RESOLVABILITY, SCREENING AND RESPONSE MODELS IN RDD SURVEYS: UTILIZING GENESYS TELEPHONE-EXCHANGE DATA

Ronghua (Cathy) Lu, John Hall and Stephen Williams
Mathematica Policy Research, Inc., P.O. Box 2393, Princeton, NJ 08543-2393

Key Words: Nonresponse, Weighting, Propensity Modeling, RDD Telephone Survey

Abstract

Answering machines, call waiting and other screening and call-blocking services are increasingly being used by households. These popular innovations, however, may be significantly influencing how telephone surveys are conducted. For random digit dialing (RDD) household surveys, for example, the resolution of whether or not a number identifies a household is becoming more challenging. Likewise, obtaining household participation in a survey is becoming more difficult. Hence, not only are survey methods being revised, but more effective methods are being sought to adjust for the lower rates of successfully screened numbers and completed responses.

This paper investigates the feasibility of using logistic regression models to predict for each sampled RDD number in a recently completed household survey setting, the likelihood of resolving the number and the likelihood, if the number is a household telephone number, that the household will complete the screening questions, and finally that the household will complete the questionnaire. The RDD numbers in this example are generated through the GENESYS system, which contains a substantial amount of auxiliary demographic information about households at the telephone-exchange level. We seek to use this auxiliary data (in addition to sample design information) in "propensity" models to adjust the sampling weights for nonresponse.

1. Introduction

We use GENESYS Sampling Systems to generate RDD numbers for many of our telephone surveys at Mathematica Policy Research, Inc. Hundred-blocks (i.e., NPA-NXX-00xx, NPA-NXX-01xx, … NPA-NXX-99xx) containing one or more listed telephone numbers are used to generate candidate telephone numbers. These numbers are then checked using an auto-dialer, so business and some nonworking numbers can be removed. The remaining numbers are assigned to the telephone centers for conducting the survey. Three issues are of concern: (1) some numbers cannot be resolved as to whether or not they are assigned to a household (assuming our survey is a household survey); (2) for those identified as households, the screening cannot be completed for some; (3) and for those screened in households, some do not complete the interview.

The setting for our study is a statewide household survey recently conducted by MPR, an RDD survey of New Jersey residents, age 18 and older. The sample was selected within five geographic strata in New Jersey. A total of 35,909 RDD numbers were assigned to telephone interviewers, but about 28 percent could not be resolved. This results in uncertainty about the number of eligible households in the study population and in important inference domains. Further, about 35 percent of identified households did not complete a screener and 20 percent of those screened in households did not complete an interview. These results, while not unusual for telephone surveys, are potential sources of survey bias.

The information available to us in the GENESYS sample files includes demographic characteristics at the telephone exchange level, such as age, income, race, education, and home ownership. Each exchange contains 10,000 potential telephone numbers. While many of these are not assigned to households, the area covered by an exchange is still large, diverse, and non-dense. This type of information is not ideal for developing propensity models, but considering its availability and the lack of comparable person level data, we decided to explore its merit for use in adjusting for missing information.

In this paper, we investigate the feasibility of using auxiliary demographic information available on the GENESYS sample files to predict "resolvability", "screening" and "response". We note that the level of screening response was low due to a relatively complex set of screening questions. The results of the survey and the GENESYS data will be the basis for constructing these three models. In addition to the GENESYS data, geographic stratification and the field outcome codes also will be used as candidate independent (explanatory) variables. We seek to use these auxiliary data in "propensity" models to adjust the sampling weights, in order to reduce the potential for bias that can result from the missing information.

2. Methodology

Propensity modeling -- We use propensity models, which are an increasingly popular method for adjusting for non-response: that is, creating a logistic regression model that predicts the likelihood of response versus non-response. Three propensity models need to be developed in our case: (1) a resolvability model for all 35,909 RDD numbers, to predict the likelihood of resolving the number as a household, non-household or undetermined; (2) a screening model for those known households (14,933 records in our survey), to predict the likelihood, if the number is a household telephone number, that the household completes the screening questions; and (3) a response model for those screened in households (4,965 cases in our survey), to predict the likelihood, if the household is screened in, that the eligible household completes the interview questions.

