2 Survey Data Collection and Data Quality

Ioannis Andreadis

The main aim of the DataPopEU project, as it has already been described in the introduction of this volume, is to use innovative methods to produce high-quality data for populism and Euroscepticism research. Innovative methods both in the phase of data collection and in the phase of data processing have been applied, along with the preparation for the subsequent analyses. In this chapter, the methods used to collect the data for the Hellenic National Election Voter Study in 2019 and the methods used to check the quality as well as clean the data of all surveys conducted in this project are presented.

2.1 The Hellenic National Election Voter Study 2019 Data Collection

The Hellenic National Election Voter Study includes a common module of survey questions provided by the Comparative Study of Electoral Systems (CSES), a collaborative program of research among election study teams from around the world. The 2019 Hellenic National Election Voter Study data (ELNES 2019) was collected using a mobile-friendly web survey (Andreadis, 2015a, 2015b) and the target population was Greek citizens aged 17+. According to the attached report, the vast majority of people in Greece has a mobile phone which is able to receive text messages (SMS). The sample was selected randomly via a Random Digit Dialing (RDD) approach for mobile phone numbers. In Greece, it is very easy to generate probability-based samples of mobile phone numbers through Random Digit Dialing procedures: all mobile phone numbers start with 69, followed by one of [0,3:5,7:9] and seven other digits. After the random generation of the mobile phone numbers, and subsequently checking that these numbers exist, sample members were recruited by sending text messages (SMS) to their mobile phones.

The design of a probability-based web survey of the general population, optimized for mobile users, using SMS as the main contact mode is a very innovative approach, which was implemented for the first time in the 2019 Hellenic National Election Voter Survey. Furthermore, it was based on a novel interaction between the server used for the ELNES web surveys with an SMS gateway service. Our novel method is based on the “push-to-web” method, i.e. a data collection method in which offline contact modes are used to encourage sample members to go online and complete a web questionnaire. The push-to-web method has been tested with postal mail as the contact mode (Dillman et al., 2009). In our innovative design we have replaced postal mail notifications with SMS notifications. A similar push-to-web approach (using landline calls as our contact mode) had already been applied successfully in the 2015 Hellenic voter study (Andreadis et al., 2015).

The selected respondents have received various text messages. We have started with a pre-notification procedure which informed the randomly selected mobile phone owners that they would soon receive an invitation to a survey organized by Aristotle University of Thessaloniki (AUTH). A telephone number at AUTH was also provided so that respondents could call in case they had any questions. The second text message was an invitation with a short URL to participate in the survey. The survey could be completed either directly on the mobile phone that received the message (if the phone had Internet access) or by copying the short URL and pasting in a browser of a tablet/laptop/desktop. The following (if necessary) text messages were reminders including once again the short URL. For details about the procedure of sending pre-notifications, invitations and reminders via text messages (SMS) and the content of these messages are available in other published articles (Andreadis, 2020).

In addition to sending text messages, we called one third of the respondents who had not clicked the link to the survey, and we asked them if they needed any help and if they preferred to participate in the survey via a telephone interview. Almost half of them (49.1%) did not answer our calls, almost 3 out of 10 (28.6%) responded that they did not need any help, because they did not want to participate in the survey, and approximately 2 out of 10 (18.7%) responded that they did not need any help and that they planned to complete the survey later. Only 0.9% of them responded that they would be willing to participate in the survey using another mode, 1.1% were ineligible and 1.8% answered that they would like us to re-send the text message with the link to the questionnaire.

However, telephone interviews would increase the response rate by much less than 0.9%, because (as our experience from other data collection projects shows) some of the respondents, who answered that they would participate in the survey if a different survey mode was offered, would not actually complete the survey if we had tried to arrange an appointment for a telephone interview with them (answering: “yes, I would participate if a different mode was available”, and actually participating when the different mode is available, are in fact two very different things). Given that we had already overachieved our goal to gather a representative probability-based sample of 1500 completed questionnaires, the expected small number of new completed questionnaires from telephone interviews were unnecessary.

Regarding the population coverage , according the report of the Special Eurobarometer 510[5] that was published in June 2021, 99% of Greek citizens have access to a personal mobile phone. This means that practically everyone in Greece has a mobile phone that accept text messages (SMS). The same report shows that access to mobile phone had increased by 8% since 2017 (indicating a significant growth rate of mobile phone access). This growth may be a result of a series of transformations initiated by the Greek government on the channels used to communicate with Greek citizens. Since 2019, and after the 2018 extreme wildfires in the area of Athens - during which many people died from these fires - SMS has become the main way of the Greek Civil Protection to alert citizens to evacuate areas in danger. In addition, due to the COVID-19 pandemic, SMS has been accepted as the most important government-citizens communication channel. The automatic SMS alerts that have been used by the Greek government to notify and remind Greeks about their COVID-19 vaccine or the paperless prescriptions that are send as as text messages to the mobile phone of the Greek citizens, are just some of the many examples of the wide use of SMS by the Greek government. In brief, Greek citizens without a mobile phone may confront large difficulties, when they want to use the services provided by the Greek state. Moreover, if they do not pay attention to their incoming text messages, they could be in danger of not being informed about serious alerts

[5] https://europa.eu/eurobarometer/surveys/detail/2232

regarding situations that could threaten their lives. These may be just some of the reasons of the significant increase in mobile phone access in Greece during the last years.

