CHAPTER 5

Comparison of Response Times between Desktop and Smartphone Users

Ioannis Andreadis

Aristotle University of Thessaloniki, Greece,
john@auth.gr

Abstract

This chapter offers a precise and thoroughly tested estimate of the impact of using a smartphone on item response times. The comparison is made between desktop and smartphone users when they use a voting advice application that was specifically designed to be used on smartphones. The analysis shows that i) after taking into account item and user characteristics that are known to affect response times and ii) using the most suitable statistical models, using a smartphone instead of a desktop is expected to increase by 17% the geometric mean of item response times.

Keywords

web surveys, item response times, smartphones, multilevel models, endogeneity, generalized structural equation models

Introduction

The aim of this chapter is to test if web survey item response times differ between desktop and smartphone users. Item response times and total 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 near the end of the questionnaire due to respondents' fatigue. In addition, during the last few years, web survey researchers have observed that the number of people who use mobile devices to participate in web surveys is increasing rapidly. Therefore, many recent publications study the implications of responding to web surveys while using mobile devices.

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 three times longer than the mean time for computer web surveys, and she presents three possible reasons for this large difference: i) slower Internet connection, ii) limited functionality of the cell phone (smaller screen size and lack of mouse and keyboard) and iii) greater probability of facing distractions for respondents completing the survey outside of their home. On the other hand, Toepoel and Lugtig (2014), offering a mobile-friendly option to respondents to 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 and that the mean values differ only by five seconds (245s on desktop, 250s on mobile). These contradictory findings cannot be attributed to country-specific characteristics only (e.g. differences of mobile Internet speed between the Netherlands and Russia), because de Bruijne and Wijnant (2013), after running an experiment with CentERpanel participants (also in the Netherlands and also with a mobile-friendly environment), compare the completion time between groups and find that there is a significant difference, i.e. the respondents required more time to finish the survey on a mobile device than on a computer, but they also find mixed results when they compare item response times between devices.

Couper and Peterson (2015) use both server- and client-level times in order 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, and 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 users' browser before these pages are requested, thereby eliminating any transmission delays. Thus, the focus of this chapter is on the time that the respondent really spends interacting with the questionnaire, reading and answering questions, and excludes transmission times.

Some of the respondents' characteristics that are known to affect response times, such as age and education level (Couper & Kreuter 2013; Yan & Tourangeau 2008), have been reported to also affect mobile web access (de Bruijne & Wijnant 2013; Fuchs & Busse 2009; Gummer & Roßmann 2014). Since mobile web access is not randomly distributed across the population, for the data analysis presented in this chapter, I employ advanced models where

completing the survey using a mobile device is treated as an endogenous variable while taking into account, in addition to the aforementioned respondent characteristics, some item characteristics that are known to have an impact on the response time, such as the length of the question text (see Andreadis 2012 and Andreadis 2014a).

Data

The findings presented in this chapter are based on the analysis of the paradata collected in May 2014 by the Greek Voting Advice Application (VAA) HelpMeVote – VoteMatch Greece (Andreadis 2013), which is the Greek part of the multi-national European project VoteMatch (votematch.eu) used for the elections for the European Parliament. Voting advice applications are special types of opt-in web surveys that help users find their proximities with political parties. In the period before an election, these applications can become very popular, and they attract thousands or even millions of users. HelpMeVote is a web application based on jQuery Mobile. As a result, HelpMeVote is compatible with all major mobile platforms and all major desktop browsers. It is able to run both on PCs and on mobile devices; it automatically scales to any screen size and it supports both touch and mouse events. The user interface follows the most common features of designing for mobile devices, e.g. large font size and large buttons. Finally, the question texts are short and the number of response options is limited and displayed vertically to eliminate the need for horizontal scrolling.

HelpMeVote for the European Elections 2014 includes 31 questions, and each question is displayed on a separate page, but it is built as an AJAX application and all pages are downloaded from the beginning to the users' browser. This means that there is no lag time between answering one question and viewing the next one. Consequently, the time between clicks can be counted accurately. The response times are recorded in hidden input fields. Communication with the server is done in the end, when all questions have been answered and the user clicks the 'Submit' button. When the respondent submits the web page, the content of the hidden fields (i.e. response timestamps) are transmitted to the server and are stored in a database along with the User-Agent header of the user's browser. Thus, it is possible to compare between desktop and mobile device users using accurately measured response times and a very large dataset (consisting of tens of thousands of cases).

