COMP 150-4: Human Factors in Security and Privacy

Quantitative Methods

Lecture 7

Prof. Daniel Votipka
Spring 2021

(some slides courtesy of Prof. Adam Aviv and Prof. Michelle L. Mazurek)

Administrivia

- Please email me when you've selected a team
  - Team member names
  - 1-2 sentence description of problem
- Added templates and an example to the project description page
- HW 2 is out
  - Working on getting anonymized data to you
  - Email me interview questions after pilot
  - Instructions for MaxQDA install on Piazza

What we did last time!

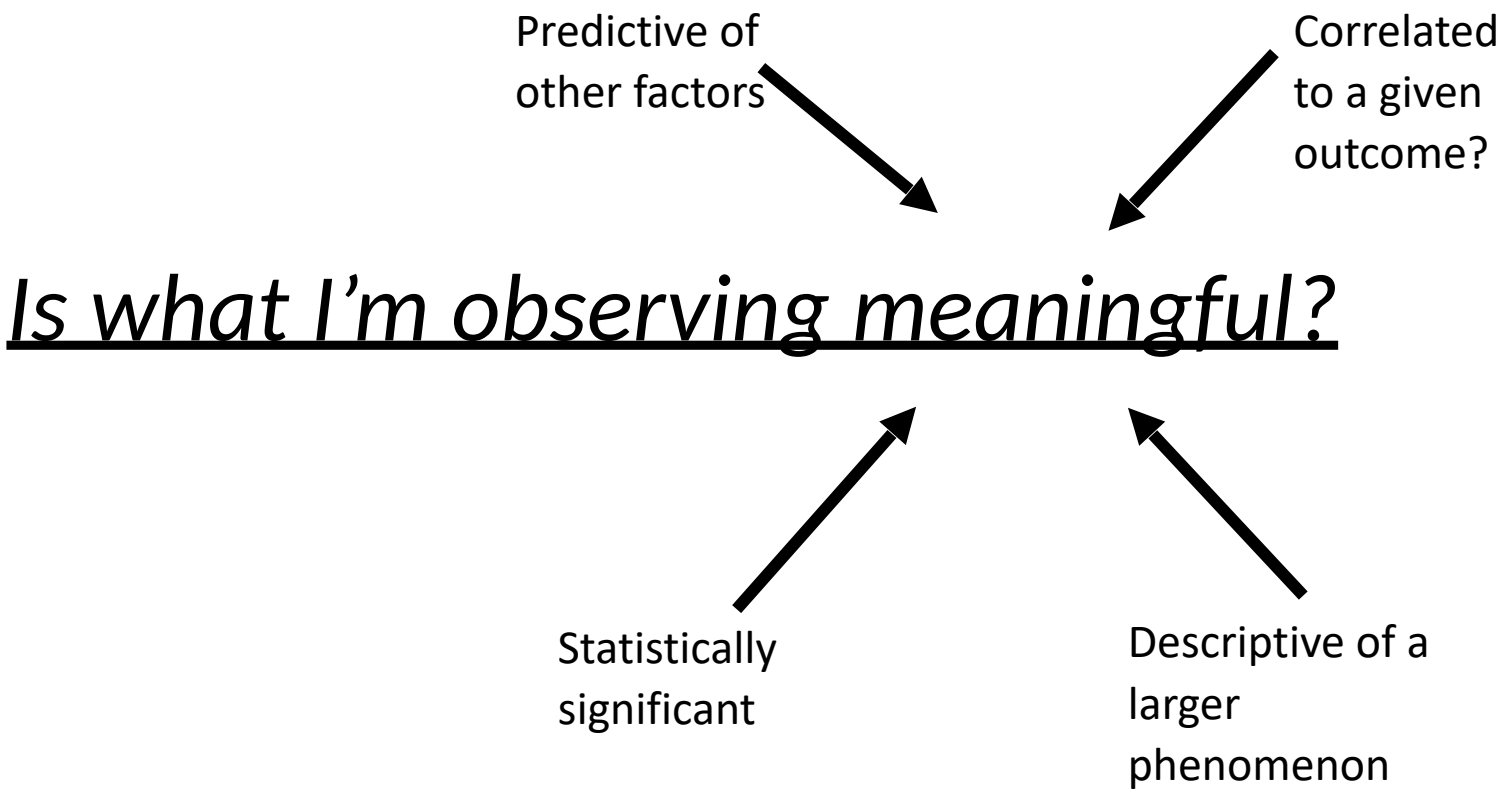
- Improve the metaphor
- Key exchange transparency
  - Secure communications
    - Adoption challenges
    - Key verification protocols
  - HTTP vs. HTTPS

