IDENTIFYING MEASUREMENT ERROR IN CONSUMER EXPENDITURE REPORTS

Clyde Tucker, Bureau of Labor Statistics

I. Introduction

The estimation of sampling error in a survey is relatively straightforward even in the case of a complex sample design. Nonsampling error is more difficult to measure because of the multitude of factors which can produce it, an underdeveloped theory, and the absence of knowledge about population parameters. This paper, a condensed version of one available on request, describes an attempt to measure at least a portion of the nonsampling error in the expenditure reports from the 1980-81 Consumer Expenditure Diary Survey conducted by the Bureau of the Census for the Bureau of Labor Statistics.1/ Specifically, surveyed consumer units, who were asked to record their daily expenditures over a two-week period, are classified according to their level of response error as a result of analyzing the latent structure in the associations among several response pattern indicators. These indicators are derived from data within the survey and provide information about the validity of each unit's expenditure report.

II. Approach to the study of nonsampling errors

There are three types of nonsampling error--systematic measurement error, measurement fluctuation, and representation error. Unlike sampling error, these errors are associated directly with the individual units in the sample. The first two are measurement error which result either from the respondent giving an incorrect response or from the recording of a correct response incorrectly. Representation error occurs when units in the population are assigned incorrect sampling rates. We will be concerned only with that part of measurement error attributable to respondents.

All nonsampling errors, and response error in particular, might better be analyzed as micro variables rather than in an aggregated form. Since nonsampling errors occur at the individual level, the only way to rid ourselves of them is through work at that level. As a first step, studies are needed in which nonsampling error is the dependent variable in models designed to determine its causes. Once these causal factors for nonsampling error are identified, we can develop new survey methods for overcoming their effects.

In our analysis, response error is defined as the dependent or outcome variable in the survey situation. Consumer unit and environmental characteristics combine with the characteristics of the survey procedure (including interviewer characteristics) to determine both the respondent's attitudes toward the survey task and the respondent's method of performing the task. These attitudes and behaviors, in turn, lead to response error.2/

Before we can specify the contributions made to response error by the various elements in the survey situation, we need measures of all the variables. Measuring the independent variables is relatively straightforward, but response error is another matter. The purpose of the analyses described in this paper was the development of a measure of response error.

III. The design

a. Assumptions underlying the measurement of response error in the CE Diary Survey

In this paper, the total response error in a consumer unit's expenditure report was measured. The interconnected assumptions which underlay this measurement are:

1. Patterns found in the information reported by respondents are related to the level of response error in their expenditure reports.
2. Various indicators of these response patterns can be developed.
3. Reasonable judgments can be made about the substance of the relationships between the indicators and response error.
4. The associations among the pattern indicators can be used to model a latent variable which represents response error.

The central assumption here is that the level of response error in a consumer unit's report of expenditures can be determined from the manner in which the respondent reports information, both expenditure information and other information. Traditionally, there have been two types of data used to identify response error in surveys--reinterviews and independent sources (Madow, 1973; Sudman and Bradburn, 1974; Hubbard, et.al., 1981; Marquis, et.al., 1981; Groves and Magilavy, 1984; Corby and Miskura, 1985). With both reinterviews and independent sources, new data are compared to the original survey measurements.

Although widely used, reinterviews and independent sources may not be the best ways to identify response errors for several reasons. Reinterviews can be quite expensive and still produce the same or different errors. Often respondents are not willing to undergo a second survey, and those that are may not agree to more rigorous procedures. Since reinterviewing is seldom done on the entire sample, inferences about those not reinterviewed must be made. Finally, when the phenomena under study are transient, reinterviews may not be appropriate.

The use of independent sources also has shortcomings. In the first place, these sources do not always exist or are not always accessible. Even when they are, accessibility may be limited to a self-selected or at least an unrepresentative subset of the sample. There may also be questions concerning the accuracy and comparabilty of these independent sources.

An alternative to the above methods which is explored in this paper is the use of information from the survey itself to identify response error. Patterns in an individual's responses can indicate the extent of response error for the variables of interest. There are several advantages to this approach. Little cost is incurred. No new interviewing procedures need be developed nor independent sources found. Generalizations from a subset of the sample is avoided; and, perhaps most importantly, the problem of self-selection is eliminated.

This approach does have its drawbacks. Identifying and measuring particular response patterns related to response error are difficult. Presumably, one at least would have a notion as to how these indicators and response error are related. Given that several pattern indicators are likely to be used, a method for combining the information from each would be needed. And we also need a way to evaluate the results of this process.

