

## Statistically Speaking

# An Introduction to Bayesian Data Analysis for Correlations

Regina L. Nuzzo, PhD

## Introduction

Bayesian approaches to data analysis can be a good alternative or supplement to traditional hypothesis testing. Unlike *P* values, simple Bayesian analyses can provide a direct measure of the strength of evidence both for and against a study hypothesis, which can be helpful for researchers for interpreting and making decisions about their results. This article provides an introduction to Bayesian methods for exploring the relationship between 2 quantitative variables, illustrated with a dataset from the literature and analyzed with free, easy-to-use software.

## Background

In recent years, statistical experts have increasingly warned researchers in the medical sciences and other fields about problems stemming from the misuse and misinterpretation of null hypothesis significance testing and *P* values [1]. For example, many researchers mistakenly interpret the *P* value to be the probability that the study's results were due to random chance or believe the 95% confidence interval to be the range of values that has a 95% chance of containing the true population parameter. One oft-suggested alternative that can sidestep some of these problems is a Bayesian approach to data analysis, which provides researchers with results in a form that can be more intuitive and less prone to misinterpretation than *P* values [2].

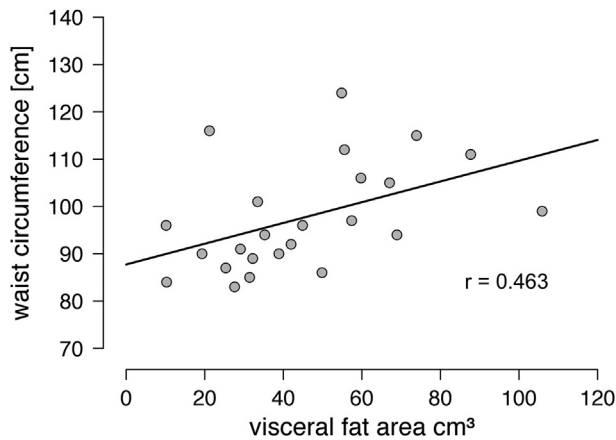
Although researchers are increasingly adopting Bayesian methodologies in many fields, lack of training in both the methods and its software packages has perhaps discouraged widespread use. To help overcome this problem, researchers at the University of Amsterdam have developed a free, open-source statistical software program called JASP that allows researchers to use a graphical interface to analyze data with both classical hypothesis testing and its Bayesian analogs [3]. Although still in preliminary release at the time of writing, the software provides enough power for many

types of analyses and is increasingly being used by educators as an alternative to other data analysis software, such as SPSS, R, and WinBUGS. This article will use JASP version 0.8.3.1 to illustrate the advantages of a Bayesian analysis of correlation and to compare the results to information provided from traditional methods.

The example data used here consist of 2 measurements taken from 24 overweight/obese adolescents: waist circumference (cm), and total area of visceral fat (cm<sup>2</sup>) as obtained from magnetic resonance images. The data were retrieved from the supporting information of an article by Eloi et al [4] on the investigation of magnetic resonance imaging as a tool for quantifying abdominal fat. A scatterplot of the data suggests a moderate, positive relationship between waist circumference and total visceral fat area (Figure 1).

For a traditional correlation analysis, we can use JASP or any standard statistical software. Table 1 shows the correlation table output from JASP through its "Regression → Correlation Matrix" menu selection. The observed correlation coefficient for Pearson *r* was 0.46, suggesting a moderate effect for the relationship between the 2 variables. In a classical 2-sided hypothesis test, this correlation coefficient is statistically significant at the .05 level (*P* = .023), with a 95% confidence interval of (0.074, 0.730).

Because this article will be comparing Bayesian results with traditional results, a quick refresher about the interpretation of these results might help cast the Bayesian results herein in sharper contrast. Recall that a *P* value tells us how likely it would be to observe results at least as extreme as what we saw in our study if the null hypothesis is true (ie, that the true population correlation coefficient is equal to zero;  $H_0: \rho = 0$ ). From our example *P* value of .023, we can say that if in fact there no correlation existed between visceral fat area and waist circumference, we would see a sample displaying such a strong linear pattern as ours (or even stronger) only about 2.3% of the time, which is a fairly rare occurrence. In



**Figure 1.** Scatterplot of waist circumference and visceral fat area. Data from Eloi et al [4].

other words, if we were to take random samples from a null situation with no correlation, we would still get sample correlations as at least as high as ours about one time out of 50. Note that this does not imply that the probability that these results were due to random chance is 2.3%, nor does it imply that there is a 97.7% probability that our hypothesis is true, nor even that the evidence favors the alternative hypothesis  $H_1$  over  $H_0$ .

Recall also that a 95% confidence interval is interpreted the range of values produced by a method which, if the study were to be exactly repeated many times, would contain the true population parameter 95% of the time. Note that this does not mean that there is a 95% chance that the true population correlation lies within these range of values.

