Spring 4-9-2014

Characteristics Associated With The Lack Of Interest In Use Of The Internet In Brazil

Emerson Gomes dos Santos
Brazilian Network Information Center, Brazil, emerson@nic.br

Characteristics associated with the lack of interest in use of the Internet in Brazil

Emerson Gomes dos Santos
Brazilian Network Information Center (NIC.br), Brazil
Email: emerson@nic.br

Abstract

International organizations cooperate to design methodologies, indicators and metrics capable of measuring access and effective use of new technologies by different sectors of society. The use of the Internet offers many benefits to everyone who is willing to use it, mainly because it makes an enormous amount of information easily available. The many uses one can have through use of the internet have made it a valuable tool in various settings of a person's social life. However, some people do not use the Internet and do not care about using the internet. Therefore, the motivation of this research is to study the characteristics of those who do not have interest in the use of internet, despite the many known advantages.

Keywords: Digital Divide, ICT Households Survey, Analysis of Complex Sample Survey Data.

1.0 Introduction

International organizations cooperate to design methodologies, indicators and metrics capable of measuring access and effective use of new technologies by different sectors of society, including The United Nations (UN), the Organization for Economic Cooperation and Development (OECD), the United Nations Educational, Scientific and Cultural Organization (UNESCO), and the World Bank, among others [International Telecommunications Union; Eurostat].

In Brazil, since 2005, the Center of Studies on Information and Communication Technologies (Cetic.br), a department of the Brazilian Network Information Center (NIC.br) - executive branch of the Brazilian Internet Steering Committee (CGI.br) - has been producing statistics on the use of information and communication technologies (ICT) in Brazil. The CGI.br's ICT Households Survey, for which the main aim is to measure the use of ICT by the Brazilian population aged 10 years old or older, shows solid and continuous progress in use of the Internet in Brazil. The results of 2005 indicate that 24% of the Brazilian population accessed the Internet in less than three months (Internet users) – i.e. they had connected to the Internet in the three months preceding the survey - whilst this proportion increased to 49% in the 2012 [Brazilian Internet Steering Committee].

The use of the Internet offers many benefits to everyone who is willing to use it, mainly because it makes an enormous amount of information easily available. The many uses one can have through use of the internet have made it a valuable tool in various settings of a person's social life. In addition, use of the Internet has made jobs easier and simplified tasks.

However, some people do not use the Internet and do not care about using the internet. Therefore, the motivation of this research is to study the characteristics of those who do not have interest in the use of internet, despite the many known advantages. The target population of our study is all the Brazilian citizens aged 10 years old or older living in whole country in 2012. The objectives are to describe some characteristics of those people and assess the relationship between those selected variables and the dependent variable indicating those specific people who reported never having accessed the Internet because of no need or no interest. In the following sections will be described the methods, the complex sample data used, the results based on the hypothesis via the specific aim, and lastly, the discussion and conclusion.

2.0 Methods

2.1 Data source and description of the complex sample design

The ICT Households Survey used for this analysis was conducted by Cetic.br in 2012. The sample plan was designed to implement a probabilistic sampling procedure at all stages of selection. The sample plan design included a stratified cluster sample in multiple stages and selected with probability proportional to population size (PPS). The target population are the citizens aged 10 years old or older [Brazilian Internet Steering Committee].

The initial allocation of municipalities considered 32 geographic strata. These strata were defined considering state, capital, and countryside. Different numbers of clusters were allocated to each stratum. The plan used to obtain the enumeration area samples may be described as a stratified cluster sample in one or two stages, depending on the stratum. The number of stages in the sampling plan depended primarily on the role given to the selection of municipalities. Several municipalities were sampled with probability equal to one (self-representative municipalities). In this case, these locations functioned as strata for enumeration areas sample selection and, later, for

household and resident selection. For this reason, this was not considered the first stage of selection. Hence, the other municipalities not included in the certainty sample function as primary sampling units (PSUs) for the initial stage of the survey. At this stage, there were two stages in the probability sample – i.e. the first was the selection of municipalities, and the second involved selection of enumeration areas. In the first two stages of sample selection (i.e. selection of municipalities and enumeration areas), the sample units were drawn based on probabilities proportional to size, also called Selection with Probabilities Proportional to size Measures (PPS). After that, the enumeration areas samples were used to select households and residents. Thus, there are "base weights of the design" for each unit surveyed, households and residents. In addition, each one was adjusted considering marginal distribution of the calibration variables for Brazilian 2010 Census [Brazilian Institute of Geography and Statistics].

