Performance of Weighted and Non-weighted Estimators in a Cell-phone based Electoral Poll: An Academic Study of the 2012 Presidential Elections in Mexico.

[1]Olivia Carrillo-Gamboa, [1]Rosa Isela Hernández-Zamora, [2]Jesús Cantú-Escalante

[1]Tecnológico de Monterrey.
Mathematics Department
Ave. E. Garza Sada 2501, Monterrey, N.L. MEXICO 64849
ocarrillo@itesm.mx; isela.hz@itesm.mx;

[2]Tecnológico de Monterrey
EGAP Gobierno y Política Pública
Av. Eugenio Garza Lagüera y Rufino Tamayo,
San Pedro Garza García, N.L. MEXICO 66269
jce@itesm.mx

The findings and conclusions in this report are those of the authors and are not necessarily a point of view of *Tecnólogico de Monterrey*.

Abstract

Electoral polls and surveys applied to political purposes have being used widely in Mexico for the last two decades. Although most of the methodologies applied are based on multiple-steps cluster sampling procedures and face-to-face interviews, these procedures are becoming of difficult application due to safety related issues.

On the other hand, methodologies based on landline phone sampling frames are known to yield samples with demographic characteristics substantially different from those of the target population –and potentially biased results- due to low coverage of the sampling frame as well as to the presence of other non-sampling errors tied to the logistic of the methods applied.

An alternative telephone methodology was applied in a series of academic studies in the context of the past 2012 Mexico's presidential elections which were based on samples drawn from a cell-phone sampling frame. The cell-phones interviews yielded demographic characteristics such as gender, geographical distribution and urban/non-urban composition comparable to the target population. The only concern from the cell-phone sample representativeness could be raised by the age-group composition since it can be easily understood older citizens -60 years and older- might have lower cell-phone coverage and could become underrepresented by the sample and at the same time younger age groups would be overrepresented.

This paper focuses on the performance of non-weighted estimators and of estimators weighted by age group when compared against official electoral results of the 2012 federal election. Both estimators were found to have an excellent practical performance and, although the weighting procedure could be the best theoretically recommended from a statistical point of view, the non-weighted estimator was slightly closer to the official results. A plausible explanation for this finding is also provided in the context of the 2012 Mexico's presidential elections.

Keywords: mobile-phone surveys; sampling bias; landline telephone; political surveys; sample selection for telephone surveys.

Introduction

Survey studies are widely used in Mexico as is the case of survey studies on electoral processes or in marketing research with the purpose of improving the commercial practices for all kinds of goods and services. In particular, electoral surveys have been used extensively for more than two decades in Mexico and the use of statistically planned methodologies keep improving with time.
This work focuses on the comparison of the performance of two estimators in the context of an electoral survey in Mexico based on a mobile phone sampling methodology. The two estimators are the proportion non-weighted estimator and a weighted estimator when adjusting for age-group representation in the target population.

The electoral survey aimed to estimate the 2012 Mexico's presidential elections' results. The survey methodology was based on a sample of national mobile phones. This methodology has proven to be useful in a series of regional studies in the Monterrey Metropolitan Area since 2011 yielding sample characteristics as gender and geographical distribution similar to those of the adults' target population.

Federal election had four presidential registered candidates, here identified by their name initials in the order of chronological appearance of their political party, which is the order accustomed by the media: JVM, EPN, AMLO and GQT.

The performance of each of the two sets of estimations is obtained by contrasting them to official electoral results of the 2012 federal election taking into account only the votes casted in favor of any of the four registered presidential candidates. For a better appreciation of the estimators' performance their closeness to official results are set along to other sets of estimations published by the media and by the *Instituto Federal Electoral (IFE)*, the federal elections official organism.

Methods

Two alternative interviewing methodologies are of common use for studying adult wide open population: landline based telephone interviews or household face-to-face interviews where household samples are drawn through multiple-steps cluster sampling techniques. Other methodologies like surveys based on postal mail or internet based questionnaires are not used in general. The postal mail is not of practical use and the later one has not enough coverage of the open adult population, although it can be the most appropriate method for covering very specific target populations of internet users.

As mentioned before, electoral surveys directed to the open adult population in Mexico often depend on methodologies based on household sampling procedures and *face-to-face* interviews. However these methodologies confront more obstacles every day to operate. For instance, it can be mentioned that nowadays in major cities in the country there are numerous household communities with restricted or no public access. These household communities have become common not only in wealthy zones but also in the middle class socioeconomic level. At the same time, women in modern and young families have more active roles in the formal economy than they did in the past and often they work out of their house. Empty households, with no persons available for interviews during day time are therefore more common and as a consequence bigger discrepancies are to be found between the target and the sampled population.

