Advanced Survey Design (S11)
18 – 22 August 2025

Course Director: *Dr. Peter Lugtig*
E-mail: *ms.summerschool@uu.nl*

This 5-day course in survey design takes student beyond the introductory courses offered in BA and MA programmes, and discusses the state-of-the-art of one of the most important data collection techniques: surveys. The course focuses on the methodology of how to do surveys, and the use statistical techniques to analyse and correct for some specific survey errors. It combines short 1-hour lectures with exercises on most of the topics discussed. We assume course participants are proficient in working with R. Most of the exercises can also be done with STATA or SPSS, but answers will be provided in R

The course assumes basic knowledge of:

-   Basic knowledge of social science research methodology
-   Multivariate statistics up to the General Linear Model
-   The basics of survey methodology (the basic of sampling questionnaire design, collecting and processing data)

Background readings for the course are:

Aggarwal, C.C. (2018) *Machine learning for text*. Springer. ISBN: 978-3-319-73530-6, doi: 10.1007/978-3-319-73531-3 **(day 4)**

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis*, *31*(3), 337–351. doi:10.1017/pan.2023.2 **(day 4)**

Antoun, C., Katz, J., Argueta, J., & Wang, L. (2018). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review*, *36*(5), 557-574 **(day 2)**

Biemer, P.P., de Leeuw E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L., Tucker, N.C., West, B., eds. (2017) *Total Survey Error in Practice*, Wiley, especially chapters 2 and 7 **(days 1, 3)**

Boeschoten, L., Ausloos, J., Möller, J. E., Araujo, T., & Oberski, D. L. (2022). A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research*, *4*(2), 388-423. (day 3)

Brunsdon, C. & Comb, L. (2019) *An introduction to R for spatial analysis and mapping (Spatial analysis and GIS)*. (2nd edition). Sage, London. ISBN-13: 978-1526428509 **(day 5)**

Dillman, D.A., J.D. Smyth, and L.M. Christian (2009) *Internet, Mail and Mixed-Mode: The Tailored Design Method*, 3rd Edition. Wiley and Sons, chapters 4 and 5 especially **(day 1,2)**

Dillman, D. A. (2017). The promise and challenge of pushing respondents to the web in mixed-mode surveys. *Survey Methodology*, *43*(1), 3-31.

Foster, Ian, et al., eds. (2016) *Big data and social science: A practical guide to methods and tools*. CRC Press **(day 3,4)**

Fowler, F.J. (1996) Improving survey questions – design and evaluation. London, Sage, Chapters 1-6 **(day 2)**

Fritz, J., Piccirillo, M., Cohen, Z. D., Frumkin, M., Kirtley, O. J., Moeller, J., ... & Bringmann, L. (2023). So you want to do ESM? Ten essential topics for implementing the Experience Sampling Method (ESM). *Advances in Methods and Practices in Psychological Science*. **(day 2)**

Groves, R.M. et al. (2009), *Survey Methodology,* 2nd edition. New York: Wiley **(background reading)**

Hox, J.J. (1997) From theoretical concept to survey question. In: Survey Measurement and Process Quality Ed. By L. Lyberg, P. Biemer, M. Collins, E. D. De Leeuw, C. Dippo, N. Schwarz, D. Trewin. Wiley, p. 47-69. **(day 2)**

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., ... & Usher, A. (2015). Big data in survey research: AAPOR task force report. *Public Opinion Quarterly*, *79*(4), 839-880**. (day 3)**

Keusch, F., Wenz, A., & Conrad, F. (2022). Do you have your smartphone with you? Behavioral barriers for measuring everyday activities with smartphone sensors. *Computers in Human Behavior*, *127*, 107054. **(day 4)**

De Leeuw, E. D., J. J. Hox, and D. Dillman (2008). *International Handbook of Survey Methodology.* New York, chapters 17 & 19. **(background reading)**

De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys*. Journal of official statistics,* 21(5), 233-255. **(day 1)**

Liu, R., Geng, J., Peterson, J. C., Sucholutsky, I., & Griffiths, T. L. (2024). Large language models assume people are more rational than we really are. *arXiv preprint arXiv:2406.17055.* **(day 4)**

Lohr, S. L. (2021). *Sampling: design and analysis.* Chapman and Hall/CRC. **(background reading)**

Lynn, P. (2020). Evaluating push-to-web methodology for mixed-mode surveys using address-based samples. In *Survey Research Methods* (Vol. 14, No. 1, pp. 19-30). **(day 1)**

Meng, X. L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *The Annals of Applied Statistics*, *12*(2), 685-726 **(day 2)**

McCool, D., Lugtig, P., Mussmann, O., & Schouten, B. (2021). An app-assisted travel survey in official statistics: Possibilities and challenges. *Journal of Official Statistics*, *37*(1), 149-170. **(day 4)**

