Research Seminar

The unconditional information in P-values, and its refutational interpretation via S-values

Sander Greenland
Professor Emeritus of Epidemiology and Statistics
University of California, Los Angeles

Jointly organized by
Swiss TPH (N. Probst-Hensch & P. Vounatsou)
University of Basel, Department of Environmental Sciences (V. Amrhein)
in collaboration, with
PhD Program Health Sciences (PPHS), Faculty of Medicine, University of Basel
Swiss School of Public Health (SSPH+)

February 26 2018, 17:15-18:30, Vesalianum, Kleiner Hörsaal, Vesalgasse 1

Please register at https://pphs.unibas.ch/curriculum/ws-bayesian-and-penalized-regression-methods-for-nonexperimental-data-analysis/

Summary. After a century of controversy, we still debate the evidential meaning of *P*-values, whether they are useful or distortive for inference, and whether to ban them entirely. Among the less-discussed problems are that (1) *P*-values force test statistics into the unit (0 to 1) scale, which render them highly noninuitive functions of data information, and which creates the illusion that *P*-values measure support, and (2) as with all inferential statistics, *P*-values depend not only focal “null hypotheses” but also on assumptions that are uncertain, leading to overconfident inferences when these assumptions are overlooked. These problems can be addressed by interpreting the observed *P*-value *p* unconditionally, in terms of the information it supplies against the entire model used to compute it. This *test model* or *total model* includes explicit assumptions such as random sampling, allocation, and parametric specifications. But it also includes implicit and often doubtful assumptions such as no selective reporting (e.g., no *P*- hacking), no data-base errors, and no outright data tampering (fraud), *whether those assumptions are recognized or not*.

An equal-interval measure of the information in *p* against this total model is the *self-information*, *surprisal* or *S*-value $s = -\log(p)$. The *S*-value is zero (unsurprising) when $P=1$, increases exponentially as *P* approaches zero, and can be intuitively yet correctly understood via a simple coin-tossing experiment. It provides a way to focus interpretation

on the tested hypothesis or model instead of on data probabilities. Because the *S*-value is not a probability, it does not invite equation with alpha levels, hypothesis probabilities, or other quantities often confused with *P*-values. I thus recommend that teaching and practice reinterpret *P*-values as compatibility measures and confidence intervals as compatibility intervals, whose information content is gauged by *S*-values rather than by criteria for "significance."