

## Syllabus: Survey Research Spring 2017

*Last updated: March 1, 2017*

---

### Course information

- Course title: Survey Research
- Course number: SRPP-AD 120
- Credits: 4
- Term: Spring 2017 (14 weeks)
- Lecture Location: Building Comp Research (A2), Room 002
- Lecture Hours: Tuesday and Thursday: 1150-105pm
- Office Hours: By appointment
- Course Prerequisites: Basic understanding of statistics

### Instructor information

- Instructor: Dr. Peter van der Windt
- Email: [petervanderwindt@nyu.edu](mailto:petervanderwindt@nyu.edu)
- Website: [www.petervanderwindt.com](http://www.petervanderwindt.com)
- Office: Building A5, Office 147
- Assistance for Stata and pre-analysis plan: Dr. Jonathan Rogers
  - Email: [jdr7@nyu.edu](mailto:jdr7@nyu.edu)
  - Office hours: Thursday: 2-3pm
  - Office: A5, 125

### Course description

This course will teach students how to design and implement a survey. We discuss a wide set of topics: question design, sampling, non-response bias, survey-experiments, etc. You will create your own pre-analysis plan for a survey. We will spend about a quarter of the course learning Stata (among others to do sampling). You will also learn how to use Qualtrics (to collect data online). At the end of this course you will be able to design and implement your own, high quality survey. Moreover, you will question much of the data that is collected by others because you know all the things that can go wrong in the process. For this course I assume that you are comfortable with basic statistics.

### Course material

- Reading: We will use “Survey Methodology”, 2nd Edition, by Robert M. Groves, Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, Roger Tourangeau. I will also assign additional readings.

- Software: We will use the statistical software package Stata (make sure you have access to it throughout the course). You learn Stata by using it, and all questions you might have are answered online. As a result, we will not use a book to learn Stata, but a Stata manual I created for this course and the internet. We will also make use of the program “Qualtrics”, which is available via the NYU website.

### Course setup

There are two types of classes:

- Lectures [“L”]: You will be expected to have done the required readings before class to the point where you can be called on to critique or defend the reading. During the class, please question me and the material. If something does not make sense to you then stop the class and say so — it probably doesn’t make sense to others either.
- Practical class [“P”]: These classes are hands-on. Have your computer with you! During these classes each student will sit in front of his/her own laptop and learn Stata, ODK or Qualtrics.

### Course grading and assignments

Your final grade consist out of four parts. Somewhere in the middle of the semester, I’ll give you an overview of your grades thus far.

- Class participation (20%): The grade for this component is based on your preparation and participation in the classes (L and P). If you are always present in class you receive a ‘B+’. If you also participate in class this grade will go up. If you cannot be present during a class, send me an email *before* the class and include the reason. You are allowed to miss one class. After that it will lower your grade at my discretion.
- Mid-term exam (20%): In the middle of the semester (see schedule below) you will make a short midterm exam in class. The exam will measure your skill with Stata, and your knowledge of some key concepts that we learned in class the weeks before. The exam will consist of two parts: 1) theory, and 2) Stata. The exam will be open book and open-internet.
- Final exam (30%): There will be an exam at the end of the course (during the exam week). The exam will measure your skill with Stata, and your knowledge of key concepts that you learned during the course. The exam will consist of two parts: 1) theory, and 2) Stata.
- Pre-analysis plan (30%): During the course you will develop your own pre-analysis plan (PAP). You will create a research design for a survey and the necessary protocols for implementation. The PAP will be on a topic of your own interest. The PAP should address all the major topics that we cover in this course. More information is given in the appendix. At one point in time during the class I want you to hand in a draft of your PAP, so that we can give advise and a letter grade (that will count as 1/3 of your final PAP grade).

### Learning R

R ([www.r-project.org/](http://www.r-project.org/)) is another, very popular statistical software package. In contrast to Stata it is open access, and more flexible (and if done well, also more powerful). However, it is yet another computer language to learn, and also a bit more difficult than Stata. If you’re interested, send me an email and I can organize a class outside of the schedule to introduce you to this program. It will not be obligatory to attend.

### Course schedule

The next part provides the course schedule by week. This is the course in broad lines: readings and topics can change as we go along (among others based on your input). There are four parts to this course: “Measurement”, “Design”, “Administration”, and “Afterwards”.