What are we doing today?

- Overview of quantitative data
- Surveys
    - Crowdsourcing

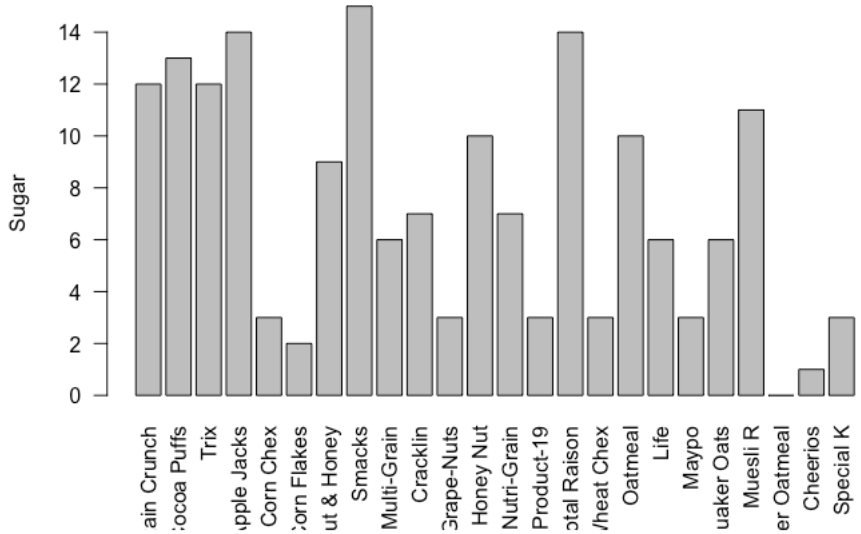
Basics of Quantitative Methods

To evaluate your hypothesis, you may ask…



Ways to use Quantitative Data

- Descriptive
  - Percentages
  - Mean, Median, Standard Deviation
  - Bar Charts
- Statistical
  - Significance Testing
  - Regression Analysis




```
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     67.84513   11.46089   5.920 2.16e-05 ***
Shelf.displayed.at.storeMiddle -1.83585    4.14710  -0.443   0.6639
Shelf.displayed.at.storeTop     0.07115    2.89293   0.025   0.9807
Calories.per.serving           -0.08665    0.10383  -0.835   0.4163
Sugar..grams.per.serving.      -1.94445    0.32008  -6.075 1.61e-05 ***
Weight..Ounces.per.serving.     2.36395   10.38943   0.228   0.8229
Fiber..grams.per.serving.       1.22238    1.14818   1.065   0.3028
Sodium..milligrams.per.serving. -0.04585    0.01727  -2.656   0.0173 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.471 on 16 degrees of freedom
Multiple R-squared:  0.8745,    Adjusted R-squared:  0.8195
F-statistic: 15.92 on 7 and 16 DF,  p-value: 4.075e-06
```

Quantitative Methods

- Surveys (closed option)
- Field measurements
- Controlled experiments
- Artifact analysis

Quantitative Methods

- Surveys (closed option)
- Field measurements
- Controlled experiments
- Artifact analysis

Survey Design

What is a survey?

