

## EINLADUNG

- Zeit: Freitag, 5. Februar 2016, 15:30 Uhr
- Ort: Hauptgebäude, Raum 2222, Ahornstraße 55
- Referent: Pierre Dragicevic, Ph.D.  
Permanent Research Scientist at Inria, France.
- Titel: Bad Stats are Miscommunicated Stats

### Abstract:

When reporting on user studies, we often need to do stats. But many of us have little training in statistics, and we are just as anxious about doing it right as we are eager to incriminate others for any flaw we might spot. Violations of statistical assumptions, too small samples, uncorrected multiple comparisons—deadly sins abound. But our obsession with flaw-spotting in statistical procedures makes us miss far more serious issues and the real purpose of statistics. Stats are here to help us communicate about our experimental results for the purpose of advancing scientific knowledge. Science is a cumulative and collective enterprise, so miscommunication, confusion and obfuscation are much more damaging than moderately inflated Type I error rates.

In my talk, I argue that the most common form of bad stats are miscommunicated stats. I also explain why we all have been faring terribly according to this criteria—mostly due to our blind endorsement of the concept of statistical significance. This idea promotes a form of dichotomous thinking that not only gives a highly misleading view of the uncertainty in our data, but also encourages questionable practices such as selective data analysis and various other forms of convolutions to reach the sacred .05 level. While researchers' reliance on mechanical statistical testing rituals is both deeply entrenched and severely criticized in a range of disciplines—and has been so for more than 50 years—it is particularly striking that it has been so easily endorsed by our community. We repeatedly stress the crucial role of human judgment when analyzing data, but do the opposite when we conduct or review statistical analyses from user experiments. I believe that we can cure our schizophrenia and substantially improve our scientific production by banning p-values, by reporting empirical data using clear figures with effect sizes and interval estimates, and by learning to provide nuanced interpretations of our results. We can also dramatically raise our scientific standards by pre-specifying our analyses, fully disclosing our results, and sharing extensive replication material online. These are small but important reforms that are much more likely to improve science than methodological nit-picking on statistical testing procedures.

Es laden ein: Die Dozenten der Informatik