According to Hypertext Transfer Protocol (HTTP) (see Fielding, Berners-Lee & Frystyk 1996) the User-Agent header field contains information about the user agent originating a HTTP request. For the purposes of this chapter, I have used PHP[21] to retrieve the HTTP_USER_AGENT element of the $_SERVER array,

---

[21] PHP is an open source scripting language especially suited for web development that can be embedded into HTML: http://php.net/.

and I have linked this information with each respondent in the database. Then, using the PHP function get_browser() I have determined the capabilities of the user's browser.[22] By using the aforementioned PHP function I can get the type of the device that the user has used to access the web server. At the time of writing this chapter using the Browser Capabilities Project, the device type field can get one of the following values: Mobile Phone, Mobile Device, Tablet, Desktop, TV Device, Console, FonePad, Ebook Reader, Car Entertainment System or Unknown.

Dealing with extremely short response times

In a previous paper, I provide a formula that can be used to flag responses which were given so quickly that the response is probably not valid (Andreadis 2014a). The method uses the decomposition of the survey response process into four major tasks given by Tourangeau, Rips and Rasinski (2000): 1) comprehension of the question, 2) retrieval of relevant information, 3) use of that information to render the judgment and 4) selection and reporting of an answer. For the estimation of the minimum time needed for Task 1 I used the table provided by Carver (1992) connecting reading speed rates and three types of reading: rauding, skimming and scanning. Bassili and Fletcher (1991), using an active timer, have found that on average, simple attitude questions take between 1.4 and 2 seconds to answer, and more complex attitude questions take between 2 and 2.6 seconds. In their experiment, time counting starts when the interviewer presses the spacebar after reading the last word of the question. Time counting stops with a voice-key (the first noise that comes from the respondent's side triggers the computer to read the clock). For VAAs and web surveys time counting stops when the user clicks on one of the available buttons that correspond to answer options. This additional step requires some extra time. Thus, the minimum time reported by Bassili and Fletcher (1991) for simple attitude questions (1.4 seconds) can be used as the minimum time for Task 4 (selecting and reporting the answer).

If all questions included in a VAA have similar complexity, then the most significant factor that affects the time spent on Task 1 is the length of the question. These two quantities (length and time) are proportional, and their ratio defines the reading speed. VAA users need time to read the sentence using a reading speed suitable for the comprehension of the ideas in the sentence. Andreadis (2014a) calculates a threshold that can be used to flag items that

---

[22]   The function looks up the browser's information in a large file that includes a list of all known browsers and bots, along with their default capabilities and limitations. The file is provided by the Browser Capabilities Project, also known as 'browscap' or 'BCP'. The file is provided in several formats, but the most commonly used is named browcap.ini and is available at: http://browscap.org/

were responded to in an extremely short time using the following formula: threshold = 1.4+[number of characters in the item]/39.375.

Using the same formula I have flagged the answers of HelpMeVote 2014 that have been given in less than the time given by the threshold as extremely short response times. Then I have counted the number of extremely short response times for each user. If more than one third of the response times of a user were extremely short, I removed the corresponding case from the dataset. The reason for the decision to eliminate the complete records of these users is that these users were found to give extremely fast responses so many times that there is strong evidence that they are not using the VAA in a normal way, but they are probably just testing or playing with the application. Thus, the rest of their answers, although they have not been flagged as extremely fast, are probably invalid, and it is better to remove them.

Dealing with extremely long response times

By observing the cases with extremely short response times we can find users who display a more or less stable speeding behavior while responding to a large number of items. The picture for extremely long response times is very different. It is very rare to observe a user spending extremely long times to answer the majority of questions. In most cases a user has spent extremely long times on a very limited number of items. This difference between extremely long and extremely short times has a very good explanation: extremely short times are the result of a decision made by users who decide to respond without paying too much attention (or even any at all) to the questions; these users usually maintain the same attitude throughout the questionnaire. On the other hand, extremely long times are the result of an interruption that usually occurs after an external distraction (e.g. an incoming email, a phone call, someone knocking at the door, etc). Thus, the occurrence of extremely long response times is associated neither with a user nor with an item. Of course, longer items require longer response times, but a typical questionnaire would not include an item which is so long that it could require an extremely long time to read. Thus, the occurrence of extremely long response times is random and it can be identified both by looking for extremely long times per item and by looking for extremely long times per user. Taking into account that a typical VAA includes about 30 items and is used by thousands or even millions of users, it is easier to look for extremely long times within each user.

