View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repository of the Academy's Library

Periodica Polytechnica Electrical Engineering and Computer Science

61(3), pp. 246-252, 2017
https://doi.org/10.3311/PPee.9898
Creative Commons Attribution

RESEARCH ARTICLE

Experiences with Using Bayes Factors for Regression Analysis in Biostatistical Setting

Tamás Ferenci[1*], Levente Kovács[1]

Received 17 August 2016; accepted after revision 16 July 2017

Abstract

Null hypothesis significance testing dominates the current biostatistical practice. However, this routine has many flaws, in particular p-values are very often misused and misinterpreted. Several solutions has been suggested to remedy this situation, the application of Bayes Factors being perhaps the most well-known. Nevertheless, even Bayes Factors are very seldom applied in medical research. This paper investigates the application of Bayes Factors in the analysis of a realistic medical problem using actual data from a representative US survey, and compares the results to those obtained with traditional means. Linear regression is used as an example as it is one of the most basic tools in biostatistics. The effect of sample size and sampling variation is investigated (with resampling) as well as the impact of the choice of prior. Results show that there is a strong relationship between p-values and Bayes Factors, especially for large samples. The application of Bayes Factors should be encouraged evenin spite of this, as the message they convey is much more instructive and scientifically correct than the current typical practice.

Keywords

Bayes Factor, p-value, null hypothesis significance testing, linear regression model

1 Introduction

The application of *p*-values – and null hypothesis significance testing in general – remains a controversial topic in many applied statistical fields, including biostatistics. The currently most widely used (frequentist) apparatus of biostatistics does-not – as readers, clinical researchers and sometimes even textbooks seem to believe – represent a straightforward logical construct, but rather an incompatible hybrid of the Fisherian and the Neyman-Pearson tradition [1-4], which is itself problematic, and an application and interpretation routine that is often deeply flawed. The most important typical errors, fallacies, misunderstandings and misuses include [5-11]:

- Confusing clinical significance (whether the effect size is meaningful in the domain, in this case, medically) with statistical significance (whether the effect is assumed to be larger than what can be attributed to sampling variation).
- Application of the apparatus in non-sampling situations or for extremely large samples.
- Forgetting that *p*-values and the related inferential apparatus only capture sampling error, but say nothing of the potential non-sampling sources of error (i.e. biases).
- Forgetting whether the null hypothesis is – medically – meaningful at all or not (especially point nulls).
- Assuming that *p*-value is an error probability, i.e. the probability that the null hypothesis is true, given the sample.

Many believe that these errors are major contributors to the "replicability crisis" that is often discussed nowadays in medicine [12, 13].

These problems are so profound, despite that so prevalent [14], that there have been memorable attempts which implemented the most radical solution: banning the apparatus completely or almost completely. Perhaps most notable is the case of the Epidemiology journal [15] (with the rather strict policy removed in 2001 when founding editor Kenneth Rothman stepped down [16]) and the more recent example of the journal Basic and Applied Social Psychology [17].These decisions, in particular the question whether they are effective or needed, led to a widespread controversy, with American

[1] Physiological Controls Group, John von Neumann Faculty of Informatics, Óbuda University

[*] Corresponding author, e-mail: kovacs.levente@nik.uni-obuda.hu

Statistical Association (ASA) issuing a statement in mid-2016, formulating the views of the world's leading scientific body and gathering many relevant paper in the topic [18].

The most important is perhaps the last fallacy from the above list: many readers are tempted to beleive that *p*-values can convey information (evidence) *on their own*, without reference to any external information. This is, of course, not true: *p* value is not the probability of the null given the sample, but the other way around, probability of obtaining the sample (or more extreme) given the null. To reverse it, we have to use the Bayes' theorem:

$$P(H_0|\mathcal{S}) = \frac{P(\mathcal{S}|H_0)\cdot P(H_0)}{P(\mathcal{S})} \tag{1}$$

where $\mathcal{S}$ symbolizes the sample. (P means either probability or density (i.e. likelihood), depending on whether the variable is discrete or continuous.) One can now immediately see that we need $P(H_0)$, that is, the *prior probability* of the null hypothesis to obtain the probability that is thought by many to be given by the *p*-value. (Forgetting this is identical to the base rate fallacy.) Its effect can be dramatic: it is quite easy to see that in the most simple situation, a *p*-value of 0.05 might very well mean 36% probability that the null is true (no effect found) if the prior probability is only 10% [19, 20]. (We assumed 80% power, a typical value.) With more advanced tools, it it even possible to show that for $p = 0.05$ the probability of the null being true *cannot be smaller* than 28.9% no matter what situation we presume [21, 22].

Many attempts have been made to replace or at least supplement *p*-values with analytical methods that are less prone to these errors, and help correct interpretation. The already mentioned ASA statement is rather vague from this aspect: "[t]hese includemethods that emphasize estimation over testing, such as confidence, credibility, or prediction intervals; Bayesian methods; alternative measures of evidence, such as likelihood ratios or Bayes Factors; and other approaches such as decision-theoretic modeling and false discovery rates" [18.