Week 1: Introduction, course overview & stats refresher

- Tue, Jan24, [L]: Why you should not trust data. The four key errors: 1) coverage error; 2) sampling error; 3) non-response error; 4) measurement error. Your pre-analysis plan. Research questions and hypotheses.
  - Watch: <https://www.youtube.com/watch?v=Pf70Hg1UMYA>
  - Read: chapters 1 and 2 of Groves et al. (2009)
- Thu, Jan26, [L]: Statistics refresher.

---

## Measurement

---

Week 2: Question design

- Tue, Jan31, [L]: From hypotheses to variables and questions. How not to ask questions.
  - Read: chapter 7 of Groves et al. (2009); Payne (1950)
- Thu, Feb2, [L]: How to measure stuff that is hard to measure? Crowdsourcing. List randomization. Endorsement experiments. Random-response method.
  - Read: Blair (2015); Windt and Humphreys (2016)

Week 3: More question design

- Tue, Feb7, [L]: Survey experiments. Causal inference.
  - Read: <http://egap.org/methods-guides/10-things-you-need-know-about-causal-inference>
  - Optional: chapter 14 in Shadish et al. (2002); Barabas and Jerit (2010); Holland (1986); Gaines et al. (2007)
- Thu, Feb9, [L]: Evaluating question quality
  - Read: chapter 8 of Groves et al. (2009)

Week 4: A week full of Stata

- Tue, Feb14, [P]: Stata. Introduction. Importing Data.
  - Read: We will distribute a ‘Stata manual’ before class.
  - Make sure to have your laptop with you
- Thu, Feb16, [P]: Stata. Descriptive analysis. Manipulating data.
  - Read: The Stata manual
  - Make sure to have your laptop with you

Week 5: Even more on question design, and an example survey

- Tue, Feb21, [L]: How to measure female empowerment?
  - Read: Van der Windt (2016)
- Thu, Feb23, [L]: Let’s check the Arabbarometer
  - Check: [www.arabbarometer.org](http://www.arabbarometer.org)

Week 6: Ethics, and Stata again

- Tue, Feb28, [L]: Ethics. Publication bias.
  - Read: Gerber and Malhotra (2008)
- Thu, Mar2, [P]: Stata. Manipulating data
  - Read: The Stata manual

Week 7: More Stata, and review

- Tue, Mar7, [P]: Stata. Manipulating data
  - Read: The Stata manual
- Thu, Mar9, [L/P]: Mid-term exam
- PAP: By the end of this week (before Thursday night 23:59), please hand in Sections 1-4 and Appendix A via email.

Week 8: Break

- Tue, Mar14: No class (Final exams for 7-week courses)
- *Thu, Mar16: No class (Final exams for 7-week courses)*

---

## Design

---

Week 9: Break

- *Tue, Mar21: No class (Spring break)*
- *Tue, Mar23: No class (Spring break)*

Week 10: Sampling

- Tue, Mar28, [L]: Population. Sampling frame. Sampling. Blocking/stratification. Snowball sampling.
  - Read: chapter 3 of Groves et al. (2009)
- Thu, Mar31, [“L”]: Multistage sampling. Selection weights.
  - Read: chapter 4 of Groves et al. (2009)

Week 11: Another week of Stata

- Thu, Apr4, [P]: Stata. Sampling
  - Read: The Stata manual
- Thu, Apr6, [P]: Stata. Sampling
  - Read: The Stata manual

Week 12: “Power”

- Tue, Apr11, [L]: Number of observations. Power. Power calculation. Budget.

- Read: TBD
- Thu, Apr13, [P]: Stata. Power calculation in Stata.
  - Read: The Stata manual
  - Make sure to have your laptop with you

Week 13: Qualtrics

- Tue, Apr18, [P]: Qualtrics (in class).
  - Check: <https://nyu.qualtrics.com/>
  - Make sure to take your laptop with you
- Thu, Apr20, [L/P]: Review session

---

### Administration

---

Week 14: Issues related to data collection

- Tue, Apr25, [L]: Non-response. Monitoring data collection.
  - Read: chapter 6 of Groves et al. (2009)
- Thu, Apr27, [L]: Survey sponsorship. Interviewer effects.
  - Read: chapter 9 of Groves et al. (2009)
  - Read: Corstange (2014)

Week 15: Ethics again, and “protocols”

- Tue, May2, [L]: Ethics. IRB. Informed consent.
  - Read: chapter 11 of Groves et al. (2009)
- Thu, May4, [L]: How everything will go wrong during data collection. Protocols. Dictionary.
  - Read: chapter 5 of Groves et al. (2009)

---

### Afterwards

---

Week 16: Did you collect high quality data?

