

## 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.](#)

19 Comments

Statistical Thinking

Disqus' Privacy Policy

Frank Harrell ▾

Favorite

Tweet

Share

Sort by Best ▾



Join the discussion...



**Frank Harrell** • 6 years ago

There are several issues with that including my not believing that the problem is discrete, i.e., I want to put a smooth prior on a parameter associated with H, not a point mass of probability for P(H).

^ | ▾ • Reply • Share >



**Siddhartha** • 6 years ago

We know by Bayes rule that  $P(H|D) = P(D|H) \cdot P(H) / P(D)$ . So assuming that P(D) is reasonably high (so the data is "typical"), and given that  $P(H) \leq 1$  always, wouldn't knowing P(D|H) basically provide an upper bound on P(H|D)?

^ | ▾ • Reply • Share >



**Frank Harrell** • 6 years ago

Coming from you that's wonderful Gerd. Lots of help needed, especially educational strategies, workshops, papers, clinical examples, open source clinical trial data that can be re-analyzed the Bayesian way, etc. My initial attack is showing the type I errors were a poor choice of what to emphasize all along.

^ | ▾ • Reply • Share >



**Gerd Rosenkranz** • 6 years ago

Would be glad if I could help in this matter.

^ | ▾ • Reply • Share >



**Frank Harrell** • 6 years ago

How strange the results are if nothing is going on to me is useful only as a last-ditch method when Bayesian results are not available or we are in a big hurry. "How strange" is context dependent, not usually well defined, and is not a real evidentiary measure. Also, formally speaking the p-value is not a measure of how strange our results are but is the chance of results being stranger than ours if nothing is going on (and if 2-tailed is not even that simple). This is a fine distinction but helps to point out limitations in the overall approach.

^ | ▾ • Reply • Share >



**Martin** • 6 years ago

It is very clear now. Thank you.

^ | ▾ • Reply • Share >



**Michael Lew** • 6 years ago

Surely we are mostly using statistical methods for inductive inference. In that case the "when certainty holds" is an irrelevancy. Deductive inference (where certainty holds) needs no statistics and yields non-probabilistic outcomes. The example seems to be mixing them together in an inappropriate manner.

You should remember that all statistical analyses take place within a statistical model of some sort. That model is an important context. Any relevant context outside the model is important for the step between statistical analysis and scientific inference.

I do not disagree with the notion that estimation of effect sizes is more generally useful than a P-value. However, I note that the two are not mutually exclusive and I assert that a P-value helps the investigator to know how strange the result would be if there was really nothing going on. That is sometimes useful.

^ | ▾ • Reply • Share >



**Frank Harrell** • 6 years ago

I'd like to hear Cohen's take on that. I think that Pollard, Richardson, and Cohen are trying to make a general point that proof by contradiction is only completely well founded in logic when certainty holds.

To the case in statistical inference, we need to talk about the probability of getting results more extreme than that obtained if  $H_0$  were true, i.e., of the evidence we've assembled against  $H_0$  and why that would be of any interest to someone who wants to quantify evidence about a nonzero effect. And the notion of "getting results more extreme" is actually ill-defined and completely context-dependent.

^ | v • Reply • Share >



**Michael Lew** • 6 years ago

The Pollard and Richardson example is very misleading as it implies that a procedure that will occasionally yield a mistaken inference is a bad procedure. By that standard ALL statistical procedures are flawed.

The logic of P-values as evidence that Fisher explained includes the step that either a rare event has happened or the null is false. In the example the 'observation' is precisely that rare event. I would be concerned if the procedure yielded false inferences when any of the common observations were made, but it doesn't. The example is misleading and should not be used.

^ | v • Reply • Share >



**Russ Lyons** • 6 years ago

Hi, Martin.

I'm not sure what your question is, but I'll respond to what I think it may be.

The example started with "If a person is an American, then he is probably not a member of Congress." Note that it uses the word "probably". What does that mean? I am making it explicit by rephrasing this as "If a person is selected (uniformly) at random from among all Americans, then s/he is probably not a member of Congress." In fact, the probability would be the number of Americans not a member of Congress divided by the number of all Americans.

Does that help?

--Russ

^ | v • Reply • Share >



**Martin** • 6 years ago

Hi.

I'm trying to follow your logic but I don't see how...

"This person is selected (uniformly) at random from among all Americans."

is the (null) hypothesis in this case.

Could you please elaborate some more?

Thanks.

^ | v • Reply • Share >



**Frank Harrell** • 6 years ago

I'll have more to report about that in the next year. I am developing material to motivate the use of Bayesian methods and to show how the indirect evidence provided by NHST and p-values is not what clinicians think it is and does not provide the needed protection.

^ | v • Reply • Share >



**Gerd Rosenkranz** • 6 years ago

I basically share the viewpoint that hypothesis tests may not deliver what would be desirable. I wonder what can be done from a practical perspective to convince scientists to stop using them. You mention in your bio that you are also working with the office of biostatistics at FDA. I would be curious whether you had a chance to address the NHST issue with them and if yes what their thoughts were.

^ | v • Reply • Share >



**Frank Harrell** • 6 years ago

This is a common misconception about Bayesian methods. When data are inadequate but prior information is available, Bayesian methods may be necessary to use. When there is no prior information, Bayesian approaches work just as well as frequentist methods in terms of precision and power but still have an interpretation advantage. The only situation I can think of where NHST might be OK is an "existence experiment", e.g., trying to gather evidence that extrasensory perception exists. This involves showing for example that one can guess which card I'm holding better than chance. But in the vast majority of situations I want to quantify evidence for effects, and Bayesian posterior probabilities are best at that. For example what is the probability that a drug lowers blood pressure more than control therapy? What is the probability that it beats control by more than 3

mmHg?

^ | v • Reply • Share >



**Marcos** • 6 years ago

It seems to me that the exemple is good to show that the logic does not work, but it does not represent how p-value is used because the alternative hypothesis would be "this person is not an American" and nobody would gather data on Congress membership to test such hypothesis.

^ | v • Reply • Share >



**Shravan Vasishth** • 6 years ago

Could you also post a comment on situations where NHST might be adequate? I have heard statisticians say the following: Bayesian methods are better when you have relatively little data but a lot of prior knowledge. When you have a lot of data, there isn't much gain in switching to a Bayesian approach.

PS I only use Bayesian approaches in my own work once I realized that p-values were not useful for anything I needed to do.

^ | v • Reply • Share >



**Russ Lyons** • 6 years ago

First, let me edit my comment after the fact: I should have written, "If I tell you that this person is a member of Congress." Second, I agree with your conclusion insofar as  $H_0$  should involve an introduction of randomness, whereas in practice there is often no such randomness (as in observational studies). That does make it hard to apply frequentist theory.

^ | v • Reply • Share >



**Frank Harrell** • 6 years ago

Thanks for the comment Russ. I think that Pollard & Richardson were trying to put this in plainer language. And perhaps the fact that you need to think so hard about what  $H_0$  is has something to do with the difficulties of frequentist inference.

^ | v • Reply • Share >



**Russ Lyons** • 6 years ago

Hi, Frank.

Your example from Pollard & Richardson, 1987, is a good one, but I think it does not illustrate the lesson the way they say. In this case,  $H_0$  should be "This person is selected (uniformly) at random from among all Americans." If I tell you that this person is American, do you believe that  $H_0$  is an accurate and complete description of how I chose this person? I think not; you would deem  $H_0$  unlikely, wouldn't you? I think the lesson here is that one needs to think carefully about what  $H_0$  means.

Best,  
Russ

^ | v • Reply • Share >