ON THE USE AND ABUSE OF P-VALUES

Giorgio Bedogni

www.giorgiobedogni.it

http://www.giorgiobedogni.it/tutorials/metodologia/pmetodologia.html

CARGO-CULT STATISTICS

Cargo-cult statistics and scientific crisis

Philip B. Stark is professor in the Department of Statistics and associate dean in the Division of Mathematical and Physical Sciences at the University of California, Berkeley.

Andrea Saltelli is an adjunct professor in the Centre for the Study of the Sciences and the Humanities at the University of Bergen, and a researcher with the Open Evidence Research group at the Universitat Oberta de Catalunya, Barcelona.

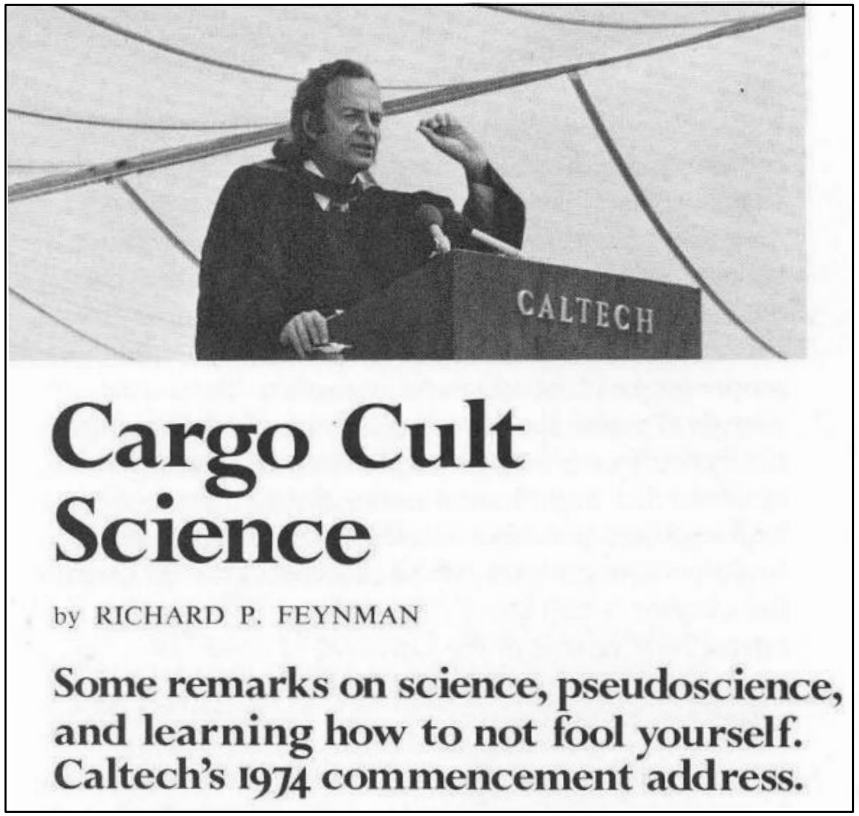
The mechanical, ritualistic application of statistics is contributing to a crisis in science. Education, software and peer review have encouraged poor practice – and it is time for statisticians to fight back. By **Philip B. Stark** and **Andrea Saltelli**

CARGO-CULT STATISTICS

«The problem is one of cargo-cult statistics – the ritual miming of statistics rather than conscientious practice»

Stark PB & Saltelli A. Significance. 2018;15:40.

CARGO-CULT SCIENCE



http://calteches.library.caltech.edu/51/2/CargoCult.htm

CARGO-CULT STATISTICS

«In our experience many applications of statistics are cargo-cult statistics: practitioners go through the notions of fitting models, computing p-values or confidence intervals»

Stark PB & Saltelli A. Significance. 2018;15:40.

CARGO-CULT STATISTICS

«They invoke statistical terms and procedures as incantations, with scant understanding of the assumptions or relevance of the calculations, or even the meaning of the terminology»

Stark PB & Saltelli A. Significance. 2018;15:40.

CARGO-CULT STATISTICS

«Statistical software enables and promotes cargo-cult statistics»

Stark PB & Saltelli A. Significance. 2018;15:40.

CARGO-CULT STATISTICS

- «The misuse of p-values, hypothesis tests, and confidence intervals might be deemed frequentist cargo-cult statistics. There is also Bayesian cargo-cult statistics»

Stark PB & Saltelli A. Significance. 2018;15:40.