The ELNES 2019 questionnaire had some additional questions that enabled us to fulfill various requirements. First, we included two filter questions at the very beginning of the questionnaire: i) Permission (consent form) and ii) Eligibility (asking people if they were eligible to vote in the 2019 Greek elections) as a screening instrument. In addition, there was a question asking whether the respondent had multiple mobile numbers. If the answer was yes, there was a follow up question asking how many mobile numbers he/she had. The final sampling probability (and consequently the sampling weight) was obtained after the process of data collection was completed, as both of these estimates need to be adjusted according to the number of distinct mobile phone numbers owned by each respondent. For instance, the sampling probability (or weight) for someone who owns two mobile phone numbers should be double (or half) of the sampling probability (or weight) of someone who owns one mobile phone number only. Thus, we have used the number of distinct mobile phone numbers owned by each respondent to calculate the sampling probability (and consequently the sampling weight) of each respondent. As a result, our data includes a variable named "sample_weight" that is calculated as one divided by the number of mobile telephones owned by the individual, to compensate for disproportionate probability of the selection of individuals with many mobile phone numbers.

Finally, we have invited via email a small number of ELNES 2015b participants, who had agreed to participate in the next election study. A total of 26052 sample members have received an invitation to participate in the survey. After reading the consent form on the first page of the questionnaire, many of the people who had been invited to the survey, did not give their consent. As a result, their questionnaire was submitted with all other (except the consent question) questions unanswered. We do not know if these sample members were eligible, because the eligibility questions were displayed on the questionnaire after the consent form, and sample members who had not given their consent were unable to answer the eligibility questions. In addition, some participants had been administered a version of the questionnaire that was split into a very short first part and a second longer part (for a similar design see

Andreadis & Kartsounidou, 2020). Other questionnaires have failed to meet data quality criteria developed in the past (Andreadis, 2012, 2014) and quality criteria that have been developed in DataPopEU and are presented later in this chapter. After completing all the aforementioned procedures, we had a final, clean dataset of 1537 completed questionnaires, corresponding to a response rate of circa 6%.

2.2 Data Quality and Data Cleaning

To achieve the first objective of DataPopEU of producing high quality data, we had to develop and apply response quality indicators and data cleaning methods. Response rates of surveys are declining over time (Beebe et al., 2010; Curtin et al., 2005). Especially, self-administered questionnaires suffer from low response rates and there are concerns about response quality (i.e. satisficing).

A large number of respondents become less motivated or less engaged at some point during the survey, especially when it is online (Chen, 2011; Fang et al., 2014). This may result in lower response quality in terms of speeding, providing non-substantive answers (i.e. no opinion, "don't know", "none of the above"), non-differentiation in grid questions or choosing midpoint responses in scales (i.e. “neither/nor”). Especially, surveys that offer rewards to participants may attract respondents who may not be interested in providing their best response. Instead, their motivation may be to complete the survey as quickly as possible (e.g., when they are after rewards/ incentives). In these cases, we have to deal with careless respondents or even bots (automatic survey-takers) resulting in meaningless, careless, or fraudulent responses, (i.e. responses of lower quality) that we need to identify and probably remove, in order to get a final cleaned dataset of high quality.

Furthermore, the length of the survey instrument considerably affects the data quality. There is evidence that lengthy online questionnaires lead to lower response rates and responses of lower quality (see for instance Crawford et al., 2001; Galesic, 2006; Galesic & Bosnjak, 2009; Marcus et al., 2007). In addition, the increasing number of mobile users has urged web survey designers to optimize their surveys for mobile devices (see Andreadis, 2015b; Antoun et al., 2017; Lugtig et al., 2016). Given that users of mobile devices spend more time on completing a web survey than desktop or laptop users (Andreadis, 2015a; Cook, 2014; Couper et al., 2007; Lambert & Miller, 2015), the need to create shorter questionnaires is becoming more urgent.

