

Surveys

Research Methods for Human-Centered Computing



Today's goal:

Cover surveys as a research method

- Outline:
 - Goals of surveys
 - How to conduct a survey
 - Analyzing survey data
 - Examples



Goals of surveys When (not) to do a survey



A study in which participants answer a number of questions

- Perceptions, attitudes, and intentions
- Behaviors (self-reported)
- Demographics/socio-economic data (self-reported)

Sampling is similar to an experiment

- Typically a stronger focus on representativeness
- But generally no random assignment



To learn about a certain population and their characteristics Answer univariate questions

Reasons?

- Learn about overall attitudes
- Quantify certain behaviors
- Learn how people do accomplish certain tasks
- Find pain points for design or opportunities for research
- Learn how people use a system/technology



To learn about differences and relationships

- Answer multivariate questions
- Reasons?
 - Discover relationships between variables
 - Find differences between groups of participants
 - Help identify a target population
 - Study a population over time



Don't do a survey if you want to establish causality

- Multivariate analyses of surveys are strictly correlational
- There are no manipulations!
- What causes what?
- Third variable problem

No ceteris paribus

Hard to get rid of confounding variables

Do an experiment instead!



Don't do a survey if you want deep qualitative insights

- You rely on participants to write down enough
- You cannot ask follow-up questions
- People tend to over-analyze their answers

Do an interview instead!

- For confidentiality, a phone interview may be better
- Potential new paradigm: Al-driven interviews

Exception: if you target population is very difficult to reach, a qualitative survey may be your only choice!



Don't do a survey if you want to complex behavioral data Self-reported data is usually pretty good But some behaviors are difficult to remember and report

Reporting biases:

Satisficing, acquiescence bias, desirability bias

Use logging/observation instead!

Get accurate and unbiased



Conducting a survey Steps involved



Administering a survey is easy

Just a bunch of questions... No system, no manipulations

Administering a **good** survey is hard

You must carefully determine the objectives Identify questions that can accomplish these objectives Think about sampling the right participants



What do you want to learn about? Formulate research questions

What are the variables involved in those questions? Create a table of relevant constructs and instruments

What other variables should you include

Make sure you cast a wide net... there is no ceteris paribus!

What is your target population

Sample carefully... again, there is no ceteris paribus!



Careful sampling is very important

- Try to match the target population
- Use a recruitment firm
- Employ stratification techniques

For multivariate questions, use a typical power analysis

For univariate questions, think about the acceptable margin of error



Keep it short

Provide proper context

Make questions easy to answer

Avoid objectionable questions

Break down questions into smaller parts You may use branching logic

Avoid check-all-that-apply questions



Give a proper introduction Including time to complete

Include attention checks

- And if needed, comprehension checks
- Pilot-test your survey
 - Think-aloud
 - Small sample



Analyzing survey data Graphs and statistics



For univariate data:

Values: frequencies, means, histograms Statistics: confidence intervals



Figure 1: Frequency distribution of the extent to which participants in our study used each platform. 9: almost constantly, 8: several times a day, 7: once a day, 6: 3-5 days a week, 5: 1-2 days a week, 4: every few weeks, 3: less often, 2: never but I have an account, 1: I do not have an account.



For multivariate data, split things into sub-groups

- Values: frequencies/ means per group, bar charts
- Statistics: chi-square test, ANOVA



Figure 1. Privacy Attitudes across Platforms



	Privacy Maximizers	Selective Sharers	Privacy Balancers	Time Savers	Self- Censors	Privacy Minimalists	
Experts							
Laperts	13 (5.6) +	6 (3) +	28 (20.4) +	4 (9.4) -	1 (6.1) -	5 (12.6) –	
Near-							
Experts	11 (6.9) +	8 (3.7) +	31 (25.4) +	7 (11.8) -	4 (7.6) -	10 (15.7) -	
Some							
Expertise	1 (4) –	0 (2.1) -	17 (14.6) +	9 (6.8) +	4 (4.4) -	10 (9.1) +	
Mostly					12 (4.2)		
Novice	1 (3.8) –	0(2)-	1 (13.9) –	4 (6.5) -	+	21 (8.6) +	
Near-							
Novices	1 (5.8) –	0 (3.1) –	1 (21.4) –	22 (9.9) +	7 (6.4) +	19 (13.2) +	
Novices							
	3 (3.9) -	2 (2.1) -	22 (14.3) +	5 (6.6) -	5 (4.3) +	3 (8.8) -	

Note: Values represent actual representation in each cluster versus (expected). + indicates larger

class-to-class membership than expected; - indicates smaller than expected. Color gradient

corresponds to the magnitude of this discrepancy (Red is smaller than expected; Green is larger than expected)

Table 8: Class-to-Class Membership Table Showing Discrepancies between Privacy

Proficiency Level and Management Strategy Class Memberships.



