

The “frequency divide” on web surveys: differences of opinion, behaviour and demographics among internet users

Vicente, Paula

Instituto Universitário de Lisboa (ISCTE-IUL), UNIDE-IUL, Lisboa, Portugal

Av. Forças Armadas

1649-026 Lisbon, Portugal

E-mail: pbcv@iscte.pt

Reis, Elizabeth

Instituto Universitário de Lisboa (ISCTE-IUL), UNIDE-IUL, Lisboa, Portugal

Av. Forças Armadas

1649-026 Lisbon, Portugal

E-mail: ear@iscte.pt

1. Introduction

Concerns about internet access and coverage error have driven much of the investigation on the potential of the internet as a survey mode. Although internet coverage still poses some legitimate concerns for web-based research, the potential for wider deployment of web surveys is growing dramatically as the incidence of household computer ownership and internet accessibility continues to rise. Although the scenario of internet coverage in some countries is poor, e.g. Greece, Spain, Hungary and Czech Republic where the percentage of households with internet access does not reach 45%, internet coverage in others such as Netherlands, Denmark or Sweden is already over 75% (European Commission 2009). In the US, nearly 65% of households have an internet connection (Callegaro and Wells 2008).

The concept of the digital divide has changed over time. At first, it basically referred to connectivity problems, i.e., the possibility/difficulty of having computers available that are connected to the worldwide web. The relationship of demographics and internet access is one of the major research questions in the field of the digital divide. Empirical investigation on the relationship of socio-demographics and internet access indicate that those accessing the internet are significantly different from those who do not in terms of age, education, income, ethnicity and region of residence (e.g Couper 2000). There are also differences between internet users and non-users in substantive variables such as political opinions and sociable attitudes (Robinson et al. 2002), perceptions (Valliant and Lee, 2005), product and service consumption (Pierkarski et al. 2008) or personal health characteristics (Dever et al 2008). Later, concern started to be shown for the development of the necessary capacities and skills to use information and communication technologies. The concept of digital literacy related to the digital divide began to be developed; it addressed questions about different abilities to use the internet (Robinson et al 2003). More recently, the concept of the digital divide has incorporated the quality of usage, based on the differences between users of the internet (Camacho 2005), in other words, internet users use the internet in different ways and for different purposes. Therefore, the digital divide concept no longer refers only to the problem of having access or not; it now incorporates the differences among those who are already connected.

The frequency of internet use is likely to affect the probability of a potential respondent seeing requests to participate in web surveys. As frequent users are connected more time and are more likely to respond to a survey, a web survey sample is likely to be biased towards frequent users. If the differences between frequent and non-frequent users are relevant to the survey, estimates are likely to suffer from bias.

This study examines differences between frequent and non-frequent users of the internet in terms of

demographics and responses. Specifically, the study aims to (1) examine demographic characteristics of frequent and non-frequent internet users, (2) assess whether frequency of internet use is associated with different opinions and behaviors and (3) whether the “frequency divide” still makes a difference when controlled by relevant demographic variables.

2. Methodology

Data for this study come from the Eurobarometer Wave 72.1 on Poverty, Social Exclusion, Climate Change and the National Economic Situation (2009) relative to Portugal collected at the request of the European Commission. The study covers the population aged 15 years and over. A multi-stage random sample design was applied. Face-to-face interviews were conducted in people's homes.

The sample size is 1051 interviews. The analysis was restricted to the respondents who said they use the internet at home (n=378). Table 1 shows its distribution across categories of frequency of internet use at home:

Table 1: Frequency of internet use at home of respondents

Frequency of use at home	%
Everyday/Almost every day	49.5
2 or 3 times a week	27.2
About once a week	9.0
Once or twice a month	6.1
Less often	8.2

The frequency of internet use at home was recorded into a binary variable – internet usage profile – coded as 1–frequent user and 0–non-frequent user; the former corresponds to daily, or almost daily, home internet usage, and the latter to usage of 2 or 3 times a week or less. This new variable is used in the subsequent analysis.

3. Demographics of frequent and non-frequent users

The analysis begins by evaluating the demographic characteristics of frequent and non-frequent users of internet. We start with a bivariate analysis in which an evaluation is made of the association between each demographic and internet use profile. Table 2 shows the percentage distribution of eight demographic characteristics for each subgroup of internet use and the results of the chi-square test performed.

A significant association ($p < 0.05$) was found between internet usage profile and age, education, employment status, marital status and presence of children in the household. Overall, the associations can be described as follows:

- Age – frequent users of the internet tend to be young as this group had a greater proportion of the 15-24 year age group (28.9%) than the non-frequent users (16.8%).
- Education – non-frequent users include a higher proportion (37.3%) of early school leavers (under 15 years); frequent users included a higher percentage of people left school at the age of 20 or older or were still studying than the non-frequent users subgroup.
- Employment status – whereas frequent users have a higher percentage of housewives/househusbands (26.7%) non-frequent users have a higher percentage of employed persons (71.2%).
- Marital status – a greater proportion of single people was found in the frequent users group (37.1%) while non-frequent users have a higher percentage of married people (71.7%).
- Presence of children in the household – there is a higher percentage of households with children under 10 years old among non-frequent users (30.4%).