Survey methodology scholars have used many methods to measure response quality. Based on the theory of satisficing, we use a series of indicators of lower quality responses to estimate the level of engagement of survey participants in answering the questionnaires: Item-nonresponse, Mid-point responses, Straight-lining and Speeding. Item-nonresponse is a problem that in many cases has been related with the length of a web survey. Longer web questionnaires suffer from greater amounts of missing data on individual questions (Galesic, 2006; Galesic & Bosnjak, 2009; Peytchev & Tourangeau, 2005) Choosing a mid-point response in scales is also an indicator of low interest or low effort (Weems & Onwuegbuzie, 2001). Respondents may choose mid-point responses when they do not process a question with the required effort. In addition, there is evidence that mid-point responses are similar to "No opinion" answers (Blasius & Thiessen, 2001). Non-differentiation in the answers to grid questions, the so-called straight-lining, is another indicator of satisficing behaviour and low response quality (Greszki et al., 2014; Schonlau & Toepoel, 2015), as it is assumed that respondents who straight-line do not pay the required attention to the questions. In our effort to obtain data quality, we start by focusing on the time spent on questionnaire items (speeding) and we have further developed the "scanning" threshold method that provides the minimum time needed to read and answer an attitudinal question given the length of the question text (Andreadis, 2014).

2.2.1 Response times

Survey response times belong to a special type of data called paradata (Heerwegh, 2003, 2004). These data are non-reactive, and they do not provide information about the respondent's answers. Instead, they provide information about the process of answering the questionnaire, i.e., paradata provide information on how the respondents have interacted with a survey. Item response times and overall survey completion response times of web surveys have attracted the attention of many researchers recently, because longer web surveys suffer from larger break-off rates and greater probability of lower quality responses. Shorter response times can be a sign of burden and an indicator of low response quality. For instance, very short response times may indicate that the respondents have not read the question carefully or that they have even completely skipped question reading. Not surprisingly, it has been shown that very fast respondents (i.e. when their item response times are below specific thresholds) appear to give random answers, introducing noise to the final dataset (Andreadis, 2012, 2014).

Response times have been measured in various ways in the survey literature. For instance, there are two types of proposed timers: active timers and latent timers. Active timers are used when an interviewer is present; the interviewer begins time counting after reading aloud the last word of the question and stops time counting when the respondent answers. This approach assumes that the respondent starts the response process only after hearing the last word of the question. Latent timers are preferred when the questions are visually presented to the respondent e.g., web surveys. This approach assumes that the respondent starts the response process from the first moment the question is presented to him/her.

Another variable of collecting response times refers to the side where time is measured (server-side vs client-side). Measurement on the server side is done by taking advantage of the timestamp that is recorded each time a respondent visits a web page. By calculating the difference between the timestamps of two consecutive pages, we can obtain a measure of the time spent on the first page, i.e., with this method the end time of the first page is the same as the start time of the second page. This means that in order to count time spent on each question, we need to keep each question on a separate web page. However, there is another problem with server-side time counting. Server-side response time is the result of the sum of the clear response time spent within the page plus the time spent between the pages. The time spent between the pages is the sum of the transmission time (from the moment the respondent submits the answer and the moment the answer is recorded on the server) and server processing time (from the moment the answer is recorded on the server until the next page is requested). The time spent between pages depends on the type and bandwidth of the respondent's internet connection, but also on unpredicted, temporary delays due to network load, etc. On the other hand, client-side time measurement is done at the level of the respondent's (or client's) computer itself. This time data comes from JavaScript code embedded in each page and it is a more accurate estimate of the time spent on answering a question, since it does not include the additional between-pages time.

Couper and Peterson (2017) have used both server- and client-level times to disentangle between-page (transmission) times from within-page (response) times and they report that mobile respondents took significantly longer to complete the survey than PC respondents. Also they report that most of this difference is due to within-page times. In compliance with their finding, I argue that transmission times are less important than response times for two reasons: i) issues related to the speed of mobile Internet will eventually be eliminated as mobile Internet providers improve their services and ii) new technologies enable web survey designers to download the next pages of the questionnaire to the respondents' browser before these pages are requested, i.e. eliminating any transmission delays.

Some scholars use the time of the whole questionnaire instead of response times per question/page. The main problem in this case is that there are web survey respondents who temporarily stop answering the questionnaire (e.g., they may receive a telephone call, or they may interact with their social media accounts). As a result, a "normal" time for the whole questionnaire may be the sum of very short response times for many items plus one (or more) very long response time(s) due to break(s). Other scholars use percentiles of the time spent on each question or the whole questionnaire. These percentiles are arbitrarily selected, and using the same percentile for all web surveys would have unpleasant consequences, because the percentage of speeders depends on many factors (e.g., the age and education distribution of the sample, if incentives are offered or not, etc).