With regard to telephone surveys, landline sampling frames are known to have low population coverage. Just as it is the case in other countries (Lavrakas, Shuttles, Steeh and Fienberg 2007; Arthur 2007; Vicente and Reis 2010), the landline telephone coverage in Mexico is dropping in time while the mobile coverage is rising. Statistics from INEGI, the government official statistics organism, reported landline telephone coverage of 53% of households in 2007 (INEGI 2007) and the coverage dropped to 44% in 2011 (INEGI 2011). Similar tendencies show statistics from the Federal Telecommunications Commission (COFETEL), reporting 19% of the population with a landline phone in 2007 and only 17% in 2011 (COFETEL 2013).

On the contrary, cell phone coverage in Mexico is increasing in time. The Federal Telecommunications Commission reported 64 subscriptions per 100 habitants to December 2007 and 87 is the corresponding figure up to June 2012 (COFETEL 2013). Additional to the population coverage, mobile phones permit reaching citizens in the target population from rural as well as form urban zones and interviews can be made to citizens from practically all

demographic profiles. However, despite the advantages of mobile phone sampling mentioned above, coverage is not perfect. People in the age group of 60 yrs. and older are found less likely to be represented in the effectively interviewed sample and could become underrepresented by the interviewed sample and at the same time younger age groups would be overrepresented. That fact gave rise to this work in order to compare the actual performance of the two estimators, the age-weighted and the non-weighted estimator for proportions, since they are applied to a phenomenon where the population parameters can be observed (or nearly observed) as it is the case in an electoral process.

Although this work does not aim comparing methodologies, relative performance of the mobile phone sampling methodology applied in this electoral survey is also viewed by contrasting the absolute differences of each of the two estimators to the official results to the corresponding differences for estimations published by the media based in other methodologies.

Sample Design

The target population for the electoral survey is that composed by all the 18 years and older citizens in Mexico voting in the presidential elections on July the 1st 2012. The country is divided in 31states and a Federal District. The 32 federal entities are very diverse in their demographic characteristics as well as in political practices and preferences and hence geographic sample representation is an important concern in the sample planning phase. As mentioned before, the methodology based on mobile phone sampling was elected taking into account previous experiences in several metropolitan area studies where phone samples drawn from all possible cell numbers in a *Calling Party Pays (CPP)* modality represented well the demographic and geographic population characteristics.

As a first national experience, sampling from all possible CPP mobile numbers was done in several steps, by generating a random national subsample in every step and mutually exclusive subsamples among the steps. A first random subsample was handed in to the call center to be worked until exhausted. Recalls were to be made at least twice to not contacted persons in a ringing phone. Persons not able to answer in the moment were offered a scheduled call at a more appropriate time. An additional disjoint and randomly ordered subsample was added when a subsample was exhausted. The number of subsamples needed would depend on the effective contact rate and interviewing results, which were also subject to meet time and budget restrictions.

Data Collection

Telephone interviews were carried out between June 25th and 28th from 9:00 to 21:00 hours central time. Data collection took place just a few days before the presidential elections day: July 1st, 2012. Contacted citizens were informed *Tecnológico de Monterrey* was carrying out an *electoral survey* for academic purposes. They were assured of the anonymity of the call (no personal data was available and no personal data was to be registered in the interview). An expected length of the interview was also given (6 to 8 minutes).

It is important to point out that cost was not a concern for the interviewed persons since calls were cost-free for those interviewed in their local area since all cell users are in the *CPP* modality.

After discarding calls classified out of the target population and those who asked to be called in another time, a total of 6648 calls were accounted. From these, 3856 (58%) agreed to be interviewed and 2792 (42%) were persons not willing to answer the electoral interview. Note that since most of the latter gave no information in order to be classified as belonging or not to the target population, this cannot be considered as a nonresponse rate from the target population. Calls to be considered as a true nonresponse from the target population would have to be from the count of the Mexican citizens, 18 yrs. and older, in possession of their voting credential and planning to vote. The 42% could be understood as an upper value for the true, but unknown nonresponse rate.

From the 3856 calls to citizens willing to answer the interview, 3479 (90.2%) counted with the voting credential required to cast their vote and only 3304 (85.7%) were planning to vote or had not completely discarded to do so. Also note that this figure does not constitute an estimate of the percent of voting citizens since the percent is over the number of persons willing to answer an *electoral survey* and many of those not able to vote or not planning to vote had discarded themselves from the beginning by not acceding to the interview. The actual official voting participation in the presidential election is reported as 63.08%; (IFE 2012b).