Out of these, perhaps the Bayes Factors are the – relatively – most well-known. The basic idea is rather simple: take the same equation as (1) but for H_1 (instead of H_0), and divide the two; thus we obtain

$$\frac{P(H_0|\mathcal{S})}{P(H_1|\mathcal{S})} = \frac{P(\mathcal{S}|H_0)}{P(\mathcal{S}|H_1)}\cdot\frac{P(H_0)}{P(H_1)} \tag{2}$$

as the term $P(\mathcal{S})$ fortunately cancels. Noting that $P(H_1) = 1 - P(H_0)$ (and likewise for the conditional probability) we actually have

$$\frac{P(H_0|\mathcal{S})}{1-P(H_0|\mathcal{S})} = \frac{P(\mathcal{S}|H_0)}{P(\mathcal{S}|H_1)}\cdot\frac{P(H_0)}{1-P(H_0)}, \tag{3}$$

but a probability divided by one minus that probability is odds, so we can write

$$\text{odds}(H_0|\mathcal{S}) = \frac{P(\mathcal{S}|H_0)}{P(\mathcal{S}|H_1)}\cdot\text{odds}(H_0). \tag{4}$$

The remaining factor on the right-hand side is called *Bayes Factor* [23, 24]:

$$BF_{01} = \frac{P(\mathcal{S}|H_0)}{P(\mathcal{S}|H_1)}. \tag{5}$$

In other words, *this is the factor with which we have to multiply the prior odds to obtain the posterior odds*.

In practice, if the two hypotheses represent restrictions on a – not necessarily one-dimensional – parameter θ, i.e. $H_0{:}\theta \in \theta_0$ and $H_1{:}\theta \in \theta_1$ ($\theta_0 \cap \theta_1 = \varnothing$) then we have

$$BF_{01} = \frac{\int_{\vartheta\in\theta_0} P(\mathcal{S}|H_0,\vartheta)\pi(\vartheta|H_0)\vartheta}{\int_{\vartheta\in\theta_1} P(\mathcal{S}|H_1,\vartheta)\pi(\vartheta|H_1)\vartheta} \tag{6}$$

where $\pi(\vartheta)$ is the prior distribution of the parameter. This is similar to the likelihood-ratio that is very well-known in frequentist statistics too, but instead of the supremum of the likelihood being taken, practically a weighted average is formed, weighted by the prior.

This definition can be substantially simplified in the practically very important scenario of the null hypothesis being a point null (i.e. $\theta = (\xi,\eta)$, where $\dim\xi = 1$ with $H_0{:}\xi = \xi_0$ and $H_1{:}\xi \neq \xi_0$, thus η represents the nuisance parameters). If we assume that the prior for ξ is continuous at ξ_0 (conditional on the nuisance parameters) then the numerator can be written as $\int P(\mathcal{S}|\xi = \xi_0, H_1, \eta)\pi(\eta|\xi = \xi_0, H_1)\mathrm{d}\eta$ instead of $\int P(\mathcal{S}|H_0,\eta)\pi(\eta|H_0)\mathrm{d}\eta$. However, $\int P(\mathcal{S}|\xi = \xi_0, H_1, \eta)\pi(\eta|\xi = \xi_0, H_1)\mathrm{d}\eta = P(\mathcal{S}|\xi = \xi_0, H_1)$, and by Bayes' theorem we have $P(\mathcal{S}|\xi = \xi_0, H_1) = \frac{P(\xi = \xi_0|H_1,\mathcal{S})P(\mathcal{S}|H_1)}{P(\xi = \xi_0|H_1)}$. As the denominator is $P(\mathcal{S}|H_1)$ (see Eq. (5)), the Bayes Factor is simply

$$BF_{01} = \frac{P(\xi = \xi_0|H_1,\mathcal{S})}{P(\xi = \xi_0|H_1)} \tag{7}$$

in this case. This is called the Savage–Dickey density ratio [25].

A characteristic of Bayes Factors is the need for prior information on the investigated parameter's distribution. This is generally true for Bayesian methods; whether it is a drawback or not, and how the prior should be selected is a matter of vast, decade-long debate [26, 27]. Alternatively, some have proposed the usage of the so-called "Minimum Bayes Factor", i.e. the smallest Bayes Factor that is possible (over all priors) [28, 29, 30], which is therefore no longer dependent on the prior (but may be dependent on certain assumptions). And, of course, one has to be willing to accept the fact that this metric is no longer a "context independent" measure, but rather the prior belief is needed to be incorporated later on (which is just an advantage, i.e. that Bayes Factors make this fact explicit).

As Bayes Factor has many further advantages, and corrects many misuses that are often apparent with *p*-values, its

wider application been endorsed by Goodman [31, 32] and Wagenmakers [33], among others.

Despite this, Bayes Factors are seldom used in practice in medicine, especially in "ordinary" clinical papers – their appearance is mostly limited to papers that specifically demonstrate or investigate their usefulness (e.g. [34]), but they almost never appear as regular apparatus in the investigation of usual clinical questions.

The aim of this paper is investigate the real-life applicability of Bayes Factors by comparing the results obtained with them to that of null hypothesis significance testing in a simple, but realistic medical scenario on individual patient data. The paperwill be purely descriptive, i.e. no in-depth attempt is made to give theoretical (mathematical) explanation to the observed phenomena.

