

ESTIMATING THE COSTS OF CONDUCTING ONE-TIME SURVEYS USING LOTUS 1-2-3

Doug Doyle, Jane H. Larson, and Mark Pierzchala, National Agricultural Statistics Service
Mark Pierzchala, NASS, USDA, room 4168, Washington, DC 20250-2000, 202-447-5333

KEY WORDS: cost modeling, spreadsheet, special surveys.

THE SPREADSHEET

The necessity of using a spreadsheet to estimate the costs of conducting one-time surveys derives from the authors' experiences in the Wisconsin office of the National Agricultural Statistics Service (NASS). Wisconsin is a state in which many requests to conduct special surveys are received. Clients include university agricultural researchers, trade associations, and the Wisconsin State Department of Agriculture, Trade and Consumer Protection. Surveys that are requested cover a range of topics including pesticide use, soil conservation practices, hog raising practices, damage to crops caused by deer, stray voltage on dairy farms, and farm finance. Costs of these special surveys range from a few thousand to tens-of-thousands of dollars and from tens to hundreds of hours of staff time. NASS state offices receive these requests because of the in-house statistical and subject area expertise present in each office, and because each state office maintains a confidential list of farmers. The list contains information which can be used to design the survey and select the sample.

The Lotus 1-2-3 spreadsheet template was developed to make the task of cost estimation easier and to reduce or eliminate the occurrence of mis-estimation of survey costs. This mis-estimation has four sources: forgetting some tasks, underestimating the complexity of some tasks, under- (or over-) estimating

certain rates (e.g., response rates), and not defining well those tasks to be completed by the office and those to be completed by the client. NASS state offices also operate under severe resource constraints and usually conduct special surveys with already fully committed personnel. The spreadsheet, if fully utilized, will give the state office director a good idea of expected resource commitment both in terms of staff time and money.

DESCRIPTION

The layout of the spreadsheet is shown in the schematic on the next page. Several areas are distinguished, indicating where a particular function is performed. Windows 1 and 2, (the screens one sees after pressing 'Home' and 'PgDn' successively) display a summary of estimated survey costs, sample sizes, expected list attrition (deadwood), response rates, and a special summary of out of pocket expenses. In the Survey Plan Questionnaire, (the next several screens further down in the template), the user details the survey plan, including the mode of collection in any combination of mail, phone, and personal visit. Rates specific to the office such as telephone, mileage, pay, and per diem rates, and rates specific to the survey such as expected mail response rates are also filled in here.

The large area called the Heart of the Spreadsheet is where item-by-item entries are made for survey tasks and expenses. These specific items are organized and summarized within major

Window 1
Window 2
Survey Plan Questionnaire Details of the survey plan and some global costs are entered here.

These two windows give a synopsis of estimated survey costs. Sample size and response rates are entered here.

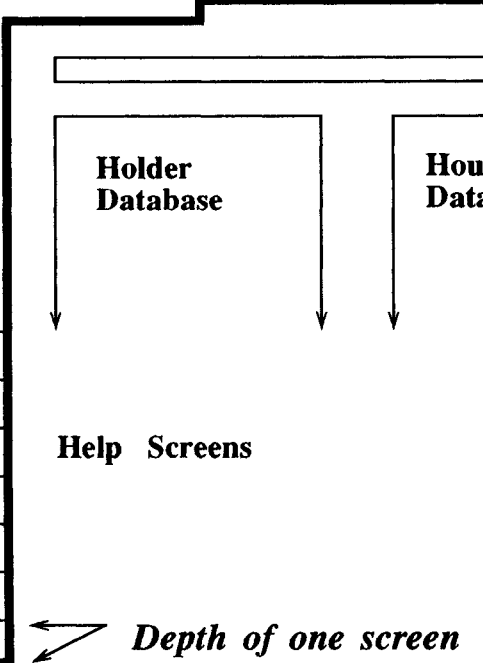
Heart of the Spreadsheet

Itemized entries are made here.

Planning	Head1
ADP	Head2
Printing	Head3
Data Collection	Head4
Edit	Head6
Analysis Summary	Head7 Head8
Dissemination	Head9
Cleanup	Head10

Database area for the nine major cost areas. Itemized costs are recast into a resource by resource format.

MACROS AND WORK AREA
<i>Width of one screen</i>
Help1
Help2
Help3
Help4
Help5
Help6
Help7



Variable holding area.