- Well defined and well written set of questions to which an individual is asked to respond

- What's the difference with an interview?
  - Surveys are self-administered and self-contained
  - There is no opportunity to ask follow up questions

Why might you want to use a survey?

- Can recruit a lot of participants rather quickly
  - Get a little bit of data from each…
  - … but from a lot of people!
- Easy and relatively cheap
  - For example, depending on the survey (of course), recruiting a few hundred participants on MTurk/Prolific could be completed in a day or two.
  - Compare to surveys, interviewing a few hundred participants could be months (or even a year!) of work.
- Quantitative Results
  - Generalizes (if done right)
    - likely to get the same results with a similar sample taking the survey
  - Easier and faster to analyze
    - Doesn't require coding (but sometimes it does …)
  - Statistical and Regression analysis (perform analysis based on quantifiable outcomes)
    - How many selected X or y?
    - Which of these set of factors (dependent variable) most correlate with the outcome (independent variable)?

Why might you *not* want to use a survey?

- Data can be shallow
  - Multiple choice questions don’t allow a ”deep dive”
  - Short-answer, free-response may not offer a full explanation without requisite follow ups
- Biases
  - Self-reported answers
    - No observations or encouragement from the research
  - Questions could be confounding or confusions
    - Only get one shot!
    - Low investment will just mov on
  - Order of the questions matter
- Easier to “cheat” and verify responses
  - Hard to tell when participants are being dishonest
  - Particularly important for crowdsourced data collection
    - Attention checks help but are not a panacea

Survey best practices

- Pilot, pilot, pilot
  - Ensure question are neutral, non-ambiguous, and not confusing
  - Try many different wordings
  - Check answers to your expectations
- Ask others for input
  - Don’t be shy sending your survey to a colleague or a friend
- Consider your sample
  - Is the survey designed for a particular subset that know key terms?
  - What if they don’t?
- No double-barreled questions!
- Keep it short and too the point
  - Longer surveys lead to fatigue, and fatigue leads to inconsistent answers
- Include attention checks
  - Make sure participants are reading the questions
  - But cannot always exclude based solely on the attention check
- Offer options not to answer (avoid lying or unclear choices)
  - Not all questions will be obvious to all participants, given them a way out so you don’t end up with bad data
  - E.g., Avoid “yes” or “not” response – include a “maybe” or “unsure”

Types of Survey Questions

- Closed-Responses
  - Select from a set of choices

How frequently do you consider the smartphone app developer before installing that app on your personal samartphone?

- *Always*
- *Usually*
- *Sometimes*
- *Rarely*
- *Never*
- *Unsure*

- Open-Ended
  - Free form text

Describe your process for evaluating the smartphone app developer before installing an app on your personal smartphone?

Likert Response Items

- Renis Likert (/lɪk.ərt/LIK-ərt)
  - Commonly mispronounced as /ˈlaɪ.kərt/ LY-kərt
    - I get this wrong!!!!
- Produce a *scale* of responses for a given item
- Example: Agreement with a statement
  - (1) Strongly disagree
  - (2) Disagree
  - (3) Neither agree nor disagree
  - (4) Agree
  - (5) Strongly agree
- Likert scales for many kinds of responses
  - Agreement
  - Frequency
  - Importance
  - Quality
  - Likelihood
  - Priority
  - Concern
  - Desirability
  - … many many more
- Points of a Likert
  - Odd: 3-point vs. 5-point vs. 7-point
  - Even: removes neutral response

Example Likert responses: https://www.peru.edu/oira/wp-content/uploads/sites/65/2016/09/Likert-Scale-Examples.pdf

Writing a “good” Likert question

- Make sure the items of the scale clearly match the question being asked
  - E.g., if agreement, ask if they are in agreement with a statement.