2 Material and Methods

2.1 Investigated questions

The aim will be to investigate the applicability of Bayes Factors in regression analysis with – standard, normal – linear models by comparing them to traditional means (i.e. *p*-values). It was selected as an example because regression analysis is one of the most fundamental tools in biostatistics, thus this will be a relevant example. However, as a preliminary investigation, it will be confined to the most simple question within regression analysis: assessing a single explanatory variable's impact (in itself) on the response variable. (Although this should be done with caution when multicollinearity is present, but is nevertheless a very basic analytical question.)

Within the null hypothesis significance testing framework, this question can be addressed by the *t*-test, as discussed in any standard textbook [35, 36]. The Bayes Factors approach in its most popular form for this case [37, 38] will be now briefly outlined.

Consider the following regression model:

$$y_i = \alpha + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_p x_{i,p} + \varepsilon_i, \tag{8}$$

where ε_i is assumed to be independent normal variate with zero mean and constant σ variance. Our research question can be formulated as $H_0{:}\beta_j = 0$ versus $H_1{:}\beta_j \neq 0$, therefore by 6 we have

$$BF_{01} = \frac{\prod_{i=1}^{n} \phi\left(\frac{y_i - \left(\alpha + \beta_1 x_{i,1} \ldots + \beta_{j-1} x_{j-1,i} + \beta_{j-1} x_{j+1,i} + \ldots + \beta_p x_{i,p}\right)}{\sigma}\right)}{\int_b \prod_{i=1}^{n} \phi\left(\frac{y_i - \left(\alpha + \beta_1 x_{i,1} \ldots + \beta_{j-1} x_{j-1,i} + b x_{j,i} + \beta_{j-1} x_{j+1,i} + \ldots + \beta_p x_{i,p}\right)}{\sigma}\right) \pi(b) b}, \tag{9}$$

where ϕ is the standard normal density. Assuming we know every regression coefficient apart from β_j and the error variance σ (these assumptions can be relaxed, or we can consider the analysis to be conditional on them) all we need is $pi(b)$, the prior distribution of a regression coefficient. The most popular choice is Cauchy-distribution, which is equivalent to a hierarchical normal/inverse gamma model (but this latter can be more easily generalized to this multivariate case):

$$\beta | g \sim N(0, g\sigma^2 (\mathbf{X}^T \mathbf{X}/n)^{-1}) \tag{10}$$

$$g \sim \text{InvGamma}(1/2, s^2/2), \tag{11}$$

where $\beta = [\beta_i]_{i=1}^{p}$, $\mathbf{X} = [x_{i,j}]_{i=1,j=1}^{n,p}$ and s is a new (hyper) parameter. This choice is usually called weekly informative, fulfilling location and scale invariance, consistency and consistency in information (objective or default prior). This is usually attributed to Jeffreys, with an expansion from Zellner and Siow (JZS prior) [23, 39].

Now that the methods are clarifed, the questions of interest will be more specifically:

- How Bayes Factors compare to *p*-values?
- How is this relationship affected by certain parameters, particularly the applied prior (s) and the sample size?

