

**Advanced Survey Design**  
**30 August – 03 September 2021**  
**Week 35**

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 of 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 5)**
- 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, 2)**
- Brunsdon, C. & Comb, L. (2019) *An introduction to R for spatial analysis and mapping (Spatial analysis and GIS)*. (2<sup>nd</sup> 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*, 3<sup>rd</sup> Edition. Wiley and Sons, chapters 4 and 5 especially **(day 2)**
- Foster, Ian, et al., eds. *Big data and social science: A practical guide to methods and tools*. CRC Press, 2016 **(day 2, 4)**
- Fowler, F.J. (1996) *Improving survey questions – design and evaluation*. London, Sage, Chapters 1-6 **(day 2)**
- Groves, R.M. et al. (2009), *Survey Methodology*, 2<sup>nd</sup> edition. New York: Wiley **(days 1-3)**
- 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)**
- Japac, 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.
- 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 1,4)
- Kreuter, F. (Ed.). (2013). *Improving surveys with paradata: Analytic uses of process information* (Vol. 581). John Wiley & Sons **(day 1,2, 4)**
- De Leeuw, E. D., J. J. Hox, and D. Dillman (2008). *International Handbook of Survey Methodology*. New York, chapters 17 & 19. **(days 1-3)**
- 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 2)**
- Lohr, S. (2009). *Sampling: design and analysis*. Nelson Education **(day 1 and 3)**

Lynn, P. (1996) Weighting for non-response. In Totman et al *et al.* Survey and statistical computing, available on:

<http://iserwww.essex.ac.uk/home/plynn/downloads/Lynn%201996%20Weighting.pdf> (day 3)

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)

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York: Springer (day 3)

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:

*Please note that the schedule is in Central European Summer Time*

Monday, 30 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 will be announced later
10:00-11:00	Lecture	Types of data and their relation to Total Survey Error: - Designed data - Organic data - Designed big data	
11:00-12:00	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	Choosing an appropriate sampling frame and sampling design. Registers, geodata and digital trace data.	
14:00-15:00	Lecture	Sampling designs: statistical efficiency, survey costs and survey practice	
15:00-16:00	Exercise	Working out a sampling design (computer exercise )	

Tuesday, 31 August			
Time	Activity	Description	Location
09:00-10:00	Lecture	Advanced questionnaire design	Utrecht Science Park, exact location will be announced later
10:00-11:00	Lecture	Mixing the modes	
11:00-12:00	Exercise	Designing for mixed mode surveys	
12:00-13:00	Lunch		
13:00-14:00	Lecture	Mobile and mixed-device surveys	
14:00-15:00	Lecture	Questionnaire design for mixed-device surveys	
15:00-16:00	Exercise	Exercise: questionnaire design for mixed-device surveys	

Wednesday, 01 September			
Time	Activity	Description	Location
09:00-10:30	Lecture	Weighting to correct for survey nonresponse	Utrecht Science Park, exact location will be announced later
10:30-11:00	Lecture	Paradata: what is it and how to use it?	
11:00-12:00	Exercise	Creating poststratification weights (computer exercise)	
12:00-13:00	Lunch		
13:00-14:00	Lecture	Sampling, coverage and nonresponse weights	
14:00-15:00	Exercise	Raking, combining weights (computer exercise)	
14:00-16:00	Exercise	Imputation or weighting (computer exercise)	

Thursday, 02 September			
Time	Activity	Description	Location
09:00-09:30	Lecture	Surveys and big data	Utrecht Science Park, exact location will be announced later
09:30-10:30	Lecture	Passive data collection using mobiles (sensors)	
10:30-11:00	Lecture	Ethics, consent, willingness	
11:00-12:00	Exercise	Introduction to working with geo-data or accelerometer data (choose 1) (computer exercise)	
12:00-13:00	Lunch		
13:00-14:00	Lecture	Sampling revisited: design-based vs. model based inference and effects on Total Survey Error?	
14:00-15:00	Exercise	Exercise on model-based inferences from (non) probability samples	
15:30-16:00	Exercise	Continue exercise from morning or afternoon	

Friday, 03 September			
Time	Activity	Description	Location
09:00-10:00	Lecture	Working with text or picture data	Utrecht Science Park, exact location will be announced later
10:00-11:00	Exercise	Object recognition, text recognition, text exercises (introductory exercise)	
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 laptops to the course with R (version 3.6 or higher) installed and the ability to install packages in R. You may also bring 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!](#)