MEASURING CATI EFFECTS ON NUMERICAL DATA

Carol C. House, U.S. Department of Agriculture
Betsy T. Morton, University of California-Berkeley

Summary

This paper discusses the potential advantages for CATI in survey data collection and presents background and analysis of a two phase comparison of CATI verses non-CATI telephone interviewing. The study is conducted by the U. S. Department of Agriculture's Statistical Reporting Service in cooperation with the Center for Computer Assisted Survey Methods at the University of California - Berkeley's Office of Computer Affairs. It uses the interviewing facilities of USDA with the CATI software developed at UCB.

The comparison of CATI versus non-CATI telephone interviewing to collect cattle inventory data shows approximately 75 percent fewer errors in the data collected by CATI. The estimates of total cattle deaths are significantly different between the two groups (a=.10). Other variables show differences between 20 and 25 percent, but the test is not powerful enough to show these differences as significant. The early phase of this study measures response errors from one percent in reporting milk cows to 58 percent in reporting steers.

Introduction

Computer assisted telephone interviewing (CATI) refers to the use of computer systems for telephone interviewing and related forms of data collection, data entry, editing and coding. An interviewer sits in front of a computer terminal with a cathode ray tube display that looks like a television screen and speaks with a respondent through a telephone head-set. The computer is programmed to display each question in turn on the screen. After reading the question to the respondent, the interviewer records each answer by depressing the keys on the terminal keyboard. The computer performs any desired checks for consistency and, if needed, displays a question requesting clarification. The computer stores the information, selects the next appropriate question and and displays it on the screen.

In addition to these basic functions, advanced CATI systems are capable of facilitating sample control, survey coding, and many additional survey operations concurrently with survey interviewing. When compared with traditional telephone data collection, CATI surveys have a number of potential advantages. A review of these identified seven areas with the greatest potential benefits to the Statistical Reporting Service (SRS) of the U. S. Department of Agriculture. These are: 1) improvements in data quality, 2) reduction in processing time, 3) efficiencies in survey management, 4) improvements in interviewer training techniques and capabilities, 5) flexibility in questionnaire design, 6) expanded pretesting capabilities, and 7) potential sample size reductions through sequential estimation. These issues are discussed in more detail in the current literature on CATI.

The need to measure the actual gain from CATI rather than merely to discuss the potential becomes apparent when one reviews the additional costs and complexities imposed by most CATI systems. Startup costs can be imposing. These include hardware procurements, development or procurement of basic CATI software that is flexible enough to fill the need of the survey organization, adaptation of questionnaires to computerized formats, and the training of staff to program and direct CATI surveys and to run new ADP equipment. Although a detailed discussion of all of these issues are beyond the scope of this paper, they provide a better understanding of the need to investigate fully this new survey methodology.

This paper presents the joint work of SRS and the Center for Computer Assisted Survey Methods (CSM) of the University of California-Berkeley who are cooperating in a research effort to examine and quantify some of these potential benefits. Most of the joint work in this area is centered around the use of CATI to collect numerical data to produce agricultural statistics.

Background

The Statistical Reporting Service began experimentation with CATI in 1980 through a cooperative agreement with the University of California - Berkeley's Center for Computer Assisted Survey Methods. This agreement provides SRS with access to a sophisticated CATI system and the opportunity to support its continued development.

The major SRS interest in CATI centers around its potential to improve the data quality and to reduce the processing time involved in data collection. The SRS mission is to collect numerical data to produce agricultural statistics, and publish these statistics in a timely manner. Much of this data is collected on recurring surveys in which a sample of respondents is rotated into the survey for a specified period of time. Thus, there frequently is a historical data base available for each sample unit. In addition many of the questions on a given survey are interrelated in quantifiable ways. Since both of these situations lend themselves to the use of on-line edit and consistency checks during interviewing, it was decided to concentrate on quantifying the improvements in data quality resulting from such checks.