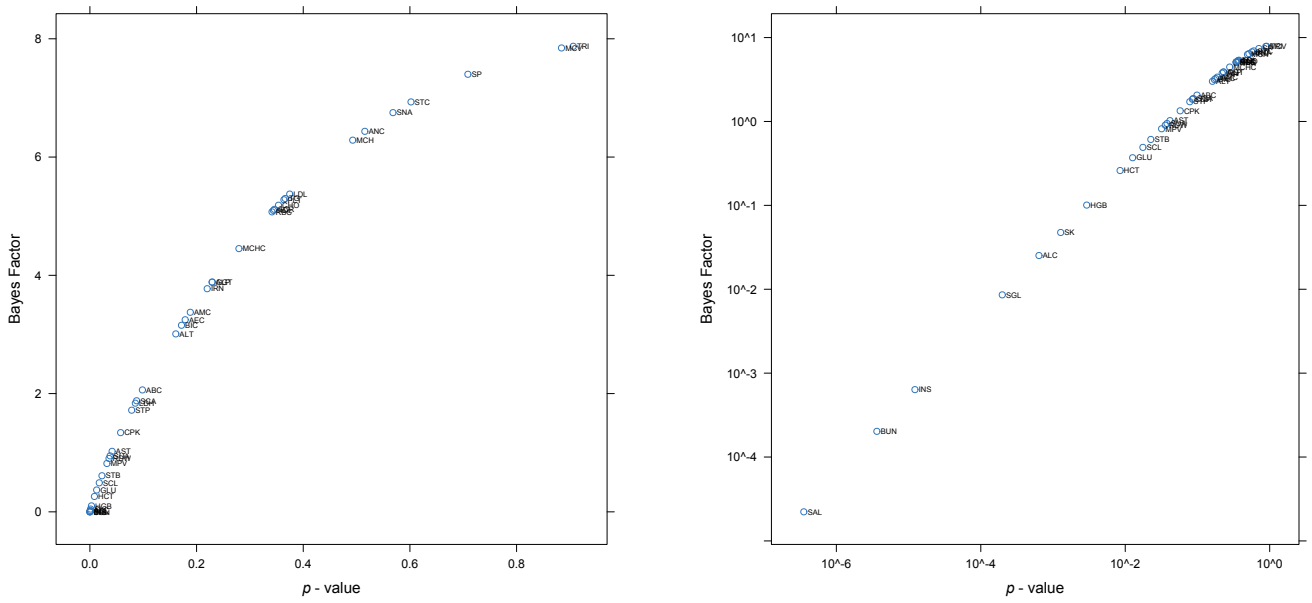
2.2 Patient data

To present a realistic example, real-life data from the representative US survey National Health and Nutrition Examination Survey (NHANES) will be used. NHANES is now a continuous public health program, with results published in biannual cycles [40]. It is a nation-wide survey aimed to be representative for the whole civilian non-institutionalized US population, by employing a complex, stratified multi-stage probability sampling plan. The amount of collected data is tremendous (although sometimes varying from cycle to cycle), including demographic data, physical examination, collection of clinical chemistry parameters, and a thorough questionnaire concentrating on anamnesis and lifestyle. Now $p = 43$ clinical chemistry parameters[1] from the 2013/14 cycle – the most recent available – will be used [41]. To make the database more homogeneous, it was filtered to males aged 18 years or more. For simplicity, subjects with any missing value were left out. Although for precise analyses it is important to take the survey structure into account by weight, now – as the focus of the study was elsewhere – this was neglected for simplicity.

On this database, regressions can be carried out by regressing one of these variables against the rest. These are clinically meaningful and based on real-life data. As we have a number of variables, this database also makes it possible to investigate regressions of very different nature (as variables have a very diverse distribution, and correlational structure).

The final sample size was $n = 1190$; this is large enough so that subsamples can be also used when studying smaller samples (with having results for the full sample).

1 Data files used: `HDL` (cholesterol – HDL), `TRIGLY` (cholesterol – LDL and Triglycerides), `TCHOL` (cholesterol – total), `CBC` (Complete Blood Count with 5-part Differential – Whole Blood), `GHB` (Glycohemoglobin), INS (Insulin), `GLU` (Plasma Fasting Glucose) and `BIOPRO` (Standard Biochemistry Profile).



(a) Linear scale

(b) Logarithmic scale

Fig. 1 *p*-values and Bayes Factors of the explanatory variables in the regression of glycohemoglobin.

2.3 Programs used

All analysis was carried out under the `R` statistical program package, version 3.3.1 [42] with a custom script developed for this purpose that is available at the corresponding author on request. The Bayes Factors were calculated with package `BayesFactor`, version 0.9.12-2 [43]. Data visualization is performed with the `lattice` package, version 0.20-33 [44].

3 Results

A comparison of the *p*-values and Bayes Factors of the predictor variables in a regression is shown on Fig. 1 for the example of glycohemoglobin.

The relationship is almost perfectly linear between the logarithm of the *p*-value and the Bayes Factor. This is no exception: Fig. 2 shows the same scatterplots for all variables (all variable selected as response, one at a time, and the remaning being predictors) in logaritmic scale. Indeed, even the smallest linear correlation coefficient between the logarithms is over 0.99.

Next, the role of the sample size will be investigated. The same analysis as on Fig. 1 was repeated, but with smaller samples. These were randomly sampled from the whole database (with replacement); sample sizes 50, 100, 200 and 500 were used. Actually, the aim of this investigation is twofold: this method makes it possible not only to investigate the effect of sample size, but also the sampling variation as now many samples could be investigated. (1000 random samples were now drawn.) Results are shown for the example of serum glucose (as explanatory variable): Fig. 3 shows the univariate distributions, Fig. 4 shows the joints distribution.

One can see that both *p*-values and Bayes Factors get smaller as sample size increases (logically), and also their variability decreases (note the logarithmic scale).

The joint distribution reveals that the relationship between *p*-values and Bayes Factors gets stronger with increasing sample size. (Thus it is no surprise that we have seen an almost perfect relationship for the whole sample.) Again, note the shifting to lower *p*-value/Bayes Factor with increasing sample size, as expected. The other observation that is very clear from the scattergram is the strong relationship in this sense too, and – more importantly – it is now apparent that this gets stronger withsample size.

Finally, the effect of the used prior was investigated. As it was already discussed, ”used prior” now means the selection of the s hyperparameter; in addition to the default $\sqrt{2}/4$ (”medium”, this was used everywhere up to here), the alternatives $1/2$ (”wide”) and $\sqrt{2}/2$ (”ultrawide”) were now investigated. Results are shown on Fig. 5 (again for the example of glycohemoglobin). One can see that the pattern is similar, with the points shifted upwards as the value of s increases; this is again logical.

4 Discussion and conclusion

p-values and Bayes Factors are strongly related. Their relationship comes as no surprise as they measure related characteristics; the strength of the connection is what can be surprising at first glance.

However, it should be noted that in simple cases it might even happen that there is a *deterministic* relationship between the two [45]. Even when not, such strong relationship has been already described in the literature [46, 47]. The reason can be best seen for point null hypotheses (as in the present case) by considering the Savage–Dickey ratio presented in Eq. (7): the BF is the ratio of two densities under the same model, while *p*-value is related to the posterior density, and they are changing roughly proportionally when S is changing [48].