Constructing covariates for the propensity models -- The variables we had for each RDD number in the GENESYS sample files were a few dozen demographic characteristics at the telephone exchange level in which the telephone number resided, and one variable at the working bank level which was household listings per working block. They are mostly continuous variables. Continuous variables in the models, however, use only the linear relationship (unless higher-order terms are used). So if we simply used the GENESYS variables as supplied, this would only allow us to capture linear relationships. We therefore decided to transform most of the continuous variables into a set of binary variables in order to capture nonlinearity.

First we partitioned each GENESYS variable into ten equal-sized ranges (deciles) based on its cumulative frequency. The reason we used ten levels was to avoid masking extremes that would be useful in the analyses. Then we computed the resolvability rates (or screening rates in the screening model, or response rates in the response model) for each level of each variable. We then combined levels that had similar rates into one group and created dichotomous variables for these groups; the ranges that were near the overall average rates were used as the base. To clarify, we present an example to demonstrate how we created the variables. The example relates to a variable used in the resolvability model: Total Households served by this exchange. The resolvability rates (as percentages) for the deciles were:

1. 70.61
2. 70.78
3. 68.38
4. 70.64
5. 72.18
6. 73.89
7. 73.57
8. 71.48
9. 74.20
10. 71.69

The average rate for the resolvability model is 71.75. Using the above rule, three binary variables can be created for this variable:

The base: levels 5, 8, 10 (since these levels are close to average rate: 71.75)

X1: level 3 (low)
X2: levels 1, 2, 4 (medium)
X3: levels 6, 7, 9 (high)

After creating binary variables for most of the continuous variables, we had 96 variables in the resolvability model, 122 variables in the screening model, and 122 variables in the response model. These variables are the initial candidates for the three models.

Using weighted stepwise logistic regression in SAS -- Once the initial candidate variables for the three models were identified, we used stepwise weighted regression to reduce the list of variables in the three models before we considered possible interactions. Stepwise logistic regression in SAS was used, setting the significance level at 0.05.

Sampling weights, adjusted at each stage, were used to obtain unbiased estimates. That is, we started with the resolvability model using the unadjusted basic weights, which were the inverse of the probability of selection. Then, after the resolvability model was finalized, we used it to adjust the weights for unresolved numbers by dividing the sampling weight for each resolved record by the estimated probability of resolution from the resolvability model. These adjusted weights were then used for the screening model. And finally, after adjusting the weights again, this time for screening non-response, we had the weights for developing the response model.

After running three weighted stepwise logistic regression models, we reduced the list of variables to 22 for the resolvability model, 9 for the screening model, and 11 for the response model.

Next we ran these reduced sets of variables and potentially important 2-order interactions for three models. Again, we used weighted stepwise logistic regression with rejection level set at 0.05 to identify the final set of variables and interaction terms, 15 main effects and 9 second-order interactions were retained in the final resolvability model, 9 main effects and 2 second-order interactions were in the screening model, and 11 main effects and 3 second-order interactions in the response model.

Running the final models in SUDAAN -- The final stage was to enter the reduced sets of variables into the weighted logistic procedure in SUDAAN, and calculating the final parameter estimates using the design features (primarily stratification). The models were run using this specialized software so that the sample design could be accounted for in the parameter estimates and the variance. Since the design being used in this survey was a simple stratified random sample (no clustering), the SAS and SUDAAN runs are essentially the same: the parameter estimates were equivalent, having only slightly different P values for the individual parameter estimates.

3. Results

Table 1 shows the variables that were retained in the final three models. Fifteen main effects and 9 interactions were retained in the resolvability model, 9 main effects and 2 interactions were retained in the screening model, and 11 main effects and 3 interactions were retained in the response model.

Table 2 shows some statistics we used to test the goodness of fit and predictive powers of the models. The first one tests the global null hypotheses (Likelihood ratio statistics) that all the explanatory variables have coefficients of 0, that is beta=0. We can see from the table that P-values are very small for all three models, so we reject the null hypothesis and conclude that at least one of the coefficients for explanatory variables in each model is not zero.