A good way to look for extreme response times within a user is to use the methods of exploratory data analysis, and more specifically the statistics used for boxplots (Hoaglin, Mosteller & Tukey 1983; McGill, Tukey & Larsen 1978; Tukey 1977). Boxplot statistics can identify outliers, i.e. values between the inner and the outer fences of the boxplot, and extreme values, i.e. values outside the outer fences. As outer fences, I use the values: $Q_1 - 3 \times IQR$ and

$Q_3 + 3\times IQR$, where IQR is the interquartile range and Q1 and Q3 are the first and the third quartiles, respectively. The problem of applying this method on the response times themselves is that it would flag as extreme too many values that are not extreme.

The distribution of response times is a semi-bounded function with zero as its lower bound. Usually it is highly skewed to the right. The logarithmic function is a good way of transforming a highly skewed distribution into one that is closer to normal distribution. Thus, in order to flag the real extremely large response times, I have applied the logarithmic function to the response times and then I have applied the aforementioned exploratory data analysis method to identify extreme values on the logs of the response times.

After flagging the extremely long response times, there is one last decision to be made: How should they be treated? I argue that they should be recoded as missing values. The logic behind this argument is very simple. We cannot leave them intact, because the recorded time is not the actual time spent on the question but the sum of the time spent on answering the question, plus an unknown amount of time due to some external distraction. We should not remove the whole record, because we do not have a user giving invalid answers (as was the case with extremely short response times). Thus, the best way of dealing with these values is to consider them as missing, because the external distraction that interrupted the user has prevented us from recording the actual time spent on the item. By recoding the extremely long response times as missing, we do not allow them to distort the average response times estimated by the sample. At the same time we do not have to disregard the whole row, because we can use these records with statistical methods that do not require list-wise deletion of cases with missing values or we can impute the missing values using the response times of the same user on the rest of the items.

Other data preparations

HelpMeVote users answer 31 questions in order to get their proximity with the Greek political parties. Before being given the output, users are asked to fill in a form with their personal information (mostly demographics, i.e. Sex, Age Group, Education Level, but also information related to their voting behavior, i.e. Vote Choice, Political Interest). Although it is not mandatory (users can click 'continue' and move on to the output without answering) the vast majority responds to most of these questions,[23] probably because they are in a responsive mood or because they consider this form as part of the VAA procedure[24].

---

[23] Vote Choice is the only item in this form that displays a large number of non-useful answers because many users either give no answer or indicate that they have not decided yet.

[24] HelpMeVote offers an 'info' page where users are informed that their responses are stored in a database anonymously to be used for academic research.

For the analysis presented in this chapter, I have kept only the cases where the demographic variables have valid values. There are three reasons which support a decision to remove the cases with missing values on demographic variables: i) the percentage of missing values is very small, ii) these variables will be used as predictors for the models in the following sections and iii) imputing the missing value of demographic variables from the answers to the rest of the questions is difficult.

More than 80,000 HelpMeVote/VoteMatch 2014 questionnaires have been completed by Greek citizens during the period before the elections for the European Parliament. In order to work with a sample that can be handled by the computational resources of a strong workstation, I had to randomly select a subsample corresponding to 10% of the total sample. In order to ensure that the findings presented in this chapter are the same as the findings that I would present if I had used the total sample I have done the following tests: i) I have checked and I have verified that the distributions of the main variables in the subsample are not different from the corresponding distributions in the total sample and ii) I have replicated the presented analysis with other 10% subsamples and I have got very similar findings. The used sample is available from OpenICPSR (Andreadis 2014b).

Finally, the distribution of the used devices is as follows: 80.7% desktops, 13.5% smartphones and 5.7% other mobile devices (mostly tablets). The focus of this chapter is on the comparison between smartphone and desktop users. Therefore, the users of other mobile devices have been excluded from the analysis.

Variables

In the following models the logarithm of the response times is used as the dependent variable (i.e. the outcome). As the main task of this chapter is to compare the response times between smartphone users and desktop users, the binary variable 'mobile' is included in the model as the main treatment under study.

As control variables from the item characteristics, I use the length of the statement and a dummy variable that takes the value of 1 when the statement is about an EU issue and 0 when the statement is about a national issue. The inclusion of the latter variable is justified by the fact that Greek voters are presumed to be less informed about EU policy issues than they are about national issues, and they are expected to need more time to express their opinion about EU issues.

From users' characteristics I use as control dummy variables taking the value of 1 for male respondents, for people aged over 49 years old (over49), for users who are interested in politics (polint) and for citizens who had already made their vote choice when they used the VAA (decided). According to the

literature my hypotheses about these predictors are as follows: older people (> 49) are expected to spend more time than younger people. Citizens interested in politics and voters who have already decided their vote choice should be more familiar with the major issues of the electoral competition, so they are expected to have clear, pre-formulated opinions about the statements, and they are expected to need less time than people not interested in politics and people who had not decided about their vote choice when they used HelpMeVote. Finally some studies have found that female respondents spend more time on web surveys, thus I expect a similar finding from the present analysis. As a final user characteristic, I use the education level as a categorical variable, and I compare all other education categories with the category of primary education (used as the reference category). The expectation here is that as we switch to higher education levels, the response time should decrease.