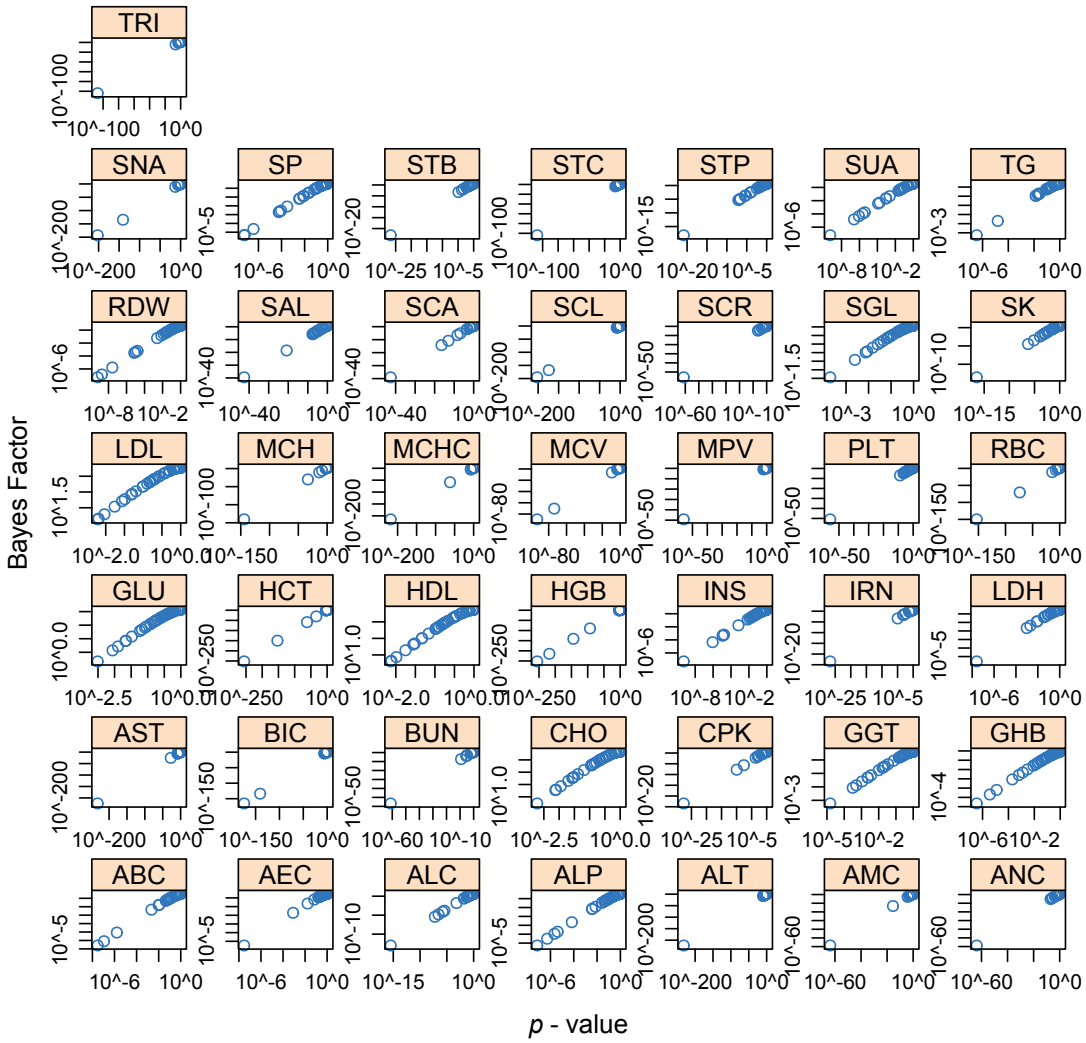
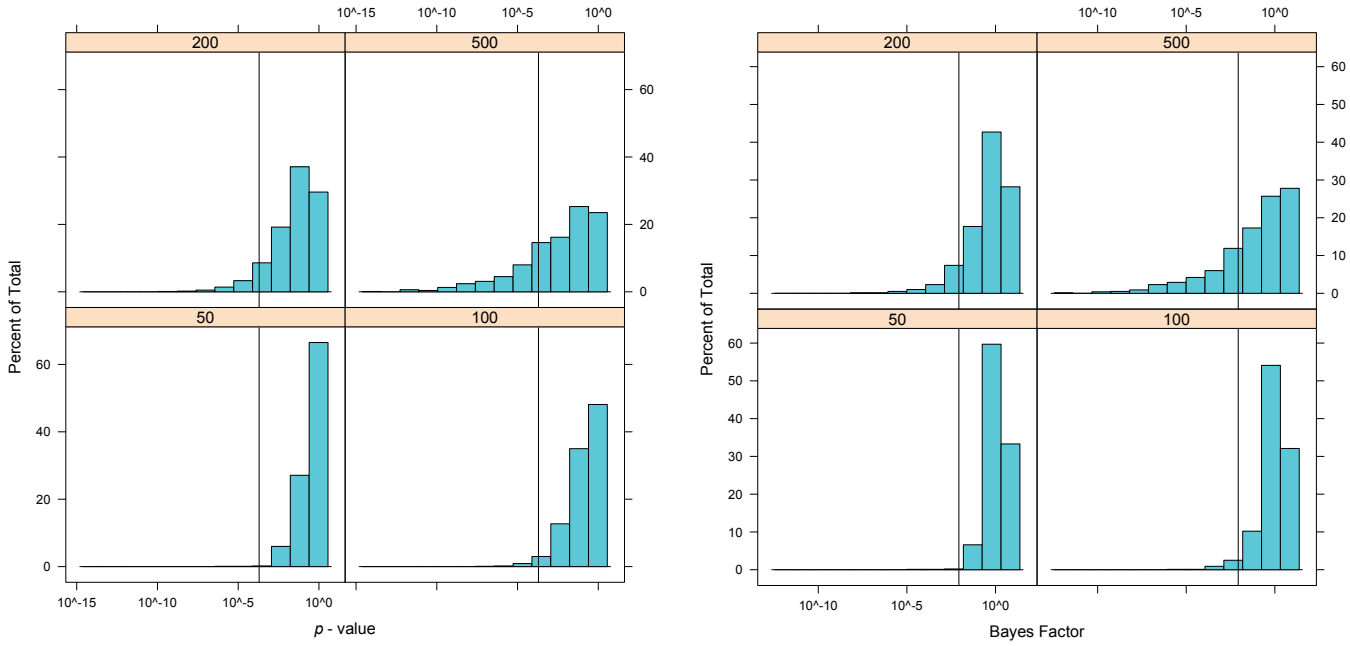


