Convincing Researchers to Transition to Bayesian Statistics

Bayes@Lund, 2017-04-20

Robert Feldt, Richard Torkar, Francisco Gomes, Lucas Gren

CHALMERS

BLEKINGE TEKNISKA HÖGSKOLA · BTH ·

DEPARTMENT OF SOFTWARE ENGINEERING

Many research areas are statistically immature

- Software Engineering is the same (and our case):
  - SE = Applied research on making SW Dev more efficient
  - Grown in size & importance since late 1960’s
  - Research has really taken off last 15 years
  - Typical data:
    - Small sample sizes
    - Mixed types of data (continuous, discrete, …)
    - Time series or measurements of processes
- Statistically immature:
  - Slowly increasing use of frequentist statistics
  - Non-parametric stats proposed last 5 years
  - Very little talk about Bayesian stats

The sad current state of statistical analysis in SE

- We analysed Statistical Maturity in SE:
  - 300 research papers from 2015
  - Published in top 4 journals: TSE, EMSE, TOSEM, IST
  - Which tests, type of data, statistical argument, data availability, practical significance, …

TOTAL		
Num papers	300	
Empirical & Quantitative	210	70%
& No statistical test	88	29%
& Parametric test	31	10%
& Non-parametric test	49	16%
& NP + Parametric	24	8%
& Bayesian analysis	0	0%

!

1 “Bayesian” Paper of 300!

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

Using Bayesian regression and EM algorithm with missing handling for software effort prediction

Wen Zhang[a,*], Ye Yang[b], Qing Wang[c]

[a] School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, PR China
[b] School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030, USA
[c] Laboratory for Internet Software Technologies, Institute of Software, Chinese Academy of Sciences, Beijing 100190, PR China

A R T I C L E   I N F O

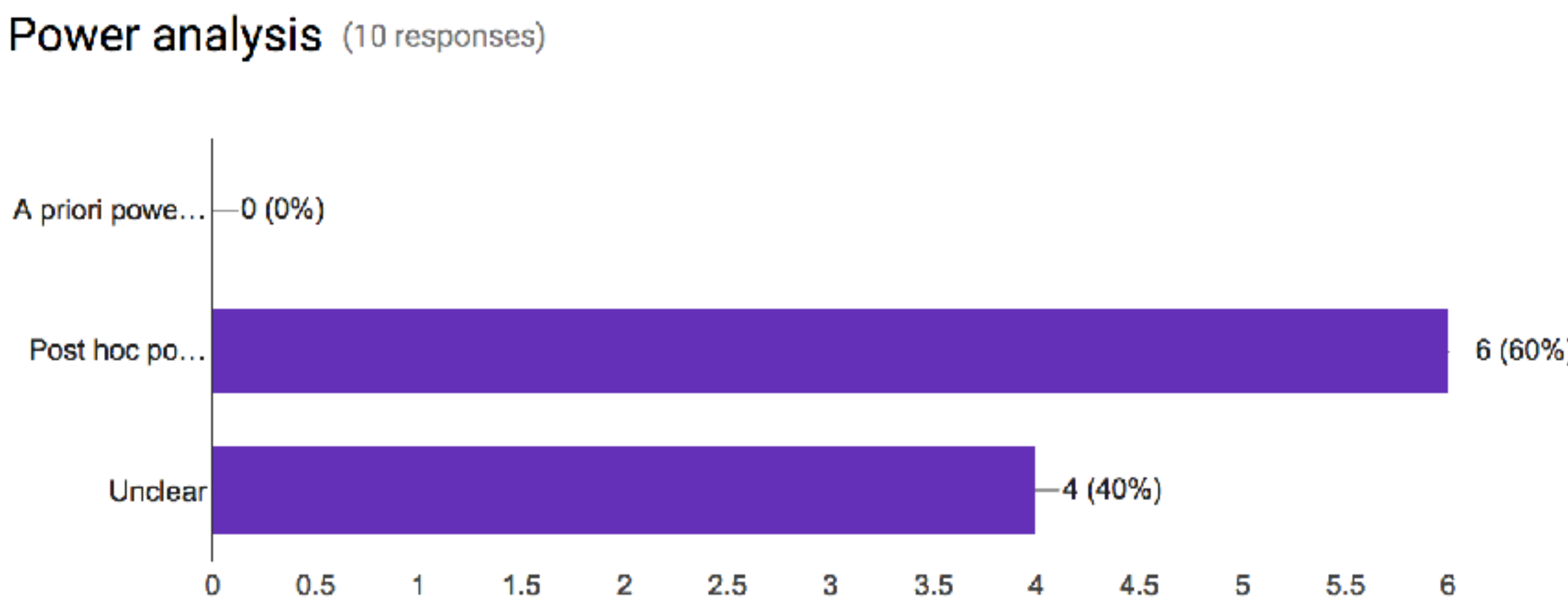
Article history:
Received 25 April 2014
Received in revised form 5 September 2014
Accepted 11 October 2014
Available online 18 October 2014