SRS set up an initial test site in the State Statistical Office (SSO) in Sacramento. This sight consists of four interview stations, each equipped with a Concept 100 CRT terminal, 1200 baud high speed modem, and two telephone line (one for data transmission and the other for interviewing). The interviewing phone is equipped with a headset. The data lines connect the interviewing stations in Sacramento to the DEC PDP 11/44 mini-computer on the UCB campus.

SRS chose a semiannual cattle inventory survey to be the first adapted to CATI interviewing. It is an important survey to the Agency and to many users of agricultural data. It also has potential for a full use of the on-line edits to check the consistency of the inventory counts. SRS staff programmed the questionnaire using the UCB QISB questionnaire development language that is a part of the UCB CATI software.

Care was taken to minimize wording differences between the CATI instrument and the original paper questionnaire. During an interview, when data fails to pass a consistency check, an additional question appears on the screen. This question informs the interviewer that an edit has failed, clearly states what check it is that has failed, and gives the data under suspicion. The interviewers use their standard probing techniques to ascertain if reporting errors were made, and then correct any misreported data. If the interviewer resolves the inconsistency without changing the data, then the program directs the interviewer to type in a short note explaining the resolution.

Phase One

SRS and CSM jointly conducted the interviewing for the first phase of the study during January 1982. The primary objectives of this phase are to pretest the CATI instrument and to allow the interviewing and professional staffs an opportunity to acquire experience in CATI interviewing. A secondary objective is to provide preliminary indications of data differences resulting from on-line editing. These objective were satisfied. The analysis shows indications of major differences in several important variables. The following paragraphs describe the analysis in more detail.

A special research sample was selected from two strata of a list of California-based cattle operations. Four interviewers with experience on this survey (but not with CATI), conducted the interviews. They placed calls for five evenings, and completed 132 interviews.

The computerized questionnaire produces two sets of data. The first consists of the answers to questions as they were originally recorded by the interviewers. The second data set consists of the answers to the same questions, but after the program runs through the edit logic and the interviewer resolves any inconsistencies. Together, the two data sets provides a pair of answers for each question: 1) the answer as first provided by the respondent, and 2) the answer finally agreed to by the respondent and interviewer as the best answer to the question. The authors perform a paired analysis on these data to measure the impact of the edit on estimates of the number of head of cattle in each of several inventory groups. The analysis treats the sample units as an unstratified simple random sample. No inference is made to the population of cattle operations in California, but merely to the sample itself.

Differences between the data due simply to keystroke error are eliminated when detected. The authors examine each pair individually. If they determine that the difference is clearly produced by transposed numbers (or similar obvious keystroke problems), they would set the difference between the pair to zero, and include the zero difference in the analysis. On several occasions it was difficult to assess whether the difference was due to keystroke error, or if, in fact, the respondent had decided to change the answer. In these specific cases, the paired answers are completely removed from the analysis.

The difference between the final answer to a question and the initial answer to the same question is defined by: diff=last-initial. The distribution of these differences is very skewed, consisting mainly of zeros, with various larger values thrown in. Some of these values are very large. This type of distribution makes the confidence intervals extremely large and any statistical tests of the mean differences unsatisfying. Therefore, analysis from phase one of this study consists of descriptive statistics of the differences found in the livestock numbers.

The descriptive statistics (for each inventory category and for the overall sample) are presented in table 1 and include: the total cattle (based on the edited data), the sum of the differences discussed above, the sum of the absolute value of those differences, the average difference per sample unit and the average absolute difference per sample unit. The authors place particular emphasis on the estimates of total absolute difference in number of animals reported. By not allowing the errors to cancel each other out from sample unit to sample unit, we get a measure of the total response error that is being corrected during the CATI interview and not just a measure of any bias that is being eliminated. The variable "Percent Abs Change" is calculated to give a relative indication of this number.

$$\text{Percent Abs Change} = \frac{\text{Total Abs Diff}}{\text{Total}} \times 100\%$$

Table 1--Results from paired comparisons in phase 1.