Relationships between continuous variables

- Values: scatterplots (with trend lines)
- Statistics: correlations, regression models (with caution)





Do not wait until your data has been collected

- Already start thinking early on what you are going to present in your paper
- It helps you uncover missing questions/constructs
- Are any of your existing questions unnecessary?



Examples Some examples of survey studies



McGregor et al. "Would You Slack That? The Impact of Security and Privacy on Cooperative Newsroom Work"

https://doi.org/10.1145/3134710



Comorio Morrehon	energy of a	Centre 19	Contraction of the second	or o	Centro 4	contraction of the second	Centro e
Scenario Number	ۍ ۲	ۍ ۲	5	2	ي ک	5	ر ب
Email	7	6	4	7	1	10	2
Phone	4	-	4	8	-	3	8
Slack/chat	6	3	4	3	-	1	-
Shared storage	-	10	-	-	8	1	-
F2F	7	-	2	-	-	2	3
Other	-	-	1	1	8	8	-

Table 3. Participants' use of communication/collaboration technologies, by scenario



Page et al. "Don't Disturb My Circles! Boundary Preservation Is at the Center of Location-Sharing Concerns"

http://bit.ly/icwsm2012



"What consistently predicted the absence or presence of privacy concerns turned out not to be the relationship type itself. Rather, it was whether or not the situation would change existing offline relationship boundaries. However, boundaries (and thus the ensuing activities) change when the associated relationships change, even when the 'who' stays the same: when acquaintances become good friends sharing may increase, while sharing may slowly decrease when relationships dry up. In turn, what was once a privacy concern may no longer be, and new concerns may appear where they were absent."



BPC	I'm worried LSS will change my relationship with others					
C1	I am bothered that others share so much information with me					
C2	I am concerned that if I share too much information, I would bother others					
C3	I worry that I might share information with more people than I intend to					
C4	I worry about feeling compelled to interact with others online					
C5	I worry that what my friends share will reflect badly on me					
C6	I'm worried about knowing the social etiquette of using LSS (e.g. who to friend, what to share, etc.)					
C7	I'm concerned about being able to control who sees my location					
C8	I'm worried others would join me at an inappropriate time if I share my location					

Table 1 Questionnaire items considered in our current analyses



Figure 1: The path model, in which boundary preservation concern (BPC) serves as a cause of the other privacy concerns (C1-C8). The numbers on the arrows represent the standardized effect sizes; *** indicates a significance level of p < .001.



Page et al. "How Communication Style Shapes Relationship Boundary Regulation and Social Media Adoption"

https://doi.org/10.1016/j.ijhcs.2013.06.003



Differences between platforms in terms of

- Boundary preservation concern
- Boundary enhancement Platform usage



Figure 3: Model of structural relationships in our study. Factor scores are standardized and fixed to zero for Facebook (FB). Error bars and values in parentheses are SEs.



Namara et al. "Emotional and Practical Considerations Towards the Adoption and Abandonment of VPNs as a Privacy-Enhancing Technology"

(in press)







Thirteen respondents are wary of their ISP or their government covertly monitoring their online activities. As such, they reported being compelled to use VPNs to safeguard their privacy and also to quell their fears of online surveillance ("I wanted to be safer from government and ISPs level logging and profiling, and disliked the impingement on my privacy and freedoms. Snowden and the UK 'Snooper's Charter' firmed this resolve."; "I do not like when my data is tracked."; "[I used a VPN for] Privacy from ISP and forums.")



Knijnenburg et al. "Dimensionality of Information Disclosure Behavior"

https://doi.org/10.1016/j.ijhcs.2013.06.003







"Who?" = Five disclosure profiles 1 Act Loc Con Inter

159 pps tend to share little information overall (LowD)
26 pps tend to share activities and interests (Act+IntD)
50 pps tend to share location and interests (Loc+IntD)
65 pps tend to share everything but contact info (Hi-ConD)
59 pps tend to share everything