### Bayesian Analysis

We can use JASP to do a Bayesian analysis of the same data using the “Regression → Bayesian Correlation Pairs” menu selection with the default options. Table 2 and Figure 2 show some results from this analysis.

The Bayesian Pearson correlation coefficient was calculated to be 0.46, the same as the classical value found previously. The other results may look strange to researchers accustomed to traditional hypothesis testing, but they succinctly summarize a great deal of information, as described to follow [5].

**Table 1**  
Traditional Pearson correlation

	$r$	$P$ Value	95% Confidence Interval	
			Lower	Upper
Visceral fat area, $\text{cm}^3$ – waist circumference, cm	0.463	.023	0.074	0.730

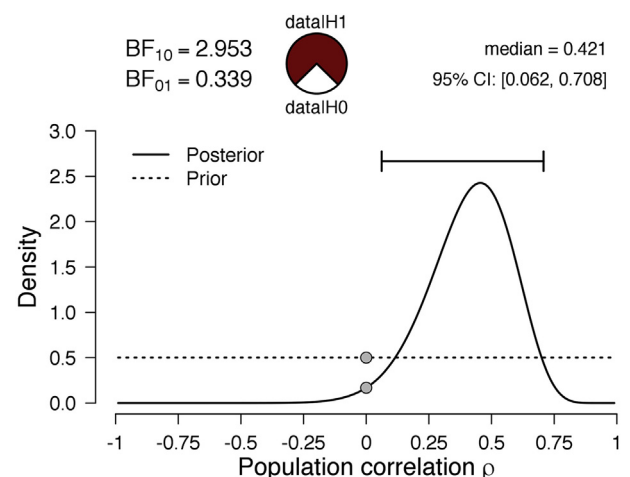
**Table 2**  
Bayesian Pearson correlation

	$r$	$\text{BF}_{10}$	95% Credible Interval	
			Lower	Upper
Visceral fat area, $\text{cm}^3$ – waist circumference, cm	0.463	2.953	0.062	0.708

### $\text{BF}_{10}$ and $\text{BF}_{01}$ : Strength of Evidence for the Hypothesis

$\text{BF}_{10}$  and  $\text{BF}_{01}$  in Table 2 and Figure 2 are Bayesian results not found in traditional hypothesis testing. They are both forms of what is known as the Bayes factor (BF), a measure that compares 2 things: the likelihood of the data under the alternative hypothesis and the likelihood of the data under the null hypothesis. It essentially answers the question, “How much more consistent are these data with the hypothesis that the effect in the population is *something* rather than *nothing*?”

$\text{BF}_{10}$  is the form of the BF that gives likelihood of the data under the null hypothesis divided by the likelihood of the data under the alternative, so that  $\text{BF}_{10}$  values greater than 1 signal more evidence in favor of the alternative and values less than 1 signal more evidence in favor of the null. In our example, the  $\text{BF}_{10}$  is 2.95, which means that our analysis found the data to be about 3 times more likely under the alternative hypothesis than under the null. This can be considered “anecdotal/borderline-moderate” evidence for the study hypothesis that the 2 variables are correlated. (See Table 3 for a list of common labels for the strength of evidence in BFs [6].) The  $\text{BF}_{01}$  is simply  $1/\text{BF}_{10}$ , that is, the likelihood of the data under the null compared to the alternative. As they both contain



**Figure 2.** Graphical results from the Bayesian analysis in JASP showing the prior and posterior distributions of the true population correlation.

**Table 3**  
Evidence categories for the Bayes Factor [6]

Bayes Factor $BF_{10}$	Label
$>100$	Extreme evidence for $H_1$
$30-100$	Very strong evidence for $H_1$
$10-30$	Strong evidence for $H_1$
$3-10$	Moderate evidence for $H_1$
$1-3$	Anecdotal evidence for $H_1$
$1$	No evidence
$1/3-1$	Anecdotal evidence for $H_0$
$1/10-1/3$	Moderate evidence for $H_0$
$1/30-1/10$	Strong evidence for $H_0$
$1/100-1/30$	Very strong evidence for $H_0$
$< 1/100$	Extreme evidence for $H_0$

BF = Bayes factor.

the same information, in practice it is usually easiest to report and interpret whichever form of the BF is greater than 1.

In Figure 2, the BF is also presented graphically with the unit circle in the output. The shaded area corresponds to the evidence in favor of the alternative hypothesis (indicated in the graphic by “Data |  $H_1$ ”), and the unshaded area corresponds to evidence in favor of the null (“Data |  $H_0$ ”). In our example, the ratio of the shaded area to the unshaded area can be seen to be about 3:1, which is the value of  $BF_{10}$ .