On the use and abuse of p-values

Taylor & Francis
Taylor & Francis Group

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: http://amstat.tandfonline.com/loi/utas20

The ASA's Statement on p-Values: Context, Process, and Purpose

Ronald L. Wasserstein & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein & Nicole A. Lazar (2016) The ASA's Statement on p-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108

To link to this article: http://dx.doi.org/10.1080/00031305.2016.1154108

ASA position statement

- "The ASA Board was also stimulated by highly visible discussions over the last few years"

- "The ASA has not previously taken positions on specific matters of statistical practice"

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement

- “That turned out to be relatively easy to do [to find points of agreement], but it was just as easy to find points of intense disagreement”

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – What is a p-value?

- "Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data (e.g., the sample mean difference between two compared groups) would be equal to or more extreme than its observed value"

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 1

- "P-values can indicate how incompatible the data are with a specified statistical model. A p-value provides one approach to summarizing the incompatibility between a particular set of data and a proposed model for the data"

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 1

- "The most common context is a model, constructed under a set of assumptions, together with a so-called null hypothesis"

- "Often the null hypothesis postulates the absence of an effect, such as no difference between two groups, or the absence of a relationship between a factor and an outcome"

Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 1

- "The smaller the p-value, the greater the statistical incompatibility of the data with the null hypothesis, if the underlying assumptions used to calculate the p-value hold"

- "This incompatibility can be interpreted as casting doubt on or providing evidence against the null hypothesis or the underlying assumptions"

Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 2

- "P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone"

- "Researchers often wish to turn a p-value into a statement about the truth of a null hypothesis, or about the probability that random chance produced the observed data"

- "The p-value is neither. It is a statement about data in relation to a specified hypothetical explanation, and is not a statement about the explanation itself"

Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 3

- "Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold"

- "Practices that reduce data analysis or scientific inference to mechanical 'bright-line' rules (such as 'p<0.05') for justifying scientific claims or conclusions can lead to erroneous beliefs and poor decision making"

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 3

- “A conclusion does not immediately become “true” on one side of the divide and “false” on the other”

- “Researchers should bring many contextual factors into play to derive scientific inferences, including the design of a study, the quality of the measurements, the external evidence for the phenomenon under study, and the validity of assumptions that underlie the data analysis”

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 3

- "Pragmatic considerations often require binary, "yes-no" decisions, but this does not mean that p-values alone can ensure that a decision is correct or incorrect"

- "The widespread use of "statistical significance" (generally interpreted as 'p<0.05') as a license for making a claim of a scientific finding (or implied truth) leads to considerable distortion of the scientific process"

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 4

- "Proper inference requires full reporting and transparency. P-values and related analyses should not be reported selectively"

- "Conducting multiple analyses of the data and reporting only those with certain p-values (typically those passing a significance threshold) renders the reported p-values essentially uninterpretable"

Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 4

- "Cherry-picking promising findings, also known by such terms as data dredging, significance chasing, significance questing, selective inference, and 'p-hacking', leads to a spurious excess of statistically significant results in the published literature and should be vigorously avoided"

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 4

- "One need not formally carry out multiple statistical tests for this problem to arise"

- "Whenever a researcher chooses what to present based on statistical results, valid interpretation of those results is severely compromised if the reader is not informed of the choice and its basis"

- "Researchers should disclose the number of hypotheses explored during the study, all data collection decisions, all statistical analyses conducted, and all p-values computed"

Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 4

- "Valid scientific conclusions based on p-values and related statistics cannot be drawn without at least knowing how many and which analyses were conducted, and how those analyses (including p-values) were selected for reporting"

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 5

- "A p-value, or statistical significance, does not measure the size of an effect or the importance of a result"

- "Statistical significance is not equivalent to scientific, human, or economic significance"

- "Smaller p-values do not necessarily imply the presence of larger or more important effects, and larger p-values do not imply a lack of importance or even lack of effect"

Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 5

- "Any effect, no matter how tiny, can produce a small p-value if the sample size or measurement precision is high enough, and large effects may produce unimpressive p-values if the sample size is small or measurements are imprecise"

- "Similarly, identical estimated effects will have different p-values if the precision of the estimates differs"

Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 6

- "By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis"

- "Researchers should recognize that a p-value without context or other evidence provides limited information"

Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – Principle 6

- "For example, a p-value near 0.05 taken by itself offers only weak evidence against the null hypothesis"

- "Likewise, a relatively large p-value does not imply evidence in favor of the null hypothesis; many other hypotheses may be equally or more consistent with the observed data"

- "For these reasons, data analysis should not end with the calculation of a p-value when other approaches are appropriate and feasible"

Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – other methods

- “In view of the prevalent misuses of and misconceptions concerning p-values, some statisticians prefer to supplement or even replace p-values with other approaches”

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – other methods

- “These include methods 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”

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – other methods

- "All these measures and approaches rely on further assumptions, but they may more directly address the size of an effect (and its associated uncertainty) or whether the hypothesis is correct"

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

ASA position statement – conclusion

- No single index should be a substitute for scientific reasoning.