These two databases summarize resource expenditures for the whole survey.

groupings and according to whether the expense is fixed or variable. A variable expense is one that is a function of the sample size. All other expenses are fixed. To the right of the Heart of the Spreadsheet are areas reserved for database use. Macro programs, work areas, and help screens appear below the Heart of the Spreadsheet.

The template was distributed to all NASS state offices with the schematic and complete documentation including a tutorial. The tutorial can be completed in about one hour. In its first use within a state office, the use of the spreadsheet may take from one to two days as users gain familiarity and as basic information such as telephone charges wage rates are gathered. It is recommended that each state office alter the spreadsheet to reflect its own practices and then use the altered spreadsheet as a template for its specific surveys. Subsequent uses for estimating specific survey costs are expected to take from one half to one day to complete for each survey.

FEATURES

The spreadsheet is designed to estimate costs of list frame surveys. Any combination of three modes of collection is allowed including the use of one mode of data collection to follow-up nonresponse from another mode. Survey costs are initially calculated by nine major task divisions such as Planning, Edit, and Summary. Any resource, such as hours worked by an individual, postage, and telephone charges can be listed repeatedly throughout the spreadsheet. Database functions recast itemized estimates by resource both within major groupings and for the survey as a whole. Thus the office director can be given an estimate of the number of hours that each person will work on the survey, and the expected out-

of-pocket expenses such as telephone charges. This statement of expected resource expenditures is intended to be comparable to printed output from the existing agency accounting system, thus allowing a gross validation of estimated expenses. Costs are considered to be either variable expenses or fixed expenses. It is a feature of special surveys that fixed expenses tend to account for a larger share of survey expenditures than for repetitive surveys. As a result, changes in sample size or changes in mode of collection do not always have a profound effect on survey costs. If required, the program can be used to provide an extensive itemized list of costs to the clients.

While the template includes an extensive listing of tasks that might be performed in any survey, it would be impossible and undesirable to list all such tasks. Therefore, the spreadsheet has been designed to be alterable. The user can insert new rows to accommodate more items. The addition of new rows does not destroy the database capabilities or the summary features. The spreadsheet also allows one to determine how changes in one part of the survey plan affect resources in another part. For example, costs saved by mailing a survey instead of telephoning may be offset if more resources are spent on editing and summarization as a result of poorly completed questionnaires.

The spreadsheet also has other uses. It can be used as an educational tool for inexperienced survey statisticians. By providing a list of survey tasks, the new employee is introduced to the complexity of conducting a survey. Additionally, a dynamic listing of survey tasks of this type can be a very useful conceptual tool when considering technological innovation for productivity improvements.

FUTURE IMPROVEMENTS

The utility of the spreadsheet in estimating costs of one-time surveys must be further validated in NASS's state offices on larger surveys. The present accounting system accumulates actual operational costs by major resource. As the spreadsheet summarizes estimated resource expenditures, a gross comparison of estimated versus actual costs is possible. Comparison of estimated versus actual costs for specific items such as hand editing of questionnaires is not currently possible. This could be accomplished if simple logs were kept by survey personnel of the time spent on each task. This sort of item-by-item cost comparison may be useful if the agency is looking for ways to increase productivity through technological innovation.

The spreadsheet is lacking in some areas. Little provision is made for estimating costs for dual frame surveys that are widely used in NASS's national programs and it cannot easily estimate differential collection costs by stratum.

THE WISCONSIN EXPERIENCE

EXPERIENCE IN ONE SURVEY

Early in 1988, the Wisconsin office of NASS generated cost estimates for participation in a multi-state survey of farm operators and their spouses. The project was initiated through the North Central Regional Center for Rural Development (NCRCD). Several pieces of information pertinent to the sampling design and survey methodology were unknown during the initial discussions. Despite this, it was necessary to estimate survey expenses as closely as possible. This is not atypical of special one-time surveys. Clients frequently request expense estimates before they develop an

acceptable sampling design or survey instrument. Specifically missing for this project were a questionnaire, sample sizes, mailing procedures, and editing and data entry procedures. The spreadsheet can be used in these situations to generate costs for several different survey plans as an aid to the client in developing the survey methodology.

SURVEY MANAGEMENT

Questionnaires were mailed to 1600 farm couples (3200 questionnaires) on February 13, 1989. Two weeks later, 1470 reminder cards were mailed to non-respondents. Twelve hundred non-respondents were reminded a second time by telephone during the week of March 14, 1989. Some respondents were also recontacted by telephone to obtain permission to use their data in the analysis, if they had not returned a permission waiver. This procedure provides a legal basis to release data to outside researchers after identification information has been removed. Considerable time was spent coordinating telephone prompts and follow-up of non-response. Tracking questionnaires was labor intensive and done without the aid of a computerized check-in system. On April 12, 1989, a subsample of non-respondents was enumerated in order to assess nonresponse bias. Twenty-five non-respondents were randomly selected and telephoned. A shortened version of both the operator and spouse questionnaire was administered. All of the questionnaires, minus identification information, were then turned over to the University of Wisconsin - Madison for a final edit prior to data entry.

