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.

17 Comments Statistical Thinking Disqus' Privacy Policy Frank Harrell

Favorite Tweet Share Sort by Best

Join the discussion…

I_love_han_hye_jin • a year ago

Can you suggesst some good book on probabilistic forcasting?

• Reply • Share ›

Frank Harrell Mod I_love_han_hye_jin • a year ago

Start with Nate Silver's The Signal and the Noise then https://hbiostat.org/rms

1 • Reply • Share ›

David Rosen • 2 years ago

Instead of the Brier Score itself, why not use R-Squared, which is just a scaled and offset version of the Brier Score (and so retains its property of being a strictly proper scoring rule) but it makes it more interpretable as a percentage where 100% is perfection and 0% means you have the same Brier Score as would a naive baseline model that always merely predicts the overall proportion of class 1 (and negative means even worse than this baseline model)

• Reply • Share ›

Frank Harrell Mod David Rosen • 2 years ago

Frank Harrell Mod ➤ David Rosen • 2 years ago

There have been some papers studying R^2 in the binary Y case and it does have a lot of advantages. Here's one good paper: http://hbiostat.org/papers/.... I think that DR Cox had a paper on the subject which I can't find right now.

Reply • Share ›

David Rosen ➤ Frank Harrell • 2 years ago

I'm talking specifically about what that paper calls R_{res}^2, which is 1 - Brier/Brier0 where Brier0 is the Brier score for a constant model that always predicts the overall proportion of class 1. The paper advocates something different -- a coefficient of discrimination. Similarly for the logarithmic loss L, I would instead use 1 - L/L0 where L0 is the logarithmic loss of the model that predicts the proportion of class 1. Do you know of any work that proposes or uses these?

Reply • Share ›

Frank Harrell Mod ➤ David Rosen • 2 years ago

There are many papers on the subject and this relates to an old manuscript of ours: http://hbiostat.org/papers/... which was used by http://hbiostat.org/papers/... . Many more papers are here: http://hbiostat.org/papers/...

Reply • Share ›

Ravi Kalia • 3 years ago

This is amazing. I think I came across something similar, based on Leo Breiman's work in his probabilist days.

https://github.com/project-...

Reply • Share ›

Shira • 3 years ago

Hi,

Thanks for the great post!

Just a question though, when using models which output probabilities that aren't calibrated (let's say Random Forests), do I need to calibrate them before applying the scoring rule. Or, can I just relate to the ranking of the outputs, and use without calibration?

Thanks!

Reply • Share ›

Frank Harrell Mod ➤ Shira • 3 years ago

This depends entirely on your goal. If you are making individual-level decisions then you typically need absolute accuracy (calibration curve = line of identify). If you are on the other hand trying to use resources wisely across units you can use the business marketing idea: create a *lift curve* so you can get the "biggest bang for the buck" by marketing to those most likely to purchase a product, for example. This does not require calibration accuracy but only validated predictive discrimination ability. Of course if the model is poorly calibrated you may not know the best cut on the lift curve. But if you have $100k to spend, the lift curve will tell you who to spend it on.

Reply • Share ›

Shira ➤ Frank Harrell • 3 years ago

Thanks so much for the quick and detailed answer!

Reply • Share ›

Frank Harrell • 5 years ago

Being improper accuracy scoring rules (being easily fooled by saying that a bogus model is optimum) precision, recall, sens, spec are all very problematic and should never have been used except in special cases of pattern recognition with exceedingly high signal:noise ratios. The fact that these measures reverse time/information flow is even more of a problem. For more see http://www.fharrell.com/201...

Reply • Share ›

Unknown • 5 years ago

I've used precision and recall (positive predictived value and sensitivity) and think it gives you a lot better idea about your positive cases.

Reply • Share ›

Lilly • 5 years ago

nice

Reply • Share ›

Donna M. Kruse • 5 years ago

Awesome post

Reply • Share ›

Frank Harrell • 5 years ago

The fact that sens and spec are used all the time does not have much to do with whether they were ever a good idea. My best advice for changing practice is to send clinicians Chapter 19 of Biostatistics for Biomedical Research available from http://biostat.mc.vanderbil... .

Reply • Share ›

顧進裕 • 5 years ago

Sensitivity and specificity are used in medical diagnoses all the time. Additionally, there are practically no Brier scores, pseudo-r-squared or c-indices for any diagnostic procedures for any diseases. How do we change this practice to help the physicians in diagnostic tasks by using the probability measures? Are there any practical and simple tools that we can use?

Reply • Share ›

顧進裕 • 5 years ago

This comment has been removed by the author.

Reply • Share ›

Related

- Is Medicine Mesmerized by Machine Learning?
- Clinicians' Misunderstanding of Probabilities Makes Them Like Backwards Probabilities Such As Sensitivity, Specificity, and Type I Error
- Classification vs. Prediction
- How Can Machine Learning be Reliable When the Sample is Adequate for Only One Feature?
- Statistical Errors in the Medical Literature