- Use language not numbers, whenever possible
  - BAD: How strongly do you agree? 1-9 where 1 is disagree and 9 is agree
    - What does 3 vs. 4 mean in such a scale?
  - OK: How frequent does this occur? Once a month, once a week, twice a week, every day

Exercise: Produce a short survey

- Research Question: *What are students security and privacy concerns when attending class online via tools like Zoom, Webex, etc?*

- Come up with a five-question survey
  - Three closed questions, two open

- 10-minutes

Exercise: Report Out

- Have you thought about security related to attending online?
  - Yes | No
- Are you required to download any specific software?
  - Yes | No | Don’t know
- What steps do your professors take to ensure privacy?
- What steps do you take to secure meeting?

Surveys as Scales of Usability

- Recall some usability metrics
  - Success/Error rate
  - Completion Time
  - Perceptions
    - How do we measure a perception?!
- Conduct a quantitative survey
  - Convert responses to a number
- Survey -> Number = Scale
  - Much more to this, actually. Should also be validated and shown to correlate with some known outcome.

- Example Usability Scales
  - System Usability Scale (SUS)
  - Computer System Usability Questionnaire (CSUQ)
  - Perdue Usability Testing Questionnaire (PUTQ)
  - Software Usability Measurement Inventory (SUMI)
  - Website Analysis and MeasureMent Inventory (WAMMI)

Computer System Usability Questionnaire (CSUQ)

			1	2	3	4	5	6	7		NA
1.	Overall, I am satisfied with how easy it is to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
2.	It was simple to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
3.	I can effectively complete my work using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
4.	I am able to complete my work quickly using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
5.	I am able to efficiently complete my work using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
6.	I feel comfortable using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
7.	It was easy to learn to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
8.	I believe I became productive quickly using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
9.	The system gives error messages that clearly tell me how to fix problems	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
10.	Whenever I make a mistake using the system, I recover easily and quickly	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
11.	The information (such as online help, on-screen messages, and other documentation) provided with this system is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
12.	It is easy to find the information I needed	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
13.	The information provided for the system is easy to understand	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
14.	The information is effective in helping me complete the tasks and scenarios	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
15.	The organization of information on the system screens is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
16.	The interface of this system is pleasant	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
17.	I like using the interface of this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
18.	This system has all the functions and capabilities I expect it to have	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
19.	Overall, I am satisfied with this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
			1	2	3	4	5	6	7		NA

List the most **negative** aspect(s):

1.
2.
3.

List the most **positive** aspect(s):

1.
2.
3.

System Usability Scale (SUS)

- Five-point Likert agreement
  - I think that I would like to use this system frequently.
  - I found the system unnecessarily complex.
  - I thought the system was easy to use.
  - I think that I would need the support of a technical person to be able to use this system.
  - I found the various functions in this system were well integrated.
  - I thought there was too much inconsistency in this system.
  - I would imagine that most people would learn to use this system very quickly.
  - I found the system very cumbersome to use.
  - I felt very confident using the system.
  - I needed to learn a lot of things before I could get going with this system.
- Replace “system” with item being evaluated

- Average score is 68
  - 68 is OK

Grade	SUS	Percentile range	Adjective	Acceptable	NPS
A+	84.1-100	96-100	Best Imaginable	Acceptable	Promoter
A	80.8-84.0	90-95	Excellent	Acceptable	Promoter
A-	78.9-80.7	85-89		Acceptable	Promoter
B+	77.2-78.8	80-84		Acceptable	Passive
B	74.1 – 77.1	70 – 79		Acceptable	Passive
B-	72.6 – 74.0	65 – 69		Acceptable	Passive
C+	71.1 – 72.5	60 – 64	Good	Acceptable	Passive
C	65.0 – 71.0	41 – 59		Marginal	Passive
C-	62.7 – 64.9	35 – 40		Marginal	Passive
D	51.7 – 62.6	15 – 34	OK	Marginal	Detractor
F	25.1 – 51.6	2– 14	Poor	Not Acceptable	Detractor
F	0-25	0-1.9	Worst Imaginable	Not Acceptable	Detractor