2.2.2 Factors affecting response times

Some of the respondents' characteristics are known to affect response times. For instance, most of the studies in the literature (Andreadis, 2015b; Couper & Kreuter, 2013; Yan & Tourangeau, 2008) tend to agree that age (older people spend more time) and education level (respondents with lower education levels spend more time) have a significant impact on response times. In addition, some studies have found that respondents with clear, pre-existent opinion/position, interested in the survey topic and male respond faster than respondents with more uncertain attitudes, less interest in the topic and female, respectively (Andreadis, 2015a, 2015b; Bassili & Fletcher, 1991). Even amongst people who have an attitude, time will depend on the attitude strength, because people with unstable positions need more time to finalise their answer than people with a stable position, who do not need to spend more time than the time to retrieve their already processed opinion from their memory. Finally, it has been shown that attitudes expressed quickly are more predictive of future behaviour than attitudes expressed slowly. Bassili (1993) has provided logistic regression evidence supporting the hypothesis that response latency is a better predictor of discrepancies between voting intentions and voting behaviour than self-reported certainty about their vote intention.

Response times are also affected by question characteristics. The type of question is one of these characteristics. For instance, prior research has identified grid questions to increase response times, especially for mobile users, due to the additional scrolling required for grid questions on mobile devices (Couper & Peterson, 2017). In addition, previous results indicate that response times are longer when the negative, rather than the positive end of the scale is presented first. Furthermore, response time is also related to the complexity of the question. As Bassili and Scott (1996) have shown, badly expressed questions (e.g. double-barrelled questions or questions containing a superfluous negative) take longer to answer than nearly identical questions without these problems. More generally, response time is longer for questions and formats that are difficult for respondents to process (Christian et al., 2009). Finally, even for the most simple question type the length of the question text is known to have an impact on the response time (Andreadis, 2012, 2014).

Other factors are related to the overall setting and environment of survey participation. Heerwegh and Loosveldt (2008) argue that web surveys respondents might have a number of programs running concurrently with the web survey and they might devote their energy to multiple activities (multitasking). This multitasking could increase the response times. In addition, the device used to participate in the survey may have a significant impact on response times. Mavletova (2013) analyzing an experiment with two survey modes conducted using a volunteer online access panel in Russia, reports that the mean time of questionnaire completion for mobile surveys was 3 times longer than the mean time for computer web surveys. On the other hand, Toepoel and Lugtig (2014) offering a mobile-friendly option to respondents of an online probability-based panel organized by a research consultancy agency in the Netherlands, find that the total response times are almost the same across devices. Finally, Andreadis (2015a) estimates that the geometric mean of response times is expected to increase by 17% when switching from desktop to smartphone..

2.2.3 Response times thresholds

The main ideas on minimum response times used in this section were first published almost twenty years ago (Andreadis, 2012). In this section, I will present the main ideas, and the threshold that we can develop based on these ideas. Before answering a survey question, a respondent needs to spend some:

- **T**ime to **R**ead and **C**omprehend the question and the available response options (TRC), and
- **T**ime to **S**elect and **R**eport an answer (TSR).

The time spent on reading and comprehension depends on respondent characteristics (e.g., age, education level) as well as the length and complexity of the question. The time spent on selecting and reporting an answer is affected by question type and the number of response options offered. For single choice items, the reporting procedure is very simple. Thus, it is reasonable to expect a fixed time spent on reporting and it should be short (clicking on a radio button is one of the simplest and fastest ways to report the answer).

Much of the time spent on the first task involves reading and interpreting the text. Survey respondents need time to read the sentence using a reading speed suitable for the comprehension of the ideas in the sentence. The unit used to measure reading speed in the related literature is “words per minute” (wpm). This unit may be suitable to measure reading speed on large texts, but it is inappropriate to measure reading speed on texts of limited size, such as the sentences used in a survey, because it is possible to have a sentence with a small number of lengthy words that is longer and requires more reading time than another sentence with more but shorter words. To avoid similar problems, I have decided to use the number of characters instead of using the number of words.

Based on a table connecting reading speed rates and types of reading provided by Carver (1992) and the average word length, which is 4.5 letters for English texts (Yannakoudakis et al., 1990), we can use classification of reading rates based on the characters read per second. A reading rate of less than or equal to 28 characters per second can be classified as a reading process, known as rauding, which is suitable for comprehension of a sentence. A reading rate of more than 28 characters per second and less than or equal to 40 characters per second can be classified as skimming, i.e., a type of reading that is not suitable to fully comprehend the ideas presented in the text. Finally, a reading rate of more than 40 characters per second can be classified as scanning, which is suitable for finding target words. Thus, if we want to classify a reading rate to one of the three aforementioned categories, we can use the following rule:

- reading rate $\leq$ 28 cps $\rightarrow$ rauding,
- 28 cps < reading rate $\leq$ 40 cps $\rightarrow$ skimming
- 40 cps < reading rate $\rightarrow$ scanning

If we divide the number of characters (without spaces) in each sentence with the number 40, we can get the minimum time (in seconds) that is absolutely necessary to read the whole sentence. Thus, the Minimum Time to Read and Comprehend (MTRC) can be calculated as **MTRC=NC/40**, where NC is the number of characters (without spaces) of the question text (including response options). The above formula corresponds to the assumption that even the fastest reader would need at least 10 secs to read a question of 400 characters.