Fig. 2 *p*-values and Bayes Factors for all variables in all regressions, logarithmic scale.



(a) *p*-value (b) Bayes Factor

Fig. 3 Effect of sample size – shown in the panel titles – and sampling variation on *p*-values and Bayes Factors (univariately), with the glycohemoglobin being the response variable and serum glucose being the investigated predictor variable; vertical black lines indicates the estimates for the full sample (logarithmic scale).

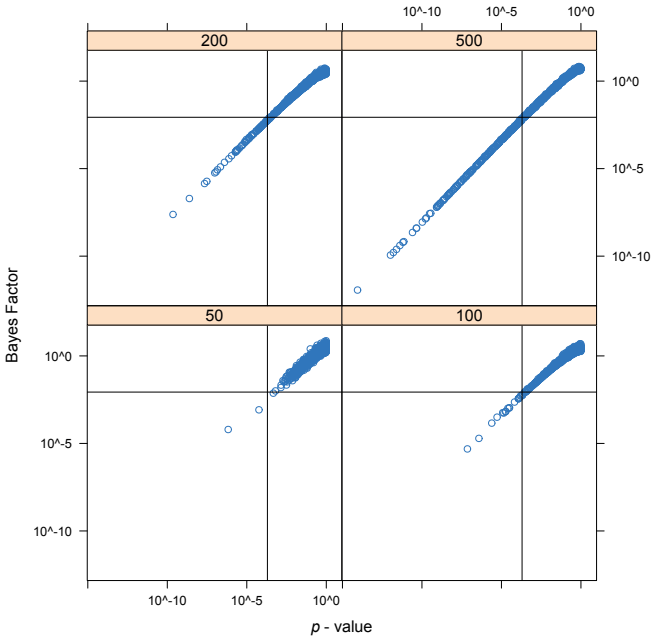


Fig. 4 Effect of sample size – shown in the panel titles – and sampling variation on *p*-values and Bayes Factors (jointly), with the glycohemoglobin being the response variable and serum glucose being the investigated predictor variable; vertical black lines indicates the estimates for the full sample (logarithmic scale).

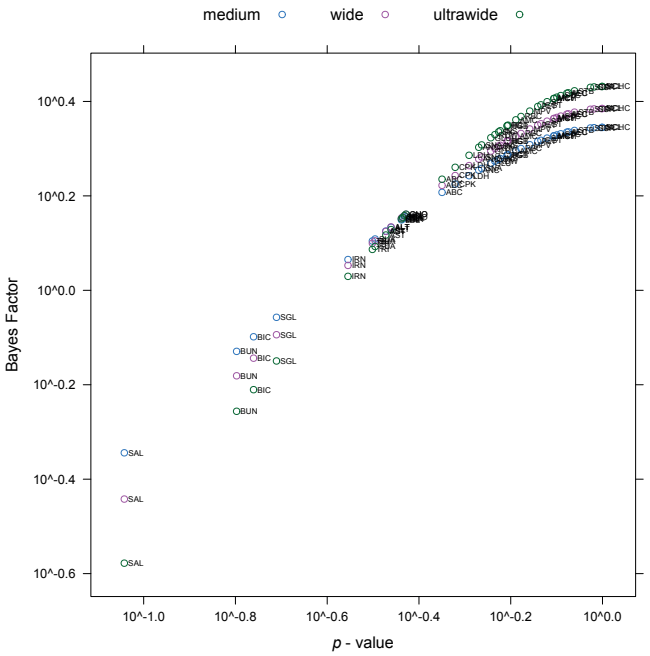


Fig. 5 Effect of the choice of prior on *p*-values and Bayes Factors, with the glycohemoglobin being the response variable (logarithmic scale).