Notice that there is no direct analog to BFs in traditional hypothesis testing. A  $P$  value considers the evidence only under the null hypothesis but not the alternative hypothesis; by contrast, a BF considers both, directly pitting  $H_0$  and  $H_1$  head-to-head. The BF, then, conveys different information than the  $P$  value and cannot be determined from knowing the  $P$  value alone. For example, it is possible for 2 sets of data to both have a  $P$  value of .02 but have very different BFs.

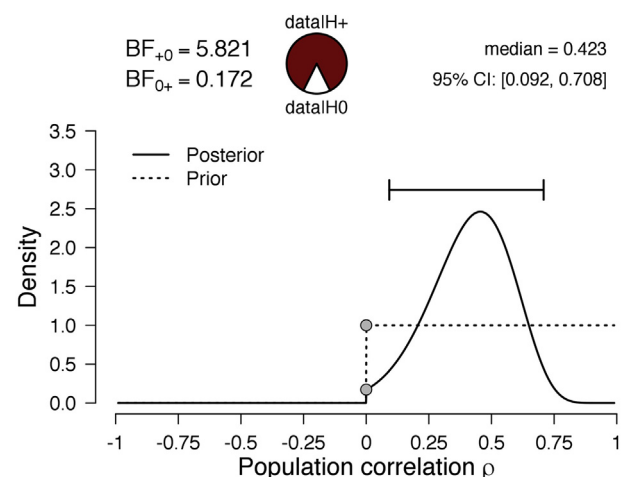
### Prior and Posterior Information About the Correlation

The density plot in the lower half of Figure 2 captures an important and interesting perspective on our study results. It graphically summarizes how our knowledge about the unknown population correlation has changed as a result of information gathered in our study. The dashed line represents our knowledge prior to the study. (This is called the “prior density,” or sometimes just “the prior,” from the Latin term for “coming before.”) Here our prior (chosen from the default option) is a flat line stretching between  $-1$  and  $1$ . This means that before the study all possible values of the correlation coefficient were considered equally likely. The solid line represents our knowledge after including study data (“posterior density,” or “the posterior,” from the Latin term for “coming after”). We can see that the study added greatly to our prior knowledge, because the posterior is now not

flat but instead places more weight on positive correlation values. The weights are especially high for those between 0 and 0.80, with a peak around 0.40. This means we are much more confident than before that the true population correlation is positive and of moderate strength, and is unlikely to be higher than about 0.75. The curve also shows which correlation values are now less plausible than before: For correlation values of about 0.70 and greater and about 0.10 and less, the solid posterior curve is now substantially below the dashed prior line, meaning that those tail values are much less plausible now than they were before the study. Notice also that the posterior density line is not completely flat for correlation values between 0 and  $-0.25$ . This means that although we are fairly confident the true correlation is positive, there is still a small possibility it is 0 or at most mildly negative.

Above the posterior density curve is a short solid line with butted ends. This is the “95% credible interval,” showing the span where 95% of the plausibility weights lie, which is where we are most confident the true correlation lies. The exact endpoints are given on the graphical output above the density curve. In our example, the 95% credible interval is (0.062, 0.708), which means that based on our prior information and study data, we can say there is a 95% chance the true correlation is between 0.062 and 0.708.

Notice again that the Bayesian analysis provides different information here than traditional testing. For instance, with a 95% confidence interval we cannot talk about the chance that we have captured the true population parameter. With a 95% credible interval we can. Furthermore, the density graphs give even more detailed information about the relative plausibility of particular values, such as negative values or very high values, which cannot be obtained from a classical confidence interval.



**Figure 3.** Bayesian analysis with a prior density of non-negative values only.

### A Bayesian Example with Stronger Evidence

In this example, the measurements from 24 obese/overweight adolescents are combined with those from 33 healthy-weight adolescents (Figure 4A). With the use of traditional testing, the Pearson correlation coefficient of 0.746 was statistically significant ( $P < .001$ , 95% confidence interval 0.602-0.842) (results not shown). When Bayesian testing with a default flat prior between  $-1$  and  $1$  is used, the Pearson correlation coefficient was 0.746. The Bayes factor reveals that the data are roughly 400 million times more likely under the 2-sided alternative than under the null, which provides extreme evidence in favor of there being a true non-zero correlation ( $BF_{10} = 380$  million), and we are 95% confident the true correlation lies between 0.588 and 0.837 (Figure 4B). This example shows how a Bayesian analysis allows researchers to report a useful estimate of the exceptionally high strength of evidence (380 million to 1 in favor of the alternative hypothesis) that would not be possible with a  $P$  value.

