More Quotes

- Bayes: Distributional + prior assumption
  Freq: Distributional + sampling dist assumption
  You don't need a prior to be 'true', you need it to be defendable. "Given this prior uncertainty, what do the data suggest?"
  Can you defend the existence of a sampling distribution? - Stephen Martin
- The thing is-Both frameworks can operate w/in counterfactual reasoning. "Assuming I am an extreme skeptic, this is what the data suggest", for example. The nice thing about Bayes is that the counterfactual reasoning is immediate, rather than dependent on samples you'll never see. - Stephen Martin

inference p-value likelihood RCT bayes multiplicity posterior drug-evaluation principles evidence hypothesis-testing 2017

Frank Harrell

Professor of Biostatistics

My research interests include Bayesian statistics, predictive modeling and model validation, statistical computing and graphics, biomedical research, clinical trials, health services research, cardiology, and COVID-19 therapeutics.

ALSO ON **STATISTICAL THINKING**

Road Map for Choosing Between …
4 years ago • 28 comments
When we raise money it's AI, when we hire it's machine learning, and …

A Comparison of Decision Curve …
8 months ago • 2 comments
Andrew Vickers Department of Epidemiology and Biostatistics Memorial …

Violation of Proportional Odds is …
2 years ago • 1 comment
Many researchers worry about violations of the proportional hazards …

Bayesian Address
2 years ago
The COVIE has elevate for designir

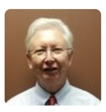
Statistical Thinking Comment Policy

Comments are welcomed. Be informed, informative, respectful, and criticize ideas, not people. Please read our Comment Policy before commenting.

Comments on this entire site are premoderated (only moderators can see this message). Change site settings.

43 Comments **Statistical Thinking** **Disqus' Privacy Policy** **Frank Harrell**

Favorite 1 Tweet Share **Sort by Best**



Join the discussion…

person • 2 years ago • edited

Hi! At the end of your article, you provide a summary of frequentist & bayesian analysis as follows:

> Frequentist = subjectivity1 + subjectivity2 + objectivity + data + endless arguments about everything
> Bayesian = subjectivity1 + subjectivity3 + objectivity + data + endless arguments about one thing (the prior)

Is it incorrect to conceptually think of subjectivity_2 & subjectivity_3 as the same thing? That is, is the *sampling distribution* technically the same thing as a *prior*?

The reasoning is: the sampling distribution describes the likelihood of obtaining a particular sample mean. A bayesian prior is often a distribution that describes a particular parameter (one of which could be the sample mean, right?).

Why is it incorrect to think of Frequentists' NHST as a prior distribution centered at some Null value?

Reply • Share ›

Frank Harrell Mod → person • 2 years ago

Thanks for the thoughts. I think that the sampling distribution (how data arose) is drastically different from a prior distribution (how parameter values, e.g., effects, arose). And we think of a sampling distribution as something that emanates from a single value of a parameter. To your last question, NHST does not use a prior but is related to the use a prior with a discontinuity (spike) at the null. Frequentists place special emphasis on zeros. Bayesians tend to not do that.

1 • Reply • Share ›

person → Frank Harrell • 2 years ago

Thank you for the response; that's a lot more clear.

Awesome article!

Reply • Share ›

Christian Hinze • 5 years ago

Could you please give an example of a clinical study analysis in a frequentist manner vs. a Bayesian? Also, could you show examples of "I found that with MCMC simulation of Bayesian posterior draws I could quite simply compute probabilities such as P(any efficacy), P(efficacy more than trivial), P(non-inferiority), P(efficacy on endpoint A and on either endpoint B or endpoint C), and P(benefit on more than 2 of 5 endpoints). " ? Thanks again!

Reply • Share ›

Frank Harrell • 5 years ago

No, posterior probabilities factor in all uncertainties and are self-contained. But if you want to discuss uncertainties in a parameter, that is represented by the entire posterior distribution.