Keywords:
Bayesian regression
EM algorithm
Missing imputation
Software effort prediction

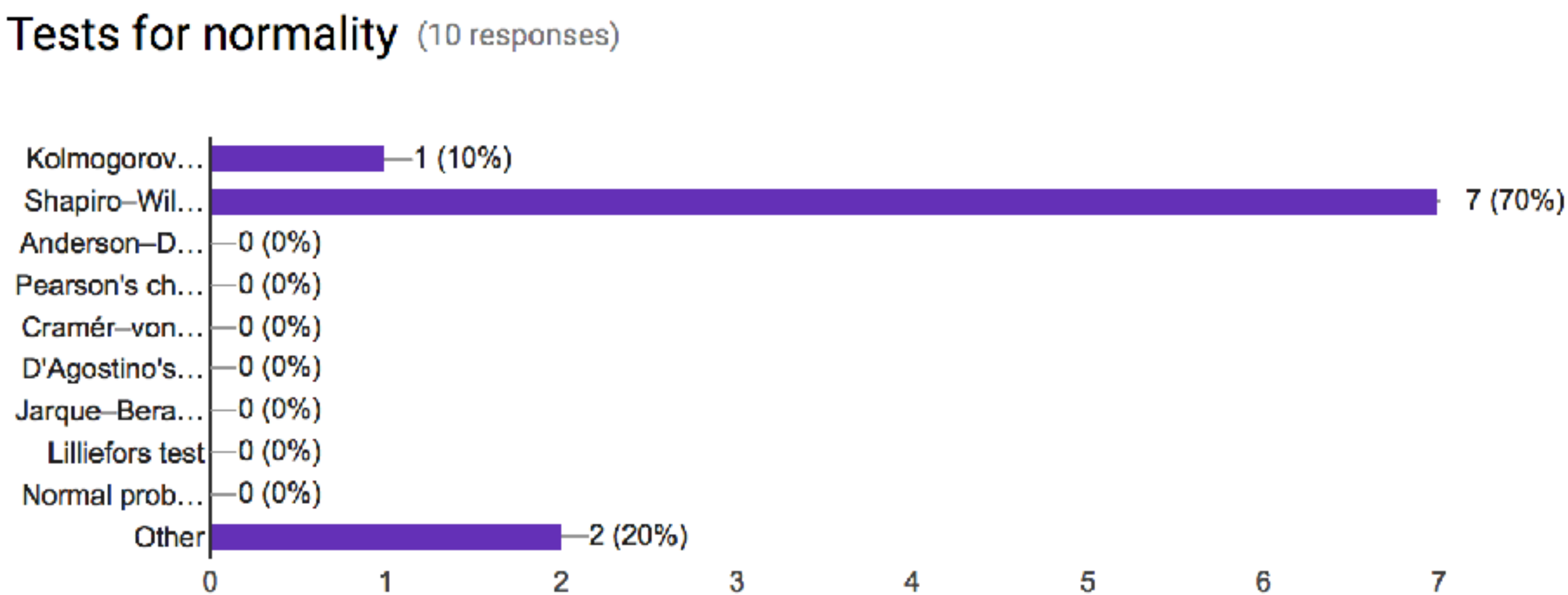
A B S T R A C T

Context: Although independent imputation techniques are comprehensively studied in software effort prediction, there are few studies on embedded methods in dealing with missing data in software effort prediction.
Objective: We propose BREM (Bayesian Regression and Expectation Maximization) algorithm for software effort prediction and two embedded strategies to handle missing data.