INVENTORY GROUP	TOTAL	TOTAL DIFF	TOTAL ABS DIFF	AVE DIFF	AVE ABS DIFF	PERCENT ABS CHANGE
Beff Cows	3,064	-341	451	-3	4	15
Milk Cows	24,021	10	230	< 1	2	1
Bulls	519	5	5	< 1	<1	1
Beff Heifers	583	-58	62	< 1	<1	11
Milk Heifers	9,252	549	649	5	6	7
Other Heifers	331	71	131	< 1	1	40
Steers	2,661	1,556	1,556	13	13	58
Calves	8,027	-679	959	-6	8	12
TOTAL CATTLE	48,459	1,210	2,918	10	25	6

The percent of absolute change ranges from one to 58 percent. The changes are the largest in animals for slaughter market (other heifers and steers), and, not surprisingly, the smallest in milk cows and bulls. (Operators are expected to be more certain of the total number of milk cows than other animals on hand simply because they milk them twice a day. When bulls are kept, they are fewer in number and generally more valuable than other animals on the operation.) The changes in sign from negative (beef cows) to positive (milk cows) of the total differences for cows, and a similar change in the heifer group (beef, milk and other) present evidence for possible question order bias. In each grouping, the first question (beef) appears to be over-reported initially, and subsequent questions of the same type (milk) (milk and other) under-reported. Similar results for livestock inventory surveys were found earlier by Steiner [6].

As an additional measurement of the value of the on-line edit, the authors run both data sets through the SRS standard batch edit programs for the cattle inventory survey, and count the number of critical and noncritical errors generated for each data set. Critical errors are those that will prevent the data from being summarized while they are present, while noncritical errors indicate data are outside ranges generally perceived as normal. Table 2 presents the total number of errors identified through these edit runs.

Table 2 -- Summary of edit errors, Phase 1

		Type of Error	
	Total	Critical	Non-critical
Number of error messages on initial responses.	47	16	31
Number of error messages on final responses	20	6	14
Number of corrections before summary	7	6	1

Fifty-five percent of the non-critical errors from the original data set are removed during the on-line edits. Of the 14 non-critical errors remaining, only one requires correcting before summarization, and it is accompanied by a note from the interviewer which reveals a misunderstanding. Six critical errors remain after the CATI on-line edit checks, all of which must be corrected for summary. Two of those are caused by a respondent's refusal to answer specific questions. The other four are accompanied by interviewer notes which indicate definitional problems and give adequate information to make the corrections. Thus of the 47 errors and discrepancies flagged by the edit programs on the data originally reported by the respondents, only fifteen percent required intervention after the close of the interview.

Phase Two

The primary purpose of the second phase was to have a controlled test of CATI verses non-CATI telephone interviewing by comparing the estimates generated from two half samples during an operational survey period. The operational sample in nine selected strata for the cattle inventory survey in California was split randomly into two subsamples. After eliminating sample units without telephone numbers, the effective sample size was 614 on the CATI sample and 609 for non-CATI.

Interviewers were subjectively assigned to one of two groups, each group of nearly equal experience and ability. All interviewers had worked before on this study, and two interviewers on the CATI team and one interviewer on the non-CATI team had previous CATI experience. Interviewing was conducted during a nine day period of time during the last week of December 1982 and the first week of January, 1983. The cattle instrument received only minor modifications between phases.

The analysis plan calls for a comparison of estimates between the two half-samples at two different stages of the processing. One comparison is of the data exactly as it come from the completed interview, before any operational SRS batch or hand editing is done. A second comparison is of the data after they have proceeded through the full editing procedures and are ready for summary. Results from the second comparison are presented below. Comparisons of the unedited data are underway, and results will be available at a later time.

Seven variables are chosen for the comparisons. Strata totals for each variable are expanded by the appropriate strata expansion factors to produce overall totals, and are presented in table 3. Note that these are not estimates of state totals for California, but represent only the nine selected test strata.