Of course, respondents need some time for the second task. Bassili and Fletcher (1991), using an active timer, have found that on average, simple attitude questions take between 1.4 and 2 seconds, and more complex attitude questions take between 2 and 2.6 seconds. Thus, the minimum time reported by Bassili and Fletcher (1991) for simple attitude questions (1.4 seconds) can be used as the Minimum Time to Select and Report an answer (MTSR).

Consequently, the minimum response time (MRT) for a simple attitude question is:

$$MRT=MTSR+MTRC=1.4+NC/40 \quad (1)$$

This means that a question of 120 characters would take at least 1.4+120/40=4.4 seconds. Scanning respondents would spend less than MRT on a question. If a respondent spent less than MRT on a sentence, the dedicated time would not be ample for a valid answer. The answer would be given by randomly clicking on any of the available response options. Only extremely capable readers would be able to comprehend the exact meaning of a statement by just scanning the text. This method has been used as one of the data quality indicators in various studies for the cleaning of various datasets (e.g. Andreadis & Kartsounidou, 2020; Hameleers et al., 2018, 2019).

The aforementioned MRT is suitable for simple attitudinal questions. Within the DataPopEU project, we have advanced this method so that it can be applied to other types of questions. For instance, Matrix/Grid/Array questions include a few sub-questions (or items) that share the same response options. In this case, the respondent needs to spend time selecting and reporting an answer for each sub-question. Consequently, the formula should be adapted as follows:

$$MRT=MTSR*NS+MTRC=1.4*NS+NC/40 \quad (2)$$

where, NS is the number of the sub-questions in the matrix question.

The aforementioned formula for speeding detection has been compared with another formula developed by Zhang and Conrad (2014) for the same purpose. Although they follow similar ideas and the same source for the classification of reading speeds, their approach has two significant differences from the method presented here: i) their reading speed threshold is set to 300 milliseconds per word, which corresponds to 15 characters per second i.e., much slower than the typical reading speed and more suitable for learning, and ii) they do not differentiate between TRC and TSR. As a result, according to their method, a simple question and a matrix question of the same length should have the same response time, despite the fact the respondents have to think and report their answers for many items in the matrix question.

These two significant differences between the method proposed by Zhang and Conrad (2014) and the method developed in DataPopEU have a significant impact on the output of speeding detection. Andreadis (2021) shows that the method of Zhang and Conrad works well and gives similar results to the method developed by DataPopEU when it is applied on matrix questions with five or more sub-questions. On the other hand, when the method of Zhang and Conrad is applied to single questions almost 8 out of 10 respondents are detected as speeders (which is much higher than the number of speeders detected by the application of the same method on matrix questions).

2.3 The SurveyDataQuality R package

This section provides a series of R functions that can be used to flag responses of lower quality and the steps we have taken to implement them in the R package SurveyDataQuality (Andreadis & Andreadis 2022). We have used four different indicators. Firstly, we use item-nonresponse (skipping) and we calculate the ratio of missing answers for each respondent. Then, we use the ratio of mid-point responses in Likert-type scale items (e.g., “neither/nor”), because respondents may choose mid-point responses when they do not process a question with the required effort. We also use non-differentiation in grid questions (straight-lining). Finally, we use the minimum time needed to read and answer an attitudinal question depending on the length of the question text.

While implementing the item nonresponse quality indicator, we have to take into account that some web survey packages offer the option to include a non-substantive response (such as “I do not know”) and set it as the default response option (i.e., it is recorded when the respondent has not selected any of the response options). In this case, a “don’t know” is equivalent to item nonresponse. If “don’t know” has been used as the default response option, we might need to merge them with item-nonresponse (e.g. transform “don’t know” to missing).

In addition, when we count the number of missing values for each respondent, we should exclude items where a missing value would not indicate a low-quality response, because sometimes there is a very good reason for a missing value (e.g., conditional questions). For the item-nonresponse indicator, we need to calculate the ratio of missing answers for each respondent and flag the case if this ratio is greater than a specified threshold (e.g., 0.33). The SurveyDataQuality function flag_missing uses three arguments: data, vars and ratio (if not provided, its default value is 0.33) and it creates a binary vector. Cases with a ratio of missing values greater than the provided (or the default) ratio, are flagged with ones, while cases with a smaller ratio of missing values have zeros.