Method: The MDT (Missing Data Toleration) strategy ignores the missing data when using BREM for software effort prediction and the MDI (Missing Data Imputation) strategy uses observed data to impute missing data in an iterative manner while elaborating the predictive model.
Results: Experiments on the ISBSG and CSBSG datasets demonstrate that when there are no missing values in historical dataset, BREM outperforms LR (Linear Regression), BR (Bayesian Regression), SVR (Support Vector Regression) and M5′ regression tree in software effort prediction on the condition that the test

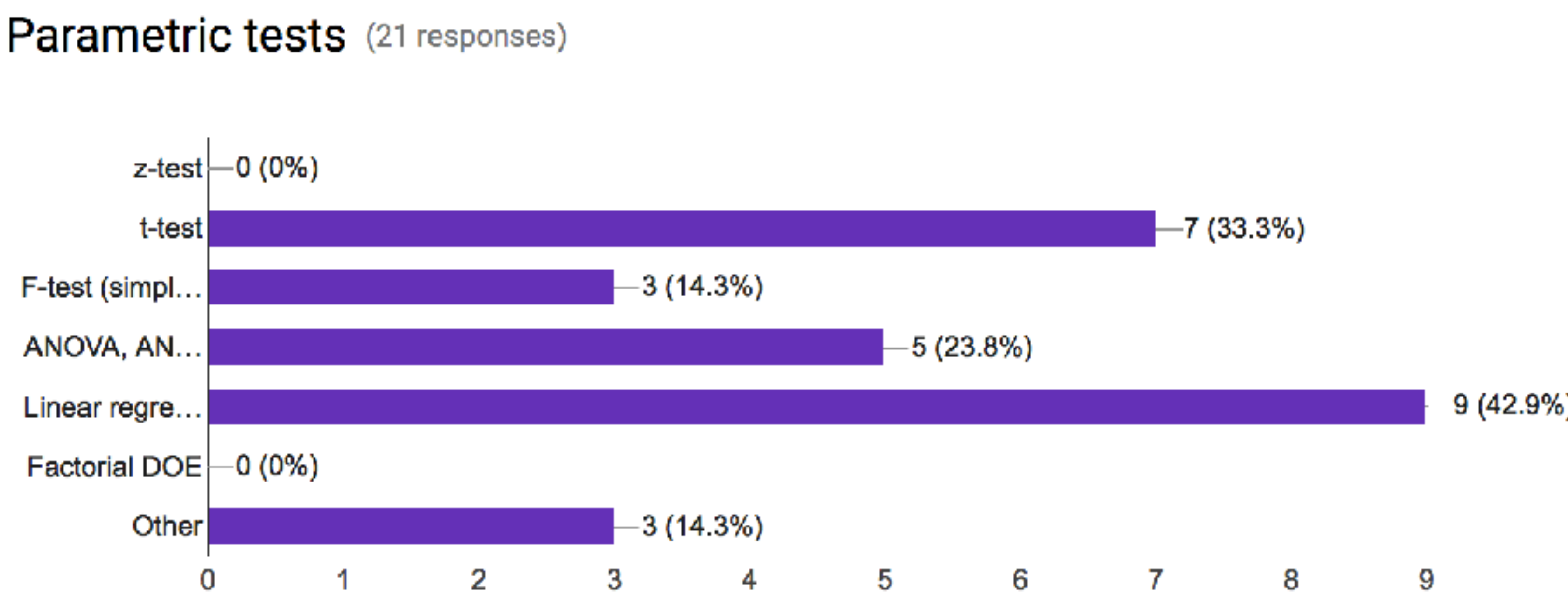
Sample of 30 papers in more detail: Power analysis



