OPTIONAL RECALL PERIOD FOR DISCRETE EVENTS

Dhirendra N. Ghosh, Howard University
Consultant to Human Sciences Research, Inc.

In many sample surveys respondents are asked to report frequency and other data related to certain events occurring within a fixed recall period. Now events are basically of two different types:

1. Events that happen with more or less regular time interval; events that have an established pattern for a respondent. Examples are grocery shopping, drinking alcohol, lovemaking, taking vacation, withdrawing or depositing money in a bank. For these type of events when a respondent is asked to report the frequency for a specified period of time, last week, last month or last year, the respondent may tend to visualize a typical week, month or year rather than the exact last week, month or year and report the frequency. Because of the established pattern a typical week may not differ much from the last week and thus the data comes out all right. And in most cases, the survey researcher is not interested in the last week either but wants to estimate for a year or six months anyway.

2. The second type of events, I am defining as discrete events, when the time interval elapsing between any two such occurrences is irregular, when there is no established pattern. Examples are: sickness, automobile accidents, being victims of crime, furniture shopping, recreational fishing. In these events, there is hardly any typical week, month or year. Historically, household surveys started with events of the first type where a fixed recall period works out all right and we tend to use the same method for discrete events, too. An alternative method is to ask the respondent to recollect event by event, starting with the most recent one, going backwards in time and stop whenever memory fails. However, this procedure poses a fundamental problem—the problem of optional recall period. Different respondents will cover periods of different lengths of time and they cannot be standardized using a simple extrapolation technique. For example, suppose a respondent counting backwards stops on the 47th day and the 5th event, it will not be correct to use the principle of proportionality to estimate the number of events for say a 60-day standardized period by 5 x 60/47 events because there was a gap of unknown length between the 5th and the 6th event.

In 1977, Human Sciences Research, Inc., of McLean, Virginia, did a methodological study on collecting data from recreational fishermen for the National Marine Fisheries Service. We tested this methodology and examined how the two components of systematic response error, namely telescoping and omission behave with discrete events, and also solve the problem of the optional recall period.

A total of 138 fishermen were first asked to estimate the number of recreational fishing trips they took in the last 30 days. Then they were asked to list the dates as far back as they could go beginning with the most recent. There were several possible outcomes:

If m = number of trips reported in last 30 days;
and n = number of trips reported during the last 30 days as the fishermen gave actual dates by counting backwards;
then the outcomes could be:

a. m = n
b. m > n
in this case the fishermen have forgotten the dates of the trips and stopped before the 30 days were over.
c. m > n
in this case the fishermen kept on going beyond the last 30 days; they apparently overestimated the number of fishing days in the last 30 days when answering the first question; this is telescoping error.
d. m < n
the fishermen underestimated the number of trips when they gave the first estimate, but in counting back they were able to recall additional trips.
e. the fishermen cannot remember anything.

Table 1 provides the frequency distribution of these five possibilities.

Table 1
Types of Recall Error

Possible Outcome Type	Frequency	Percent
a	65	47.1
b	22	15.9
c	27	19.6
d	14	10.1
e	10	7.2
Total	138	99.9

Then the total number of trips reported by the two methods for categories c and d separately were calculated. Table 2 shows these results.

Table 2
Type c and d Errors in Recall

Outcome	Total m	Total n	Difference	Mean
c	221	116	105	3.9 (105/27)
d	35	52	- 17	-1.2 (-17/14)

These data indicate that, in giving estimates of the total number of trips in a 30-day time period, fishermen's magnitude of telescoping error may be more than the omission error.

Further examination was made, however, involving editing of the data in the following way:

In the case of (b), take (m)
In the case of (c), take (n)
In the case of (d), take (n)

After these corrections were performed, the mean and variance for 128 fishermen were calculated. A comparable group (randomly assigned), was called biweekly during the same time period and asked to report the number of fishing trips. The mean and variance for the number of trips for the last 30 days were calculated for this group as well. The results were:

Original Group (N = 128)	Mean = 2.95	S.D. = 3.37
Biweekly Group (N = 41)	Mean = 2.93	S.D. = 3.78

The comparability of these data suggest that the editing procedures for the original group were useful in making more valid estimates possible.

Also examined was the length of time for which dates are supplied by fishermen. Using the time periods for which such dates were obtained, a frequency distribution was developed. This is shown as Table 3. This table indicates that if 30 days were chosen as the reference period then a great deal of information about fishing trips will be lost. With a two-month reference period, one would obtain information from 51.5 percent more fishermen.

Table 3
Frequency Distribution of Fishermen Reporting Dates for Selected Reference Periods

Length of Time	Number of Fishermen	Percent
0-15 days	23	18.0
16-30 days	20	15.6
31-45 days	19	14.8
46-60 days	15	11.7
61-90 days	21	16.4
More than 90 days	30	23.4
	128	99.9

Mean = 72.1 days
Standard Deviation = 74.0

The problem of non-uniform recall period can be solved in the following way. The objective is to estimate the frequency for everybody for a standardized recall period. The respondents who covered 60 days form the anchor group. We fit two regression lines with data obtained from this group.

1. The frequency of the events for the first 30 days (counting backwards) as the independent variable (predictor) and the frequency of the events for the 31st to the 60th day as the dependent variable (predicted).

2. The frequency of the event for the first 45 days (counting backwards) as the independent variable (predictor) and the frequency of the events for the 46th to the 60th day as the dependent variable (predicted).

For an individual respondent who does not cover 60 days, the following procedure is used. The date is noted when the respondent stops. If the last date falls on a less-than-30 days counting backwards, he is asked to estimate the number of trips during the period between his last reported date and 30 days. For these people, the first regression is used to estimate the number of trips between the 31st and 60th day (counting backwards). Similarly for respondents stopping between 30 and 45 days, we take the data for the first 30 days and treat it the same way as above. For respondents stopping between 45 and 60 days, we take the frequency for the first 45 days, and use the second regression above. However, one may decide not to assume any super population model and instead treat the whole problem as a problem of imputation for missing data. Then instead of fitting any regression line, we treat the data from the anchor group as the hot deck and supply the frequency for the missing period.