The next step tests model goodness of fit, using two statistics: Hosmer-Lemeshow (HL) test and the Deviance. HL test is based on grouping predicted probabilities into ten cells. The value can range from zero to one, with larger values indicating better fit. We can see from the table that all three models pass this test. The deviance compares the fitted model with a saturated model, larger P values indicating better fit. The screening and response models are good fits according to the deviance test. For the resolvability model, we do not present the value in the table, since there are many explanatory variables in this model. As Paul Allison mentioned in his book Logistic Regression Using the SAS System, the deviance can be unreliable if there are many explanatory variables in the model, because too many profiles allow small cell counts. We have 1364 unique profiles in the resolvability model, with only 152 unique profiles in the screening model and 268 unique profiles in the response model.

The last statistics describe how well we can predict the dependent variable based on the values of the independent variables. The model concordance shows that the models are marginal but useful, with the response model being most effective based on this measure. Three measures of association: Sommer's D, Gamma and Tau-a, can all range from zero to one, with larger values indicating better association between the predicted and observed values. These statistics are all based on the concordance/discordance numbers. Tau-a tends to be closest to the generalized R-square, which is a measure that has much different meaning than the R-square we are familiar with in linear regression with a continuous dependent variables. We see the association measures are relatively low. For predictive power, the response model appears to be slightly better than the other two models.

4. Conclusions and Limitations

The results of our analysis indicate that aggregate demographics from the GENESYS file show promise for adjusting RDD sampling weights; the response model is strongest of three models based on most tests. All models are somewhat weaker than experienced for propensity models based on un-aggregated data. The significant variables in the resolvability model were economic demographic information such as education, home ownership, income, race and age; demographics such as education, income and telephone listings per "working hundred block" are important in the screening model; while income, age, race and number of listings in working hundred blocks are used in the response model.

References

Potter, F.J., McNeill, J.J., Williams, S.R. and Waitman, M.A., (1991) "List-Assisted RDD Telephone Surveys", Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA. pp 117-122.

Hosmer, D.W. and Lemeshow, S. (2000) Applied Logistic Regression 2nd Edition, John Wiley & Sons: New York

Allison, P. (1999) Logistic Regression Using the SAS System: Theory and Applications, SAS Institute, Inc., Cary, NC

Marketing Systems Group. Genesys Sampling Systems Methodology, Philadelphia, P.A. MSG 2000

Marketing Systems Group. Genesys Sampling Systems Methodology, Philadelphia, P.A. MSG 1994

TABLE 1. LOGISTIC REGRESSION RESULTS FOR VARIABLES RETAINED IN THE THREE PROPENSITY MODELS

1. Resolvability Model

Variable	Coefficient	P-value
Intercept	0.8620	<0.0001
Owner4: High percentage in Owner Occupied	0.0791	0.0249
Totalhh3: High percentage in Total Households	0.0976	0.0012
Age11: Low percentage in Age 0-17	0.1918	<0.0001
Yearedu4: High percentage in Median Years Education	0.2357	<0.0001
Black3: High percentage in Black	0.1554	0.0003
Ncntysiz1: Nielson County size coded as B**	-0.2321	<0.0001
Age13: High percentage in Age 0-17	0.1266	<0.0001
White1: Low percentage in White	-0.0819	0.0071
Owner1: Low percentage in Owner Occupied	-0.1041	0.0100
Totalhh1: Low percentage in Total Households	-0.1130	0.0050
Medrent3: High percentage in Median Rent	0.0696	0.0279
Homev1: Low percentage in Median Home Value	-0.2059	<0.0001
Homev2: Median percentage in Median Home Value	-0.0990	0.0027
Inc53: High percentage in Income 35K-<50K	0.0816	0.0286
Yearedu3: High percentage in Median Years Education	0.1075	0.0122
Owner4*age11: High in Owner Occupied &Low in Age 0-17	0.4469	<0.0001
Totalhh3*yearedu4: High in Total HHs& High in Median Years Educ.	–0.2193	0.0197
Totalhh3*inc53: High in Total HHs&High in Inc. 35K-<50K	0.2053	0.0035
Age11*ncntysiz1: Low percentage in Age 0-17&Nielson County size B	-0.3317	0.0127
Yearedu4*inc53: High in Median Years Edu & High in Inc 35K-<50K	-0.3089	0.0030
Black3*medrent3: High percentage in Black& High in Median Rent	-0.1463	0.0207
Ncntysiz1*medrent3: Nielson County Size B& High in Median Rent	0.4810	0.0004
Owner1*homev1: Low in Owner Occupied& Low in Median Home Value	-1.0666	0.0021
Medrent3*homev1: High in Median Rent &Low in Med Home Value	0.2653	0.0027