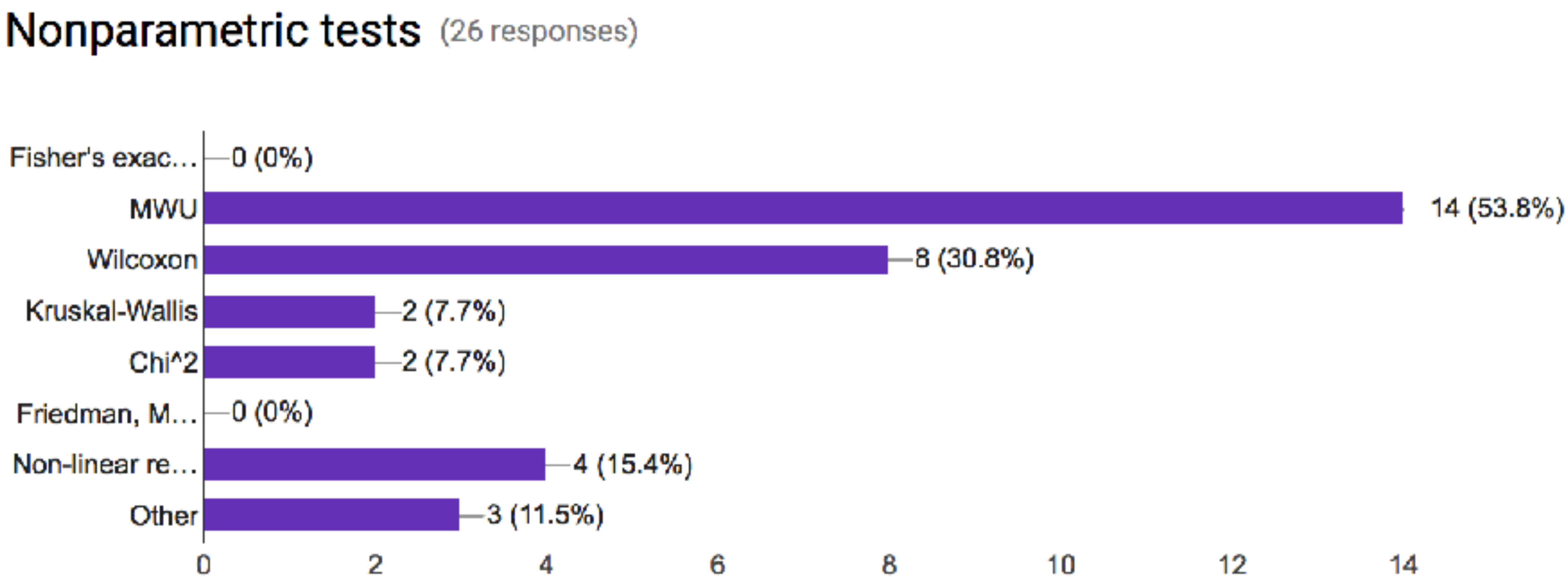
Sample of 30 papers in more detail: Normality tests