Unfortunately, the treatment variable of the model (mobile) is endogenous, and it depends on variables that also affect the outcome (e.g. age). In order to correctly estimate the treatment effect, I employ advanced statistical methods (described in the following section). In order to model the endogeneity of the treatment I use as its predictor the age dummy variable 'over49', but I do not use the education level because I have not found the education level to have an impact on the treatment variable. I also use a variable named 'scorex' which indicates the position of the user on the political left/right axis, because I have found that it is a good predictor of using a smartphone (as users move from the left to the right of the axis, they tend to use smartphones more), while it does not have an impact on response times.

Methods

Smartphone web access is not randomly distributed across the population. Thus in order to study the impact of using a smartphone on item response times, I had to employ a constrained endogenous-switching model (also known as endogenous treatment-effects model), i.e. a model where the treatment (completing the survey using a smartphone) is considered as an endogenous variable (Greene 2012; Heckman 1978; Maddala 1983; Wooldridge 2010). In these models, instead of having a single linear equation for the prediction of the outcome, I have two equations. The first is the linear equation for the outcome. The treatment is a binary variable that is considered to take the values 0 or 1 when a latent variable is smaller or larger than 0, respectively. This latent variable is also given by a linear equation. The error terms of these two linear equations follow a joint bivariate normal distribution, and they are allowed to be correlated. The coefficient for this correlation can be estimated by the endogenous treatment-effects model. If the estimated correlation between the treatment errors and the outcome errors is significant (i.e. if we reject the null hypothesis of no correlation) then the impact of the treatment on the outcome cannot be estimated

correctly by a simple model and we have to use the estimates provided by the advanced model. On the other hand, if the advanced model indicates that the correlation coefficient is not statistically significant, we can use the estimates of the simple, single equation model.

At the same time I had to take into account other factors that are known to have an impact on item response time. These factors are characteristics of the respondent, e.g. gender, age, education, interest in the theme of the survey, knowledge about the survey topics; and characteristics of the items, such as the length or the difficulty of the item. There are two levels in the model: the respondent level and the item level. The usual approach is to consider the items as the lower level and the respondents as the higher level, i.e. to consider a hierarchical linear model where the items are nested within the respondents (van der Linden 2008), but there are example of reversed roles, i.e. where the hierarchical model is built on the basis that respondents are nested within items (Swanson et al. 2001). The item response times within the same user may be correlated (intraclass correlation) due to individual characteristics (e.g. education) that affect reading speed. As a result, the assumption of independence of the observations is violated. Using a non-hierarchical model would underestimate the standard errors of regression coefficients – especially for the coefficients of the user level predictors – resulting in non statistically significant coefficients to appear as significant (Gelman & Hill 2006; Hox 2002). Another advantage of using a multilevel model is that the residual variance is partitioned into a between-user and a within-user component. Consequently, by using a multilevel model, it is possible to study the effects of both user level and item level characteristics, get better estimates of the standard errors of the regression coefficients and compare the between-user with the within-user variance.

For the data analysis of this chapter I needed an endogenous treatment-effects multilevel regression model. To my knowledge, there are not any out-of-the-box regression procedures that can be used for the estimation of this complicated model in any of the statistical (either commercial or open source) software packages. According to Skrondal and Rabe-Hesketh (2004) a way to deal with this problem is to use generalized Structural Equation Modeling (SEM). Structural models are able to show causal dependencies between endogenous and exogenous variables. This means that structural equation models can be used as alternatives to the systems of regression equations (such as the endogenous treatment-effects model) used by Heckman (1978) and other econometricians. With generalized structural equation modeling we can generalize Heckman models (both selection and endogenous treatment models) to include multilevel effects. The corresponding structural equation model includes two equations, one linear regression (to model the outcome) and a censored regression (for the treatment selection model). By adding a common latent variable in both equations we can model the correlation between them. By constraining the latent variable to have variance and

coefficient in the selection equation both equal to 1 and the variance from the censored regression equal to the variance of the linear regression we can have an identified model.

The multilevel structure can be modeled in SEM by including a random intercept at the user level. This is done by adding a latent variable that is constant within users and varies across users and a path from this latent variable to the outcome variable. For details on estimating multilevel linear models as structural equation models the interested reader can consult the related literature by Bauer (2003) and Curran (2003). For the technical details see also the book by Skrondal and Rabe-Hesketh (2004).