ESTIMATED VS. ACTUAL EXPENSES

Although a complete accounting of final survey expenses is not possible, much of

the actual expense information had been recorded. As mentioned earlier, some of the information necessary for generating accurate expense estimates was unknown during planning. The survey instrument was not in a final draft stage, sample sizes had not been determined and mailing alternatives were not known with certainty. In addition, the number of "callbacks" for each respondent had not been determined and a method had not been accepted for follow-up. Postcard reminders had been discussed early in the planning, but it was not known if a telephone follow-up would be conducted. Despite the uncertainty, the spreadsheet was useful in casting scenarios and generating expense estimates for each scenario. An early sampling design included initial mailings to 700 farm couples followed by telephone calls to each non-respondent. Estimated expenses for this design showed that collecting data by telephone during a follow-up contact was costly. A second design included an increase in the sample size to 1600 couples. Telephone prompts could be made to each non-respondent to encourage mail response, but not to collect data. A tentative agreement was reached by NASS and NCRCD regarding design and survey administration.

A cost estimate of \$4,700 was generated for this survey. Included in the agreement were categorical estimates for sample selection, mailing, telephone and postcard follow-up of non-respondents, manual questionnaire editing, and other related survey administration. In addition, the final expense estimate included telephone interviews of a sub-sample of non-respondents to assess nonresponse bias. Figure I shows some cost comparisons. Estimates for the federal government portion of the project were well under actual federal expenses. All of

these underestimates are for human resources. The capital resources were estimated more precisely, suggesting that our efforts to plan a project need improvement where human time is considered for a task.

One example of this underestimation was in mail room preparation. It was anticipated that the questionnaires would be addressed by machine. Instead, this intensive process was done by hand for 1600 pairs of questionnaires by a federally paid statistical assistant. Mail preparations took nearly three days instead of a few hours. As a result of this shift of work from the mailroom account to the federal account, actual mailroom expenses were less than the estimated mailroom expenses. However, this shift contributed to the underestimation of the federal expenses. Secondly, the time necessary to edit each questionnaire was underestimated, primarily because final versions of the instrument were not available early in the planning process. Nearly fifty respondents were called back because they had not returned permission waivers. This had not been factored into the data collection expenses. Administrative changes also occurred during the survey. When cost estimates were generated, NASS reimbursed telephone enumerators for portal to portal mileage. Halfway through the survey, this reimbursement was discontinued. A significant change in telephone enumerator pay scales was also implemented during the survey. Although such changes were anticipated, they exemplify problems with any expense forecast.

ADAPTABILITY

One important feature of the spreadsheet is a capability of adapting to a survey research center's specific needs. By inserting rows for alternative survey

activities in appropriate categories, the spreadsheet was modified to the Wisconsin office procedures. The Wisconsin office operates under a cooperative agreement with the state Department of Agriculture. In this environment it is essential to cast expenses within both federal and state budgets. Each has separate accounting years and procedures. In WINDOW 2 of the spreadsheet, extra rows were inserted to break out the federal and state expenses for data collection. Further, a line was included to delineate expenses for telephone and field enumerators. Expenses for telephone prompts were also incorporated into the spreadsheet in the data collection cost area. The number of telephone prompts was determined by carrying forward the response rates from waves of questionnaire mailings (in the Survey Plan Questionnaire) to the data collection area (in the Heart of the

Spreadsheet). Telephone connect-time estimates were then used to estimate expenses with an additional row in this latter area for telephone prompt expenses.

REFERENCES

Ford, B. L. and Tortora, R. D. (1978): A Consulting Aid to Sample Design. The Consultant's Forum, *Biometrics* 34, 299-304.

Boen, J. R. (1982): A Self-Supporting University Statistical Consulting Center. *The American Statistician*, November 1982, Vol.36, No. 4, 321-325.

Zahn, D. A. and Isenberg, D. J. (1983): Nonstatistical Aspects of Statistical Computing. *The American Statistician*, November 1983, Vol. 37, No. 4, 297-302.

This paper was reviewed by Mary Anne Cummins, National Agricultural Statistics Service.

FIGURE I