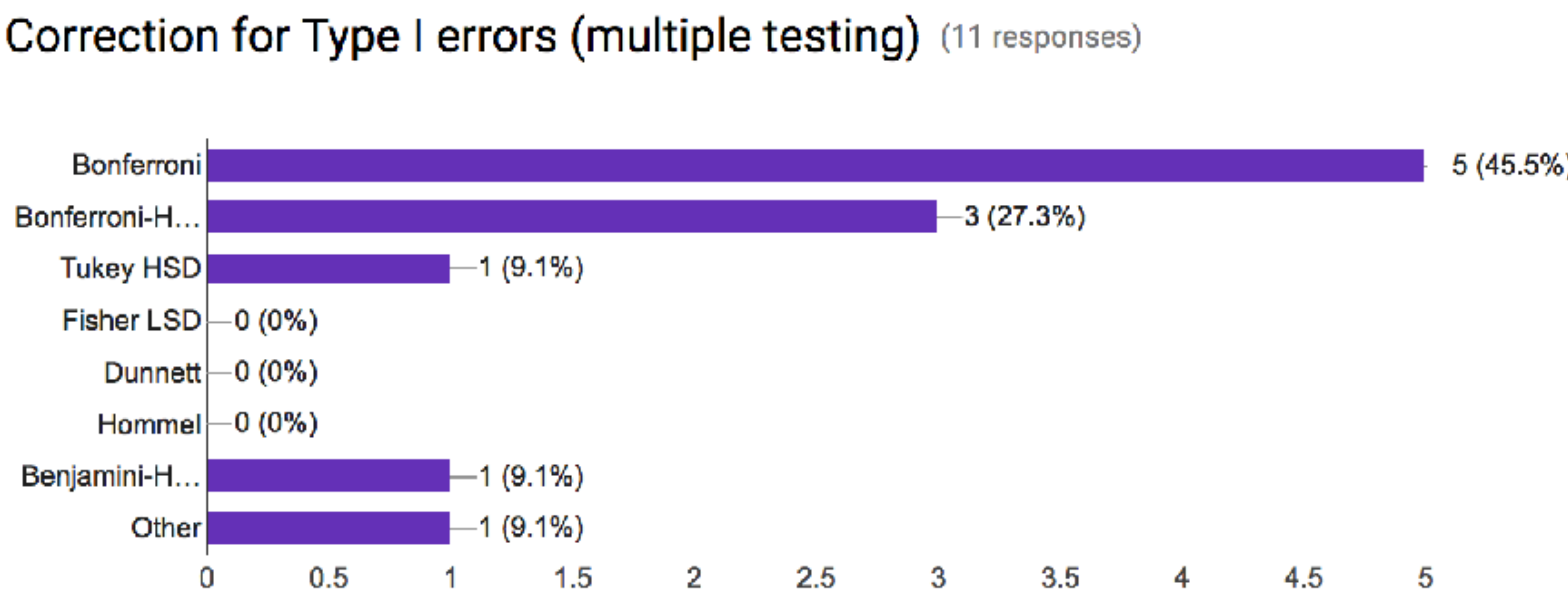
Sample of 30 papers in more detail: Parametric tests