In the following section I present a generalized SEM. This is a very complicated model that takes into account both the endogeneity and the multilevel structure of the dataset, i.e. it is a generalized structural equation model that represents an endogenous treatment-effects multi-level regression. This model requires a tremendous amount of computer resources (both CPU power and memory), and I had to randomly select a subset of the data to run this complicated analysis. As mentioned before, I have verified the findings presented in the next section by running the analysis again on additional random subsamples.

Findings

As I have already explained in the previous section, since the treatment is endogenous, we need a generalized structural equation model that represents an endogenous treatment-effects multi-level regression. This model is presented in **Figure 1**. The main question in these models is whether the correlation between the error terms of the equations is significant. This question is important because if the correlation is not significant, we can forget about the endogeneity of the treatment variable and we can use a simpler model, such as a multilevel linear regression. As **Table 1** indicates, the value of the correlation coefficient ρ is estimated at 0.011 and the corresponding test shows that it is not significantly different from 0 (the p-value of the test is 0.937). This means that we do not need the censored regression and we can use the estimates of a simpler model.

Figure 2 and **Table 2** show the generalized structural equation model that is equivalent to a multilevel regression. **Figure 2** includes the estimated coefficients and the estimated values for the error terms. **Table 2** shows the exponential values of the coefficients. Since I have used the logarithm of the response times as the outcome of the model, the interpretation of the estimated regression coefficients is the following: if the estimated coefficient for an independent variable X is b, when X is increased by one unit the logarithm of the outcome is expected to increase by b units. In terms of the outcome itself, its expected value is multiplied by e^b.
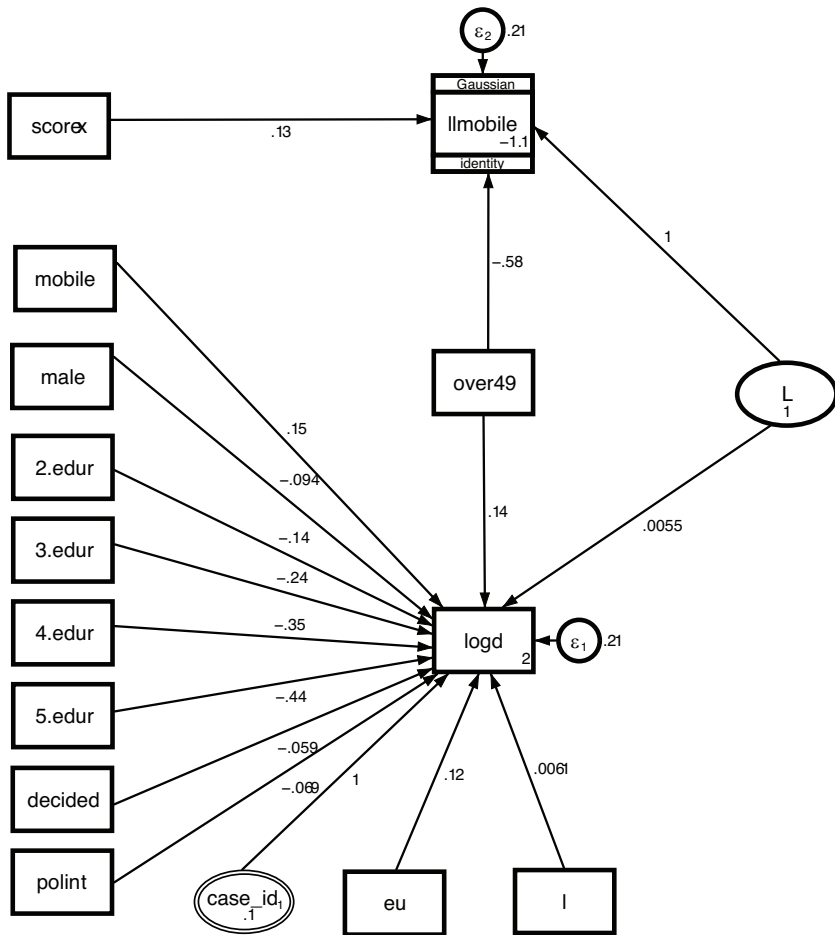
Figure 1: Generalized structural equation model.

	Coef	Std. Err.	z	P > z	[95% Conf Interval]	
ρ	0.011	0.140	0.08	0.937	−0.264	0.286

Table 1: Significance of the correlation.