2.2 Measures and Objectives

2.2.1 Dependent variables, independent variables and objectives

I wish to measure the relationship between a categorical dependent variable that indicates those who do not have interest in the use of internet with selected independent variables. The independent variables include: gender; groups of age; level of education; marital status; economic activity status; social status; characteristics of housing (number of televisions, telephones and mobile phones) and divisions geographical (type of area and region). The independent variables were classified as the following values: Gender: 1 (Male), 2 (Female); Groups of Age: 1 (10 to 15 years old), 2 (16 to 24 years old), 3 (25 to 34 years old), 4 (35 to 44 years old), 5 (45 to 59 years old), 6 (60 years old or older); Level of Education: 1 (Up to Elementary), 2 (High School and over); Marital status: 1 (Single), 2 (Married), 3 (Widowed or Separated); Economic activity status: 1 (Economically active population), 2 (Non-Economically active population); Social status (takes into account ownership of durable goods for household consumption, plus the level of education declared by the head of the family): 1 (A or B), 2 (C), 3 (D or E); Total TV: 1 (up to 1), 2 (2 or over); Total phone: 1 (none), 2 (1 or over); Total mobile phone: 0, 1, 2, 3 or 4 units; Area: 1 (Urban), 2 (Rural); Region: 1 (North), 2 (Northeast), 3 (Southeast), 4 (South), 5 (Center-West).

Therefore, in order to evaluate this association, we defined the following research Aim: Assess the relationship between those selected variables, that represent the

characteristics of those people, and the variable "No Internet Users" which was categorized as: "1" if people who have never accessed the Internet, but have used computers and said that never accessed the Internet because they saw no need or had no interest and "0" otherwise.

The central hypothesis is that people never having accessed the Internet because of no need or no interest are associated with group of age, level of education, and social status. In addition, people who live in more developed regions may have more interest to use the Internet.

2.2.2 Sampling error variables

In the ICT Households Survey, there were one hundred thirty-two stratum identifiers, two weighting variables (households and residents) and one primary sampling unit (PSU) variable. There are one thousand fifty nine PSUs. The resident weight was used in our analyses since the unit of analysis of this study was the person and most variables we assessed were in the interview file. This weight variable was calculated to adjust for couple of things including nonresponse, post stratification, incomplete interviews, and calibration.

3.0 Statistical analyses

In this study, the complex sample design effects were incorporated into the analyses, including weights, clustering and stratification. We utilized the Taylor series linearization method for variance estimation, which accounted for the sampling weights and the characteristics of the complex sample design. In order to estimate the effects of the complex sample design features on variance estimates SAS commands were utilized.

Initial descriptive summaries were performed for the key demographic and analysis variables that will be included in subsequent bivariate analyses (Wald chi-square statistic). For the research aim, since the dependent variable "No Internet Users" was a binary variable, a multivariate logistic regression analysis of survey data was utilized to assess the relationship between this variable and some selected variables. Preliminary bivariate association analyses were used to find out potential predictors of "No Internet Users", it means predictors of people who never having accessed the Internet because of no need or no interest. The predictors that have a bivariate

association with “No Internet Users” at a significance level of $p < 0.25$ were considered as candidates for main effects in a multivariate logistic regression model, following Hosmer and Lemeshow’s (2000) incremental process for specifying the initial model, refining the set of predictors and then determining the final form of the logistic regression model. Therefore, those were included in multivariate logistic regression models, in which was evaluated the categorical predictors in the model by using the design-adjusted Wald test then all significant predictors and variables of interest were retained in the model.