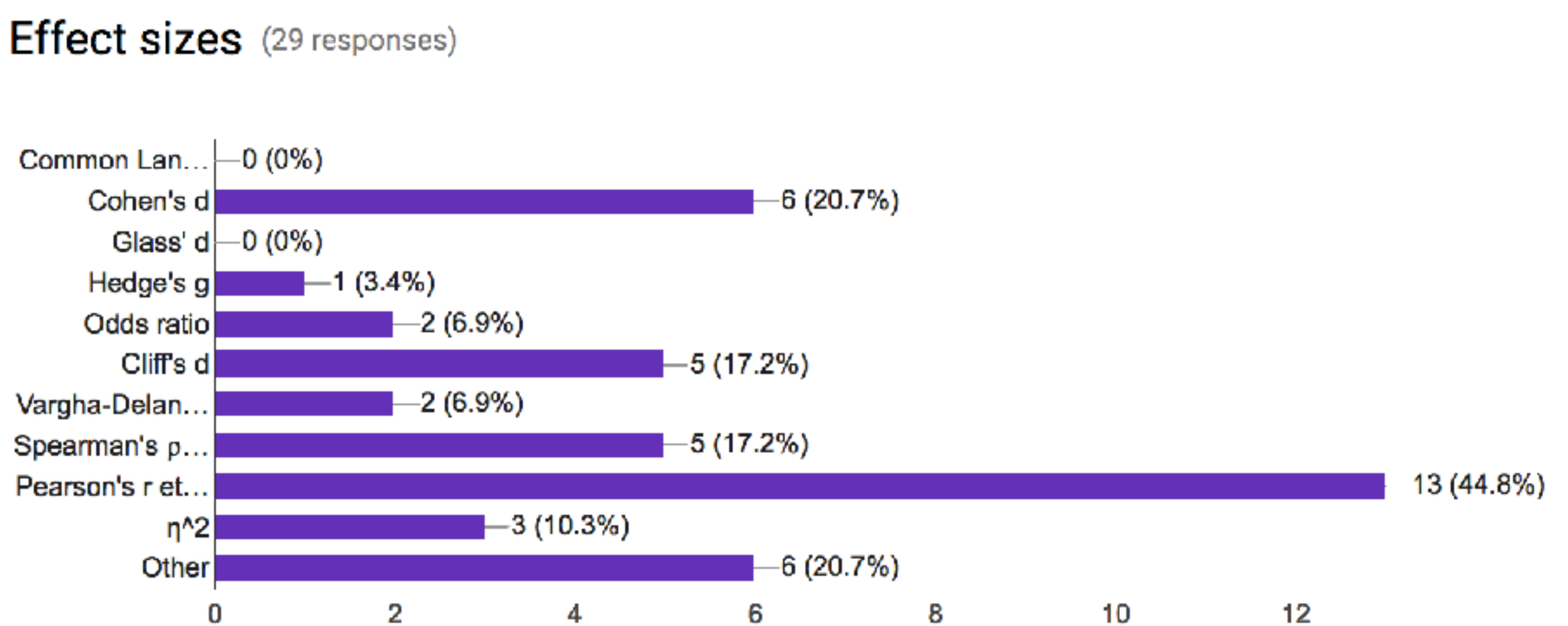
Sample of 30 papers in more detail: Non-Parametric tests



Sample of 30 papers in more detail: Corrections



Sample of 30 papers in more detail: Effect sizes



SE should transition to Bayesian Statistical Analysis

- Distributions are rarely normally distributed
- Sample sizes often small & data often missing
- Statistical knowledge is limited; might as well learn Bayesian from the get-go
- Needs flexible modelling; today people adapt their data/ problem to their “hammers” (NSHT, Frequentist stats)
- Effect sizes sometimes used but still unknown
  - **Unintuitive**: Some of them don’t really measure effect size
  - **Ad hoc**: Subjective scales to judge Cohen’s d, for ex.
  - **Different scale**: Not measured on same scale as underlying measure
- Bayesian Analysis adress a majority of these! SE should transition!

But how can we convince the SE community?

- Researchers mainly just want:
  - Draw as clear conclusions as possible from their data
  - Be up to statistical standards of their field
- Arguments we could make:
  - "Bayesian is philosophically better"
    - We doubt they will care much
  - "Bayesian is mathematically better"
    - We doubt they will understand it
  - "Bayesian is better in practice"
    - Better, but can we really show it

But how can we convince the SE community?

- Our current 4-step plan:
    - 1. Summarise current state and one illustrative re-analysis of frequentist SE analysis/data
    - 2. Guideline paper that maps SE research questions to parametric <-> non-parametric <-> bayesian analyses
    - 3. Apply guidelines in our own papers to create set of “model”/example papers
    - 4. Gradually use/introduce more flexible/specific models
- **We need your input!** Examples of similar transitions, re-analysis etc???

Questions and/or ideas!?

robert.feldt@chalmers.se

richard.torkar@cse.gu.se

@drfeldt

@rtorkar