According to **Figure 2**, the constant term is estimated at 2.01. This is the expected mean of the logarithm of the response times. According to **Table 2**, the exponential value of the constant term is 7.47. This is the geometric mean of response times.

In order to answer the main research question of this chapter, i.e. the impact of using a smartphone on the response time, I focus on the interpretation of
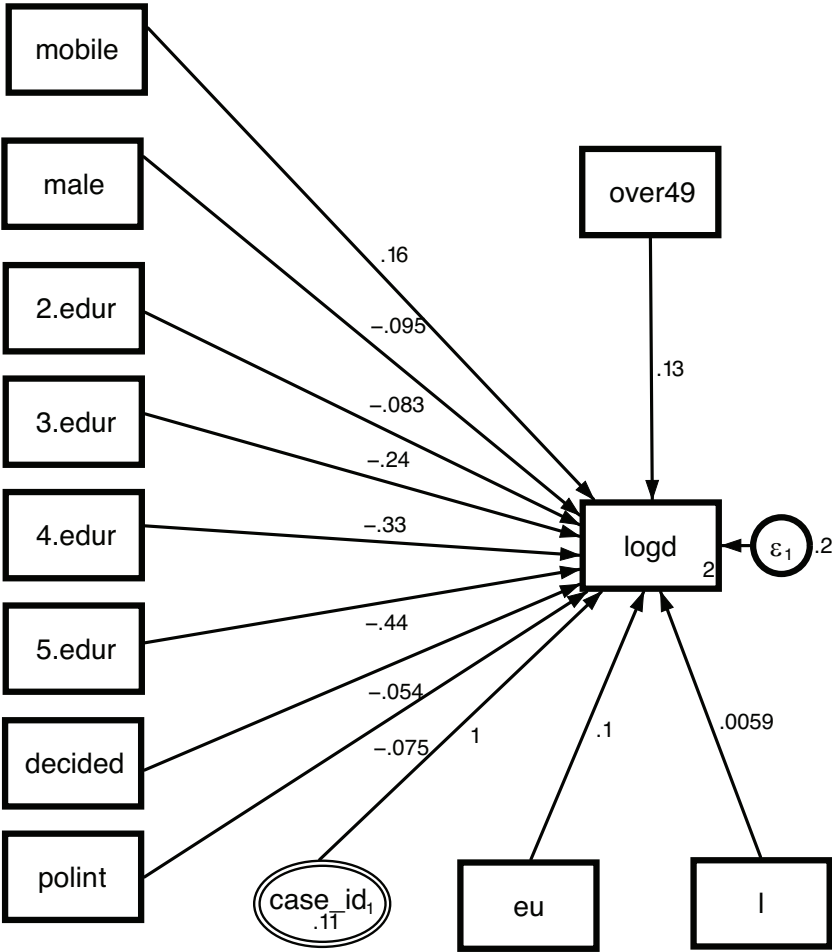
Figure 2: Generalized structural equation model equivalent to a multilevel linear regression model.

the coefficient of the mobile variable: the coefficient is 0.16 and the exponential value is 1.17. This means that when switching from desktop to smartphone the geometric mean of response times is expected to increase by 17%. To provide an estimate of the treatment effect in seconds, I calculate the increase on the overall geometric mean: 7.47*17% = 1.27 seconds per item. The impact on response times of using a mobile device is significant even after taking into account the impact of the control variables that were included in the model.

Moving on to the interpretation of the coefficients of the item characteristics, we can observe that the coefficient for the length of the statement (l) is 0.059 and its exponential value is 1.0059. This means that, while holding all other

Log of item response time	exp(b)	Std. Err.	z	P > z	[95% Conf.	Interval]
Mobile	1.175	0.015	12.43	0.000	1.145	1.205
Length of item (l)	1.006	0.000	111.77	0.000	1.006	1.006
EU issue (eu)	1.110	0.002	48.86	0.000	1.105	1.115
Male	0.909	0.009	−9.91	0.000	0.893	0.927
Over 49 years old (over49)	1.133	0.013	11.10	0.000	1.109	1.159
Vote choice (decided)	0.947	0.009	−5.94	0.000	0.930	0.964
Political interest (polint)	0.928	0.010	−7.29	0.000	0.909	0.947
Education (reference: Primary)						
Lower secondary (2.edu)	0.921	0.058	−1.30	0.193	0.813	1.043
Upper Secondary (3.edu)	0.784	0.043	−4.45	0.000	0.704	0.873
Tertiary (4.edu)	0.716	0.039	−6.15	0.000	0.644	0.797
Postgraduate studies (5.edu)	0.645	0.035	−8.02	0.000	0.579	0.718
M1[case_id]	2.718	.	.	.	.	.
Constant	7.471	0.412	36.48	0.000	6.706	8.323

Table 2: Generalized structural equation model equivalent to a multilevel linear regression model (exponentiated coefficients).

predictors constant, for every additional character in the statement the geometric mean of response times increases by 0.59%. According to the model, if a statement refers to an EU policy issue the respondents need more time to give their answer. The corresponding coefficient is 0.1 and its exponential value is 1.11, indicating an 11% increase in the geometric mean of response times when switching from a national issue to an EU issue.