Presser, S. , M.P. Couper, J.T. Lessler, E. Martin, J. Martin, J.M. Rothgeb, and E. Singer (2004) "Methods for Testing and Evaluating Survey Questions", *Public Opinion Quarterly*, 68 (1): 109-130. **(day 2)**

Schouten, B., Calinescu, M., & Luiten, A. (2013). Optimizing quality of response through adaptive survey designs. *Survey methodology*, *39*(1), 29-5 **(day 2)**

Schouten, B., Peytchev, A., & Wagner, J. (2017). *Adaptive survey design.* Chapman and Hall/CRC. **(day 2)**

van Driel, I. I., Giachanou, A., Pouwels, J. L., Boeschoten, L., Beyens, I., & Valkenburg, P. M. (2022). Promises and pitfalls of social media data donations. *Communication Methods and Measures*, *16*(4), 266-282. **(day 3)**

Waal, T. D. (2016). Obtaining numerically consistent estimates from a mix of administrative data and surveys. *Statistical Journal of the IAOS*, *32*(2), 231-243. **(day 5)**

De Waal, T., van Delden, A., & Scholtus, S. (2020). Multi-source statistics: basic situations and methods. *International Statistical Review*, *88*(1), 203-228. **(day 5)**

Wrzus, C., & Neubauer, A. B. (2022). Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment*, 10731911211067538. **(day 2)**

More specific reading materials will be references in the course slides, which will be available to participants at the start of the course. These more specific readings are recommended if students want to go into more depth into specific issues.

Day-to-day program:

Monday, 28 August			
Time	**Activity (topic and lecturer)**	**Description**	**Location**
09:00-10:00	Lecture	Introduction to the Total Survey Error Paradigm	Utrecht Science Park, Exact location to be announced
10:00-11:00	Lecture	Mixed mode surveys	
11:30-12:00	Exercise	Exercise: study design and minimizing Total Survey error (in groups)	
12:00-13:00	Lunch (included)	Lunch is at the University mensa daily (included)	
13:00-14:00	Lecture	Push-to-web surveys	
14:00-15:00	Lecture	Sampling frames, time, costs, and recruitment strategies	
15:00-16:00	Exercise	Designing a recruitment strategy	

Tuesday, 29 August			
Time	**Activity**	**Description**	**Location**
09:00-10:00	Lecture	Advanced questionnaire design	tba
10:00-11:00	Lecture	Designing for mixed mode surveys	
11:00-12:00	Exercise	Designing for mixed-device surveys	
12:00-13:00	Lunch		
13:00-14:00	Lecture	Adaptive survey designs	
14:00-15:00	Lecture	Ecological Momentary assessment (experience sampling)	

15:00-16:00	Exercise	Exercise: questionnaire design for a modern survey	

Wednesday, 30 August			
Time	**Activity**	**Description**	**Location**
09:00-10:30	Lecture	Big data and TSE: - Digital trace data - Organic data - Administrative data	**tba**
10:30-11:00	Lecture	Designed big data	
11:00-12:00	Exercise	Exercise on selection bias in (designed) big data	
12:00-13:00	Lunch		
13:00-14:00	Lecture	Data donation	
14:00-15:00	Exercise	Data donation using your own data	
14:00-16:00	Exercise	Exercise: Analyzing digital trace data (computer exercise)	

Thursday, 31 August			
Time	**Activity**	**Description**	**Location**
09:00-09:30	Lecture	Passive data collection using mobiles (sensors)	**tba**
09:30-10:30	Lecture	Conducting surveys with apps, sensors: IT, ethics, consent, willingness	
10:30-12:00	Exercise	Exercise: geo-data or accelerometer data (choose 1) (introductory exercise)	
12:00-13:00	Lunch		
13:00-15:00	Lecture	Large Language Models and surveys	
15:30-16:00	Exercise	Exercise: Object recognition, text exercises, LLMs (choose 1)	

Friday, 1 September			
Time	**Activity**	**Description**	**Location**
09:00-10:00	Lecture	Data integration at level of individual respondents. Probability and non-probability integration	**tba**
10:00-11:00	Lecture	Data integration at level of sample	
11:00-12:00	Exercise	Data integration	
11:00-16:00	Exercise	Your own project. Consultations with teachers of the course to discuss your survey questions in more depth. You may bring your own dataset, questionnaire or study design to discuss. Alternatively, there is time to finish some of the exercises earlier or read specific literature	
12:00-13:00	Lunch		

Please bring your laptop to the course with R (version 4.0 or higher) installed and the ability to install packages in R. You may also bring Pyhon, STATA or SPSS on your computer if you want to (also) do exercises in SPSS or STATA, but answers will be provided in R.

Coffee, tea will be provided throughout the course

For information about the Social Programme, please visit the Utrecht Summer School website!