2. Screening Model

Variable	Coefficient	P-value
Intercept	0.5074	<0.0001
Shisp2: High percentage in Hispanic	0.1177	0.0106
Sttlpop4: High percentage in Total Population	0.1407	0.0001
Sinc45: High percentage in Inc. 25K-<35K	0.114 0	0.0105
Sinc51: Low percentage in Inc.35K-<50K	-0.1975	0.0029
Sinc52: Med percentage in Inc. 35K-<50K	-0.0885	0.0635
Shhlist1: Low percentage in HH List Per Working Bank	-0.0541	0.4338
Syearedu2: Med percentage in Med Years Educ	-0.1694	0.0041
Shomev4: High percentage in Med Home Value	0.0787	0.0417
Sage43: High percentage in Age 35-44	0.0873	0.0246
Shisp2*syearedu2: High in Hisp &Med in Med Years Edu	0.9698	0.0176
Sinc52*shhlist1: Med in Inc. 35K-<50K &Low in HH list per Working Bank	-0.3690	0.0119

3. Response Model

Variable	Coefficient	P-value
Intercept	1.4234	<0.0001
Rcollege1: Low percentage in College Graduate	-0.5726	0.0003
Rinc32: Med percentage in Income 15K-<25K	0.5612	<0.0001
Variable	Coefficient	P-value
Rblack1: Low percentage in Black	-0.2084	0.0874
Rtotalhh3: High percentage in Total HH	0.3348	0.0048
Rage71: Low percentage in Age 65+	-0.4419	0.0001
Rhhlist1: Low percentage in HH List Per Working Bank	-0.2699	0.0179
Rinc23: High percentage in Income 10K-<15K	0.1591	0.0744
Rmedrent3: High percentage in Med Rent	0.1850	0.0292
Rage51: Low percentage in Age 45-54	-0.2260	0.0150
Rhisp1*rblack1: Low percentage in Hisp & Low in Black	-0.5098	0.0076
Rblack1*rage71: Low in Black & Low in Age 65+	0.7702	0.0027
Rage71*rinc23: Low in Age 65+ & High in Income 10K-<15K	0.3445	0.0484

** B means in a metropolitan area that is not in the 21 most population, but that had a 1990 population of at least 85,000 households.

TABLE 2. LOGISTIC REGRESSION RESULTS FOR SUMMARY STATISTICS OF THREE MODELS

Global Chi-Square Test (Likelihood Ratio)

Statistics	Resolvability Model	Screening Model	Response Model
χ-Square	298.06	100.40	85.26
Degree of Freedom	24	11	14
P-Value	<0.0001	<0.0001	<0.0001

Model Goodness of Fit

Statistics	Resolvability Model	Screening Model	Response Model
Hosmer-Lemeshow Test	0.65	0.80	0.58
Deviance	N/A	0.41	0.80

Statistics Measuring Predictive Power

Statistics	Resolvability Model	Screening Model	Response Model
Percent Concordant	54.40	52.10	57.90
Percent Discordant	43.60	43.30	39.60
Percent Tied	2.00	4.60	2.50
Somers'D	0.11	0.09	0.18
Gamma	0.11	0.09	0.19
Tau-a	0.04	0.04	0.06