TELEPHONE DIALING SYSTEMS - SOME ISSUES AND ANSWERS

Gary M. Mullet, University of Cincinnati

INTRODUCTION

Marketing researchers are well aware of the controversy about whether or not they should attempt to contact households with unpublished telephone numbers in telephone surveys. While Rich (6) suggests that for some telephone surveys, no attempt need be made to include households with unpublished numbers, Blankenship (1, 2) suggests random digit dialing be used to minimize possible biases in survey research.

In their excellent survey paper, Frankel and Frankel (4) discuss several dialing methods. Among these is generating a (random) sample of directory published numbers and then adding a constant to the last digit to produce the final sample to be called. They state that for this method "it is impossible to determine the exact probability of selection for any selected number."

In what follows, using some simplifying assumptions and applying Bayes' theorem, the probability of selection for any number, published or unpublished, is derived. Also looked at is the companion problem: what is the probability that the number you are about to dial in the survey is listed (unlisted)?

NOTATION AND ASSUMPTIONS

The following notation will be used subsequently:

N = the total number of telephone households in the survey area of which

L = the number of listed telephone households and

U = the number of unlisted telephone households.

Obviously, U + L = N. One simplifying assumption is that no household has more than one telephone number. Hence, N is also the entity of telephone numbers in the survey area.

A sample of size n will be randomly generated from the L listed numbers. These numbers, in turn, will be suitably incremented to generate the final sample of n numbers on the final dialing sheets. n is generally much, much smaller than L and will be so assumed here. Essentially, this means the probabilities can be treated as though a binomial rather than a hypergeometric model.

If all of the telephone numbers in a given survey area were known, they could be arranged in increasing numerical order, either among the L and U subsets or in total. Hence, we will define

$P(L_i)$ = probability that the i^{th} listed household is included in the final set to be dialed, $i = 1, 2, \ldots, L$

$P(U_j)$ = probability that the j^{th} unlisted household is included in the final set to be dialed, $j = 1, 2, \ldots, U$

Once the final sample of n telephone numbers is generated, we next define

$P(DL_i)$ = probability that the i^{th} number to be dialed belongs to a listed household, $i = 1, 2, \ldots, n$

$P(DU_i)$ = probability that the i^{th} number to be dialed belongs to an unlisted household, $i = 1, 2, \ldots, n$.

Clearly, $P(DL_i) + P(DU_i) = 1$.

Also assumed is that all listed telephone househoulds are included in the sample frame and, more importantly, that unlisted numbers are assigned with no discernable pattern vis-a-vis published numbers.

With n much smaller than N, the assumption that no household has more than one telephone number is not restrictive, since households with more than one are quite few and, hence, the probability of selecting and dialing two or more numbers from the same household is small. However, all who desire to have their telephone number published are not in the sampling frame. Hence, the discussion refers to the numbers which are published in the current directory and both voluntary and involuntary unlisted. Finally, the assumption of no pattern of listed numbers vis-a-vis unlisted is more fact than opinion on the basis of private discussions with both telephone company personnel and others in the survey research industry.

It cannot be emphasized too strongly that the probability results are dependent upon the assumptions stated above. If it is known, for example, that one exchange has relatively more or fewer unlisted household numbers than another, this must be taken into account when deriving the sampling probabilities. Other obvious exceptions must also be accounted for, if they exist.

SELECTION OF DIRECTORY LISTINGS

Here, the list of household telephone numbers to be dialed is randomly selected from the most recently published telephone directory. Clearly, each household listed at the time of publication of the directory has a probability of n/L of being in the sample; each unlisted household has probability 0.

The major advantages of this method are the relatively low cost and the pre-screening for business numbers. The major drawback is that unlisted households are not specifically included (although they may be included when a household that was originally listed has its number assigned to some other household as unlisted after the first household moves.)

PLUS-ONE SAMPLE GENERATION

A sample size n is generated from the L listed numbers. Clearly, at this stage, each listed household has probability n/L of inclusion and each unlisted household has probability of 0, given the assumptions. The digit "one" is added to each number to generate a final list of n numbers to be dialed.