Whether or not useful response patterns can be identified depends on the particular survey. Fortunately, the CE Diary Survey collected a large body of information on both consumer unit characteristics and expenditures. Furthermore, the expenditure information covered a period of time which was long enough to ascertain patterns in the reporting of expenditures.

Most response errors in the CE Diary Survey will be in the form of underreports. It is difficult to imagine an individual recording more items than were purchased or even consistently overreporting the price of items, but the failure to report all items is quite likely given the time and effort required to fill out the diary. There is substantial information to support this assertion (Sudman and Ferber, 1971; Pearl, 1979; and U.S. Department of Labor, 1983).

Multiple indicators of response error are desirable since any one indicator is unlikely to provide complete and accurate information. Thus, a number of response pattern indicators were examined. The associations between the several response pattern indicators were used to model a latent variable which represented response error.

An evaluation of the results was a necessary final step. The question to be answered was not simply how well the latent variable described the associations among the pattern indicators but also how well did this approach identify response errors.

b. Achieving comparability

If response error is going to be studied at the micro level, the data must be comparable from case to case. We employed three methods to achieve this comparability in the CE Diary data. We selected for analysis goods bought by most consumers on a frequent basis and ones on which the diary was designed to collect information. These goods were food and non-alcoholic beverages for home consumption and food and non-alcoholic beverages consumed away from home. To arrive at a comparable group of consumer units, we selected those which completed two diary weeks, 8991 of 10319 urban consumer units. We adjusted the expenditures for two factors affecting their comparability from unit to unit -- sales tax and inflation.

c. Latent structure analysis

Latent structure analysis, a technique similar to factor analysis, is used when only qualitative data are available. A latent variable which is not observed directly is derived from associations among at least two manifest (observed) qualitative variables. This latent variable is taken to explain the relationships between the manifest variables. There can be any number of manifest variables and also more than one latent variable just as factor analysis often produces more than one factor. The response pattern indicators served as the manifest variables in this study. A single latent variable was interpreted to be an ordinal scale of response error.

Let us consider latent structure analysis in mathematical terms. When variables A and B are not independent, the following relationship will not hold:

$$\pi_{ij}^{AB} = \pi_i^A \cdot \pi_j^B \qquad (1)$$

where i indexes the classes of A,j indexes the classes of B, π_{ij}^{AB} is the probability an individual is in cell (i,j), π_i^A is the probability an individual is in class i and π_j^B is the probability an individual is in class j.

For the expression in (1) to be true,A and B must be independent. The purpose of the latent variable X is to achieve this independence. That is,we want to arrive at the following latent class model:

$$\pi_{ijt}^{ABX} = \pi_t^X \cdot \pi_{it}^{AX} \cdot \pi_{jt}^{BX} \qquad (2)$$

where t indexes the classes of X, π_{ijt}^{ABX} is the probability of being in cell (i,j,t) of the unobserved ABX table, π_t^X is the probability that an individual is in one of the mutually exclusive and exhaustive classes of X, π_{it}^{AX} and π_{jt}^{BX} are the conditional probabilities that an individual is in a particular class of A and B , respectively, given that person is in a certain class of X. Equation (2) states that, within a class of X, A and B are independent of one another.

Goodman (1974) describes the procedure to be followed for identifying the classes of the latent variable X and, thus, estimate the parameters (probabilities) on the right-hand side of (2). Clogg (1977) has developed a computer program (MLLSA) which uses Goodman's procedure to identify the latent structure model for polytomous manifest variables. After the model has been estimated, the parameter estimates are used to generate expected frequencies (F_{ij}) for the manifest cells. With these expected frequencies and the observed ones (f_{ij}), two Chi-square tests can be performed to determine the fit of the model. They are the Pearson's Chi-square and the Chi-square based on likelihood ratios with the degrees of freedom associated with the model computed in the following way for q manifest variables with the number of classes for each labeled I_k and the number of latent classes labeled T:

$$DF = \left(\prod_{k=1}^{q} I_k\right) - 1 - \left\{\left(\left(\sum_{k=1}^{q} I_k\right) - (q-1)\right) T - 1\right\} \qquad (3)$$

The ultimate purpose for using latent structure analysis in this study was to assign the individual consumer units to the classes of the latent variable or the points on the scale of response error. The units in a cell (i,j) were assigned to the class which received the modal proportion of the cases in that cell. This method of assignment and its associated error is analogous to the lambda measure of association. Lambda can serve as a supplement to the Chi-square statistics.3/.