Table 2: Demographic characteristics of frequent and non-frequent internet users

	Frequent users (%)	Non-frequent users (%)	χ^2	Significance
Gender			1.620	0.203
Male	59.9	53.4		
Age			10.512	0.015
15-24	28.9	16.8		
25-39	40.6	41.4		
40-54	20.9	31.9		
55+	9.6	9.9		
Education (age at which left school)			33.177	0.000
-15	15.8	37.3		
16-20	32.1	37.3		
20+	25.0	14.5		
Still studying	27.2	10.8		
Employment status			20.213	0.000
Employed	54.0	71.2		
Unemployed	2.1	2.1		
Student	10.2	10.5		
Retired	7.0	6.8		
Housewives/househusbands	26.7	9.4		
Marital status			12.186	0.002
Married/living with partner	55.4	71.7		
Single	37.1	20.6		
Other	7.4	7.8		
Type of region of residence			4.142	0.126
Rural area or village	22.0	31.1		
Small or middle sized town	52.2	44.2		
Large town	25.8	24.7		
Children (< 10 years) in the household			3.964	0.046
Yes	21.4	30.4		
Lowest net monthly income needed for an acceptable standard of living			1.705	0.426
< 1000 euros	17.1	22.8		
1000-2000 euros	52.9	51.7		
> 2000 euros	30.0	25.5		

A multivariate binary logistic model was then estimated in order to identify which characteristics were most relevant in the explanation of the internet usage profile. The eight demographic characteristics were used as explanatory variables – gender, age, education, employment status, marital status, type of region of residence, presence of children under 10 years old in the household and lowest net monthly income – and the internet usage profile was used as the dependent variable.

Table 3 summarizes the values of the Wald statistic and respective significance value for each variable with significant effect on internet usage status.

Table 3: Wald test values and significance of the binary logistic analysis of demographic characteristics associated with internet usage profile

Explanatory variables	Wald test ⁽¹⁾	Significance
Education	17.667	0.001
Type of region of residence	11.948	0.003
Children under 10 in the household	4.478	0.034

⁽¹⁾ Stepwise estimation; the outcomes concern the last step model.

The model indicates that education is the strongest predictor of a frequent user of the internet. The type of community and presence of children in the household may be considered the set of second strongest predictors. However, gender, age, employment status, marital status and lowest net monthly income did not have a statistically significant effect. Although age, employment status and marital status were strongly associated with internet usage its effect turns non-significant when controlling for the interrelations between variables.

4. Survey estimates of frequent and non-frequent users

A series of multivariate logistic regressions was conducted to compare the answers of frequent and non-frequent internet users to survey questions, taking the internet usage profile as explanatory variable (1-frequent user; 0-non-frequent user). A subset of items was selected from the questionnaire for the analysis.

Three different models were used. The internet use profile variable was the only independent variable in Model 0; no other variables were used as controls. This model was therefore the baseline measurement, comparing answers from frequent and non-frequent users of internet. In Model 1 we controlled for the most influential variable on internet use status: education. The type of region of residence and presence of children in the household were added as predictors in Model 2. This approach gives a total of 4 explanatory variables including internet use profile.

Table 4 summarizes the results of the analysis using the three regression models for a set of 18 questions selected from the Eurobarometer. For each survey question, three estimated odds ratios are reported, corresponding to the three regression models. The odds ratio compares the odds of certain responses for frequent users to those of non-frequent users, conditional on other covariates controlled in the model. The table also indicates whether the estimated odds ratio is significantly different from 1 at the 0.05 level of significance.

It was found that the internet usage profile has a significant effect on survey responses (model 0) for all eighteen items. After controlling for education (model 1), significant differences between frequent and non-frequent users remain for four of the items and these differences continue to be statistically significant even after incorporating the effect of all three demographic characteristics influencing the internet usage profile (model 2).

Five questions in the analysis related to the goods people own. The probability of having the goods was modeled. The baseline model showed that the probability of having each of the goods is higher for frequent users than for non-frequent users. That relation holds across the multivariate models for all the items. After controlling for the demographic variables the effect of the internet usage profile becomes non-significant (model 2).

For the three variables concerning actions to prevent climate change, the probabilities of having taken each of the suggested actions were modeled on the logistic models. The estimated odds ratio for “separating waste” reflects a significant difference in all the models.

For the four items concerning satisfaction, we modeled the probabilities of being satisfied or very satisfied with the statements in the logistic models. The odds ratio of model 0 indicates that there is a greater

probability of frequent users being satisfied/very satisfied with all the statements. The difference in odds becomes non-significant when demographic characteristics were controlled in the estimation.