  Wasserstein RL & Lazar NA. Am. Stat. 2016;70:129.

You are invited to read this wonderful paper…

Eur J Epidemiol (2016) 31:337–350
DOI 10.1007/s10654-016-0149-3

ESSAY

Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations

Sander Greenland[1] · Stephen J. Senn[2] · Kenneth J. Rothman[3] · John B. Carlin[4] · Charles Poole[5] · Steven N. Goodman[6] · Douglas G. Altman[7]

Caveat

- "Analysis interpretation depends on contextual judgements about how reality is mapped onto the model and how the formal analysis results are mapped back to reality"

  Greenland S. Eur J Epidemiol. 2017;32:3-20.

ASA position statement - 1 year after

The ASA's *p*-value statement, one year on

Its aim was to stop the misuse of statistical significance testing. But **Robert Matthews** argues that little has changed in the 12 months since the ASA's intervention

38 | SIGNIFICANCE | April 2017

ASA position statement - 1 year after - Robert Matthews

- “Yet a year on, it is not clear that the ASA’s statement has had any substantive effect at all”

- “The commentaries accompanying the statement show that even at the time some participants feared there would be no lasting impact”

- “Goodman’s frustration with this resonates with mine: exactly how [are] scientists supposed to do that?”

Matthews R et al. Significance. 2017;14:38.

ASA position statement – 1 year after – – Ron Wasserstein*

- “The statement did not go as far as it should go, but it went as far as it could go”

- “To address the statement’s shortcomings the ASA is convening a Symposium on Statistical Inference this October”

*Executive Director, American Statistical Association

Matthews R et al. Significance. 2017;14:38.

ASA position statement – 1 year after – – David Spigelhalter*

- “If by this he [Matthews] means [to ban] all use of p-values than I must disagree”

- “Fortunately the Neyman-Pearson’s decision-theoretic idea of “accepting” the null has just been consigned to the overwflowing dustbin of inappropriate scientific ideas”

  Matthews R et al. Significance. 2017;14:38.

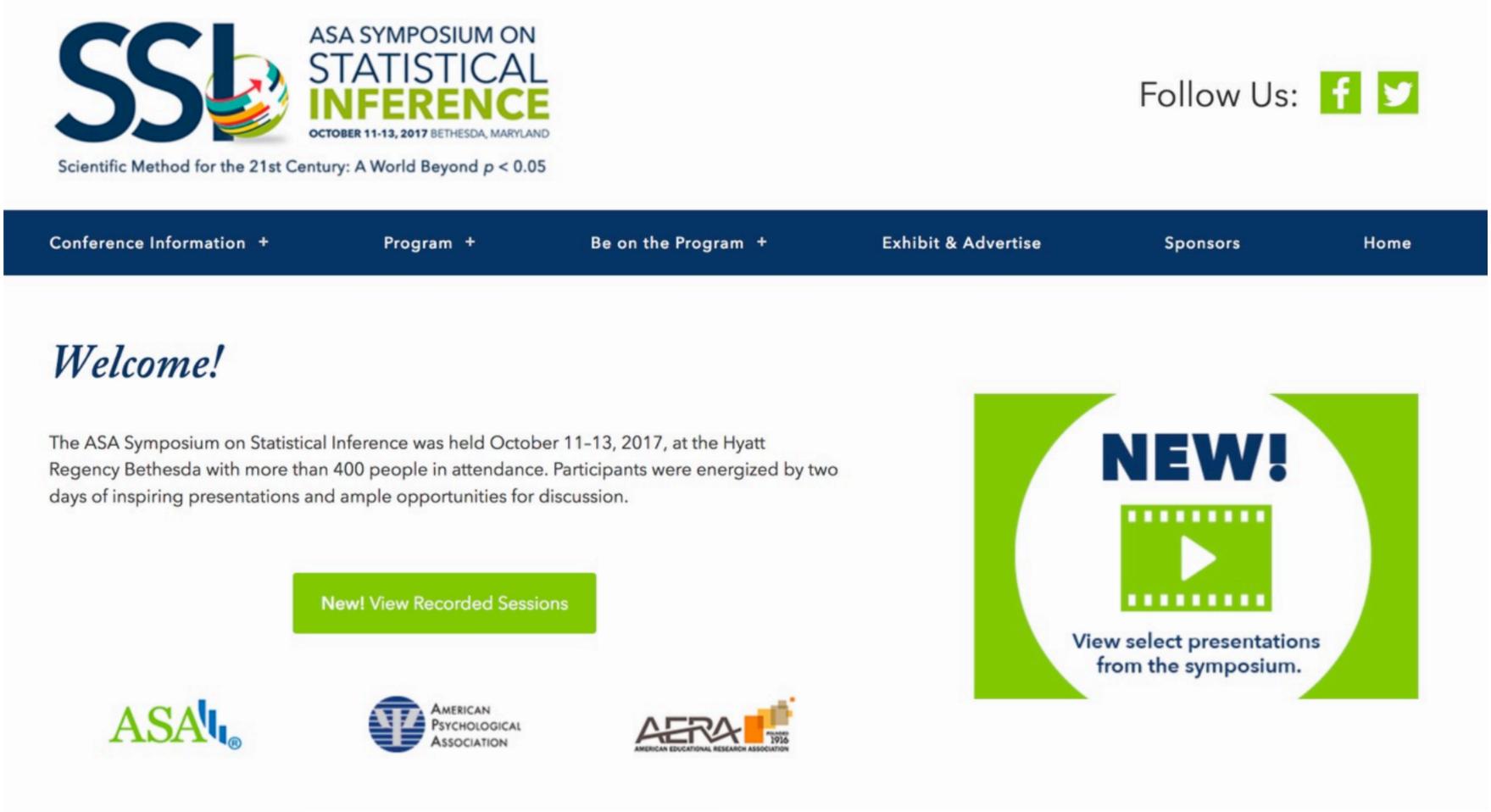
  *President, Royal Statistical Association

