JEPS BULLETIN

The Official Blog of the Journal of European Psychology Students

ABOUT

JEPS WEBSITE BLOG CONTRIBUTORS CONTRIBUTE!

BAYESIAN STATISTICS: WHAT IS IT AND WHY DO WE NEED IT?



There is a revolution in statistics happening: The Bayesian revolution. Psychology students who are interested in research methods (which I hope everyone is!) should know what this revolution is about. Gaining this knowledge now instead of later might spare you lots of misconceptions about statistics as it is usually instructed in psychology, and it might help you gain a deeper understanding of the foundations of statistics. To make sure that you can try out everything

you learn immediately, I conducted analysis in the free statistics software R (www.r-project.org; click HERE for a tutorial how to get started with R, and install RStudio for an enhanced Rexperience) and I provide the syntax for the analysis directly in the article so you can easily try them out. So let's jump in: What is "Bayesian Statistics", and why do we need it?

Frequentist vs. Bayesian Statistics: An Example

Psychology students are usually taught the traditional approach to statistics: Frequentist statistics. Assume, for instance, you want to test the hypothesis that people who wear fancy hats are more creative than people who do not wear hats or hats that look boring. You carefully choose a sample of 100 people who wear fancy hats and 100 people who do not wear fancy hats and you assess their creativity using psychometric tests.

In R we can easily simulate data for this example; just copy this syntax into R and run it (everything with a # in front is an explaining comment that is not processed by R).

#First, set a seed first for the quasi-random number generation in our data simulation #If you set the same number you will get exactly the same data as I set.seed(666)

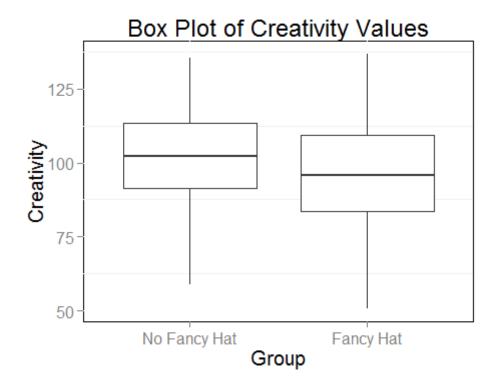
##Prepare variables for data simulation

```
n1fh = 100 # Number of people wearing fancy hats
n2nfh = 100 # Number of people wearing no fancy hats
mu1 = 103 # Population mean of creativity for people wearing fancy hats
mu2 = 98 # Population mean of creativity for people wearing no fancy hats
sigma = 15 # Average population standard deviation of both groups
n = n1fh+n2nfh # Total sample size
##Generate the simulated data
y1 = rnorm(n1fh, mu1, sigma) # Data for people wearing fancy hats
y2 = rnorm(n2nfh, mu2, sigma) # Data for people wearing no fancy hats
```

Let's look at the descriptive statistics for both groups.

```
mean(y1)
mean(y2)
sd(y1)
sd(y2)
mean(y1)-mean(y2) # Mean difference
```

The mean values are M(fancy hat) = 102.00, SD = 15.44, and M(no fancy hat) = 96.61, SD = 17.03. Hence, the mean difference is 5.39. Let's further investigate the data in a box plot.



From the boxplot it also seems that there might be a difference. How can you reliably test if this difference is not just present in your sample but indicates an actual difference for the two

underlying populations of fancy hat-users and non-fancy hat-users?

The frequentist way. In your statistics class you learned that to compare the creativity of the two groups you should compute a "*t*-test for independent samples". You conduct this test in your favorite statistics software, R.

```
t.test(y1,y2, var.equal=TRUE) #Frequentist t-test
```

The result (as you would report it according to APA-guidelines) is $t_{198} = 2.35$, p = .020. In your statistics course you learned how to interpret this magical "p-value": "The probability of obtaining a group difference of the observed magnitude or larger, given the null hypothesis (that in the population there is no difference in the two groups' creativity), is 2.0%". Now, because 2.0% is very unlikely (more unlikely than the usual, but arbitrary, cut-off of 5%), you reject the null hypothesis. You accept the alternative hypothesis which states that there is a difference in the two groups' creativity. Since the mean value of people wearing fancy hats is higher, you conclude that people who wear fancy hats are more creative than people who do not wear fancy hats. Congratulations, hypothesis supported! But, wait, is it really that easy? Is there a different way to think about these data?