The present research also makes it clear that – in the investigated scenario – the relationship gets stronger with increasing sample size: for samples larger than a few hundred observation, the relationship is almost perfect.

When using JZS prior, the choice of the *s* parameter had no major impact on the relationship between *p*-values and Bayes Factors, but uniformly shifted Bayes factors.

Finally, it is important to emphasize that these findings do not make Bayes Factors pointless: even for a perfect relationship, the message conveyed by Bayes Factors is different (and, as we have seen, much more instructive and scientifically correct thanthe current typical practice with *p*-values).

References

[1] Goodman, S. N. "Values, Hypothesis Tests, and Likelihood: Implications for Epidemiology of a Neglected Historical Debate." *American Journal of Epidemiology*. 137(5), pp. 485-496. 1993. https://doi.org/10.1093/oxfordjournals.aje.a116700

[2] Hubbard, R., Bayarri, M. J. "Confusion Over Measures of Evidence (p's) Versus Errors (alpha's) in Classical Statistical Testing." *The American Statistician*. 57(3), pp. 171-178. 2003. https://doi.org/10.1198/0003130031856

[3] Lenhard, J. "Models and Statistical Inference: The Controversy between Fisher and Neyman-Pearson." *The British Journal for the Philosophy of Science*. 57(1), pp. 69-91. 2006. https://doi.org/10.1093/bjps/axi152

[4] Lehmann, E. L. "The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?." *Journal of the American Statistical Association*. 88(424), pp. 1242-1249. 1993. https://doi.org/10.2307/2291263

[5] Goodman, S. "A Dirty Dozen: Twelve P-Value Misconceptions." *Seminars in Hematology*. 45(3), pp. 135-140. 2008. https://doi.org/10.1053/j.seminhematol.2008.04.003

[6] Stang, A., Poole, Ch., Kuss, O. "The ongoing tyranny of statistical significance testing in biomedical research." *European Journal of Epidemiology*. 25(4), pp. 225-230. 2010. https://doi.org/10.1007/s10654-010-9440-x

[7] Lew, M. J. "Bad statistical practice in pharmacology (and other basic biomedical disciplines): you probably don't know P. *British Journal of Pharmacology*. 166(5), pp. 1559-1567. 2012. https://doi.org/10.1111/j.1476-5381.2012.01931.x

[8] Perezgonzalez, J. "The meaning of significance in data testing." *Frontiers in Psychology*. 6, p. 1293. 2015. https://doi.org/10.3389/fpsyg.2015.01293

[9] Gigerenzer, G. "Mindless statistics." The Journal of Socio-Economics. 33(5), pp. 587-606. 2004. https://doi.org/10.1016/j.socec.2004.09.033

[10] Nuzzo, R. "Statistical errors." *Nature*. 506(7487), pp. 150-152. 2014.

[11] Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., Altman, D. G. "Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations." *European Journal of Epidemiology*. 31(4), pp. 337-350. 2016. https://doi.org/10.1007/s10654-016-0149-3

[12] Ioannidis, J. P. A. "Why Most Published Research Findings Are False." *PLoS Med*. 2(8), pp. e124. 2005. https://doi.org/10.1371/journal.pmed.0020124

[13] Goodman, S. N. "A comment on replication, P-values and evidence." *Statistics in Medicine*. 11(7), pp. 875-879. 1992. https://doi.org/10.1002/sim.4780110705

[14] Haller, H., Krauss, S. "Misinterpretations of significance: A problem students share with their teachers." *Methods of Psychological Research*. 7 (1), pp. 1-20. 2002.

