SoK: I have the (Developer) Power!
Sample Size Estimation for Fisher‘s Exact, Chi-Squared, McNemar‘s, Wilcoxon Rank-Sum, Wilcoxon Signed-Rank and t-tests in Developer-Centered Usable Security

Anna-Marie Ortloff¹, Christian Tiefenau², Matthew Smith¹,²
¹: University of Bonn, ²: Fraunhofer FKIE
Statistical power:
The probability of detecting an effect, if a true effect exists.
Developer-Centered Usable Security (DCUS)
DO WE HAVE THE POWER?
Power Meta-Analysis (simulated a-priori power analysis)
Power Meta-Analysis (simulated a-priori power analysis)

Is power sufficient?
- No
- Yes

Effect size to be detected (Cohen's criteria):
- Small (d=0.2)
- Medium (d=0.5)
- Large (d=0.8)

Mean power to detect effects per study:

0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
We’re not using power analysis.

In SOUPS and CHI USP publications from 2020/2021 only 8 of 74 (10.8%) quantitative papers used a priori power analysis.
HOW CAN WE GET THE POWER?
Power Analysis

Hypothesis test

N

power

α error

effect size
A Priori Power Analysis

Hypothesis test

- Power
- Effect size
- \( \alpha \) error

N
A Priori Power Analysis

Hypothesis test

- Power: 0.8
- Effect size
- \( \alpha \) error: 0.05
- \( N \): ?

Behavioural Security and Privacy Group
A Priori Power Analysis

Hypothesis test

- $N$
- $\alpha$ error: 0.05
- Power: 0.8
- Effect size: well...
Which effect size should I use?

- Use (arbitrary) guidelines for large, medium, small effects
- Literature research
- Do a pilot study
- Decide on the smallest effect size of interest
Method

Literature Collection

- SOUPS, USENIX Security, S&P, CCS, ICSE, USP Tracks of CHI
- 2010 - 2021
- Include user study
- Participants: software developers, similar expert users, or proxies
- Domain of usable security and privacy

- 54 papers
- including 64 studies, 467 hypothesis tests, 413 variables

Data Structure

- Relevant information on power and effect sizes in these studies
USING THE POWER DATABASE
I have the power!

This is the companion website for the paper *SoK: I Have the (Developer) Power! Sample Size Estimation for Fisher’s Exact, Chi-Square, Wilcoxon Rank-Sum, Wilcoxon Signed-Rank and t-Tests in Developer-Centered Usable Security.*

For more information, see About.

https://powerdb.info/
Searching the database

Click here to download the database as a SQLite file

Operator
- All selected (AND)
- Any selected (OR)

Variable

Variable category
- security

Participants type
- test
- fishers

Data collection method
# Searching the database

### Results:

1. **Test: Fisher's Exact Test**


**Participants:** professional software developer (N=36)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>DV Categories</th>
<th>Independent Variable</th>
<th>IV Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>security</td>
<td></td>
<td>prompting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;a binary variable secure indicating whether participants used any kind of security in their code&quot;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Level yes</td>
<td>Any kind of security was used in the code</td>
<td>1. Level true</td>
<td>Categories for prompting</td>
</tr>
<tr>
<td>2. Level no</td>
<td>No kind of security was used in the code</td>
<td>2. Level false</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Categories for security</td>
<td></td>
</tr>
<tr>
<td>1. security</td>
<td></td>
<td>1. study-related variable</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. prompting</td>
<td></td>
</tr>
</tbody>
</table>

- ** Effect size: odds ratio 46.33 | Cramer’s V 0.211**
Effect sizes in DCUS

- Overall: N=157, median=0.47
- Artifact related variable: N=15, median=0.63
- Behavior: N=30, median=0.49
- Security: N=33, median=0.48
- Participant judgment: N=35, median=0.47
- Functionality: N=11, median=0.39
- Efficiency: N=23, median=0.39

Effect size judgment:
- Small
- Medium
- Large
More Meta Analysis?

- How do you interpret effect sizes?

Come talk to me at SOUPS about:

- Power Analysis (this work)
- Effect sizes (on going)
- Anything else meta
- ...

ortloff@cs.uni-bonn.de
https://amortloff.github.io

Behavioral Security and Privacy Group