The Bayesian way. A basic but effective way to conduct a t-test using Bayesian statistics is the *Bayes factor*. The Bayes factor represents the ratio of the likelihoods of the data given the null hypothesis versus the alternative hypothesis. In simpler words, it answers the question "How likely are my data if the creativity of the two groups differs, in comparison to how likely the data are if there is no difference?" We compute the Bayes factor for our example:

```
##Bayesian t-test: via Bayes factor
install.packages("BayesFactor") # Install BayesFactor-package
library('BayesFactor') # Load BayesFactor-package
yx = data.frame(y,x) # Prepare data
bf = ttestBF(formula = y ~ x, data=yx) # Estimate Bayes factor
bf # Investigate the result
```

We yield a Bayes Factor of 1.98. This indicates that the data are 1.98 times more likely under the alternative hypothesis (that there is a difference) than under the null hypothesis (that there is no difference).

Here comes the interpretative stunner: If you held even odds for both hypotheses beforehand, the Bayes factor can be interpreted as the posterior odds for both hypotheses based on the data. Hence, if you did not have strong assumptions about the outcome before seeing the data, our posterior odds would now be 1.98:1 for the alternative hypothesis. You can now be a bit more confident that your assumption is true than before you collected the data.

Receiving "Free Lunch" or not: A Comparison of the Foundations of the two Statistical Schools

From the Bayesian analysis, we concluded that "the hypothesis that there is a difference between the two groups' creativity is slightly favored over the hypothesis that there is no difference". In comparison, from the frequentist analysis we concluded "the probability of obtaining a group difference of the observed magnitude or larger, given the null hypothesis (that in the population there is no difference in the two groups' creativity) is 2.0%" – we rejected the null hypothesis and accepted the alternative hypothesis.

From this comparison you can see that the Bayesian approach to statistics is more intuitive; it resembles how we think about probability in everyday life – in the odds of hypotheses, not those of data. In comparison, the frequentist conclusion sounds complex and difficult to comprehend. Where do these interpretational differences come from? To find out, let us compare the foundations of both schools.

Frequentist principles. In frequentist statistics, neither the rejected hypothesis (H0: There is no difference between the two groups) nor the accepted hypothesis (H1: There is a difference) are directly tested, but only the probability of the data. This has been compared to receiving free lunch: One does not state what the alternative hypothesis is but eventually one does accept it without testing it. This is like receiving lunch without paying (Rouder, Wagenmakers, Verhagen, & Morey, submitted)! In other, more complex, but famous words:

"What the use of p implies is that a hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred. This seems a remarkable procedure" (Harold Jeffreys, 1891–1989).

This ironical statement touches the fact that the p-value is the proportion of all possible samples one could assess that could be "at least as extreme" as the observed data if the null hypothesis is true. Hence, the rejection of the null hypothesis is based on the probability of samples that have not been observed.

Bayesian principles: The Concept of the Bayesian Prior, Likelihood, and Posterior. In Bayesian statistics, there is no "free lunch"; there are no conclusions about hypothesis that have not been tested or data that have not been observed: Rather, as in the Bayes factor example, probabilities of hypotheses can be directly tested and compared (Dienes, 2010).

The Bayes theorem, the basic rule behind Bayesian statistics, states that the posterior (the probability of the hypothesis given the data) is proportional to the likelihood (the probability of the data given the hypothesis) times the prior (the probability of the hypothesis):

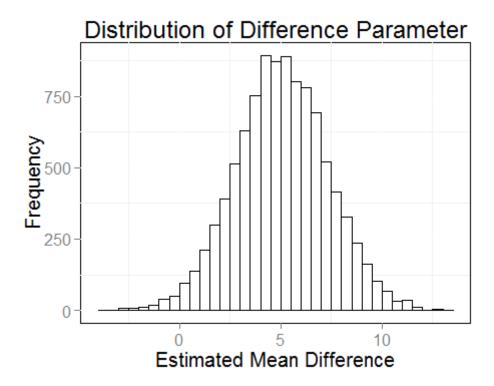
 $Pr(Hypothesis|Data) \propto Pr(Data|Hypothesis) Pr(Hypothesis)$