Moving on to the user level, we can see that the coefficient for male is −0.095 and its exponential value is 0.909. This means that the geometric mean of response times in the group of men is 90.9% the geometric mean of response times in the group of women. In other words, switching from female to male respondents, the expected response time is decreased by 9.1%. Following the same logic, we observe that when we switch from undecided people to people who have already made their choice the geometric mean of response times is decreased by 5.3%. Similarly, moving from people who are not interested in politics to people who are interested in politics the geometric mean is expected to decrease by 7.2%. On the other hand, the exponentiated coefficient for older people is 1.13, indicating a 13% increase in the geometric mean of response times when switching from younger people to users over 49 years old. Finally, when we switch from primary education to higher education levels,

the response time decreases; only the difference between primary and lower secondary education levels is not statistically significant. The largest difference is observed between the two extreme education levels: the ratio of geometric means of postgraduate studies to primary education levels is 0.64, indicating that the time spent by the most educated users is 64% the time spent by the less educated users, i.e. a decrease of 36%.

According to **Figure 2**, the variance of the random intercept is estimated to be 0.11 and the estimated error variance is 0.2. A likelihood ratio test indicates that the random intercept variance is large enough that we could not ignore it. This verifies that the decision to use a multilevel model was correct. Indeed, if a single level model had been used, non significant differences (e.g. the response time difference between primary and lower secondary education levels) would appear as significant.

Finally, I have explored whether there are any significant interaction terms between smartphone use and respondent characteristics (age, gender and education) or the length of the question. None of these interaction terms have a significant impact on the item response times at the 0.01 significance level.

Discussion

This chapter advances mobile research in various ways. Firstly, it offers a precise and thoroughly tested estimate of the impact of using a smartphone on item response times. The comparison was made between desktop and smartphone users when they use a voting advice application that was specifically designed to be used on smartphones. The analysis has shown that i) after taking into account item and user characteristics that are known to affect response times and ii) using the most suitable statistical models, when switching from desktop to smartphone the geometric mean of item response times is expected to increase by 17%.

The lack of a significant interaction between the use of a mobile device and the length of the question indicates that the longer times of smartphone users cannot be attributed to the smaller display of their devices. This finding was expected because the application was carefully designed to fit on the small screens of mobile devices. The lack of any significant interactions between smartphone use and respondent characteristics probably indicates that mobile users do not need more time because they face some difficulties while using their smartphones. If there was an issue of usability, this issue would probably be worse for older people. Thus, it seems that the most reasonable explanation for the longer times of smartphone users is that they are probably completing the survey outside of their home and their environment gives them more distractions than are available to the desktop users, who complete the survey in a more quiet room in their home or in their office.

In addition to the aforementioned finding, this chapter has presented an advanced statistical methodology to deal with the multilevel structure of the data while taking into account the endogeneity of the treatment. This is achieved by employing a generalized structural equation model that represents an endogenous treatment-effects multi-level regression. Although the data analysis performed in this chapter has shown that the correlation between the error terms was not significant and the simple multilevel model was adequate in this case, the advanced method proposed here may be necessary in other response time models with endogenous treatment variables.

Lastly, this chapter offers an innovative method to prepare a dataset of response times for statistical analysis by treating the low and the high extreme values differently. It shows how to flag users who have been answering so fast that they should be removed from the dataset. In addition, it proposes a way to deal with the extremely large response times by identifying the actual extremes instead of trimming the dataset using arbitrary selected threshold that lack any theoretical justification and lead to the removal of cases that should remain in the dataset.

I conclude this chapter with some ideas for further research on the topic. A more advanced model could compare three categories: desktop, smartphone and tablet users. Another extension could be to check the actual answers of the respondents for typical indicators of low quality (e.g. straight-lining) and try to test if there are any differences between mobile and desktop users. Other mobile/desktop comparisons could involve an analysis that would involve both response times and response patterns. In any case, since the trend shows a continuous increase of survey respondents using their mobile devices, the research community should focus on research projects that will help us build a deep understanding of the implications of this trend.