https://measuringu.com/interpret-sus-score/

CROWDSOURCING
CROWD
PARTICIPATION
ORGANIZATION
USE
ONLINE
TASK
BASED
SOCIAL
COMMUNITY
EXAMPLE
OPEN
WORK
SOURCE
DUALS
GROUP
TASKS
MOTIVATION
RESULTS
PARTICIPATE
ACTIVITY
CROWDS
SIMPLE
MODELS
MARKETING
SEVERAL
SOLUTIONS
SUCCESSFUL
PROJECT
CROWDSOURCED
CALL
INTERVIEW
COLLABORATION
OUTSOURCING
PROBLEMS
QUESTIONS
GLOBAL
PROBLEM
INCLUDE
ATTENTION
STATES
ISSUES
COLLABORATITION
SOLVE
DISTRIBUTED
BUSINESS
BETWEEN
PROCESS
WEB
COST
WAY
MADE
UNDERTAKEN
WORLD
SEE
TYPE
SATISFACTION
PARTICIPANTS
REWARDS
TEAM
SCALE
EXPLAINS
GOVERNMENT
UNKNOWN
ESSENTIALLY
POTENTIAL
QUESTION
CRITICISMS
LARGE
KNOWN
PERFORMED
LOWER
EMPLOYEE
MONETARY
QUICKLY
DIFFICULT
SPECIFIC
ECONOMIC
CROWDSOURCE
MODEL
WINNING
CREATING
MOTIVATIONS
IMPLICATIONS
PRODUCE
LABOR
USUALLY
TRADITIONALLY
CREATED
INITIATED
IMPORTANT
GENERATE
USING
SOURCING
GIVEN
WIRED
EFFORT
FOLLOWING
TOGETHER
DATABASE
HISTORY
FREE
ALLOWS
DESIGNED
ORGANIZATIONS
SOLVING
SUPPORT
ESTABLISHED
PURPOSES
WENT
CRITIC
DIFFERENCE
PLANNING
LACK
PARTICIPATING
PHYSICALLY
LANGUAGE

In our context

- Finding study participants online
- Service handles details of recruitment, payment, etc.

Why crowdsource?

- Large numbers of participants
  - Without complicated logistics
  - From around the country, world
- Easily controlled conditions
  - Easily redirect to survey software, your server, etc.
- Relatively inexpensive

Why not crowdsource?

- Limited followups
- Some participants will enter garbage
- Specific demographics participate
  - Younger, more technical than general population
  - Better than recruiting only students!
  - Usually worse than, e.g., Craigslist recruiting

Participant problems

- Attempted repeaters
  - Especially if you pay too much
- Entering garbage / not paying attention / bots
- Discussion in forums
  - What about deception?
- Terms of service may limit request types

Participant solutions

- Collect a lot of data
  - Noise distributed across conditions
- Use cookies, IP tracking, worker IDs
- Ensure there is no “shortcut”
- Use attention check questions, repeats
  - Carefully designed and placed
- Do NOT use well-known “trick” questions
- Monitor forums

More best practices

- Pay fairly!!
- Screen for quality
  - Approval rate, # of hits, etc.
- Consider ethics of the task
- Don’t reject for spurious reasons

- https://wearedynamo.fandom.com/wiki/Guidelines_for_Academic_Requesters

Useful platforms

- MTurk
- Prolific
- Crowdflower

- crowdsource.com
- Samasource (*for training AI models)*
- Google consumer surveys

Redmiles et al., S&P 2019

- Turk vs. web panels vs. RDD phone survey
- Turk + web panels: too online
- Turk: better than web panels for younger
- Neither is great for oldest, least educated
  - But at least web panels have some?

What we did today!

- Overview of quantitative data
- Surveys
    - Crowdsourcing

What’s next?

- Field studies
- Statistical analysis

Logistics + Questions?

- Please email me when you've selected a team
  - Team member names
  - 1-2 sentence description of problem
- Added templates and an example to the project description page
- Instructions for MaxQDA install on Piazza