Hence, what a Bayesian analysis does is estimating how likely your hypothesis is, from your data, weighted a little bit with your assumptions. This "little bit" depends on the certainty of your

assumptions: If you have strong assumptions and are quite sure about potential outcomes, you should specify an "informative" prior which will more strongly influence the result. Strong assumptions can for example be based on strong theory, or prior data that have been collected. If however your assumptions are weak, for example because you are the first to research a topic and there are no previous data available to base assumptions on, you should specify a "non–informative" prior which will only influence the result to a negligible extent.

The prior is a critically discussed and for many people strange facet of Bayesian statistics. In Bayesian analysis, the prior is mixed with the data to yield the result. This means if two people have different assumptions about potential effects, they might specify different priors and hence yield different results from the same data. Is it legitimate that subjective assumptions influence the results of statistical analysis? Opponents of Bayesian statistics would argue that this inherent subjectivity renders Bayesian statistics a defective tool. Proponents however see priors as a means to improve parameter estimation, arguing that the prior does only weakly influence the result and emphasizing the possibility to specify non–informative priors that are as "objective" as possible (see Zyphur & Oswald, in press).

In Bayesian statistics there are three parts of an analysis: The prior, the likelihood, and the posterior. First, you specify the prior. As explained, the prior represents your assumptions about how large a potential difference between the two groups might be and how sure you are about it, translated into a statistical distribution. The likelihood are your data. In a Bayesian *t*-test these two, your assumptions and the data, are translated into the posterior. The posterior is your result, a statistical distribution that shows you the magnitude of the difference between the two groups (the mean or median of the distribution) and how sure you can be about the difference (the variance of the distribution). Let's compute a Bayesian t-test and look at the posterior distribution. Due to huge computational complexity, a posterior distribution cannot be fully computed. Instead, we draw single values from the distribution many times. By summarizing and plotting the single draws we get a good approximation of what the distribution looks like.



As a point estimate of the group difference in creativity, we can use the mean value of the distribution.

mean(beta) # mean difference creativity score

As a result, the estimated mean difference in the groups' creativity is 5.06 points on the scale of the creativity-test.

Further Advantages of Bayesian Statistics for Students' and Researchers' Everyday Life

Apart from increased conceptual clarity, Bayesian statistics implies various further advantages over frequentist statistics.

Alpha-adjustement. In frequentist statistics, when someone conducts more than one analysis on the same data, they need to apply alpha-adjustment. This means that in order to avoid increased frequency of false rejections of the null hypothesis, data have to speak against the null more strongly in each additional analysis one applies.

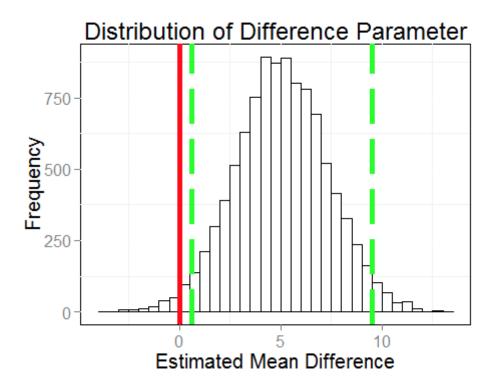
In comparison, in Bayesian testing there is no need for alpha adjustment (Dienes, 2011; Kruschke, 2010). One can conduct analysis on a data set and draw resulting inferences as many times as they want, without risking increased likelihood of false conclusions.