References

Andreadis, I. (2012). *To Clean or not to Clean?* Paper presented at the Improving the Quality of VAA Data XXII World Congress of Political Science (IPSA), Madrid, Spain. Retrieved from http://www.polres.gr/en/sites/default/files/IPSA12.pdf

Andreadis, I. (2013, September). *Voting Advice Applications: a successful nexus between informatics and political science.* Paper presented at BCI '13, Thessaloniki, Greece. DOI: http://doi.acm.org/10.1145/2490257.2490263

Andreadis, I. (2014a). Data Quality and Data Cleaning. In D. Garzia, & S. Marschall (Eds.), *Matching Voters with Parties and Candidates. Voting Advice Applications in Comparative Perspective.* ECPR Press. ISBN: 9781907301735

Andreadis, I. (2014b). *Paradata from Political Web Surveys.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. Retrieved October 2, 2014. DOI: http://doi.org/10.3886/E17816V3

Bassili, J.N. and **Fletcher, J.F. (1991).** Response-time measurement in survey research a method for CATI and a new look at nonattitudes. *Public Opin.Q., 55*(3), 331–346.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, *28*(2), 135–167. DOI: http://dx.doi.org/10.3102/10769986028002135

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

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), 271–86. Wiley Online Library. DOI: http://dx.doi.org/10.1111/j.1467-985X.2012.01041.x

Couper, M. P, & **Peterson, G.** (2015). Exploring Why Mobile Web Surveys Take Longer. Paper presented at General Online Research 2015.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*(4), 529–569. DOI: http://dx.doi.org/10.1207/s15327906mbr3804_5

de Bruijne, M., & **Wijnant, A.** (2013). Comparing Survey Results Obtained via Mobile Devices and Computers: an Experiment with a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey. Social Science Computer Review. Sage Publications, 0894439313483976. DOI: http://dx.doi.org/10.1177/0894439313483976

Fielding, R. T., Berners-Lee, T., & **Frystyk, H.** (1996). Hypertext Transfer Protocol--HTTP/1.0. Retrieved from http://tools.ietf.org/html/rfc1945

Fuchs, M., & **Busse, B.** (2009). The Coverage Bias of Mobile Web Surveys Across European Countries. *International Journal of Internet Science*, *4*(1), 21–33.

Gelman, A., & **Hill, J.** (2006). Data analysis using regression and multilevel / hierarchical models. Cambridge: Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511790942

Greene, W. H. (2012). Econometric Analysis. 7th ed. Upper Saddle River, NJ: Prentice Hall.

Gummer, T., & **Roßmann, J.** (2014). Explaining Interview Duration in Web Surveys a Multilevel Approach. Social Science Computer Review. SAGE Publications, 0894439314533479.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, *46*, 931–959. DOI: http://dx.doi.org/10.2307/1909757

Hoaglin, D.C., Mosteller, F., & **Tukey, J.W.** (1983). Understanding robust and exploratory data analysis. Wiley.

Hox, J. J. (2002). Multilevel analysis: Techniques and applications. Psychology Press.

Maddala, G. S. (1983). Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511810176

Mavletova, A. (2013). Data Quality in PC and Mobile Web Surveys. Social Science Computer Review. Sage Publications, 0894439313485201. DOI: http://dx.doi.org/10.1177/0894439313485201

McGill, R., Tukey, J. W., & **Larsen, W. A.** (1978). Variations of box plots. *The American Statistician*, *32,* 12–16.

Skrondal, A., & **Rabe-Hesketh, S.** (2004). Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. CRC Press. DOI: http://dx.doi.org/10.1201/9780203489437

Swanson, D. B., Case, S. M., Ripkey, D. R., Clauser, B. E., & **Holtman M. C.** (2001). Relationships among Item Characteristics, Examined Characteristics, and Response Times on USMLE Step 1. *Academic Medicine*, *76*(10). DOI: http://dx.doi.org/10.1097/00001888-200110001-00038

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. SAGE Publications, 0894439313510482. DOI: http://dx.doi.org/10.1177/0894439313510482

Tourangeau, R., Rips, L. J., and Rasinski, K. (2000). The psychology of survey response. New York:

Tukey, J.W. (1977). Exploratory Data Analysis. Addison-Wesley.

van der Linden, W. J. (2008). Using Response Times for Item Selection in Adaptive Testing. *Journal of Educational and Behavioral Statistics*, *33*(1), 5–20. DOI: http://dx.doi.org/10.3102/1076998607302626

Yan, T. and **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

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. 2nd ed. Cambridge, MA: MIT Press.