[15] Lang, J. M., Rothman, K. J., Cann, C. I. "That confounded P-value." *Epidemiology*. 9(1), pp. 7-8. 1998.
[16] The Editors "The Value of P." *Epidemiology*. 12(3), p.286. 2001.
[17] Trafimow, D. "Editorial." *Basic and Applied Social Psychology*. 36(1), pp. 1-2. 2014. https://doi.org/10.1080/01973533.2014.865505
[18] Wasserstein, R. L., Lazar, N. A. "The ASA's Statement on p-Values: Context, Process, and Purpose." *The American Statistician*. 70(2), pp. 129-133. 2016. https://doi.org/10.1080/00031305.2016.1154108
[19] Colquhoun, D. "An investigation of the false discovery rate and the misinterpretation of p-values." *Open Science*. 1(3), 2014. https://doi.org/10.1098/rsos.140216
[20] Sterne, J. A. C., Smith, G. D. "Sifting the evidence-what's wrong with significance tests?." *Physical Therapy*. 81(8), pp. 1464-1469. 2001.
[21] Sellke, T., Bayarri, M. J., Berger, J. O. "Calibration of p Values for Testing Precise Null Hypotheses." *The American Statistician*. 55(1), pp. 62-71. 2001. https://doi.org/10.1198/000313001300339950
[22] Berger, J. O., Sellke, T. "Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence." *Journal of the American Statistical Association*. 82(397), pp. 112-122. 1987. https://doi.org/10.1080/01621459.1987.10478397
[23] Jeffreys, H. "*The Theory of Probability*." Oxford Classic Texts in the Physical Sciences, OUP Oxford, 1998.
[24] Kass, R. E., Raftery, A. E. "Bayes Factors." *Journal of the American Statistical Association*. 90(430), pp. 773-795. 1995. https://doi.org/10.1080/01621459.1995.10476572
[25] Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., Grasman, R. "Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method." *Cognitive Psychology*. 60(3), pp. 158-189. 2010. https://doi.org/10.1016/j.cogpsych.2009.12.001
[26] Samaniego, F. J. "*A Comparison of the Bayesian and Frequentist Approaches to Estimation*." Springer Series in Statistics, Springer New York, 2010.
[27] Robert, C. "*The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*." Springer Texts in Statistics, Springer New York, 2007.
[28] Edwards, W., Lindman, H., Savage, L. J. "Bayesian statistical inference for psychological research." *Psychological Review*. 70(3), p. 193. 1963.
[29] Bayarri, M. J., Berger, J. O. "Quantifying surprise in the data and model verification." *Bayesian Statistics*. 6. pp. 53-82. 1999.
[30] Goodman, S. N. "Of P-values and Bayes: a modest proposal." *Epidemiology*. 12(3), pp. 295-297. 2001.
[31] Goodman, S. N. "Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy." *Annals of Internal Medicine*. 130(12), pp. 995-1004. 1999. https://doi.org/10.7326/0003-4819-130-12-199906150-00008
[32] Goodman, S. N. "Toward Evidence-Based Medical Statistics. 2: The Bayes Factor." *Annals of Internal Medicine*. 130(12), pp. 1005-1013. 1999. https://doi.org/10.7326/0003-4819-130-12-199906150-00019
[33] Mulder, J., Wagenmakers, E. J. "Editors' introduction to the special issue ''Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments''." *Journal of Mathematical Psychology*. 72, pp. 1-5. 2016. https://doi.org/10.1016/j.jmp.2016.01.002
[34] Ioannidis, J. P. A. "Effect of Formal Statistical Significance on the Credibility of Observational Associations." *American Journal of Epidemiology*. 168(4), pp. 374-383. 2008. https://doi.org/10.1093/aje/kwn156
[35] Sen, A., Srivastava, M. "*Regression analysis: theory, methods, and applications*." Springer Science & Business Media, 2012.
[36] Draper, N. R., Smith, H. "*Applied regression analysis*." John Wiley & Sons, 2014.
[37] Rouder, J. N., Morey, R. D. "Default Bayes Factors for Model Selection in Regression." *Multivariate Behavioral Research*. 47(6), pp. 877-903. 2012. https://doi.org/10.1080/00273171.2012.734737
[38] Liang, F., Rui, P., Molina, G., Clyde, M. A., Berger, J. O. "Mixtures of g Priors for Bayesian Variable Selection." *Journal of the American Statistical Association*. 103(481), pp. 410-423. 2008. https://doi.org/10.1198/016214507000001337
[39] Zellner, A., Siow, A. "Posterior odds ratios for selected regression hypotheses." *Trabajos de Estadistica Y de Investigacion Operativa*. 31(1), pp. 585-603. 1980. https://doi.org/10.1007/BF02888369
[40] Centers for Disease Control and Prevention, National Center for Health Statistics "National Health and Nutrition Examination Survey." 2016. [Online]. Available from: http://www.cdc.gov/nchs/nhanes.htm. [Accessed 12th August 2016].
[41] Centers for Disease Control and Prevention, National Center for Health Statistics "National Health and Nutrition Examination Survey, {NHANES} 2011-2012." 2013. [Online]. Available from: http://wwwn.cdc.gov/nchs/nhanes/search/nhanes13_14.aspx. [Accessed 12th August 2016].
[42] R Core Team "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2016. URL: https://www.R-project.org/
[43] Morey, R. D., Rouder, J. N. "BayesFactor: Computation of Bayes Factors for Common Designs." 2015. R package version 0.9.12-2, URL: https://CRAN.R-project.org/package=BayesFactor
[44] Sarkar, D. "*Lattice: Multivariate Data Visualization with R*." Springer, New York, 2008
[45] Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., Iverson, G. "Bayesian t tests for accepting and rejecting the null hypothesis." *Psychonomic Bulletin & Review*. 16(2), pp. 225-237. 2009. https://doi.org/10.3758/PBR.16.2.225
[46] Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., Wagenmakers, E.-J. "Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests." *Perspectives on Psychological Science*. 6(3), pp. 291-298. 2011. https://doi.org/10.1177/1745691611406923
[47] Rouder, J. N., Morey, R. D., Speckman, P. L., Province, J. M. "Default Bayes factors for {ANOVA} designs." *Journal of Mathematical Psychology*. 56(5), pp. 356-374. 2012. https://doi.org/10.1016/j.jmp.2012.08.001
[48] Marsman, M., Wagenmakers, E.-J. "Three Insights from a Bayesian Interpretation of the One-Sided P Value." *Educational and Psychological Measurement*. pp. 1-11. 2016 https://doi.org/10.1177/0013164416669201