Interpretation of confidence. A common misconception about frequentist statistics concerns the interpretation of confidence intervals. Confidence intervals are used to depict how sure one can be about the estimate of an effect. For the difference in the two groups' creativity, our frequentist

t-test showed us a confidence interval of $\text{Cl}_{95}[0.86, 9.92]$. Many people, even experienced researchers, think of this as implying that we can be 95% sure that the true difference between the groups is in the range of 0.86 to 9.92 points. This and other misconceptions about confidence intervals are similar to misconceptions about p values, and they are common, even amongst experienced researchers (Hoekstra, R. D. Morey, Rouder, & Wagenmakers, 2014). Indeed, the CI only tells us that "if we draw samples of this size many times, the real difference between the groups will be within the CI in 95% of cases". Hence, the CI does not tell us anything tangible. Rather, the probability that the true difference lies within these borders is either 0 or 1, because it is either in there or not!

The Bayesian equivalent to confidence intervals is the credibility interval (see Kruschke, 2010). We depict the credibility interval for our example.

```
##Credibility interval
CredInt = quantile(beta,c(0.025,0.975)) #Credibility interval for the difference betwee
CredInt
##Visualize credibility interval in the histogram
hi + geom_vline(xintercept=CredInt[1],color ="green", linetype = "longdash", size = 2)
    geom_vline(xintercept=CredInt[2],color ="green", linetype = "longdash", size = 2) + #
    geom vline(xintercept=0, color = "red", size = 2) #line at zero difference
```



The figure depicts the Bayesian credibility interval (green lines) and the zero-difference location (red line). As depicted by the green lines, of the 1000 values that we drew from the posterior, 95% lie within 0.62 and 9.50. This is the credibility interval for the difference between the two groups' creativity. In comparison to frequentist confidence intervals, the interpretation of this credibility interval is easy and intuitive: We can be 95% sure that the difference between the groups lies between 0.62 and 9.50 points on the scale of the creativity-test. Hence, while this

interval is very similar to that from the frequentist analysis, it tells a different, more satisfying story.

Conclusion

A remark regarding Bayesian statistics remains: Some aspects of Bayesian analysis are complex. Recently, some good introductions to Bayesian analysis have been published. For example, Kruschke (2014) offers an accessible applied introduction into the matter. In addition, frequentist analysis can also be complex and difficult to comprehend. At least the analyzed model is always the same: There are no "Bayesian models" or "frequentist models" in statistics, but only different ways to analyze a model. Hence, in our example we analyzed the same t-test model twice, once using frequentist analysis and then using Bayesian analysis. Therefore, there is no reason to be daunted by "Bayesian models", but as discussed there are many reasons to learn and enjoy Bayesian analysis!

For more than 50 years Bayesian statistics has been advocated as the right way to go (for a classic account see Edwards, Lindman, & Savage, 1963). The computational complexity of Bayesian statistics has been a major obstacle for its application. Modern computational power could overcome this issue several years ago but frequentist statistics used this time lag to burn into researchers' minds. Hopefully, this introduction managed to free your mind and evoke your interest in Bayesian statistics. A good way to deepen your understanding is to engage in fruitful exchange with your colleagues, read into the suggested literature, and visit some courses. We are planning to provide you with further tutorials on Bayesian data analysis in the JEPS Bulletin, to support your change to the Bayesian side!

Suggested Readings

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference.* Basingstoke: Palgrave Macmillan.

Kruschke, J. K. (2014). *Doing bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. (2013). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*, 841–860. doi:10.1111/cdev.12169

References

Dienes, Z. (2011). Bayesian versus orthodox statistics: which side are you on? *Perspectives on Psychological Science*, *6*, 274–290. doi:10.1177/1745691611406920

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference in psychological research. *Psychological Review*, *70*, 193–242. doi:10.1037/h0044139

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, 1-8. doi:10.3758/s13423-013-0572-3

Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 658–676. doi:10.1002/wcs.72

Kruschke, J. K. (2014). *Doing bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

Rouder, J. N., Wagenmakers, E.-J., Verhagen, J., & Morey, R. (submitted). The p < .05 rule and the hidden costs of the free lunch in inference. Retrieved from http://pcl.missouri.edu/node/145

Zyphur, M. J., & Oswald, F. L. (in press). Bayesian estimation and inference: A user's guide. *Journal of Management*. doi:10.1177/0149206313501200

About the author

Peter is currently doctoral student at the section for learning and instruction research of ETH Zurich in Switzerland. He graduated from Psychology at the University of Graz in Austria. Peter is interested in conceptual knowledge development and the application of flexible mixture models to developmental research. Since 2011 he has been active in the EFPSA European Summer School and related activities.





