Quantitative Data Collection

Today…

- Hypothesis testing and research questions
- A discussion of data-collection
  - Why would you opt for one data-collection protocol over another?
  - What is the role of deception in data collection?
  - What are the ethical considerations?
- Pick a Presentation & Exam date (Doodle Poll)

What is a Quantitative Study…

- Privacy & Security Algorithms

  - Usability in terms of quality metrics such as:
    - Learning time
    - Efficiency of Use
    - Memorability
    - Use Errors
    - Subjective Satisfaction
    - Protection…

  - Analysing numbers provided to draw conclusions

Good Reasons to do a Quantitative Study…

- Validating or invalidating the hypotheses formed from qualitative studies
- Analysing participants' awareness of security/privacy
- Determining the potential for success of a new security/privacy preserving system
- Validate or prioritise participant needs and expectations
- To base new designs on reliable information and avoid being guided by pre-conceptions

Risks of Quantitative Studies…

- Removing Outliers (Performance)
- Error Margins
- Determining the number of users to test
- Population Sampling
- Random Results (Many systems and platforms)
- Beware of correlations (Users are different)
- Overlooking covariants and over simplified analysis

Population, Sample, and Raw Data

- Samples – need to be representative of the target population

- Why? Important conclusions can be inferred by analyzing the sample data
  - Inductive statistic or statistical inference
  - **Example:** Predicting the weather

- What are concrete tasks users should be able to do?
  - Set realistic metrics, based on understanding of users

- **Exercise #1:**
  - Using your project topic, select a population sample for your research questions.

Hypothesis Testing…

- Statistical decisions about populations based on samples
- **Example:** Decide whether longer passwords are really effective in protecting data?
- A hypothesis is a conjecture, or informed guess, that might be true
- **Example:** “Longer passwords are more secure than shorter passwords”
- A hypothesis must be falsifiable

Hypothesis Testing…

- A research question is broader and does not need to be a statement to be disproved
  - Often the case for exploratory work
  - **Example:** "How do users make up longer passwords?"
- Null Hypotheses
  - Good for deciding when one solution is better than another
  - Formulate a hypothesis to say there is no difference between the two

Hypothesis Testing…

- Alternative Hypotheses:
  - Others besides null hypotheses e.g. p =0.7, p!=0.5…
- **Tests of hypotheses:**
  - **Example:** for every 20 long passwords created, 16 are easily guessable
  - → Reject hypothesis that longer passwords are secure
- But…we might be wrong, the sample is small for example

Errors, Significance, and Tests…

- **Type I error:** Reject hypotheses when it should be accepted
- **Type II error:** Accept a hypotheses when it should be rejected
- Reduce chances of error by increasing the sample size
- **Level of significance:**
  - Max. probability (α) of tolerable risk of a Type I error
  - In practice 0.05 or 0.01 level of significance
  - → Hypothesis has a 0.05 probability of being wrong

Studies: Purposes and Goals…

- What are you hoping to learn?
- What are your hypotheses?
  - Listed on paper, can draw on prototypes (e.g. paper)
- What are your metrics for success?
  - More secure (adversarial model), quicker to use, …
- What are you comparing to? What data might be useful?
- **Exercise #2:** Formulate a hypothesis you would like to test based on your research question(s)

Broad Types of Studies…

- In formulating your hypothesis remember to think…
  - What type of study?
    - Descriptive, Relational, Experimental…
  - How?
    - Formative (initial)
    - Summative (validate)

Types of Studies…

- What do people want/think/do overall…
  - Surveys, Interviews, Focus Groups…
- What people want/think in context…
  - Contextual Inquiry (Interviews)
  - Observations in the field
- Expert evaluation
- Usability Test: Online, Lab, Log Analysis…
- Controlled experiments to test causation

Example…

It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception

Marian Harbach, Emmanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, Matthew Swith

(SOUPS 2014)

Goal and Methodology…

- Goal:

- → Discover real world (un)locking behavior of smartphone users

- Methodology:

  - Only wanted to look at current behaviour
  - Did not require comparing conditions
    - Conditions aren’t needed

Online Survey & Removing Outliers…(n=260)

- Good for collecting data for a large number of participants
  - Across a wide geographic space
- Used QR codes to verify information about participants
- Removed 60 response sets due to incorrect completion codes
  - i.e The smartphone check failed
  - Implausible timing or wrong answers to two or more attention check questions
- How about manipulation and bias introduction?