The i^{th} number in the original list of L published numbers will be in the final list to be dialed if

i) the number which is immediately preceding it in numerical order is listed and

ii) it is selected in the original sample of n published numbers.

Therefore,

$P(L_i) = (L/N)(n/L) = n/N, \; i = 1, 2, \ldots, L$

that is, the product of the probability of any given number being listed times the probability of its being selected in the original sample list.

In order to firm this up, assume that the i^{th} number of those published is 555-1234. It will show up in our final (+1) sample if 555-1233 is both listed and selected given it's listed (or is both unlisted and selected given it is unlisted). It will be listed with probability L/N and selected with probability n/N. (It will also be unlisted with U/N and selected with probability 0. These zero probabilities are neglected above.)

Analogously, for unlisted numbers, we find

$P(U_j) = (L/N)(n/L) = n/N, \; j = 1, 2, \ldots, U$

Thus, under the assumptions of the model, all telephone households have equal probability of n/N of being selected.

Obviously, now

$P(DL_i) = L/N$ and

$P(DU_i) = U/N$

on the basis of the above results.

In actual practice, if the last digit in the original number selected is 9, the number dialed is the original number decremented by nine, i.e., if originally 555-5559 was selected then 555-5550 would be dialed.

The biggest advantage to using +1 dialing is that we now sample households with unlisted phone numbers. Also, newly issued numbers can be sampled. A minor drawback is that the interviewer must now screen for business numbers which may now be in the sample even though excluded from the original frame.

PLUS-RANDOM SAMPLE GENERATION

If instead of adding a one to each of the n numbers generated, a random digit is added to generate the sample to be dialed, we have results as follow. First we assume that the digit added is between 1 and 9, that is, no telephone number generated in the original sample will remain unchanged.

Now the i^{th} number in the original list of L published numbers will be in the final list to be dialed if (again neglecting zero probabilities)

i) the immediately preceding number is listed
ii) selected
iii) a random 1 is generated and added

or

iv) the i^{th} number minus two is listed
v) selected
vi) a random 2 is generated

or

.
.
.

or

vii) the i^{th} number minus 9 is listed
viii) selected
ix) a random 9 is generated

Therefore,

$$P(L_i) = (\tfrac{L}{N})(\tfrac{n}{L})(\tfrac{1}{9}) + (\tfrac{L}{N})(\tfrac{n}{L})(\tfrac{1}{9}) + \ldots + (\tfrac{L}{N})(\tfrac{n}{L})(\tfrac{1}{9})$$
$$= 9(\tfrac{L}{N})(\tfrac{n}{L})(\tfrac{1}{9})$$
$$= \tfrac{n}{N}, \; i = 1, 2, \ldots, L$$

and

$$P(U_j) = (\tfrac{L}{N})(\tfrac{n}{L})(\tfrac{1}{9}) + \ldots + (\tfrac{L}{N})(\tfrac{n}{L})(\tfrac{1}{9})$$
$$= 9(\tfrac{L}{N})(\tfrac{n}{L})(\tfrac{1}{9})$$
$$= \tfrac{n}{N}, \; j = 1, 2, \ldots, U$$

Again the result is satisfying: all telephone households are equally likely to be included in the final dialing list. Also, as before,

$P(DL_i) = L/N$ and

$P(DU_i) = U/N$.

The above methodology will yield identical results for using any uniformly distributed random numbers, 1, 2, ..., r added to the generated list of published numbers.

Even though this result is satisfying from a probabilistic viewpoint, there are two difficulties from a practical point of view.

a) it is more expensive to generate a separate random increment to be added to each number in the original sample and

b) there is a small probability that a given number could appear multiple times in the final sample. This probability is of the order of magnitude of $(n/N)^2$ that it appears twice, $(n/N)^3$ that it appears three times, etc. With n small compared with N, this multiple appearance is unlikely, although Murphy's Law should warn the user.