Reply • Share ›

Randy Collica • 5 years ago

One disadvantage of a Bayesian approach is that it doesn't give you an estimate of error. Can this be accomplished simply by applying bootstrap methods to obtain a confidence level to the posterior probabilities?

Reply • Share ›

Donald Williams • 5 years ago

realized previous post misspelled your name (apologies)

Hello Deborah: I am not sure what is meant by "...required "the possibility of infinitely many repetitions of identical experiments"." The usefulness of frequentist statistics, as I see it, is the emphasis on how ones model does perform over the long run given a set of assumptions (that may or may not be realistic). As such, I do not see this as much of a problem, so long as we do understand that the estimates (e.g., errors) are derived from many assumptions that probably do not generalize one-to-one to "real" world settings. That said, they can be very useful, and at minimum, we should ensure our model is has optimal calibration (whether this is obtained from error rates, coverage, parameter bias,..etc.).

The estimates, however, are indeed computed from sampling a population with a set of characteristics many (many) of times. Infinite would be nice, but really we need about 5,000 to 10,000 repetitions to get an estimate that is stable. Of course, we can then vary population values to see what could happen over the long run under different assumptions. So, I really do not think the argument against repetition is very useful, as that is just how error (type one, or bias) and power are estimated. This is just the way it is, and may or may not makes sense, but is useful (IMO) so long as we do not take our model too seriously (they are all wrong after all!).

I prefer Bayesian statistics, but I too study long run properties of my models. I do not really see any other way to see how a give model is performing. Posterior predictive checks are also useful, and often provide similar inferences as simulations. For example, exploring variances of each group can show misfit when assuming equal variances, and over the long run this misfit results in an inflated error rate.

Furthermore, if confidence and credible intervals are basically equivalent with a so-called "uniformative" prior, it follows that Bayesian have, at minimum, expected error rates. One can not have equivalent intervals without this being the case! Once said, notion probably "feels" like common sense. With a prior centered at zero, and is informative, this will just make power lower than a sense. With a prior centered at zero, and is informative, this will just make power lower than a frequentist estimate, so do not see how errors are not controlled. If anything, the Bayesian model can be considered sub-optimal (in NHST framework), since it can provide conservative estimates. This is the kind of Bayesian statistics that I use.

Now, often Bayesians do not focus on error rates, and would prefer to have a model that best describes the data generating process. This approach often leads to controlling error rates, optimal power, among other things. This occurs as a by-product, so to speak, from focusing on modeling the data.

Where Bayesian methods really shine is that we do not need to come up with different ways to estimate the standard error, or ways to approximate the degrees of freedom to obtain reasonable inferences. For example, even to accommodate unequal variances, a Welch's t-test resorts to approximating the degrees of freedom for the sampling distribution of the t-statistic. In a multilevel framework, the sampling distribution is entirely unknown, but people have figured out approximations that ensure optimal error rates. Indeed, some would even say that exact p-value do not exist (so much for exact error rates :-)) In contrast, whether comparing two group or multilevel with varying slopes and interecepts, Bayesian methods do not depend on a known sampling distribution and everything is estimated much the same (Yes, even NHST error rates are obtained!). It is actually quite elegant!

D

Delete

| • Reply • Share ›

Frank Harrell • 5 years ago

I resonate with Donald's comments on these points and don't see justification for some of Deborah's. Writing simulation pseudo-code will expose many of the issues properly. I don't need to show long-run operating characteristics to show that Bayesian methods optimize the probability of making the right inference for a given set of data. True I need a large number of simulated clinical trials to demonstrate perfect calibration of Bayesian posterior probabilities, but these simulations are made under an entire array of treatment effects not for one single effect as with frequentist methods.

| • Reply • Share ›

Donald Williams • 5 years ago

This comment has been removed by the author.