While counting the number of midpoints for each respondent, we should use items where a midpoint would indeed indicate a low-quality response e.g., excessive use of the "neither/nor" option. For the midpoint indicator, we need to take the following steps: i) select a list of survey items/variables, ii) count how many times each respondent has selected the midpoint, iii) calculate the ratio of midpoints for each respondent and iv) flag if the ratio is greater than a specified threshold (e.g., 0.5). The SurveyDataQuality function flag_ midpoint uses four arguments: data, vars, midpoint (if not provided, its default value is 3) and ratio (if not provided, its default value is 0.5). Cases with a ratio of midpoints greater than the provided (or the default) ratio, are flagged with ones, while cases with a smaller ratio of midpoints have zeros.

To check straight-lining for each respondent, we need to choose items in a grid question. We can be more confident about the low quality of the response if at least one of the items is not in the same direction as the other items of the grid. For each grid question, we need to take the following steps: i) count how many times each respondent has selected the same response and ii) if all responses are the same, we flag the corresponding case. The SurveyDataQuality function flag_ straight uses two arguments: data, and vars. Straight-lining cases are flagged with ones, while the value of the rest of the cases is zero.

In order to apply this method, we first need: a) a web survey platform that enables us to capture the time spent on the questions and b) a table with the length of the question texts and the number of sub-questions, similar to the table presented in Figure *2-1*.

```
read_csv("http://www.datapopeu.gr/question-chars.csv")
#> # A tibble: 37 × 3
#>     name  n_chars n_sub
#>     <chr>   <dbl> <dbl>
#>   1 Q4        265     1
#>   2 Q5        259     4
#>   3 Q6        212     1
#>   4 Q7        250     1
#>   5 Q8        382     1
#>   6 Q10       563     7
```

Figure 2-1 Length of the question texts and number of sub-questions

After preparing the table, we can calculate the minimum response time (MRT) for each question, according to formula (2) presented earlier in this chapter. Then, we can bring in the data of the time spent on each question by each respondent and compare these times with the calculated thresholds. If a respondent has spent less time on a question than the minimum time required to understand and respond to the question, we can flag the corresponding case. The SurveyDataQuality function flag_time uses three arguments: data, threshold_file and ratio (if not provided, its default value is 0.1) and it creates a binary vector. Cases with a ratio of extremely fast responses greater than the provided (or the default) ratio, are flagged with ones, while cases with a smaller ratio have zeros.

2.4 Discussion

In this chapter, I have presented the innovative methods we have developed in DataPopEU both for data collection and data cleaning. Regarding data collection, the DataPopEU project has shown that it is feasible to conduct a successful large-scale web survey optimized for smartphones, using SMS as the main contact mode. Regarding data quality and data cleaning, I have presented a new R package dedicated to survey data quality that can be used to apply the quality indicators developed in DataPopEU.

Using text messages to invite individuals to smartphone-friendly web surveys seems to be a method with great potential. In fact, sending text messages could be better than calling. For instance, when respondents are called on their mobile phones, they may be preoccupied, and this may have a serious impact on their decision to participate in the survey thus increasing refusal rates. Moreover, when we use mobile phone numbers, we do not need to run a selection procedure in the final stage (after the phone number is selected), because each number usually corresponds to one person only. Even business mobile phone numbers are different from business land-line phone numbers. A business land-line phone is found on the office desk unlike a business mobile phone which is typically given to a single employee, and it is used by this employee only. In any case, even if multiple people have access to the text messages of a mobile phone, the survey links are personalized (each link has a unique code and it can be used only once).

Although the findings for the use of text messages in web surveys are encouraging, it should be noted that findings can differ across countries depending on who pays the cost of the text messages (e.g., in the USA the recipient of the SMS pays for the incoming message). Another problem could be legal restrictions on contacting mobile phones. The readers who are interested in these issues can find a relevant discussion by Andreadis (2020).

Regarding data quality, this chapter provides a formula that can be used to flag responses which were given so quickly that the response is probably not valid. The method is based on the decomposition of the survey response process into components and a threshold is estimated as the sum of two minimum times: the minimum time for the comprehension of the question and the minimum time for the selection and reporting of an answer.

Finally, we have combined a variety of response quality indicators and we are able to create an innovative multidimensional model to measure response quality that can help identify respondents with indications of an overall satisficing behavior. For instance, all quality indicators can be combined to reveal the respondents who have failed to pass at least two of the four quality checks we have presented. These quality indicators and the corresponding functions have been included in the R package "Survey Data Quality" that is being developed at Github[6] and can be downloaded and used by everyone interested in using these functions to assess the quality of survey data.

Acknowledgements

The research work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: 3572).