IV. The response pattern indicators

a. Development of the indicators

Before the response pattern indicators were developed, we laid out an analysis plan. The two classes of expenditures analyzed were Food and Non-alcoholic Beverages for Home Consumption and the total of Food and Nonalcoholic Beverages Consumed At Home and Away From Home. These classes were examined for both the total sample and what we refer to as the "intact" families. The intact families are those units where no members were away and no visitors were present during the two-week period. We believed that this group of 6208 units, being more homogeneous, would provide the most conclusive results.

The first of the response pattern indicators, RECAL, is a dichotomous variable which measures whether or not recall information is contained in a consumer unit's expenditure report. While respondents are supposed to keep the diary themselves, many times the interviewer must conduct a recall interview at the end of the week for all or part of the expenditures which the respondent has failed to record. We chose this variable because we felt that the presence of recall information increases the likelihood that the expenditure report is incomplete.

Another indicator, FDDIF, measures the difference between the respondent's average weekly expenditure for food at home as reported in the diary and a prior estimate of this expenditure given by the respondent at the beginning of the two-week diary period. This difference was divided by the sum of these two values. The smaller the reported expenditure is compared to the estimated expenditure, the less confidence we have in the reporting. Because the respondent's estimate of food expenditure may be somewhat inaccurate, we have recoded the continuous variable into three discrete categories which accentuate gross differences in FDDIF from respondent to respondent. The dividing lines between categories are based on the distribution of the original variable (the 25th and 75th percentiles). FDDIF is also the new discrete variable.

A third indicator which is similar to FDDIF provides further information about the reported expenditures compared to the respondent's estimate of usual expenditures. Before the diary is placed in the home, the respondent is not only asked about the usual grocery expenditures but also about the number of trips made to the grocery store in a week. By examining the pattern of reported daily food expenditures, we estimated the number of grocery store trips per week made by the respondent during the two-week diary period.4/ TRIP is the difference between the respondent's prior estimate and our estimate of reported trips divided by the sum of these values, and it was recoded like FDDIF.

Two other indicators, FWEEK and AWEEK, were derived by comparing expenditures for the first week (either for food at home or total food purchases) to those for the second week. Previous consumer diary research (Turner, 1961; Kemsley, 1961; Sudman and Ferber, 1971; Pearl, 1979) indicated that the first week expenditures tend to be higher than those for the second week. While this could be the result of telescoping earlier expenditures into the first week, our procedures are designed to overcome this tendency. It is likely, therefore, that a decline in reported expenditures shows a loss of interest in keeping the diary and, thus, greater underreporting.

To create FWEEK we began by computing the difference between the first week expenditure for food at home and the same expenditure for the second week divided by the sum of the two expenditures. AWEEK was created in a similar manner except expenditures for both food at home and away were used. Both were recoded into three discrete categories as discussed above.

We hypothesized that the middle category of these variables identify consumer units with a low response error since they are the ones with the smallest difference between the first and second week expenditures. We were fairly certain that category three containing units having much larger expenditures in the first week, indicated a high level of response error. There was less certainty about category one in which units reported more expenditures in the second week.

b. Relationships of the indicators to expenditures

After creating the five indicators, we evaluated their probable connection to response error by examining the mean weekly expenditures for food at home and total food over the categories of these indicators. Even though a particular respondent's reported expenditure size is not necessarily related to response error, if the direction of the <u>means</u> are what we would expect, it gives us some confidence that the indicators are valid. Since the most likely response errors, as already stated, are underreports, those categories considered to have the greatest response error should have the lowest means. Categories with the least response error would have the highest means.

Table 1 displays, for all families, the weekly means for food at home (FDHOMEAV) and total food purchases (FDALLAV) which are the simple averages of the respondent's two weekly reports.5/ In the cases of TRIP, FDDIF and RECAL, category one should have the least response error and the highest numbered category the most. As you can see, the means for both FDHOMEAV and FDALLAV are in the expected direction for these variables. The category means are always significantly different from each other. These same patterns hold for intact families. While the declines in the means found when moving from category one to two in TRIP and FDDIF are certainly meaningful, it is the differences in the means of categories two and three that are most striking. The differences in the means for the two categories of RECAL are similar to those for the first two categories of TRIP and FDDIF.