- Tue, May9, [L]: Recall. Back-checking. Survey quality and data accuracy. Benford’s law.
  - Read: De Nicola and Giné (2014)
- Thu, May11, [L]: What to do with the data when you have it? Storage.
  - Read: chapter 10 of Groves et al. (2009)

## Appendix: Pre-analysis plan

During the course you will develop your own pre-analysis plan (PAP). The PAP will be on a topic of your own interest (yes, you can use a topic that you are thinking of for your capstone). The PAP should address all the major topics that we cover in this course. Your research design should consist out of the following 12 sections and 3 appendices (expected length in parentheses, single spaced sentences):

- Top of the first page: Title of the project and your name and contact details
- Section 1: Research question (max a few lines)
- Section 2: Motivation and policy relevance, and literature review (1 page)
- Section 3: Hypotheses (1/2 page)
- Section 4: Outcomes of interest. Independent variables (1 page)
- Section 5: Units under study (1/2 page)
- Section 6: Sampling strategy (1 page)
- Section 7: Sample size and power calculation (1 page)
- Section 8: Budget (1 page)
- Section 9: Survey management: survey mode, enumerators, etc. (1 page)
- Section 10: Ethics (1/2 page)
- Section 11: Timeline and deliverables (1 page)
- Section 12: Limitations and risks (1 page)
- Appendix A: Survey (2 pages)
- Appendix B: Dictionary (1 page)
- Appendix C: Surveyor Protocol (1 page)

Writing a PAP is not easy, but an important skill to have. It follows the same rules that you would use for writing an academic paper. Below I list several of them:

The Do's:

- Follow the guidelines. If I ask for you to hand in sections 1-4 and appendix A, please hand in sections 1-4 and appendix A. Nothing more. Nothing less.
- Incorporate the comments we give you.
- Use writing that people understand. No need to write more complicated than necessary. "I have set out to argue" can also be written as "I will argue".
- Be academic in your writing. Do not use words like "fascinating", "amazing", etc. That's your opinion. Readers don't care about your opinion. The data will provide evidence.
- It also means that you have to be precise. Avoid words like "a lot" or "in the beginning" if you can be more precise.
- Use references. If you use a number somewhere (e.g. in Congo unemployment is 95%), you need to add a source.
- Use the same tense throughout your document. If you start writing in the present tense, do that throughout your document.

- Hypotheses should follow from the literature/motivation section. Your hypotheses should not come out of the blue. If your hypothesis is “Women in the Congo work more than men in the Congo”, your literature/motivation section should not only contain information about poverty in the Congo, but also about differences between men and women in the Congo.
- In Section 4, for each of the variables that you’ll use, say which questions in your survey it relates to.
- If you have many hypotheses, you will likely have many independent and dependent variables. Either write things out clearly in Section 4, like “For hypothesis 1 my independent variable is X1 and my dependent variable is Y1. I’ll measure them as follows... For hypothesis 2 my independent variable is X2 and my dependent variable is Y2. I’ll measure them as follows...” If you have many hypotheses, it is probably cleanest to provide a table with the following columns: hypothesis, independent variable, dependent variable and question in the survey.
- All pages should have a page number.

Some Do’s when it comes to your survey:

- Make sure your survey has ‘starting’ and ‘ending’ questions. Starting questions are things like ‘Location’, ‘Date’, ‘Start time’, ‘Enumerator name’, etc. Ending questions are things like ‘Did the respondent pay attention’, ‘End time’, etc.
- Every question should have a question number.
- Every question, and really every question, should have the following options also as possible response: ‘not applicable’, ‘refuses to respond’, ‘don’t know’.
- Do the layout of the survey. Have it in such a state that I could print it and give to a respondent.
- Add the possible answers to your questions in your survey.

The Dont’s:

- Do not have hypotheses that are a combination of multiple hypotheses. E.g. “X leads to increases in Y and does not decrease Y.” These are two hypotheses. Write them out as two sentences.
- Do not make value judgements. We are academics.
- Avoid contractions: write “they’ve” as “they have”, “can’t” “as cannot”, etc.
- Avoid writing the full name of the authors, the title of their book or article, etc. Use referencing as is given in the example on the next page.
- No need for a coverpage and pictures. That’s for secondary school.

A (not fully worked out) example is given on the next two pages, just to give you an idea of what you have to hand in. Needless to say, your PAP will have more meat and will be more specific.

Examples of real pre-analysis plans can be found here (please have a look around): <http://egap.org/content/registration>. As you can see, they come in many different flavors but most of them share the main ingredients discussed below.