RANDOMIZATION OF THE R LAST DIGITS

As an extreme, with R = 7, this is a random digit dialing method. Since complete random digit, though giving the desired household probabilities, is very inefficient, many alternatives have been proposed within the same basic framework (3, 4). In general, these methods are variations on randomizing the last 2, 3, or 4 digits using given exchanges which contain operating numbers.

Since some of the modified methods call for eliminating exchanges which have a scarity of working numbers, one disadvantage to these methods is that certain households have zero probability of inclusion. Another, particularly for the completely random digit method, is that a large proportion of numbers which are generated will be non-working. Hence, the field cost increases tremendously with the nonproductive dialings. Another drawback to some of these methods is that the required sample weighting may introduce some inefficiencies.

The major advantage to any method of this type is that listed and unlisted households can both be reached. For R = 7, the probability that a given number is unlisted is U/N.

DISCUSSION

For some, the results above seem counterintuitive. If, for example, there are two unlisted telephone numbers in numerical sequence, the second can never be dialed under plus-one. Also, a number which follows any unlisted in sequence will have zero probability of being dialed. However, these are conditional probability arguments, whereas the above are unconditional arguments.

As an oversimplified analogy, consider the readily calculated probability of selecting an ace in a specified number of draws from a well shuffled bridge deck. This is like the unconditional probability presented in earlier sections. Now consider finding the same probability given that an unspecified 10 cards have been removed from the deck. It is an exercise in elementary probability to show that since the 10 cards are unspecified, the final result is identical to drawing from the full deck, i.e., still unconditional. If the 10 cards were next turned face up, certainly the probability of now selecting an ace is conditional on what the 10 known cards are. Essentially, this is the distinction in the previous paragraph.

For a longer look at the plus-one dialing system, see (5) which does not deal with the probability issues presented here. Also neglected, of course, are sampling non-telephone households, differing cooperation rates between the two types of household, etc. Only the theoretical probabilities have been considered.

SOME RESULTS OF +1

During late 1978 and early 1979, a private research firm asked respondents who had just completed a telephone interview whether or not the household had a listed or unlisted number. The results below are the aggregate of 17 metropolitan areas in the continental United States.

Completed interviews	29,747	
Reporting listed	19,594	(65.9%)
Reporting unlisted	7,615	(25.6%)
Don't know/refused	2,538	(8.5%)

These data show that plus-one dialing does, in fact, garner a large number of households that report their telephone as being unlisted. Before attempting to compare these numbers with any population data, we must consider differences in contact rates and completion rates, if any. Also, of course, we must remember that it is a self reported behavior. In spite of these caveats, it is really apparent that plus-one sampling succeeds in generating completed interviews with (persons in) households that have unlisted telephones.

SUMMARY AND CONCLUSIONS

This paper has shown that using either plus-one or a particular plus-random dialing system gives equal probabilities to listed and unlisted telephone households. Also the numbers on a given list to be dialed will appear in proportion to their population proportions of the two types, in probability. Either system is theoretically a valid sampling method when applied to telephone households, under appropriate conditions and assumptions. Other sampling methods were also briefly examined.

REFERENCES

1. Blankenship, A.B., "Listed versus Unlisted Numbers in Telephone Survey Samples," Journal of Advertising Research, 17 (Feb., 1977), pp. 39-42.

2. Blankenship, A.B., Professional Telephone Surveys, New York: McGraw-Hill, Inc., 1977.

3. Danbury, Thomas, "Alternative Sampling Models for Random Digit Dialing Surveys," Annual Conference of the Advertising Research Foundation, 1975.

4. Frankel, Martin R. and Lester R. Frankel, "Some Recent Developments in Sample Survey Design," Journal of Marketing Research, XIV (Aug., 1977), pp. 280-93.

5. Landon, E. Laird, Jr. and Sharon K. Banks, "Relative Efficiency and Bias of Plus-One Telephone Sampling," Journal of Marketing Research, XIV (Aug., 1977), pp. 294-9.

6. Rich, Clyde L., "Is Random Digit Dialing Really Necessary," XIV (Aug.,1977) pp. 300-5.