

TALK

HOW TO AVOID DATA CLEVER-HANSING: WHAT FINAL DECISIONS MUST BE MADE EARLY AND WHY

Petr Tureček

Faculty of Science, Charles University, Prague, Czech Republic

petr.turecek@natur.cuni.cz

ABSTRACT

When we pre-register an analysis — or even when we plan data collection outside of the preregistration framework — we should first make several crucial decisions. How will we treat outliers and influential observations? How will we handle incomplete responses? Should we log-transform the response variable before analysis? Which predictors will be included and which ignored?

If we leave these decisions unspecified and try to make them on the fly, two dangers arise: Co-authors or reviewers may prompt us to "try also this and that," leading to a cascade of redundant, unnecessary analyses. Worse, we ourselves may be tempted to favor the choices that subtly nudge the results toward our preferred hypothesis.

This workshop will demonstrate the dangers of "too many options" through a practical exercise in R and introduce ways to recognize individuals and research fields that are not fully honest or responsible with their data.