One-Month field Study (n=52)…

- Give participants in-the-moment questions at different times
- Record their answers against a given context
- Afterward, interview them (yes/no answers or numbers)

- Results (some):
  - Should surfing attacks are only perceived to be a relevant risk in 11 on 3140 … situations
  - Users spend up to 9.0% of the time the use their smartphone on dealing with unlock screens

Discussion…

- Why are they using two protocols?
- What does one methodology offer over the other?
- Is there a reason to use two methods, rather than collect more data using a single methodology?
  - To what degree is the data reliable?
  - How skewed is the sample population?
  - Did participants behave as normal?

How Many Users to Test…

- Requires many users generally about 4 times more than in simple user testing (5 X 4 =20)
- Due to individual differences in user performance
- Large numbers needed to smooth over variability and error margins
- Sometimes also task repetition required to validate results
  - Statistical distributions
    - Sample should closely resemble target population

Data to Collect…

- Performance (time, success rate, errors,…)
- Opinions and attitudes
- Actions and Decisions
- Keystrokes, mouse movements…
- Demographics (age, gender, privacy attitudes,…)
- Preferences …

Standard Deviation and Outliers

- Deviation of each observed value from the mean of the values.
- **Example:**
  - 1520 measures of user time-on-task performance for 70 privacy configuration tasks on various websites
  - Standard deviation of 52% of mean values
  - 10 mins per task → standard deviation of 5.2 mins for the metric
- Outliers of excessively slow users were removed

Logistics for a study…

- How many participants?
  - Statistical power
  - Time, budget, participant's time
- What kind of participants?
  - Skills, background, interests
  - Motivations
  - How to get a "representative sample"
- What do you need to build, if anything?
  - Prototype fidelity

Study Designs…

- Within Subjects?
  - Every participant tests everything
  - Crucial to randomise order (learning effect)
  - Fewer participants
- Between Subjects?
  - Each participant tests one version of the system
  - You compare these groups
  - Groups should be similar (verify!)
  - Still randomise!

Validity…

- Is the study ecologically valid?
  - Does it mirror real-life conditions and context?
  - Think – Sample Composition!
- To what degree can we generalise about our results (externally valid)?
- What biases does our sample introoduce?

Example…

The Emperor's New Security Indicators: An evaluation of website authentication and the effect of role playing on usability studies

Stuart E. Schechter, Rachna Dhamija, Anja Ozment, Ian Fischer

(Oakland 2007)

Study Objectives ...

- Evaluate online security indicators
  - (e.g., presence of https, presence of security pictures)
- Contrast real-life behaviour and study behavior
- How are these objectives aligned?
- Why examine both at once?

Methodology and Discussion...

- Laboratory study

- Three conditions
  - Role-playing
  - Role-playing with security priming – (Warnings)
  - Asked to use their own information

- Discussion:
  - Why was this a laboratory study?
  - Could an online survey have worked?
  - What about a field study?

Ethics and Deception...

- Why was deception used in this study?
- Why the stipulation:
  - Participants must only be deceived in ways that cause them to believe they are less secure than they actually are…
- Could this study have worked without using deception?
- How did this study use a simulated attacker?

Exercise…

- Based on your project, hypothesis, and two data collection protocols (methods):
  - List some tasks and metrics you would like to evaluate
  - How do you plan to handle outliers? (e.g. slow users, seniors,…)
  - What is your goal significance level? How does it affect numbers of participants needed?
- How might qualitative methods help?

Additional Points…

- Identity Leak Checker…
  - 25000 people submitted their credentials
- What are their password characteristics?
- How guessable are their passwords?
- How do demographic factors correlate with password strength?
- How do these real passwords compare to leaked/collected passwords?

What the data says and Ethics…


RISK ASSESSMENT / SECURITY & HACKTIVISM

It's official: Computer scientists pick stronger passwords

Landmark study says people in business school choose weakest passwords.

- What are some ethical concerns here?
  - What seemed to be done well?
  - What could have been done better?

Social Phishing and Threat Scenarios…

- Use social networking sites to get data for targeted phishing
  - …In this study we simply harvested freely available acquiantance data by crawling social network sites…
  - …we launched an actual (but harmless) phishing attack targeting students…
- **Control group:** message from stranger
- **Experimental group:** message from a friend

Ethics…

- How was consent obtained?
- Who was potentially affected by the study?
  - …The number of complaints made to the campus IT support center was small (30 complaints, or 1,7% of the participants) …
- Important to take into consideration in formulating your user study protocol

Next Week…

- 1-2 page project proposal
- Describe the systen you propose to design or evaluate
- Discuss what you hope to learn from your user study and/or the hypotheses you plan to test
- User Consent Form
- User study protocol