Share this:



Related Posts:

1. Bayesian Statistics: Why and How

Posted in <u>Discussion</u>, <u>How-to</u>, <u>Research Methodology</u> on <u>November 17</u>, 2014 by <u>Peter Edelsbrunner</u>. 16 Comments



Sort by Best ▼



Join the discussion...

YBA • 5 days ago

Thank you for this post and the example. I can run your example fine but when I tried using ttestBF on my data, I got this error:

Error in is.infinite(c(x, y)):

default method not implemented for type 'list'

In fact, I also get that error if I run this very simple code:

why <- data.frame(sensitivity = c(5, 6, 7, 7, 8, 9), perspective = c(1, 1, 1, 0, 0, 0)) ttestBF(sensitivity ~ perspective, data = why)

The usual t.test runs fine:

t.test(sensitivity ~ perspective, data = why)

What am I doing wrong?

```
∧ V • Reply • Share >
```

Maria Teresa Stella · 2 months ago

A question: is the Bayesian method already being used for data analysis in recent papers? Or are we the "new generation" to make this method more popular? I have recently read about all of the downsides of Null Hypothesis Significance Testing (NHST) but I don't know that much yet about this new approach:)

Peter A Edelsbrunner → Maria Teresa Stella • 2 months ago

Hi Maria.

there is a recent review of the use of Bayesian methods in Psychology: https://rensvandeschoot.files....

Result: It is increasing quickly!

Generally, my impression is that for example in Cognitive Psychology, Bayesian methods are not uncommon, while in other fields such as Educational Psychology I do not know any published work making use of Bayes.

I think that in about three years, in each subfield of Psychology Bayesian articles will have been published.

If you want to check out some more Bayes-literature I suggest this recent overview:

https://nicebrain.files.wordpr...

And yes, I think you and me are the generation to make this not only popular, but the standard.

Cheers,
Peter

1 ^ V · Reply · Share ›

Maria Teresa Stella → Peter A Edelsbrunner • 2 months ago

Thank you very much Peter! I am working on my final project about the negative aspects of Null Hypothesis Significance Testing and I wanted to put Bayesian methods as a better alternative but I didn't really know where to look for additional materials:)

∧ V • Reply • Share >

Peter A Edelsbrunner → Maria Teresa Stella • 2 months ago

Cool!

For a "positive" view on NHST, or frequentist approaches in general, I suggest you to look at Daniel Lakens' work! For example about sequential testing he has interesting things to say.

http://daniellakens.blogspot.c...

And:

There is a bit of a different third approach based on likelihoods, for this I suggest you to read Dienes' (2008) book "Understanding Psychology as a Science". It is a rather short book and gives nice views on all three approaches (frequentist, Bayes, likelihood).

1 ^ V · Reply · Share ›

Maria Teresa Stella → Peter A Edelsbrunner • 2 months ago

Thank you! The articles suggested by my professor reject NHST in quite a "drama-queen" style;) (for example http://ist-socrates.berkeley.e...

∧ ∨ • Reply • Share ›

Peter A Edelsbrunner → Maria Teresa Stella • 2 months ago

that's a classic:D

one more suggestion regarding NHST and Bayes: Wagenmakers 2007 "a practical solution for the pervasive problems of p-values"; a rather long article but a lot in there.

enjoy!

1 ^ V · Reply · Share ›

Maria Teresa Stella → Peter A Edelsbrunner • 2 months ago

Cool! Thanks for all of your extra info, it will be quite useful :D

∧ V · Reply · Share ›

Excellent and concise =)

Reply • Share



Peter Edelsbrunner • 2 years ago

Hi Jonathon,

thanks for your reply!

I fully agree with you! Good software such as yours (JASP) eases application of Bayesian statistics immensely for students who are not used to R, I find that really great!