4.0 Results

4.1 Descriptive analyses

Table 1 displays the characteristics of our study population and the subpopulation (No Internet Users). There are 258 people reported to have no interest in the Internet, at the interview time. We can compare the distribution of the total population with this specific subpopulation, and the variables of interest: groups of age, social status and region. About 10% of the subpopulation was 16 to 24 years old (category 2 ad 6) versus 18% in the total population and there were 19% people between 60 years old or more in the subpopulation (No Internet Users) and only 14% in the whole population. The social status shows lower percentage for the class 3 (D or E) compare to the population (17% opposite 24%). About geographical characteristics, 8% of total population were in North (category 1) but in the subpopulation, there were 15% of people.

Categories of variable		No Internet	Total Population	Categories of variable		No Internet	Total Population
Gender	1	52%	48%	Total TV	1	55%	50%
	2	48%	52%		2	45%	50%
Groups of Age	1	16%	12%	Total phone	1	57%	65%
	2	10%	18%		2	43%	35%
	3	20%	19%	Total mobile phone	0	7%	9%
	4	19%	17%		1	27%	23%
	5	15%	20%		2	28%	27%
	6	19%	14%		3	22%	20%
Level of Education	1	60%	57%		4	16%	21%
	2	40%	43%	Area	1	91%	85%
Marital status	1	88%	90%		2	9%	15%
	2	12%	10%	Region	1	15%	8%
Economic activity status	1	61%	63%		2	21%	27%
	2	39%	37%		3	43%	43%
Social status	1	28%	27%		4	15%	15%
	2	56%	49%		5	5%	7%
	3	17%	24%				

Table 1. **Weighted estimates of percentages describing the population features (%) of independent variable for people "No Internet Users" and for the total population aged 10 years old or older ICT Households 2012 (N=17380).**

4.2 Descriptive analyses

Table 2 presents the Wald Chi-square tests results, (design-adjusted), testing the hypothesis that two categorical variables are independent, for each predictor variable and the "No Internet Users" variable. Wald test fails to reject a null hypothesis of independence between "No Internet" and these variables: Groups of Age, Social Status, Total phone, Area and Region (at significance of 0.25).

This indicates that may be an association between the response variable and the selected variables. The appropriate inference in this case would thus be that there is evidence of a bivariate association between these five categorical factors in this

subpopulation. Examining additional potential predictors of "No Internet", we can see that p-values indicates rejection of the null hypothesis of independence. In other words, Gender, Level of Education, Marital Status, Economic activity status and Total TV are not significantly associated with no internet users.

Predictor	Chi-Square Test	p value
Gender	Wald Test - F(1,927)=0.4264	0.51
Groups of Age	Wald Test - F(5,927)=1.9124	0.09
Level of Education	Wald Test - F(1,927)=0.3106	0.58
Marital status	Wald Test - F(1,927)=0.1955	0.66
Economic activity status	Wald Test - F(1,927)=0.1417	0.71
Social status	Wald Test - F(2,927)=1.8891	0.15
Total TV	Wald Test - F(1,927)=0.6772	0.41
Total phone	Wald Test - F(1,927)=1.9308	0.17
Area	Wald Test - F(1,927)=3.4054	0.07
Region	Wald Test - F(4,927)=2.3551	0.05

Table 2. Initial Bivariate Design-Based Tests of Association Assessing Potential Predictors of "No Internet Users" for the total population aged 10 years old or older ICT Households 2012 (N=17380)

4.3 Logistic regression models predicting "No Internet Users" (for people who declared that no need or no interest in the Internet)

This section shows the results of the logistic regression analyses. The hypothesis tests for each of the variables in the initial model, which considering the variables: Groups of Age, Social Status, Area, Region, Total Phone and Total mobile phone. The Chi-square test statistics and associated p-values in the Table 3 indicate that three of those six variables in the initial model were excluded to the final model because they are not significantly at level of 10% (Social Status, Total Mobile Phone and Area).