ASA position statement – 1 year after – – David Spigelhalter

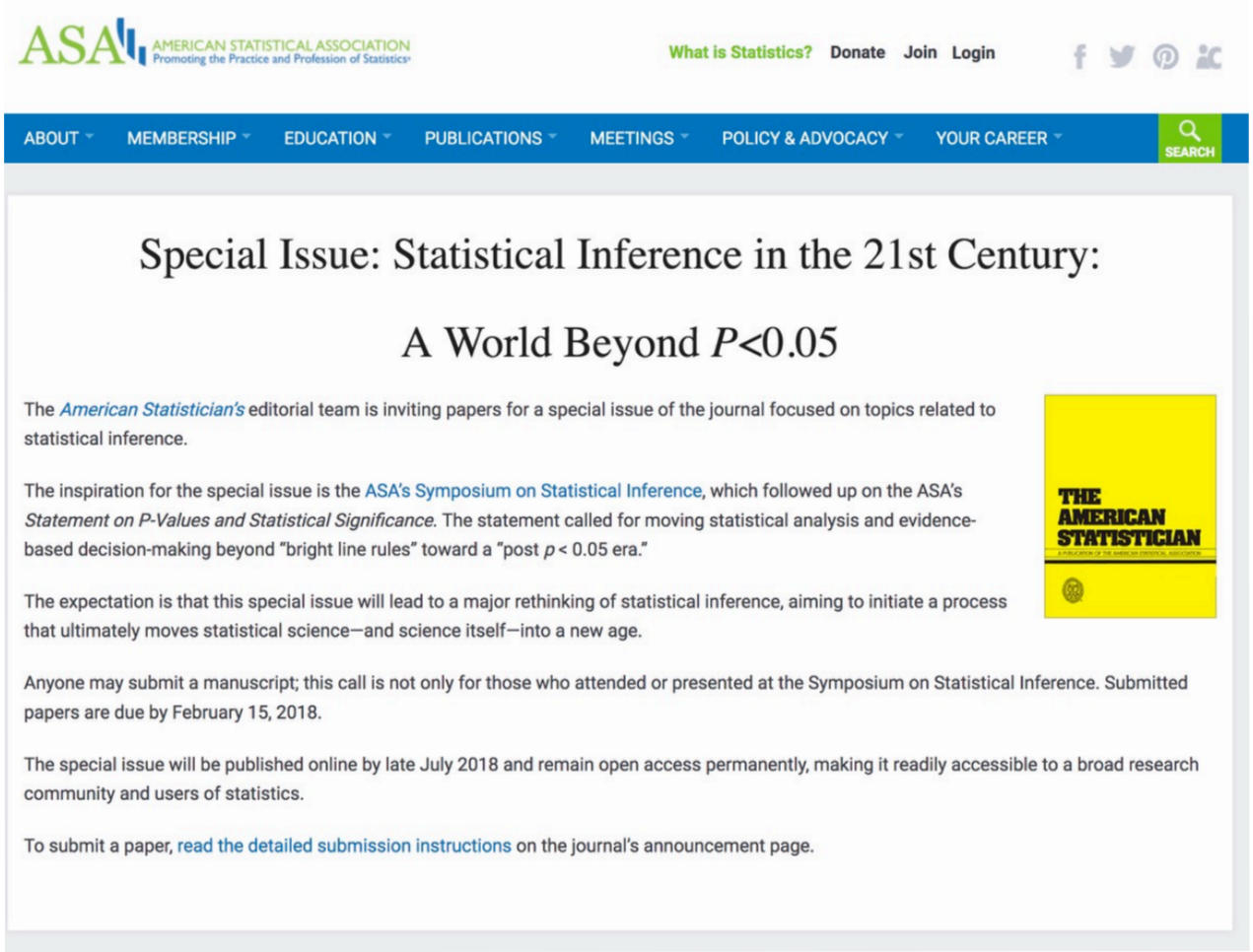
- "While it would be wonderful if every analysis would be informed by someone skilled in statistical methodology, whether a nominal statistician or not, the rise of data science means that even more practitioners will be without a full professional training and continue to do their *t*-ing and *p*-ing"