The means for AWEEK and FWEEK are particularly interesting given our earlier discussion of these variables. As predicted, the second category, where the week-to-week variation is smallest, has the greatest means. Category three, containing individuals with larger first week expenditures, has much smaller means. But category one, with respondents who have greater second week expenditures, also has smaller means which are almost identical to those for category three.

V. The latent response error variables

a. Creation of the latent variables

We conducted four latent structure analyses using MLLSA. There was an analysis for each of the four cells in the analysis plan. The input for these analyses was the weighted cell frequencies from four-way cross-tabulations of the response pattern indicators. To create the latent response error variable for food at home, we used TRIP, FDDIF, RECAL and FWEEK. TRIP, FDDIF, RECAL and AWEEK were cross-tabulated to develop the latent response error variable for total food purchases.

The results of the analyses for all families are presented in Table 2; results for the intact families are quite similar and can be found in the expanded version of the paper. The Chi-square values testing independence in the four-way tables are extremely large. The introduction of the latent variables, however, greatly reduces the size of the Chi-square values. Significant relationships still remain, but this might be expected given the large sample size.

Note that the number of degrees of freedom associated with the test of the latent model does not equal the calculated figure from (3) which is thirty. A boundary problem results from the fact that a few estimates of the conditional probabilities are close to zero. To overcome the problem, these probabilities are set to zero (never more than three in a model) creating a situation analogous to placing <u>a priori</u> restrictions on parameters.

Besides the Chi-square statistics, we can evaluate the other information produced by the MLLSA program. In each model, ninety percent of the cases were correctly classified using the modal class probabilities, and lambda is always about .80. The index of dissimilarity is so small that only about three percent of the cases would have to be shifted to achieve a dissimilarity index of zero.

b. Modeling expenditure with the latent variables

Using the modal latent class probabilities for each cell in the four-way table, respondents were assigned to one of the latent classes. The lowest numbered class is the one deemed to have the least amount of response error while the highest has the most. The labeling of these classes is based on the theoretical relationships between the response pattern indicators and the latent response error variable. The latent class variables were named CLASHOME and CLASTOTL for food at home and total food purchases, respectively. In order to evaluate these variables, each was entered as the independent variable for modeling the appropriate weekly expenditure variable (either FDHOMEAV or FDALLAV).

In all cases, the F-ratios are significant indicating there is indeed a relationship between the latent response error variable and the expenditure variable. Moreover, the R-square values are similar for both intact families and all families (.10 to .18). The fact that the R-square values are not particularly large is consistent with the idea that individual expenditure amounts should still vary widely within each of the classes. It is only response error in these expenditure reports which should be controlled by the latent variable.

In connection with this last point, we were interested in seeing whether the latent variable was orthogonal to income and family size, probably the two most important predictors of food expenditure. We modeled the expenditure variables using income, family size and the latent variable to examine this question. Looking at the sum of squares attributed to the latent variable after controlling for income and family size, we found that, in every case, two-thirds to three-fourths of the amount of variance explained by the latent variable was unique. This information, along with the relatively small R-squares, gives us confidence that we are not simply measuring expenditure size with the latent variable.

c. Analysis of expenditure variable means across the latent classes

Means of the appropriate expenditure variable for the classes of the latent variable for both all families and intact families are found in Table 3. The patterns of the means are quite similar for both sets of families. Not only are the differences always significant, but they are also large enough to be meaningful. This is especially true for the differences between the category three mean and the other two means. And, certainly to the extent that lower means indicate more underreporting, we can say category three identifies respondents who underreport the most.

Of particular interest here is the comparison of the patterns in the FDHOMEAV means and the patterns in FDALLAV means. In the first place, the actual size of expenditures for food away from home (not shown) tends to be a good deal smaller than that for food at home. Therefore, the differences in food at home expenditure dominates the differences in total food expenditure means. As it turns out, the mean food away from home expenditure for category three of CLASTOTL is somewhat smaller than the mean for the other categories, but there is much less discrimination than with food at home.

d. Relationships among the response pattern indicators and the latent response error variables

Cross-tabulations of the response pattern indicators and the latent variables were done. They provided us with an understanding of the contribution of each of the indicators to the latent response error variables. For instance, presence in the first category of TRIP largely determines membership in the category with the least amount of response error in expenditure reports for food at home (category one of CLASHOME). The situation is the same with total food purchases for all families. This is not true in the case of CLASTOTL for intact families.