2.5 References

Andreadis, I. (2012). To clean or not to clean? Improving the quality of VAA data. *XXII World Congress of Political Science (IPSA)*. http://ikee.lib.auth.gr/record/132697/files/IPSA-2012.pdf

Andreadis, I. (2014). Data Quality and Data Cleaning. In D. Garzia & S. Marschall (Eds.), *Matching Voters with Parties and Candidates. Voting Advice Applications in Comparative Perspective* (pp. 79–91). ECPR Press. http://www.polres.gr/en/sites/default/files/VAA-Book-Ch6.pdf

Andreadis, I. (2015a). Comparison of Response Times between Desktop and Smartphone Users. In D. Toninelli, R. Pinter, & P. de Pedraza (Eds.), *Mobile Research Methods: Opportunities and challenges of mobile research methodologies* (pp. 63–79). Ubiquity Press. https://doi.org/10.5334/bar.e

6 https://github.com/andreasa13/SurveyDataQuality

Andreadis, I. (2015b). Web surveys optimized for smartphones: Are there differences between computer and smartphone users? *Methods, Data, Analysis*, *9*(2), 213–228. https://doi.org/10.12758/mda.2015.012

Andreadis, I. (2020). Text Message (SMS) Pre-notifications, Invitations and Reminders for Web Surveys. *Survey Methods: Insights from the Field*, 1–12. https://doi.org/10.13094/SMIF-2020-00019

Andreadis, I. (2021). Web Survey Response Times What to Do and What Not to Do. *JSM Proceedings, Survey Research Methods Section*. 2021 Joint Statistical Meetings, Alexandria, VA. http://www.asasrms.org/Proceedings/y2021/files/1912254.pdf

Andreadis, I., & Andreadis, A. (2022). Survey Data Quality. *JSM Proceedings, Survey Research Methods Section*, 1999–2006.

Andreadis, I., & Kartsounidou, E. (2020). The Impact of Splitting a Long Online Questionnaire on Data Quality. *Survey Research Methods*, *14*(1), 31–42. https://doi.org/10.18148/SRM/2020.V14I1.7294

Andreadis, I., Kartsounidou, E., & Chatzimallis, Ch. (2015). *Innovation, an answer to lack of funding: The 2015 Hellenic National Election Voter Study* (No. WP22; Working Papers).

Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of Mobile versus PC Web on Survey Response Quality. *Public Opinion Quarterly*, *81*(S1), 280–306. https://doi.org/10.1093/poq/nfw088

Bassili, J. N. (1993). Response latency versus certainty as indexes of the strength of voting intentions in a CATI survey. *Public Opinion Quarterly*, *57*(1), 54–61.

Bassili, J. N., & Fletcher, J. F. (1991). Response-time measurement in survey research a method for CATI and a new look at nonattitudes. *Public Opinion Quarterly*, *55*(3), 331–346. https://doi.org/10.1086/269265

Bassili, J. N., & Scott, B. S. (1996). Response latency as a signal to question problems in survey research. *Public Opinion Quarterly*, *60*(3), 390–399.

Beebe, T. J., Rey, E., Ziegenfuss, J. Y., Jenkins, S., Lackore, K., Talley, N. J., & Locke, R. G. (2010). Shortening a survey and using alternative forms of prenotification: Impact on response rate and quality. *BMC Medical Research Methodology*, *10*(1), 50. https://doi.org/10.1186/1471-2288-10-50

Blasius, J., & Thiessen, V. (2001). Methodological Artifacts in Measures of Political Efficacy and Trust: A Multiple Correspondence Analysis. *Political Analysis*, *9*(01), 1–20. https://doi.org/10.1093/oxfordjournals.pan.a004862

Carver, R. P. (1992). Reading rate: Theory, research, and practical implications. *Journal of Reading*, *36*(2), 84–95.

Chen, P.-S. D. (2011). Finding Quality Responses: The Problem of Low-Quality Survey Responses and Its Impact on Accountability Measures. *Research in Higher Education*, *52*(7), 659–674. https://doi.org/10.1007/s11162-011-9217-4

Christian, L. M., Parsons, N. L., & Dillman, D. A. (2009). Designing Scalar Questions for Web Surveys. *Sociological Methods & Research*, *37*(3), 393–425. https://doi.org/10.1177/0049124108330004

Cook, W. A. (2014). Is Mobile a Reliable Platform For Survey Taking? *Journal of Advertising Research*, *54*(2).

Couper, M. P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and nonresponse in an Internet survey. *Social Science Research*, *36*(1), 131–148. https://doi.org/10.1016/j.ssresearch.2005.10.002

Couper, M. P., & Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 271–286. https://doi.org/10.1111/j.1467-985X.2012.01041.x

Couper, M. P., & Peterson, G. J. (2017). Why Do Web Surveys Take Longer on Smartphones? *Social Science Computer Review*, *35*(3), 357–377. https://doi.org/10.1177/0894439316629932

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web Surveys: Perceptions of Burden. *Social Science Computer Review*, *19*(2), 146–162. https://doi.org/10.1177/089443930101900202

Curtin, R., Presser, S., & Singer, E. (2005). Changes in Telephone Survey Nonresponse over the Past Quarter Century. *Public Opinion Quarterly*, *69*(1), 87–98. https://doi.org/10.1093/poq/nfi002

Dillman, D. A., Smyth, J. D., & Christian, L. Melani. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method.* (3rd editio). Wiley & Sons.