JASP is accessible and fun to use (everybody try it out! http://jasp-stats.org) but meant for application. The more fun a software, the more it tempts to use models without proper understanding of their nature and assumptions, supporting faulty applications and interpretations of results. This is the fault of the "lazy" user, not of the software that aims at accessibility and application rather than flexibility and deep understanding.

JASP covers the applied part on the dimension towards understanding Bayes (but also increased complexity and inaccessibility):

Application: JASP <-> BayesFactor <-> BUGS <-> MCMC-sampler: Understanding

Using the BayesFactor package was useful for this primer. For students who are

see more

∧ V • Reply • Share ›



Jonathon Love • 2 years ago

Re: statistical blackbox

i think that using a graphical user interface like JASP, and using a highly abstracted package like the BayesFactor package are pretty similar from a 'black box' perspective (indeed, many of the options in JASP map directly to options in BayesFactor, so i don't think there's too much difference specifying a function argument vs selecting a checkbox). both of them conceal everything which is going on 'under the hood' -- which is good.

good abstraction is the goal of software design, making things as simple and accessible as they can be made; and the BayesFactor package does a great job, and JASP you can think of as a layer higher than that. as long as the user can drill down to the level of abstraction that they want to see; as they can with JASP and BayesFactor because they're open source, then everyone wins; both those with simple needs, and those who want to dive deeper.

i'm saying all this, because i think science needs to value accessibility and the user-interface, not poo-poo it.

http://jasp-stats.org

∧ V • Reply • Share ›



Peter Edelsbrunner • 2 years ago

Thank you everybody for the nice comments, and Angelos for the link! In general, I am not a fan of statistics software with graphical user interface (statistical blackbox - clicking without knowing what's going on) but I guess for the end user JASP is a great alternative; in some ways more user friendly than SPSS. I heard about plans for JASP three years ago and I am glad to see that it has developed a lot since then - props to the Bayesian crowd developing it!

@Ivan: Thanks! We'll see what we can come up with...:)

∧ V • Reply • Share >



Angelos Krypotos • 2 years ago

Excellent work! There is also a new software, available in alpha, that allows you to run simple Bayesian analyses using a Gui https://jasp-stats.org It is not as powerful as R, at least not for the time being, but it is quite good for simple analyses.

∧ V • Reply • Share >



Ivan • 2 years ago

Awesome stuff Peter! Is there a chance of us getting a series of posts like this? I bet you have a lot more to teach us.

∧ V • Reply • Share >



Nick Byrd • 2 years ago

This was very helpful, especially the R tutorials. I am familiar only with basic model based descriptive statistics, so whenever I hear "Bayesian analysis" I think, "That's probably too complicated for me." This post helped me think otherwise. Hopefully I can start practicing some Bayesian analyses so that I can use it with confidence on future datasets.

Also, I appreciate the citations amd suggested reading!

∧ ∨ • Reply • Share •



Niall Bourke • 2 years ago

Good stuff! Bayesian seems to be the sexy word of statistics these days. This was a very nice overview!

∧ V • Reply • Share >

ALSO ON BLOG.EFPSA.ORG

Editor's Pick: Our favorite MOOCs

1 comment • 4 months ago•

fdabl — great list! For materials on the more technical side of neuroscience, I found the following two courses ...

Structural equation modeling: What is it, what does it have in common ...

1 comment • 9 months ago•

Peter Edelsbrunner — A colleague who is expert for electronic music pointed me towards another similarity that I ...

Python Programming in Psychology -From Data Collection to Analysis

1 comment • 2 months ago •

Gavin Cooper — Nice write-up as an introduction that to my mind quickly shows the utility of python to ...

Collaborating With Researchers in Other Fields

1 comment • 10 months ago •

ivanflis - Hi Etien. Thanks for the text, it was an interesting read. I find it particularly interesting that you've ...

CATEGORIES

APA Guidelines

Authors' experience

Discussion

How-to

Interviews

JEPS

Literature research

Manuscript analysis

Open Access

Presenting results

Publishing in scientific journals

Research Methodology

Scientific writing

Uncategorized