SAMPLING AND REPORTING IN TIME-USE SURVEYS

Edwin L. Robison, Bureau of Labor Statistics
BLS, 2 Massachusetts Ave., NE Room 4985,Washington, DC 20212

Key Words: time-use survey, sampling, reporting objectives

General

The Bureau of Labor Statistics (BLS) is planning a Time Use Survey to obtain data from adults on the duration of their activities in a 24-hour period. Many variables potentially affect the way people use their time. For example, individuals may allocate their time differently based on gender, age, income, and presence of children. The Current Population Survey (CPS) is a nationally representative sample rich in variables of this type, and can be used as a base from which to select a subsample for a Time Use Survey. To limit burden, no more than one adult in a household will be included in the Time Use Survey.

Studies of the allocation of time examine the tradeoffs between leisure, paid employment, and other activities. Presumably these tradeoffs vary by season, during holidays, and by day of week, since individuals' use of time is known to vary. Accordingly, data collection for time use will be spread throughout the year. It was decided to split the sample evenly between weekdays and weekends, but the decision is being reviewed.

Nonresponse and response issues loom larger for surveys on time use than for other surveys. In our plan sampled individuals will have a specified date to report about. To minimize recall bias problems, a response is sought the very next day only. If unsuccessful, it was decided to retain the individual in the sample, but to specify a new specified date to report about. Since time use behavior varies markedly by day of week, the new specified date will be on the same day of the week as the original specified date (e.g., both will be Tuesdays). This strategy can somewhat alleviate bias, but certain uses of time are systematically excluded. For example, only the very end of a spell away from home can be captured with this protocol.

The Current Population Survey Sample

The Current Population Survey is sponsored by BLS and conducted by the Bureau of the Census, primarily for the collection of labor force data. Monthly estimates representative of the Civilian Noninstitutional Population (CNP) are based on reported information for individuals residing in a representative sample of households. The CPS monthly sample currently includes about 52,000 "eligible" households. Of these, about 48,000 are successfully interviewed, yielding demographic and labor force data for about 100,000 adults (16 years of age or older).

The CPS is designed so that each State has an adequate sample of households for making reliable annual unemployment estimates. (Procedures assume a 6% unemployment rate and produce household samples that have coefficients of variation of 8% or less on State annual unemployment.) Less populous States need denser samples than the more populous States. The overall sampling fractions range from 1/280 households in Wyoming (sampling interval of 280) to 1/3600 households in parts of California (sampling interval 3600). (Note that if the only interest is in making national-level estimates, a State-based design like this is not particularly efficient. In relative terms, the less populous States are overrepresented. You would shift sample out of the less populous States, into the more populous States, until the sampling intervals were nearly equal.)

In most States a two-stage sample is needed, since it is expensive and impractical to sample households from every locality. In the first stage a sample of substate areas called Primary Sampling Units is selected that includes all of the larger cities and suburbs but only a portion of the rural counties. Within the selected Primary Sampling Units, clusters of 4 housing units are then selected in a second stage of sampling (reduces costs when compared to sampling individual households).

The CPS has a rotating panel design. Respondents get "exhausted" when interviewed time after time, and nonresponse gets worse. Instead, the CPS sample is divided into panels. A given panel is included in the CPS for 4 consecutive months; it is temporarily rotated out for 8 months; then it is returned for another 4 consecutive months. In any given month the CPS sample is comprised of eight panels, one in sample for the first time, another for the second time, etc. One of them is an "outgoing" panel in its eighth month-in-sample (MIS 8) that is then permanently dropped from the CPS sample, becoming eligible for subsampling in the Time Use Survey.

Each household in the CPS has a known probability of selection (ex: 1/280) and the inverse of that probability

(280) is usually referred to as a sampling interval or sampling weight. In a typical State, most households have nearly the same sampling weight, but the weights are later modified to adjust for nonresponse. The household sampling weights as modified for nonresponse could in principle be applied to each individual in a responding household, but that is not done. The household weights are substantially modified by several procedures, particularly by the "second-stage ratio estimation" weight adjustment procedure. The procedure ensures that weighted estimates of population totals that can be made from each CPS panel will exactly match a set of "known" independent population control figures produced by Census. After the adjustments, individuals in a household will generally have different weights.

Strategy for Using the CPS as a Sampling Frame

The Time Use Survey sample will be a subsample of the Current Population Survey. It will be restricted to adults from responding households that are being permanently dropped from the CPS during a single calendar year (the outgoing MIS 8 panels). CPS data can be used to assure a subsample is taken that has an adequate number of adults of each sex, age group, etc. To limit household burden, at most one adult will be sampled per household. Half of the sample will be randomly assigned to weekdays (10% to each weekday), and the other half of the sample will be assigned to weekends (25% each to Saturday and Sunday).