  Matthews R et al. Significance. 2017;14:38.

ASA position statement – 1 year after



ASA position statement – 2 years after?



ASA AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics

What is Statistics? Donate Join Login

ABOUT MEMBERSHIP EDUCATION PUBLICATIONS MEETINGS POLICY & ADVOCACY YOUR CAREER SEARCH

Special Issue: Statistical Inference in the 21st Century:

A World Beyond $P<0.05$

The *American Statistician's* editorial team is inviting papers for a special issue of the journal focused on topics related to statistical inference.

The inspiration for the special issue is the ASA's Symposium on Statistical Inference, which followed up on the ASA's *Statement on P-Values and Statistical Significance.* The statement called for moving statistical analysis and evidence-based decision-making beyond "bright line rules" toward a "post $p < 0.05$ era."

The expectation is that this special issue will lead to a major rethinking of statistical inference, aiming to initiate a process that ultimately moves statistical science—and science itself—into a new age.

Anyone may submit a manuscript; this call is not only for those who attended or presented at the Symposium on Statistical Inference. Submitted papers are due by February 15, 2018.

The special issue will be published online by late July 2018 and remain open access permanently, making it readily accessible to a broad research community and users of statistics.

To submit a paper, read the detailed submission instructions on the journal's announcement page.

In gentle praise of significance tests – Sir David Cox

Ready to put your skin in the game?

Ready to put your skin in the game?

«Skin in the game keeps human hubris in check»

Taleb N. Skin in the game. Penguin; 2017.

Ready to put your skin in the game?

«Academia has a tendency, when unchecked (from lack of skin in the game), to evolve into a ritualistic self-referential publishing game»

Taleb N. Skin in the game. Penguin; 2017.

Ready to put your skin in the game?

«Indeed, journals may reject papers that use more reliable or more rigorous methods the discipline is accustomed to, simply because the methods are unfamiliar»

Stark PB & Saltelli A. Significance. 2018;15, 40.

Ready to put your skin in the game?

- «I will not be proselitizing for a given statistical school, so you can relax. Frankly, they all have shortcomings, insofar as one can even glean a clear statement of a given statistical "school"».

  Mayo D. Statistical inference as severe testing: how to get beyond the statistics wars. Cambridge University Press; 2018.

How do I (= Giorgio) put my skin the game?

- «If they've made a mistake, correct them gently and show them where they went wrong. If you can't do that, then the blame lies with you. Or no one»

Marcus Aurelius, Ad se ipsum (X 4)

http://www.perseus.tufts.edu/hopper/text?doc=urn:cts:greekLit:tlg0562.tlg001.perseus-grc1:10.4.1

How do I (= Giorgio) put my skin the game?

«But if on my part I fail to produce yourself as my one witness to confirm what I say, I consider I have achieved nothing of any account»

(Socrates in) Plato, Gorgias

http://data.perseus.org/citations/urn:cts:greekLit:tlg0059.tlg023.perseus-grc1:472b

Thank you for your attention!

I invite you to join Qeios:

www.qeios.com/invitation-to-join/researcher/314