| • Reply • Share ›

Deborah Mayo • 5 years ago

Error statistics (neiterh Fisherian nor N-P style) never required "the possibility of infinitely many repetitions of identical experiments". That's absurd. When people complain about cherry-picking, p-hacking, optional stopping, data-dependent endpoints, etc. it's because they prevent a stringent test in the case at hand. The appeal to the "ease" (frequency) of producing impressive-looking results, even under Ho, only alludes to hypothetical possibilities (nor need they be identical). Such appeals are at the heart of criticisms of bad statistics and bad science. Unless your Bayesianism takes account of "what could have occurred but didn't" I fail to see your grounds for caring about preregistration, RCTs, etc. You seem to have boxed yourself into an inconsistent position--and I don't know what kind of priors you favor-- based on a mickey-mouse caricature of hypotheses tests.
On the other hand, if your Bayesian does consider what could have occurred--counterfactual reasoning that we can simulate on our computers today--then you can't say such considerations are irrelevant.

Frequentists also estimate, and any statistical inference can be appraised according to how well or stringently tested claims are (that's just semantics).

| • Reply • Share ›

Donald Williams • 6 years ago

I am not really sure there are deep problems with confidence intervals.

I use both Bayesian and frequentist in simulation studies, but only Bayesian for analyzing "real" data. That said, just because people misinterpret something does not mean it is bad. With this logic, almost all things in life have deep issues. I find confidence intervals useful, although I do not think that they necessarily generalize exactly to actual research situations. Exploring long-run outcomes, CI coverage, bias,..etc provide useful information, IMO. The problem as I see it, is individuals not realizing the limitations of models, frequentist, Bayesian, or agent based.

| • Reply • Share ›

Donald Williams • 6 years ago

I am not really sure there are deep problems with confidence intervals.

I use both Bayesian and frequentist in simulation studies, but only Bayesian for analyzing "real" data

I use both Bayesian and frequentist in simulation studies, but only Bayesian for analyzing "real" data. That said, just because people misinterpret something does not mean it is bad. With this logic, almost all things in life have deep issues. I find confidence intervals useful, although I do not think that they necessarily generalize exactly to actual research situations. Exploring long-run outcomes, CI coverage, bias,..etc provide useful information, IMO. The problem as I see it, is individuals not realizing the limitations of models, frequentist, Bayesian, or agent based.

Reply • Share ›

Frank Harrell • 6 years ago

I was speaking more of the problem with statisticians and stat grad students understanding the concept. If after multiple attempts at understanding a primary concept in a paradigm one has to give up, there is a problem with the paradigm.

Reply • Share ›

a.foss • 6 years ago

Thank you for a very interesting and informative post. One oft-repeated argument against the frequentist perspective that has never resonated with me is the fact that CI are hard to explain. In most (but not all) cases I'm interested in selecting the technique that will maximize my chance to deliver correct conclusions, regardless of how hard it would be for a collaborator to understand the statistical methods I used. Should I avoid using MCMC because my collaborators can't understand the technique?

I may be at risk of attacking a straw-man here, because of course there are deep philosophical/statistical problems with CI. But that's kind of my point -- isn't it best to focus on these fundamentally problematic aspects rather than didactic issues?

Reply • Share ›

Donald Williams • 6 years ago

Hi Frank:
I basically agree with everything you have said here. The trimmed means is far from intuitive. I did some simulations with the trimmed means to see how
a determined researcher can "find" significance. Basically, it reduces
to a multiple comparisons problem. However, even with the exact same data
but using different thresholds to trim, inflates the error rate almost
0.05 * the number of tests (assuming, on average, no difference between
group). Lots up researchers degrees of freedom with this approach, and others
(winsorizing).

Reply • Share ›

Frank Harrell • 6 years ago

Very interesting Donald, and do call me Frank. I would emphasize mean squared and mean absolute estimation errors, probably. I don't find trimmed means satisfactory because I'm unable to define to a collaborator what they mean prospectively. With Bayes you can estimate the mean or quantiles of the raw data distribution, or better, estimate the whole distribution.

Reply • Share ›

Frank Harrell • 6 years ago

Thanks to all 4 of you for pointing us to some excellent resources.

Reply • Share ›

Greg Gandenberger • 6 years ago

I have a set of blog posts that are intended to provide an accessible introduction:

- http://gandenberger.org/201...
- http://gandenberger.org/201...
- http://gandenberger.org/201...