Predictor	DF	Wald Chi-Square Statistic	P-value
Social Status	2	3.544	0.17
Groups of Age	5	9.133	0.10
Total Phone	1	3.291	0.07
Total Mobile Phone	1	1.043	0.31
Area	1	2.174	0.14
Region	4	21.077	0.00

Table 3. Wald Chi-square tests for the parameters in the Initial Logistic Regression Model for the total population aged 10 years old or older ICT Households 2012 (N=17380)

Next, Table 4 presents Wald Chi-square test statistics for the parameters in the final estimated model, which indicates that each of the three variables in the model significantly improve the model fit using a 0.10 level of significance. This table gives the multiple degree of freedom test for the overall effect of each variable.

Predictor	DF	Wald Chi-Square Statistic	P-value
Groups of Age	5	9.517	0.09
Total Phone	1	4.632	0.03
Region	4	15.522	0.00

Table 4. Final Estimated Logistic Regression Model for "No Internet" (no need or no interest) for the total population aged 10 years old or older ICT Households 2012 (N=17380)

The software procedures for logistic regression analysis of survey data generally offer the analyst the option to output parameter estimates and standard errors or transformed estimates of the corresponding adjusted odds ratios and confidence intervals. Therefore, Table 5 shows the final model coefficients (labeled Estimate), their standard errors (Standard Error), the Wald Chi-Square statistic, and associated p-values. The coefficient for Total Phone is statistically significant, as well as Groups of Age and Region. Table 5 also presents Odds Ratios (ψ) and their respective 95% Confidence Intervals (CI).

Predictor*	Category	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	ψ	95% CI for ψ
Intercept		-3.78	0.37	104.36	<.0001		
Groups of Age	10-15 years old	-0.05	0.43	0.01	0.91	0.96	(0.412; 2.211)
	16-24 years old	-0.83	0.40	4.30	0.04	0.44	(0.198; 0.956)
	25-34 years old	-0.21	0.32	0.45	0.50	0.81	(0.434; 1.504)
	35-44 years old	-0.11	0.46	0.06	0.81	0.89	(0.36; 2.212)
	45-59 years old	-0.57	0.32	3.29	0.07	0.56	(0.303; 1.047)
Total Phone		-0.42	0.19	4.63	0.03	0.66	(0.449; 0.963)
Region	North	0.81	0.26	9.60	0.00	2.26	(1.349; 3.779)
	Northeast	-0.11	0.26	0.18	0.67	0.90	(0.541; 1.487)
	South	0.09	0.33	0.07	0.79	1.09	(0.567; 2.106)
	Center-West	-0.23	0.36	0.40	0.53	0.80	(0.391; 1.621)

*Reference categories for categorical predictors: Groups of Age (60 years old or more) and Region (Southeast).

Table 5. Final Estimated Logistic Regression Model and Odds Ratio (ψ) for "No Internet" (no need or no interest) for the total population aged 10 years old or older ICT Households 2012 (N=17380)

The results of the logistic regression analyses for the outcome, "No Internet" were presented in Table 4 and 5, where the estimates take into account the complex design features of ICT Households 2012. The results from the final model (Table 5) suggest that there are significant disparities in some characteristics related to interest in the Internet, including groups of age, existence of phones and geographic location of their residence (Region).

In terms of age disparities, compared to oldest people (60 years old or more), the odds ratio for people who are 16 and 24 years old is 0.44, in others words, people between 16 and 24 years old presents 56% decrease in the odds of No Internet Users comparing with people are 60 or more. This group of younger people is the only group of age that presents statistical significance in odds relative of 60 or over.

The coefficient for total phones says that, holding Groups of Age and Region at a fixed value, we will see 34% decrease in the odds of not using the internet for people that lives in household without telephones relative those households where there is one or more telephones.

This fitted model also shows that, holding Groups of Age and Total Phone at a fixed value, the odds of North Region over the odds of Southeast is 2.26. In terms of percent change, we can say that the odds for North are 126% higher than the odds of

no need or no interest in use the Internet, relative to Southeast. Only the North region presents a statistically significant difference in the odds relative to the Southeast.