Limiting the Time Use Survey to respondents in outgoing CPS panels greatly simplifies the mathematics of, and computer processing for, the survey. Respondents in each CPS panel already have weights that are adjusted to sum to independent population controls. A subsample of outgoing month-in-sample 8 respondents can be drawn and assigned straightforward weights for the Time Use Survey that will at least approximately match the same control totals. Using MIS 8 nonrespondents would complicate the mathematics of weighting. Also, it would make the file manipulation processes more cumbersome. For example, the most recent CPS response for an MIS 8 nonresponse could be in any of seven files (the corresponding files for months-in-sample 1 through 7 of the panel), or there may never have been a response for the household at all.

CPS Cases Available For Subsampling

There are about 100,000 adults in the 48,000 households that are successfully interviewed each month for the CPS. Of these, about 12,500 adults in 6,000 households are in the outgoing month-in-sample 8 panel. Over the course of a year there are 12 such panels with a total of about 150,000 adults in 72,000 interviewed households. The maximum possible sample size for the Time Use Survey is about 72,000 adults in a year (6,000 per month), given the intention to subsample at most one adult per household.

From a more practical standpoint, the annual maximum sample that should be considered for a Time Use Survey is closer to 54,000 adults. The CPS has a Sate-based design that ensures adequate data in each State. State sampling intervals range from about 1/280 to 1/3600. If you are only interested in national data, then the States with relatively low populations are overrepresented in the CPS sample. Thinning out the CPS sample by 18,000 households in the less populous States (an overall 25% cut) has very little impact on national standard errors, increasing them only about 5%. [Note that every additional bit of sample helps lower standard errors, so it is impossible to set an unarguable cutoff.]

The described maximum Time Use Survey sample of 54,000 adults annually (4,500 per month) would have standard errors larger than a simple random sample of the same size with 100% response. The anticipated response rate is only about 75%. A simple random sample of about 23,000 adults with 100% response would be as effective as a 75% response with the 54,000 maximum multistage sample that is available for the Time Use Survey. (These differ by a multiplicative factor of about 2.35). The 23,000 is sometimes called an effective sample size. The 54,000 is reduced to 23,000 in the steps that follow.

- The 54,000 maximum adult sample reduces in behavior to about 42,000 when you "undo" the aspects of CPS sampling that differ from simple random sampling. The CPS sample in small States was thinned out to derive the 54,000 figure, but differences in State sampling intervals remain, and the most populous States are still a bit underrepresented compared to the others. Also, most States have a two-stage sample, and the households are selected in clusters of about four housing units each.
- A further reduction in behavior to about 36,000 arises since at most one adult is selected per household. The number of adults varies by household, so the subsampling weights differ. As a simple example: a household with only one adult may have a subsampling weight of 1 (take that adult with certainty); whereas in a household with two adults each may have a subsampling weight of 2 (take with probability .5=1/2). The different subsampling weights make the Time Use Survey

sample less efficient and are a further step away from simple random sampling.

- Reduce still further to about 31,000 since the intention is to split the sample unevenly to days of the week. Saturday and Sunday each get 25% of the sample compared to only 10% for each weekday.
- Finally, assuming a 75% response rate, this reduces to an effective sample size of about 23,000 adults.

Recommended Sample Sizes for the Time Use Survey

No fewer than 12,000 adults should be subsampled for a Time Use Survey. A projected 9,000 responses can be obtained, assuming a 75% response rate, but the net results will be comparable to a simple random sample of only about 5,000 adults. The 12,000 sample size only allows us to contrast time use distributions for a few major subpopulations of interest and to contrast weekdays with weekends. It is recommended that 12,000 adults be added to target about 15 of the smaller subpopulations of special interest that can be defined by family type, presence of children, sex, age, education, and race. In general, about 1,000 sample cases are needed to compare the most important time use characteristics of one subpopulation to another (750 responses comparable to a simple random sample of about 400 adults).

Time, Goods, and Well-Being, A book edited by F.T. Juster and F.P. Stafford includes data and associated s parameters that enable the calculation of approximate estimates, standard errors, and confidence intervals under different assumptions. The parameters were instrumental in helping us determine the needed sample size. If a simple random sample size r is known, the formula for variance is approximately $var=s^2/r$. Take the personal care variable as an example. Overall it accounts for almost half of reported time use (p=.466) with a parameter of s=.076. The numbers are not directly shown in this form by Juster, but they can be easily obtained from published tables. Without this information the temptation might have been to use the much larger parameter .5, which is the square root of .466*(1-.466).

This formula can be turned around. Suppose instead that I know how small the confidence interval is that I want for a certain data item for a certain subpopulation, from which I can easily calculate a desired variance. The formula $r=s^2/var$ tells you how large a simple random sample must be to give you the desired confidence interval. The factor 2.35 can be used to translate r to an actual sample size n for the Time Use Survey that takes into account nonresponse (75% response rate assumed) and departures from simple random sampling. If more than one data item is used to calculate a desired sample size for a subpopulation, the sample sizes won't agree. Some of the computed sample sizes may even exceed the number of cases available for sampling. Naturally you have to arrive at a single sample size for a subpopulation. Also note that it can be difficult accounting for the interactions between sample sizes among overlapping subpopulations.