Title: The Role of White Socks on School Performance  
Name: Peter van der Windt<sup>1</sup>

## 1 Research Question

This study explores the following question: Does wearing white socks increase school performance?

## 2 Motivation and policy relevance, and literature review

Socks are important. They keep the feet warm. Each morning millions of people put on socks. Over 1.7 billion individuals put on socks daily.<sup>2</sup> Exploring the impact of wearing socks and what type of socks will thus have an impact on a large number of people.

There has been a lot of work of the impact of different types of clothing on human behavior. We know that warm jackets lead to improvements in reaction time (see for example Sircar (2014)). And Freie (2007) has found that wearing a hat improves the ability to cook. We know much less about the role of socks. Edwards and Frizella (2012) finds that the color of the sock leads to improvements in driving skills. Particularly those that wear white socks score well.

Despite the importance of socks, we know little about how sock color influences school performance. We aim to explore this with this study.

## 3 Hypotheses

This study will set out to test the following two hypotheses:

H1: Wearing socks improves school performance

H2: Wearing white socks improves school performance most compared to other color socks

## 4 Outcomes of interest and independent variables

The outcome of interest (the dependent variable) for this study is school performance. We will measure this using the GPA score of a student (Q7 in the survey).

The independent variable of interest is whether an individual wears socks, and the color of the sock (Q9 and Q10 in the survey).

## 5 Population and units under study

This study aims to generalize towards the general student population.

---

<sup>1</sup>Email: petervanderwindt@nyu.edu

<sup>2</sup>See Mahler (2009).



For this study we focus on students at New York University Abu Dhabi (NYUAD). We have obtained a list of all the students at NYUAD and this will be our sampling frame.

## 6 Sampling strategy

We make use of stratified random sampling. We stratify by age and gender.

## 7 Sample size and power calculation

For our power calculation we made the following assumptions: XYZ. It follows that we will need to conduct a survey with 234 individuals.

## 8 Budget

The cost are limited because it will be conducted on campus. I will need to hire a research assistant (\$600), surveys have to be printed (\$150), and I will give a small incentive for respondents (\$150). This will add up to an estimated \$900.

## 9 Survey management

The survey will be conducted by me and my research assistant. Answer questions related to socks is very sensitive to the gender of the interviewer. As a result, the research assistant will be a women, so that she will conduct the interviews with women and I will conduct them with men.

NYUAD students are very busy. As a result, we expect a large refusal rate of the surveys (see earlier power calculation). In response, the survey takes on average twenty minutes and students receive a small incentive. Moreover, we will randomly visit half of the students in the morning and the other half in the evening. As a result we can test (experimentally) the importance of survey timing on response rates.

NYUAD students also travel a lot. Therefore I expect a lot of non-response. To avoid this, all people will be visited during weeks in which there are classes.

The survey will be conducted using tablets. I have created the survey on ODK.<sup>3</sup> The questions and question numbers are presented in the appendix.

## 10 Ethics

Asking an individual whether he or she wears socks and color socks can be a sensitive topic. However, we will explain to each individual that he/she can choose not to answer a question if he/she wants to. The complete informed consent text can be found in the appendix.

I have obtained IRB approval at New York University Abu Dhabi.

---

<sup>3</sup>The excel is available upon request.

## 11 Timeline and deliverables

I expect the project to take six months:

- January: prepare survey
- February: present survey design
- March-April: data collection
- May: data cleaning and analysis
- June: finish report

As output I will produce the following that I will disseminate widely:

- Survey
- Academic paper that I aim to publish
- Policy report that will be presented at different venues

## 12 Limitations and risks

Finally, we have to be clear about the limitations and risks of this study.

We aim to generalize to the larger population of students but we only conduct a survey with NYUAD students.

We do not learn about the causal impact of socks and color. We only obtain a correlation. There can be many other issues at play: e.g. maybe smarter people wear white socks.

On the next pages you will find the survey, the dictionary and the surveyor protocols.