Reply • Share ›

Greg Gandenberger • 6 years ago

This comment has been removed by the author.

Reply • Share ›

Greg Gandenberger • 6 years ago

This comment has been removed by the author.

Reply • Share ›

Greg Gandenberger • 6 years ago

This comment has been removed by the author.

Reply • Share ›

Donald Williams • 6 years ago

Hi Frank (I hope that is OK). I (with a collaborator) am currently working on a paper "introducing" a Bayesian heteroscedastic skew-normal model. We are basically characterizing parameter bias, error rates, and power, all while estimating the degree of skew as well as sigma for each group. (yes, error rates. I don't really agree with type one error, but still find that is is useful so long as we understand its limitations).

Interestingly, for those not wanting to learn Stan or Bayesian methods, I would loosely advocate the trimmed means approach as long as the degree of trim was not used to get a significant p-value. Counter to my original thoughts, the trimmed means approach actually does perform rather well.

As you said, the Bayesian framework is much more interpretable (IMO) and does not entail directly altering the data. Furthermore, the Bayesian approach actually provides estimates for skew, sigma, etc, which are important for prospective power analyses. Here, it also becomes clear how important priors can be! Not that they influence the final estimates all that much (although they can), but that the model needs them to converge.

Finally, if one does learn a general Bayesian approach, it will likely fulfill all their needs. This mean no jumping around packages or functions to get the "correct" estimate of standard error. Generalized estimating equations are a good example: there are many bias corrections (small N, etc.) for SE that it can make ones head spin (all resulting in slightly to drastically different p-values).

Reply • Share ›

Frank Harrell • 6 years ago

Nicely put Donald. Another example of flexibility that is worked out in detail in the Box and Tiao book is having a parameter specifying the degree of non-normality of the data, and having a prior for that. They show how this leads to something that is almost the trimmed mean but which is much more interpretable.

Reply • Share ›

Donald Williams • 6 years ago

This comment has been removed by the author.

Reply • Share ›

Donald Williams • 6 years ago

Hi John: Differences in interpretation aside, I see the greatest benefit of using Bayesian methods as flexibility. For example, fitting a skewed normal model, in which sigma and the skew are estimated for each group.This can also easily be extended to a multilevel framework. As far as I know, this is not currently possible in a frequentist framework.

Finally, I see the use of so called noninformative prior as not very Bayesian. We generally can rule out effects greater than d of 1, if not less.

In sum, the benefits of Bayesian are only fully realized, IMO, when one sees the benefits of informative prior (especially in MLM) and the great flexibility offered in Rstan and brms..etc.

Reply • Share ›

John K. Kruschke • 6 years ago

Agreed. Moreover, a Bayesian perspective at the design stage goes even further: It incorporates uncertainty into the research hypothesis, instead of assuming a specific effect size (or small set of candidate effect sizes) as is traditionally done in a frequentist approach to design.

Reply • Share ›

John K. Kruschke • 6 years ago

For a few simple side-by-side comparisons of Bayesian and frequentist, for hypothesis testing and parameter estimation, see the article linked in this blog post: http://doingbayesiandataana...

Reply • Share ›

John K. Kruschke • 6 years ago

Here's an introductory article focused exactly on putting frequentist and Bayesian side by side, for both hypothesis testing and parameter estimation. Links at this blog post:
http://doingbayesiandataana...

Reply • Share ›

John Kapson • 6 years ago

Check out this YouTube lecture series by Richard McElreath. It parallels his outstanding book, "Statistical Rethinking."

Statistical Rethinking - Lecture 01

see more

Reply • Share ›

John Kapson • 6 years ago

Ditto on "Statistical Rethinking." One of the clearest explanations of the Bayesian approach I've seen.

Reply • Share ›

Ahmed Kachkach • 6 years ago

I found these series of courses quite useful as an introduction to Bayesian statistics! There's a couple videos that show the difference between a Frequentist and a Bayesian approach.