Coming up with reasonable sample sizes is the most difficult part of sample design. It was determined that no one time use characteristic was of dominant interest. What we did was develop a spreadsheet that computed an entire distribution of time use proportions for various subpopulations and sample sizes (10 characteristics with proportions summing to 1.0). We could observe what sample sizes enabled us to point out important differences between distributions. The sample of 12,000 was adequate overall and a sample of 1,000 adequate for individual characteristics. Although the process has mathematical underpinnings, usable data for the United States is so scanty that the recommendations are only very approximate.

Subsampling Methodology

There are several ways to go about selecting a valid subsample that includes at most one adult per household. It is necessary that each household and each adult have a known nonzero probability of being included. For example, sample each household with a 1/3 chance, then select one adult at random within each household. For a household with four adults, each adult has a 1/12= (1/3)*(1/4) chance of being subsampled. Nothing may be optimized, but you have a valid subsample in the sense that all of the probabilities are known and consistent estimates can be made of the population. As long as the subsampling rates are known, they can be varied by household and person characteristics.

It is proposed that each adult be assigned a number on a calibrated scale (not necessarily integers), then the scaling can be used to assist in sampling. It is necessary to allow for households to have adults from several subpopulations that have different sampling needs, and scaling is a reasonably easy way to do this. Suppose for example that 1 in 20 white men in the CPS need to be subsampled, compared to 1 in 5 black men. If white men were assigned a value of 1 on the scale, then black men would be assigned a value of 4 since 1/5 is 4 times larger than 1/20. The relative scale values can be used to determine household subsampling probabilities, within-household subsampling

probabilities for adults, and the method can be modified to thin out the CPS sample in less populous States.

If all adults in every household belonged to the same subpopulation, subsampling would be easy without the scaling. Suppose I had 5,000 CPS cases and wanted 1/10 or 500 of them to be subsampled. Simply take 1/10 of the single-adult households; take 2/10 of the two-adult households and randomly select an adult; take 3/10 of the three-adult households and randomly select an adult; etc. for larger households. As long as households have fewer than 11 adults, each person has a 1/10 chance of being subsampled. [It is a little more difficult to allow for very dense subsamples, at adult rates such as 1/2. You can't subsample larger households at rates greater than unity such as 3/2 or 4/2. This can be partially compensated for by selecting more small households.]

Scaling can be operationalized as a one-stage operation. Sort out the various subpopulations by how dense the subsamples need to be. For example, 1/20 for the thinnest subsample, 1/10 for the next, up to 1/2 for the densest. The example uses integers in the denominators, but that is not a requirement. It doesn't matter if subsamples overlap; a particular person is associated with the densest subsample applicable. These subsampling intervals can be left as is, but are commonly changed to relative scores, the thinnest 1/20 becoming 1, 1/10 becoming 2, up to 1/2 becoming a score of 10. Take the thinnest sampling rate 1/20 and use it as a sampling interval of 1 in 20 that you apply to the scores. You sum through the scores selecting every adult that makes the sum exceed a multiple of 20. The first subsampled adult is the one that makes the sum exceed 20, the second subsampled adult is the one that makes the sum exceed 40, and so forth. Using this general method, adults in different subpopulations can be subsampled at any desired rates. When some subpopulations need very dense subsamples, the method needs substantial tailoring to properly handle the restriction of subsampling at most one adult per household.

For thin subsamples, the scaling method is commonly modified to "undo" the different sampling rates used in the original sample. Multiply a person's scale value by the person's associated weight in the original sample. Then refigure the appropriate sampling interval for the new scale. The method can be modified to be the vehicle that reduces the CPS sample in the less populous States.

Controlling Bias

Time use behavior can have seasonal variation, so it is planned to spread data collection evenly throughout most of the year. Major holidays such as New Years present particular problems in staffing for the interview process and in obtaining responses.

The quality of time use reporting is generally best when the memories are fresh. We plan to assign individuals a specific date to report about, and have interviewers contact them the following day. If a report is not obtained, lengthening the recall period is very risky, as is dropping the individual from the sample. On the other hand, accepting a later replacement date is risky, and certainly it can't be made simply at the convenience of the respondent. We decided to minimize recall bias by only seeking a response the day after the specified date. If unsuccessful, the individual will be retained in the sample and a new specified date assigned freezing the day of the week.

The strategy can help alleviate bias provided Monday's are like Mondays, Tuesdays are like Tuesdays, etc. The strategy is imperfect, especially in regard to some time use behaviors that occur in spells and are easily missed. For example, it is reasonable to assume that time use activity during vacations away from home is different from the ordinary. However, the above strategy almost entirely misses vacations away from home. When data for a replacement day is obtained, we plan to also obtain some basic information about the original day so we can compare profiles of responses and nonresponses.

Juster, F. T., and Stafford, F. P., eds. (1985), Time, Goods, and Well-Being, Ann Arbor: Institute for Social Research, The University of Michigan.