Fang, J., Wen, C., & Prybutok, V. (2014). An assessment of equivalence between paper and social media surveys: The role of social desirability and satisficing. *Computers in Human Behavior*, *30*, 335–343. https://doi.org/10.1016/J.CHB.2013.09.019

Galesic, M. (2006). Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey. *Journal of Official Statistics*, *22*(2), 313–328.

Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, *73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Greszki, R., Meyer, M., & Schoen, H. (2014). The impact of speeding on data quality in nonprobability and freshly recruited probability-based online panels. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online Panel Research: A Data Quality Perspective* (pp. 238–262). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118763520.ch11

Hameleers, M., Andreadis, I., & Reinemann, C. (2019). Investigating the Effects of Populist Communication. Design and Measurement of the Comparative Experimental Study. In C. Reinemann, J. Stanyer, T. Aalberg, F. Esser, & C. H. de Vreese (Eds.), *Communicating Populism* (pp. 168–182). Routledge. https://doi.org/10.4324/9780429402067-9

Hameleers, M., Bos, L., Fawzi, N., Reinemann, C., Andreadis, I., Corbu, N., Schemer, C., Schulz, A., Shaefer, T., Aalberg, T., Axelsson, S., Berganza, R., Cremonesi, C., Dahlberg, S., de Vreese, C. H., Hess, A., Kartsounidou, E., Kasprowicz, D., Matthes, J., … Weiss-Yaniv, N. (2018). Start Spreading the News: A Comparative Experiment on the Effects of Populist Communication on Political Engagement in Sixteen European Countries. *The International Journal of Press/Politics*, *23*(4), 517–538. https://doi.org/10.1177/1940161218786786

Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review*, *21*(3), 360–373.

Heerwegh, D. (2004). Uses of Client Side Paradata in Web Surveys. *Prispevek Na Simpoziju International Symposium in Honour of Paul Lazarsfeld, Brussels, Belgium*.

Heerwegh, D., & Loosveldt, G. (2008). Face-to-Face versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality. *Public Opinion Quarterly*, *72*(5), 836–846. https://doi.org/10.1093/poq/nfn045

Lambert, A. D., & Miller, A. L. (2015). Living with Smartphones: Does Completion Device Affect Survey Responses? *Research in Higher Education*, *56*(2), 166–177. https://doi.org/10.1007/s11162-014-9354-7

Lugtig, P., Toepoel, V., & Amin, A. (2016). Mobile-Only Web Survey Respondents. *Survey Practice*, *9*(3).

Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., & Schutz, Astrid. (2007). Compensating for Low Topic Interest and Long Surveys: A Field Experiment on Nonresponse in Web Surveys. *Social Science Computer Review*, *25*(3), 372–383. https://doi.org/10.1177/0894439307297606

Mavletova, A. (2013). Data Quality in PC and Mobile Web Surveys. *Social Science Computer Review*, *31*(6), 725–743. https://doi.org/10.1177/0894439313485201

Peytchev, A., & Tourangeau, R. (2005). Causes of Context Effects: How questionnaire layout induces measurement error. *Presented at the American Association for Public Opinion Research (AAPOR) 60th Annual Conference.*

Schonlau, M., & Toepoel, V. (2015). Straightlining in Web survey panels over time. *Survey Research Methods*, *9*(2). https://doi.org/10.18148/SRM/2015.V9I2.6128

Toepoel, V., & Lugtig, P. (2014). What Happens if You Offer a Mobile Option to Your Web Panel? Evidence From a Probability-Based Panel of Internet Users. *Social Science Computer Review*, *32*(4), 544–560. https://doi.org/10.1177/0894439313510482

Weems, G. H., & Onwuegbuzie, A. J. (2001). The Impact of Midpoint Responses and Reverse Coding on Survey Data. *Measurement and Evaluation in Counseling and Development*, *34*(3), 166–176. https://doi.org/10.1080/07481756.2002.12069033

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, *22*(1), 51–68. https://doi.org/10.1002/acp.1331

Yannakoudakis, E. J., Tsomokos, I., & Hutton, P. J. (1990). N-Grams and their implication to natural language understanding. *Pattern Recognition*, *23*(5), 509–528. https://doi.org/10.1016/0031-3203(90)90072-S

Zhang, C., & Conrad, F. (2014). Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, *8*(2), Article 2. https://doi.org/10.18148/srm/2014.v8i2.5453