5.0 Discussion

I intended to estimate a logistic regression model, which would compute the relevant odds for each pertinent independent variable. In addition, the complex sample design of the ICT Household Survey was accounted for in this analysis. In general, the adjusted odds ratio represents the multiplicative impact of a one-unit increase in the predictor variable on the odds of the outcome variable being equal to 1 holding all other predictor variables constant.

The results of the model at the specific characteristics of Brazilian population, we found that some of them do not show disparities to people declaring no need or no interest to be on the Internet. However, of those variables that could be assessed in groups, one of them is about characteristics of people (Gender, Level of Education, Marital Status, Economic activity status) and other about characteristics of households (Total TV, Total mobile phone and Status Social) all them are not significantly associated with No internet Users. Besides, people who live in North presents greater odds of no interest of the Internet compared with the more developed Region (Southeast), and the Area variable is also not significantly associated with No internet.

First, the proportion of individuals who have already accessed the Internet is 10% in the oldest group (60 years old or more) compared with 83% for the people in the group of age between 16 and 24 years old. Therefore, the final model show us that this group of younger people presents 56% decrease in the odds of no need or no interest on the Internet. In general, besides the higher proportion of users of the Internet, younger people shows more use of more complex activities, like using microblogs, such as Twitter, and creating or updating blogs and/or Internet pages. These are different activities compared to what the oldest people usually declare to do on the Internet, such as searching for information on goods and services like travelling [Brazilian Internet Steering Committee].

Second, about the geographical disparities, this study show that the odds of North Region is more than a factor of 2 or (126%) higher than the odds of no need or no interest in use the Internet, relative to Southeast. Regionally, in the Southeast, the technology is present in almost half of the households (48%), however in the North;

access to the Internet is at less than one quarter of households (21%), which represents the lowest proportion among the country's regions [Brazilian Internet Steering Committee]. If we look at the rural/urban area division, we can also see a higher proportion of households with Internet in urban areas. In addition, there is a higher proportion of Internet user among people born into a wealthy family (reflecting social status). However, in this study we do not find statistical significance for area and social status relative to no interest on use the Internet. This finding might imply the policy interventions to focus on this critical issue of equal efforts made on the distribution of infrastructure for all parts of the country.

Lastly, the results of total phones surprisingly show 34% decrease in the odds of no interest on the Internet for people that live in households without telephones, relative those households where there is one or more telephones. However, currently there are no consistent studies about this and future investigation of the appropriateness of these patterns has to be done.

6.0 Conclusion

Analyses were conducted in SAS, a software package that includes procedures capable of correctly accounting for complex sampling features such as ICT Household.

The mainly findings are: 1) younger people have decreased odds of no need or no interest on the use of the Internet compared to the oldest age group, 2) the results suggest some geographical disparities with odds in North Region more than 126% higher than the odds of no need or no interest in use the Internet, relative to Southeast, and 3) area (rural or urban), level of education, and social status were not statistically significant in explaining no interest on use the Internet. This finding might imply the policy interventions to focus on this critical issue of equal efforts made on the distribution of infrastructure for all parts of the country and implied the direction of future studies about the use of the Internet.

References

International Telecommunications Union *ICT data and Statistics*. Available at: <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>. Accessed on: January 24, 2014

Brazilian Institute of Geography and Statistics *Censo Demográfico 2010*. Available at: <http://www.ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm> . Accessed on: January 24, 2014

Brazilian Internet Steering Committee (2012) *Survey on the use of information and communication technologies in Brazil: ict Households and ict enterprises 2012*. Available at: < http://www.cetic.br/publicacoes/2012/tic-domicilios-2012.pdf >. Accessed on: January 24, 2014

Eurostat Statistics by theme. Available at: <http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes>. Accessed on: January 24, 2014

Hosmer, D.W. and Lemeshow, S. (2000), *Applied Logistic Regression*, 2d ed., John Wiley & Sons, New York.

Heeringa, S.G., West, B.T. and Berglund, P.A. (2010) *Applied Survey Data Analysis*. Chapman Hall / CRC Press: Boca Raton, FL.