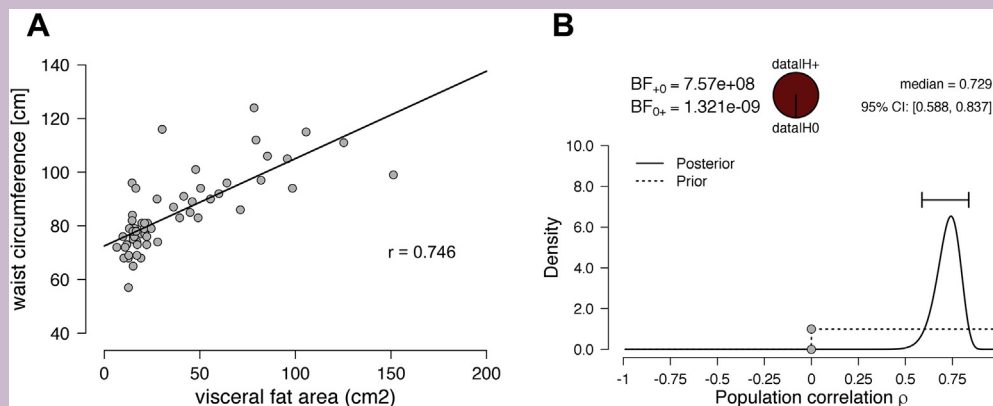


Figure 4. (A) Example data with larger sample size. (B) Bayesian analysis result for larger sample.

### Analysis With Prior Information

The aforementioned Bayesian analysis was conducted with a “noninformative” prior, where every value between  $-1$  and  $1$  was considered equally likely. It is possible, however, to have “informative” prior information about the population correlation. For example, suppose we had good scientific reasons to believe that the true correlation between total visceral fat area and waist circumference in overweight/obese adolescents could not reasonably be negative, but we have no information to suggest which positive values are most likely. We can then change our prior density from a flat line between  $-1$  and  $1$  to a flat line between  $0$  and  $1$ . In JASP we can do this by choosing the “Correlated positively Hypothesis” option in the Bayesian Correlation Pairs menu. As seen in Figure 3, this change in prior increases  $BF_{10}$  from 2.95 to 5.8, so now the evidence in favor of a positive correlation is solidly “moderate” (Table 3). The change in prior also shifted slightly the lower limit of the 95% credible interval, from 0.062 to 0.092. In other words, by including our stronger prior information with the study data, we now have stronger evidence in favor of the alternative hypothesis and can tighten the plausible values for our estimate slightly.

It is possible to further change the default flat prior to take advantage of prior information and use more detailed priors to place more weight on certain values [5,7]. This requires justification from researchers compared with using the default flat prior, however. As suggested by their name, priors must be based on information available before the study, not after. It is unethical to engage in “prior hacking,” in which a prior is chosen to produce a desired outcome. In practice, it is generally most conservative to use a default, non-informative prior.

### Conclusion

Bayesian analyses are a good alternative or supplement to classical significance testing, using posterior density graphical outputs, credible intervals, and BFs to provide additional information about the hypothesized effect. New software packages can now make it easier than in the past to run these analyses. Bayesian approaches allow researchers to directly include prior information when available, visually inspect how much the current study data have improved knowledge about the unknown parameter, report a range of values that has a 95% probability of containing the true parameter and use a single

number to summarize the strength of evidence contained in the study both for and against the study hypothesis.

## References

1. Wasserstein RL, Lazar NA. The ASA's statement on  $p$ -values: Context, process, and purpose. *Am Stat* 2016;70:129-133.
2. Gurrin LC, Kurinczuk JJ, Burton PR. Bayesian statistics in medical research: An intuitive alternative to conventional data analysis. *J Eval Clin Pract* 2000;6:193-204.
3. JASP [computer program]. Version 0.8.3.1. Amsterdam: JASP Team; 2017.
4. Eloi JC, Epifanio M, de Gonçalves MM, et al. Quantification of abdominal fat in obese and healthy adolescents using 3 Tesla magnetic resonance imaging and free software for image analysis. *PLoS One* 2017;12:e0167625.
5. Marsman M, Wagenmakers EJ. Bayesian benefits with JASP. *Eur J Dev Psychol* 2017;14:545-555.
6. Lee MD, Wagenmakers EJ. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge, UK: Cambridge University Press; 2013.
7. Kalil AC, Sun J. Bayesian methodology for the design and interpretation of clinical trials in critical care medicine: A primer for clinicians. *Crit Care Med* 2014;42:2267-2277.

---

## Disclosure

R.L.N. Department of Science, Technology, and Mathematics, Department of Psychology, HMB S340F, Gallaudet University, 800 Florida Avenue, NE,

Washington, DC 20002-3695. Address correspondence to: R.L.N.; e-mail: [Regina.Nuzzo@Gallaudet.edu](mailto:Regina.Nuzzo@Gallaudet.edu)  
Disclosure: nothing to disclose

---