by Jeffreys (1961). However, note that the factor 3.86 inspires less confidence than does the p value; this illustrates the well-known point that p values overestimate the evidence against the null, at least when the p value is misinterpreted as the posterior probability of the null hypothesis being true (e.g., Edwards et al., 1963; Wetzels et al., 2011).

Correlation example: The Facebook data

In the Facebook study, Kanai et al. (in press) investigated the relation between the number of Facebook friends and the normalized gray matter density at the peak coordinate of the right entorhinal cortex (see Fig. 2).

Entering $r_{XY} = .48$ and $n = 40$ in Eq. 13 yields a Bayes factor $BF_{10} = 17.87$, indicating that the data are 17.87 times more likely to have occurred under H_1 than under H_0 , a “strong” Bayes factor according to the coarse category scheme proposed by Jeffreys (1961).

Partial correlation example: The rapid resumption data

In the study on rapid resumption, Lleras et al. (2011) tested the partial correlation between search time (X) and rapid resumption (Y) while controlling for age (Z). The partial correlation was found to be $r_{XY|Z} = -.01$, with a p value of .95.

We can compute the Bayes factor using the coefficient of determination for both models. The null model M_0 regresses search time (X) on age (Z), containing only the regression coefficient for Y . Hence, in Eq. 15, $p_0 = 1$, and $R_0^2 = 0.6084$. The alternative model M_1 contains the regression coefficients for Y and Z . Hence, $p_1 = 2$, and $R_1^2 = 0.6084408$. The sample size n is 40.

The Bayes factor BF_{10} is 0.13, indicating substantial evidence in favor of the null hypothesis: The data are $1/0.13 \approx 7.70$ times as likely to have occurred under the null hypothesis than under the alternative hypothesis (see Table 1).

Concluding remarks

In this article, we outlined a default Bayesian test for correlation and partial correlation. Just like the default Bayesian ANOVA (Wetzels et al., in press) and t test (Rouder et al., 2009; Wetzels et al., 2009), the correlation test follows directly from the regression framework for variable selection proposed by Liang et al., (2008). We did not strive for new statistical development. Instead, our goal was to show experimental psychologists how they can obtain a default Bayesian hypothesis test for correlation and partial correlation. As we mentioned throughout this article, the Bayesian hypothesis test comes with important practical advantages, as compared with the standard frequentist test; for instance, the Bayesian hypothesis test can quantify evidence in favor of the null hypothesis and allows researchers to collect data until a point has been proven or disproven.

It should be noted that Jeffreys (1961) also proposed a Bayesian correlation test, one that differs slightly from the one outlined here. We prefer the JZS correlation test because it follows directly from the regression framework of Liang et al., (2008), incorporating modern Bayesian developments into a more general JZS testing framework. This JZS framework now encompasses linear regression, the t test, ANOVA, and (partial) correlation, and extensions to other popular statistical models are likely to follow.

By making default Bayes factors easily available to experimental psychologists, we hope and expect that the field will start to turn away from p values and move toward a Bayesian

assessment of evidence. This transition is bound to improve statistical inference and accelerate scientific progress.

Acknowledgements This research was supported by a Vidi grant from the Netherlands Organisation for Scientific Research (NWO). We thank **MacLean et al.** and **Kanai, Bahrami, Roylance, and Rees** for sharing their data. Correspondence concerning this article may be

addressed to Ruud Wetzels, University of Amsterdam, Department of Psychology, Weesperplein 4, 1018 XA Amsterdam, the Netherlands.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix Calculating the Bayes factor using *R*

```
## Run the script below to compute the Bayes factors for the examples.
## Matlab code and info on how to compute the Bayes factors
## on Rouder's website (http://pcl.missouri.edu/bf-reg) can be found at
## www.ruudwetzels.com

## Functions to compute the Bayes Factor for (partial) correlation

jzs_corbf=function(r,n){
  int=function(r,n,g){
    (1+g)^((n-2)/2)*(1+(1-r^2)*g)^(-(n-1)/2)*
    g^(-3/2)*exp(-n/(2*g))};
  bf10=sqrt((n/2))/gamma(1/2)*integrate(int,lower=0,upper=Inf,r=r,n=n)$value;
return(bf10)}

jzs_partcorbf=function(r0,r1,p0,p1,n){
  int=function(r,n,p,g){
    (1+g)^((n-1-p)/2)*(1+(1-r^2)*g)^(-(n-1)/2)*
    g^(-3/2)*exp(-n/(2*g))};
  bf10=integrate(int,lower=0,upper=Inf,r=r1,p=p1,n=n)$value/
  integrate(int,lower=0,upper=Inf,r=r0,p=p0,n=n)$value;
return(bf10)}

###MacLean et al. 2010 (correlation)

r=-0.3616346;
n=54;
jzs_corbf(r,n) #BF10=3.85

###Kanai et al. in press (correlation)

r=0.4844199;
n=40;
jzs_corbf(r,n) #BF10=17.87

### Lleras et al. in press (partial correlation)

n=40;
p0=1;
p1=2;
r0=sqrt(0.6084);
r1=sqrt(0.6084408);
jzs_partcorbf(r0,r1,p0,p1,n) #BF10=0.13
```