We were concerned that the importance of TRIP in defining the first category of the latent variables meant we were simply measuring expenditure size and not response error. Other information, however, suggested that this was not true. In the first place, FDDIF which is related to TRIP did not behave in the same way with respect to this category. Secondly, TRIP's effect is not isolated from the effects of the other pattern indicators. CLASHOME and CLASTOTL are the products of the interrelationships of all of these indicators. This is reflected in the fact that the importance of TRIP diminishes in the case of CLASTOTL for intact families. Finally, the first category of TRIP is comprised of respondents who made either the same or more trips in comparison to the number of trips they estimated prior to keeping the diary. When these two groups of respondents were examined separately, we found that the means for FDHOMEAV and FDALLAV were actually higher for respondents who made the same number of trips as estimated compared to those who made more trips.

Just as TRIP is important for determining which respondents are in the lowest response error category, FDDIF identifies respondents in the highest category. The two categories of RECAL differ from one another in that category two is more likely than category one to have members with the highest level of response error. As for FWEEK and AWEEK, their middle category contains a smaller percentage of cases in the worst error category of the latent variables than do their extreme categories.

e. Demographics and the latent response error variables

As a preview of the causal analysis described in Section II, Table 4 displays the relationships between certain CU demographic characteristics and the latent variables for all families. The results for intact families are similar. A pattern emerges from this table which is not entirely unexpected. Those consumer units most likely to be assigned to the category with the greatest response error are ones composed of young, single individuals with low income. They are more often than not renters.

The description given above is one of people who may lead somewhat unsettled lives. They may not spend much time at home or, at least, have erratic schedules. These people may not be inclined to take the time to keep the diary.

VI. Discussion

Nonsampling errors, and response errors in particular, are very difficult to measure, but there has been an increasing number of attempts to do so in recent years. Most of the attempts which have dealt with error at the micro level have involved the use of reinterviews or independent sources such as administrative records. These procedures, however, are not appropriate for many surveys, and the CE Diary Survey is one of them. This paper has presented another approach to the measurement of response error at the micro level. It appears to be a useful one although improvements certainly could be made.

The goal of this research was to produce a latent variable which could be used as the dependent variable for determining the causes of response error. It may not be desirable to use this measure of response error to adjust estimates directly, but the latent variable can be used to revise survey procedures and evaluate the revisions. Hopefully, we can determine the contributions to response error made by the various elements in the survey situation; and, with these, we can specify the nature of the causal process leading to response error.

FOOTNOTES

1/ For a description of the CE Diary Survey refer to U.S. Department of Labor (1983).

2/ The utility of viewing response error as the dependent variable was emphasized by Borus (1966). For another but similar conceptualization of the survey situation see Sudman and Bradburn (1974).

3/ See Clogg, 1977. Another measure of the goodness of fit not so dependent on the sample size is the dissimilarity index defined as

$$\sum_{ij} \frac{ABS(f_{ij} - F_{ij})}{2n} .$$

4/ To estimate the number of observed trips we first computed the average amount spent on food per grocery trip as reported by the respondent prior to keeping the diary. We then calculated the mean expenditure per trip for families of different size. We divided these means in half and used the new values to compute the number of observed trips. For every CU, we counted the number of days in which expenditures for food exceeded the amount associated with the CU's size and divided the total in half to arrive at the number of observed trips per week.

5/ All results use weighted data unless otherwise indicated. The weight is the simple average of the weights for the two individual weeks. Statistical tests assume simple random sampling.

REFERENCES

Borus, M. E. "Response error in survey reports of earnings information," Journal of the American Statistical Association, 1966 (61), 729-738.

Clogg, C. C. "Unrestricted and restricted maximum likelihood latent structure analysis--A manual for users," Working Paper No. 1977-09, Population Issues Research Office, Pennsylvania State University, July, 1977.

Corby, C. and S. Miskura, "Evaluating data quality in the economic and decennial censuses," Proceedings of the 1st Annual Research Conference, Bureau of the Census, 1985, 159-175.

Goodman, L. A. "The analysis of systems of qualitative variables when some of the variables are unobservable: Part I - a modified latent structure approach, "American Journal of Sociology, 1974 (79), 1179-1259.

Groves, R. M. and L. Magilavy, "An experimental measurement of total survey error," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1984, 698-703.

Hubbard, R., et al., "Validity of self reports of illegal activities and arrests by drug treatment clients," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1981, 329-334.