Table 3 -- Direct expansion estimators of cattle variables

VARIABLE	CATI	NON-CATI	% DIFF 1/	DIFF IS SIGNIFICANT 2/ (PROB>F)
Total Cattle	2,599,217	2,719,425	4.4%	no (.59)
Total Beff Cows	518,977	661,112	-21,5%	no (.15)
Total Milk Cows	784,156	731,937	7.1%	no (.20)
Total Other Heifers	70,990	71,766	-1.1%	no (.95)
Total Steers	138,549	184,229	-24.8%	no (.26)
Total Calves Born	1,116,723	1,175,842	-5.0%	no (.45)
Total Cattle Deaths	40,209	34,976	15.0%	yes (.08)

1/ % Diff = ((CATI - NonCATI)/NonCATI) x 100%
2/ univariate tests for a=.10 significance level

The table also presents the results from univariate tests run using SAS statistical programming package. One variable (total cattle deaths) of seven is significantly different (at a=.10) in these tests. Two additional variables, total beef cows and total steers, have differences between twenty and twenty-five percent. However, these differences are not detectable at the given significance level because the power of the test is too low. Before more powerful comparisons can be made at a univariate level for many of the cattle variables, testing must include several states. This must wait until additional states have CATI capabilities.

Two multivariate test results (Hotelling-Lawley Trace and Wilks' Criterion) fail to show significant differences for a=.10. In both tests, the alpha level for significance is a=.13. All test statistics are computed using replicate totals instead of individual data to avoid violating basic test assumptions. There are ten replicates across strata in each half-sample of data. Table 4 displays counts of critical and non-critical errors from the SRS generalized edit programs that are generated for each data set. The CATI half sample has 77 percent relatively fewer critical errors than the non-CATI half sample. Seven of the twelve critical errors from the CATI sample result from a consistency check inadvertently left out of the programming of the on-line CATI checks. This check has since been added. The remaining five indicate that the amount of feed fed to cattle-on-feed is too low to be classified as "on-feed". In all cases, the numbers were verified with the respondents, but his/her definition did not match that of SRS.

Table 4 -- Summary of edit errors, phase 2

	Type of Error		
	Total	Critical	Non-critical
Total Error Messages			
Non-CATI	245	53	192
CATI	199	12	187
% Rel Difference 1/	19%	77%	3%
Errors Corrected for Summary			
Non-CATI	84	53	31
CATI	20	12	8
% Rel Difference 1/	76%	77%	74%

1/ Percent Relative Diff = ((NonCATI - CATI) / NonCATI) x 100%

Only a three percent relative difference in non-critical error flags is found between the two data sets. However, because all inconsistencies were verified by the respondents during the interview for the CATI sample, there are 76 percent relatively fewer changes made to the CATI data than to the non-CATI data as a result of these error flags.

References

[1] Groves, Robert M., Marianne Berry; and Nancy Mathiowetz. "Some Impacts of Computer Assisted Telephone Interviewing on Survey Methods." In Preceedings of the Section on Survey Research Methods. Houston, Texas, August 11-14, 1980. Washington, D. C.: American Statistical Association, 1980. p. 519-524.

[2] House, Carol C. "The Department of Agriculture's Experience with Computer Assisted Telephone Interviewing: An Overview of the First Test Year." Unpublished Staff Report of the U. S. Department of Agriculture, Statistical Reporting Service, Statistical Research Division. August, 1981.

[3] Nicholls, William L., II, and George A. Lavender. Berkeley SRC CATI: Preliminary Interviewer Manual for Random Adult Studies. Berkeley: University of California, Survey Research Center, December, 1979.

[4] Nicholls, William L., II, George A. Lavender, and J. Merrill Shanks. Berkeley SRC CATI: An Overview of Berkeley SRC CATI Version I. Berkeley: University of California, Survey Research Center, February 1980.

[5] Rustemeyer, Anita, Gerald H. Shure, Miles S. Rogers, and Robert J. Meeker. "Computer-Assisted Telephone Interviewing: Design Considerations." In Preceedings Of The Section on Survey Research Methods, San Diego, California, August 14-17, 1978. Washington, D. C.: American Statistical Association, 1980. p. 507-512.

[6] Steiner, Michael A. "Study on the Effects of Changing the Placement and Wording of Questions in Hog Surveys. Forthcoming Staff Report. U.S. Department of Agricultural, Statistical Reporting Service, Statistical Research Division, Washington, D. C.