References

- Bartlett, M. S. (1957). A Comment on D. V. Lindley's Statistical Paradox. *Biometrika*, *44*, 533–534.
- Bayarri, M. J., & Garcia-Donato, G. (2007). Extending Conventional Priors for Testing General Hypotheses in Linear Models. *Biometrika*, *94*, 135–152.
- Berger, J. O., Ghosh, J. K., & Mukhopadhyay, N. (2003). Approximations and Consistency of Bayes Factors as Model Dimension Grows. *Journal of Statistical Planning and Inference*, *112*, 241–258.
- Berger, J. O., Pericchi, L. R. (2001). Objective Bayesian Methods for Model Selection: Introduction and Comparison (with Discussion). In Lahiri, P. (Ed), *Model Selection (135–207)*. Beachwood, OH: Institute of Mathematical Statistics Lecture Notes—Monograph Series, volume 38.
- Draper, N., & Smith, H. (1998). *Applied Regression Analysis*. New York: Wiley-Interscience.
- Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian Statistical Inference for Psychological Research. *Psychological Review*, *70*, 193–242.
- Fernandez, C., Ley, E., & Steel, M. (2001). Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics*, *100*, 381–427.
- Foster, D., & George, E. (1994). The Risk Inflation Criterion for Multiple Regression. *The Annals of Statistics*, *22*, 1947–1975.
- Good, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. Minneapolis: University of Minnesota Press.
- Hooijtink, H., Klugkist, I., & Boelen, P. (2008). *Bayesian evaluation of informative hypotheses*. New York: Springer.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford: UK Oxford University Press.
- Kanai, R., Bahrami, B., Roylance, R., Rees, G. (in press). Online Social Network Size is Reflected in Human Brain Structure. *Proceedings of the Royal Society B: Biological Sciences*.
- Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*, 377–395.
- Kass, R. E., & Wasserman, L. (1995). A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion. *Journal of the American Statistical Association*, *90*, 928–934.
- Lewis, S. M., & Raftery, A. E. (1997). Estimating Bayes Factors via Posterior Simulation With the Laplace–Metropolis Estimator. *Journal of the American Statistical Association*, *92*, 648–655.
- Liang, F., Paulo, R., Molina, G., Clyde, M., & Berger, J. (2008). Mixtures of g Priors for Bayesian Variable Selection. *Journal of the American Statistical Association*, *103*, 410.
- Lindley, D. (1980). L.J. Savage-His Work in Probability and Statistics. *The Annals of Statistics*, 81–24.
- Lleras, A., Porporino, M., Burack, J., & Enns, J. (2011). Rapid Resumption of Interrupted Search is Independent of Age-Related Improvements in Visual Search. *Journal of Experimental Child Psychology*, *109*, 58–72.
- MacLean, K., Ferrer, E., Aichele, S., Bridwell, D., Zanesco, A., & Jacobs, T. (2010). Intensive Meditation Training Improves Perceptual Discrimination and Sustained Attention. *Psychological Science*, *21*, 829–839.
- Masson, M. E. J. (2011). A Tutorial on a Practical Bayesian Alternative to Null-Hypothesis Significance Testing. *Behavior Research Methods*, *43*, 679–690.
- Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, *13*, 25–45.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t -Tests for Accepting and Rejecting the Null Hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237.
- 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*, 1359–1366.
- Stigler, S. M. (1989). Francis Galton's Account of the Invention of Correlation. *Statistical Science*, *4*, 73–86.
- Vanpaemel, W. (2010). Prior Sensitivity in Theory Testing: An Apologia for the Bayes Factor. *Journal of Mathematical Psychology*, *54*, 491–498.
- Wagenmakers, E. J. (2007). A Practical Solution to the Pervasive Problems of p Values. *Psychonomic Bulletin & Review*, *14*, 779–804.
- Wetzels, R., Grasman, R. P. P. P., Wagenmakers, E. J. (in press). A Default Bayesian Hypothesis Test for ANOVA Designs. *The American Statistician*.
- 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*, 291–298.
- Wetzels, R., Raaijmakers, J., Jakab, E., & Wagenmakers, E. J. (2009). How to Quantify Support For and Against the Null Hypothesis: A Flexible WinBUGS Implementation of a Default Bayesian t -test. *Psychonomic Bulletin & Review*, *16*, 752–760.
- Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g -prior distributions. *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, 233–243.
- Zellner, A., & Siow, A. (1980). Posterior Odds Ratios for Selected Regression Hypotheses. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian Statistics* (pp. 585–603). Valencia: University Press.