Kemsley, W. F. F. "The household expenditure enquiry of the ministry of labour," Applied Statistics, 1961 (10), 117-135.

Lazarsfeld, P. F. and N. W. Henry, Latent Structure Analysis, Boston: Houghton -Mifflin, 1968.

Madow, W. G. "Net differences in interview data on chronic conditions and information derived from medical records," Series 2, No. 57, Vital and Health Statistics, 1973.

Marquis, K. H., et al., Sensitive Topic Surveys, R-2710/2-HHS, The Rand Corporation, 1981.

Pearl, R. B. "Revaluation of the 1972-73 U.S. Consumer Expenditure Survey," Tech. Paper No. 46, U.S. Department of Commerce, Bureau of the Census, July, 1979.

SAS Institute, Inc. SAS User's Guide: Basics, 1982 Edition, Cary, N.C: SAS Institute, Inc., 1982a.

SAS User's Guide: Statistics, 1982 Edition, Cary, N.C: SAS Institute, Inc., 1982b.

Sudman, S. and N. M. Bradburn, Response Effects in Surveys, Chicago: Aldine, 1974.

and R. Ferber, "Experiments in obtaining consumer expenditures by diary methods," Journal of the American Statistical Association, 1971 (66), 725-735.

Turner, R. "Inter-week variations in expenditure recorded during a two-week survey of family expenditure," Applied Statistics, 1961 (10), 136-146.

U.S Department of Labor Consumer Expenditure Survey: Diary Survey, 1980-81, Bull. 2173, Bureau of Labor Statistics, September, 1983.

Table 1. Means of Food-at-Home and Total Food for Pattern Indicators -- All Families

	Weighted N	Mean of FDHOMEAV	Mean of FDALLAV
TRIP			
1	2380	$47.32	$62.98
2	2923	39.55	55.89
3	3688	17.55	30.27
FDDIF			
1	2120	46.55	60.69
2	4239	36.38	52.41
3	2632	15.44	28.39
RECAL			
1	6256	34.26	49.91
2	2735	28.32	42.00
AWEEK			
1	2209	---	40.24
2	4420	---	53.58
3	3362	---	42.22
FWEEK			
1	2154	26.87	---
2	4309	38.97	---
3	2528	26.46	---

Table 2. Results of Latent Structure Analyses -- All Families

	Food-At-Home			Total Food		
	Pearson	L-R	DF	Pearson	L-R	DF
Chi-squares without latent variable	3186	2913	46	2674	2579	46
Chi-squares with latent variable	165	162	32	125	125	32
Index of dissimilarity		0.05			0.05	
Lambda		0.85			0.83	
Cases correctly classified		91%			90%	
Probability for latent class						
1		.32			.32	
2		.42			.41	
3		.26			.27	

Table 3. Weekly Means of Food-at-Home and Total Food Expenditures for the Latent Classes

	All Families		Intact Families	
CLASHOME	Weighted N	Mean of FDHOMEAV	Weighted N	Mean of FDHOMEAV
1	2550	$46.43	1733	$46.64
2	3994	35.22	2675	36.12
3	2448	14.60	1800	16.89
CLASTOTL	Weighted N	Mean of FDHOMEAV	Weighted N	Mean of FDHOMEAV
1	2297	$63.74	881	$61.86
2	4095	50.37	3505	53.11
3	2599	28.08	1822	29.59

Table 4: Relationships Between Demographics and Latent Variables -- All Families

	Income			CU Size		
CLASHOME	Incomplete Reporters* (1694)	Under $15K (3404)	$15K and over (3894)	1 (2496)	2 (2593)	3 or More (3904)
1	23%	27%	32%	27%	28%	30%
2	40	43	47	36	47	47
3	37	30	21	37	25	23
CLASTOTL						
1	20	24	29	25	24	27
2	42	45	50	37	51	50
3	38	31	21	38	25	23

	Age of# Reference Person			CU Tenure	
CLASHOME	Under 25 (1053)	25-44 (3577)	45 and over (4362)	Owner (5537)	Renter+ (3455)
1	21%	29%	29%	31%	24%
2	38	44	47	46	42
3	41	27	24	23	34
CLASTOTL					
1	18	27	26	28	22
2	40	46	49	49	43
3	42	27	25	23	35

* Respondents who fail to report all income.
\# Reference person is the one who owns or rents the dwelling.
\+ Includes respondents living on property of others without paying rent.