## A Appendix: Survey

(Layout still has to be done.)

| Q   | Question                              | Response    |
|-----|---------------------------------------|-------------|
| Q0  | Surveyor                              | Peter Other |
| Q1  | Date?                                 |             |
| Q2  | Person is present?                    | 0 1         |
| Q3  | Person is present the second time?    | 0 1         |
| Q4  | Person is willing to participate?     | 0 1         |
| Q5  | Nationality?                          |             |
| Q6  | Age?                                  |             |
| Q7  | What is your GPA?                     |             |
| Q8  | How many books do you read per month? |             |
| Q9  | Are you wearing socks at the moment?  |             |
| Q10 | If yes, what color socks              |             |
| Q11 | Gender?                               | Male Female |
| Q12 | Do you like NYUAD?                    | 0 1         |
| Q13 | Time of ending of survey?             |             |

## B Appendix: Dictionary

I will print the following dictionary, and give this to the surveyor. It serves as a guide during the survey implementation.

| Individual's name | Location | Surveyor           | Time to be interviewed |
|-------------------|----------|--------------------|------------------------|
| Johny Walberg     | A5 132   | Peter              | Morning                |
| Eddy Getnam       | A4 021   | Research assistant | Morning                |
| Fred Edelberg     | A2 213   | Research assistant | Evening                |
| Albert Fres       | A2 212   | Peter              | Morning                |
| Emilia Bonard     | A3 213   | Research assistant | Evening                |

Etc.

## C Appendix: Surveyor Protocol

Enumerator please follow the following steps carefully.

- Knock on the door. Only leave after three times knocking.
- If the person is not there record this on the survey, and return the next day the exact same time.
- If the person is there tell the following consent information:
  - *“Hello. I am [Name]. Before we begin, I would like to take a minute to explain why I conduct this survey and what I will be doing with the information you provide to me. Please stop me at any time if you have any questions. After I’ve told you a bit more about our project, you can decide whether or not you would like to participate. This research is being conducted by Peter van der Windt in collaboration with New York University Abu Dhabi. We are collecting information on school performance and sock use and color. The researcher will use the information collected in articles that might be published, as well as in academic presentations. Participation should take about 20 minutes. Participation are on a purely voluntary basis. There are no risks to you from answering these questions. The information we collect today is private and confidential. We will not share any details from the survey with anyone besides the research team. If at any time and for any reason, you would prefer not to answer any questions, please feel free not to. If at any time you would like to stop participating, please tell me. We can take a break, stop and continue at a later date, or stop altogether. You will not be penalized in any way for deciding to stop participation at any time. If you have questions, you are free to ask them now. If you have questions later, you may contact me by calling [Phone number]. Are you willing to participate?”*
- If the person says ‘no’, say ‘thank you’ and leave. Fill out the first few questions of the survey.
- If the person says ‘yes’, continue with the survey. Fill out the (complete) survey.
- Make sure you are alone with the respondent when you ask the questions.

## References

- Barabas, J. and Jerit, J. (2010). Are Survey Experiments Externally Valid? *American Political Science Review*, 104(02):226–242.
- Blair, G. (2015). Survey Methods for Sensitive Topics. *Comparative Newsletter*, pages 12–16.
- Corstange, D. (2014). Foreign Sponsorship Effects in Developing-World Surveys: Evidence from a Field Experiment in Lebanon. *Public Opinion Quarterly*, (forthcoming).
- De Nicola, F. and Giné, X. (2014). How Accurate Are Recall Data? Evidence from Coastal India. *Journal of Development Economics*, 106:52–65.
- Edwards, F. and Frizella, J. (2012). Does Sock Color Influence Driving Behavior? Evidence from San Francisco. *Journal of Nonsense*, 123(2):12–27.
- Freie, J. (2007). *Socks Around Your Feet*. Doesn’t Exist, New York City.
- Gaines, B. J., Kuklinski, J. H., and Quirk, P. J. (2007). The Logic of the Survey Experiment Reexamined. *Political Analysis*, 15(1):1–20.
- Gerber, A. S. and Malhotra, N. (2008). Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals. *Quarterly Journal of Political Science*, 3:313–326.
- Groves, R. M., Fowler, F. J. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2009). *Survey Methodology*. Wiley, New York City, 2nd edition.
- Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396):945–960.
- Mahler, F. (2009). Socks in the World. *Journal of Nonsense*, 21(3):54–33.
- Payne, S. L. (1950). Thoughts About Meaningless Questions. *Public Opinion Quarterly*, 14(4):687–696.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company, New York City.
- Sircar, M. (2014). Put on a Jacket: Experimental Evidence on Reaction Time and Jacket Use. *Journal of Nonsense*, 12(2):32–33.
- Van der Windt, P. (2016). Can Development Lead to Women Empowerment? Evidence from a Field Experiment in the Congo. *Working paper*.
- Windt, P. V. D. and Humphreys, M. (2016). Crowdfunding in Eastern Congo: Using Cell Phones to Collect Conflict Events Data in Real Time. *Journal of Conflict Resolution*, 60(4):748–781.