Bayesian statistics syllabus

see more

Reply • Share ›

Frank Harrell • 6 years ago

Nicely put. A big selling point of Spiegelhalter's work is that he doesn't preach but shows the problem-solving power of Bayes with real examples. One problem with considering type I error even only pre-study is that it can be really hard to define and the sample space can be complex. I tend to come at this from the standpoint of planning around how much information will result, quantified e.g. by width of 0.95 credible interval, or by using the probability that posterior probability will exceed some high number for a distribution of true unknown effects.

Reply • Share ›

Roy Tamura • 6 years ago

Nicely written. A long time ago, I heard Don Berry say that frequentist concepts like Type 1 error, coverage probability, power, randomization, were very valuable at the design stage of experimentation. But that Bayesian summaries of the data at the end of the study were superior for analysis and interpretation of the data. That made sense to me. I think it is regrettable that there was so much vitriol and attacks made against both camps in the formative years of statistics.

Reply • Share ›

d.q. • 6 years ago

Indeed, kids should learn from Bayes pre-school.

Let them elevate their degrees of belief as they grow older :)

Reply • Share ›

Joachim Vandekerckhove • 6 years ago

Zoltan Dienes has a number of examples like that in an upcoming issue of Psychonomic Bulletin and Review. Stay tuned! :)

^ | v • Reply • Share ›

Frank Harrell • 6 years ago

Start with https://cran.r-project.org/...

But I hope that a reader will tell us of a place where there are multiple side-by-side analyses. I've been looking for that.

^ | v • Reply • Share ›

Learner • 6 years ago

Is there a tutorial or beginner level material that compares and contrasts frequentist and Bayesian methods. It will be good to take the same problem and show how the two disciplines approach it. More importantly, as you seem to have done in your own journey, get to the core of the matter and see where they are different. Sometimes use of a name in common parlance (like probability) by the two approaches may be different which may not be apparent unless looked at with a lot of rigor by a researcher like you

^ | v • Reply • Share ›

Frank Harrell • 6 years ago

Yes it's worthy of much more discussion. I try to come at this from a simpler perspective: do posterior probabilities remain well calibrated when used to trigger stopping a study where doing extremely frequent looks at the data (they do). In the future I'll blog about how to do simple simulations that demonstrate such mathematical necessities. A key issue underneath this is getting investigators and reviewers to agree on a choice of prior up front, or at least not having investigators use priors that are very inconsistent with those of reviewers.

^ | v • Reply • Share ›

Frank Harrell • 6 years ago

I just started reading it and like it very much. It's highly recommended by Andrew Gelman. Full name is Statistical Rethinking: A Bayesian Course with Examples in R and Stan.

^ | v • Reply • Share ›

Unknown • 6 years ago

I'm just an amateur. But, with regards to "defective concepts", I'm troubled by the strong likelihood principle. It looks like Birnbaum proved it a long time ago from two reasonable principles, and it is inconsistent with commonly used significance tests.

When I google for it, all I find is that Deborah Mayo disputes it, and there's a new proof of it by Gandenberger. I'm a bit surprised that it doesn't get more mention or attention since there are a lot of statisticians like you that express concerns.

^ | v • Reply • Share ›

Donald Williams • 6 years ago

Another great book is Rethinking by Richard Mcelreath.

^ | v • Reply • Share ›

Meyer Katzper • 4 years ago

link failed - books recommended for those without advanced statistics background see this Suggest fixing.

^ | v 1 • Reply • Share ›

Frank Harrell Mod → Meyer Katzper • 4 years ago

This is a current bug in citeulike.org which was reported 2018-06-14. I hope it's fixed soon.

^ | v • Reply • Share ›

Related

- [Musings on Multiple Endpoints in RCTs]
- [Bayesian vs. Frequentist Statements About Treatment Efficacy]
- [Statistical Errors in the Medical Literature]
- [p-values and Type I Errors are Not the Probabilities We Need]
- [EHRs and RCTs: Outcome Prediction vs. Optimal Treatment Selection]