Table 4: Survey questions odds ratios for the three logistic regression models

Question	Model 0	Model 1	Model 2
<i>Which of the following goods do you have? ...</i> ^(a)			
DVD player	1.906*	1.107	1.028
Music CD player	1.699*	1.189	1.146
Car	2.356*	1.591	1.766
Television	3.684*	1.125	1.390
Computer	3.833*	1.534	1.585
<i>Which of the following actions have you personally taken to fight climate change?</i> ^(a)			
Reducing energy consumption at home	2.136*	1.618	1.603
Separating waste	3.219*	1.962*	2.127*
Reducing the consumption of disposables	1.861*	1.632	1.516
<i>How satisfied are you with ...?</i> ^(b)			
Your life in general	2.162*	1.205	1.309
Your family life	1.838*	1.290	1.303
Your health	1.999*	1.150	1.425
Your present standard of living	1.938*	1.298	1.457
<i>Which of the following do you think are absolutely necessary for a person/family to have a minimum acceptable standard of living?</i> ^(a)			
Good quality and diversified food	1.693*	1.761*	1.777*
Having access to means of communication	0.417*	0.391*	0.350*
<i>Which of the following groups are at greater risk of poverty?</i> ^(a)			
People in precarious work	1.825*	2.649*	2.694*
Addicts	1.984*	1.565	1.610
<i>Do you think that it is necessary for children to have access to pre-school education before primary?</i> ^(d)			
	1.872*	1.490	1.608*
<i>To what extent do you agree/disagree that environmental protection is an obstacle to economic growth in the European Union?</i> ^(e)			
	2.592*	1.947	2.247

^(a) Probability modeled: "yes"; ^(b) Probability modeled: "satisfied/very satisfied"; ^(c) Probability modeled: "almost never/never;

^(d) Probability modeled: "absolutely necessary"; ^(e) Probability modeled: "totally disagree"

* Significant at p-value<0.05.

The odds ratio for the two items relative to conditions for a minimum acceptable standard of living are significant across the multivariate models with an upward trend.

For the remaining items, it should be noted that the effect of the internet usage profile remains significant for "good quality and diversified food", "having access to means of communication" and "people in precarious work" even after controlling for the demographic characteristics (model 2).

5. Conclusion

The study compared the answers of frequent and non-frequent home internet users on a set of attitudinal and behavioral items. The outcomes reveal a significant association between the profile of internet

usage and demographic characteristics: frequent users tend to be younger, more educated and single, and less likely to be employed and to live in households with children under 10. However, multivariate analysis revealed that, when accounting for the interrelations between demographics, only education, type of region of residence and presence of children in the household stay significant.

The multivariate logistic models also highlighted the differences between frequent and non-frequent users in attitudes and behaviors. The odds ratio values showed that most of the differences between frequent and non-frequent internet users disappeared after controlling for the relevant demographic variables; however for four of the eighteen items differences remained significant: *Actions to fight climate change to be personally taken*: “Separating waste”; *Absolutely necessary for a person/family to have a minimum acceptable standard of living*: “Good quality and diversified food” and “Having access to means of communication”; *Groups at more risk of poverty*: “People in precarious work”. This means that the differences between frequent and non-frequent internet users go beyond demographic differences in some important behavioral and attitudinal aspects; so, the “frequency divide” still makes a difference even if controlled by relevant demographic variables and is unlikely to be reduced just by sampling weighting procedures. These outcomes stress the importance of controlling the frequency of internet usage when selecting respondents for a web survey. If this is not done, the sample will tend to be formed predominantly by frequent users and this could be a cause of bias in survey estimates.

REFERENCES

- Callegaro, M. and Wells, T. (2008). Is the Digital Divide Still Closing? New Evidence Points to Skewed Online Results Absent Non-Internet Households. Available at <http://www.knowledgenetworks.com/accuracy/summer2008/calle-garo-wells.html>. Accessed 22nd February 2011.
- Camacho, K. (2005) Digital divide in Word Matters: multicultural perspectives on information societies, Alain Ambrosi, Valérie Peugeot and Daniel Pimienta (Editors), C & F Éditions.
- Couper, M. (2000) Web surveys: a review of issues and approaches, *Public Opinion Quarterly*, 64, 464-494.
- Dever, J., Rafferty, A. and Valliant, R. (2008) Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias?, *Survey Research Methods*, 2, 2, 47-62.
- European Commission (2009) Eurobarometer 72.1 Codebook.
- Pierkowski, L., Galin, M., Baim, J., Frankel, M., Augemberg, K. and Prince, S. (2008). Internet Access Panels and Public Opinion and Attitudes Estimates. Paper presented at the 63rd Annual Conference of the AAPOR, New Orleans, 15-18 May.
- Robinson, J., DiMaggio, P., Hargittai, E. (2003) New social survey perspectives on the digital divide, *IT&Society*, 1, 5, 1-22.
- Robinson, J., Neustadt, A., Kestnbaum, M. (2002) The online “diversity divide”: public opinion differences among internet users and nonusers, *IT&Society*, 1, 1, 284-302.
- Valliant, R. and Lee, S. (2005). Economic characteristics of internet users vs. non-users and implications for web-based surveys, *Webuse&Society*, 1, 34-51.