Consumer Expenditure Survey – Conference Training Projects
Paulin, Rojas, and Wilson
CEXINFO@bls.gov - (202)-691-6900

Project 1: Calculating mean food at home expenditures with summary variables

The FMLI files in the Interview Survey contain all summary variables generated from the survey. These represent the sum of the more detailed expenditures that make them up. This project will be utilizing FDHOMECQ and FDHOMEPQ along with the unique identifier NEWID in order to calculate our desired number. Select these variables out of the relevant datafile: FMLI141x. The "141" identifies the year (2014) and quarter (1) in which the data were collected. The 'x' denotes that the file is produced using the top coding and other methods from the year 2014. There exists a similar file which does not contain an 'x,' this file functions as the fifth quarter of the previous year which is necessary for producing calendar year estimates.

The FMLI file contains observations uniquely identified by a NEWID. Associated with that NEWID are a host of demographic characteristics and summary level expenditure data. Only select the variables associated specifically with *food at home*. Since the data are collected based on a three month recall and this project requires expenditures collected in the first quarter of 2014 (as opposed to expenditures which occurred in the first quarter of 2014), sum together the current quarter version of the variable and the previous quarter version of the variable. Doing so will get a *food at home* variable (FDHOME) which represents the total expenditure amount collected in the first quarter of 2014.

After computing the FDHOME variable, the mean is interpretable in two ways. Either directly compute the mean of the vector which will include all individuals who reported no expenditures in this category (85 consumer units) or remove those individuals who did not make an expenditure on *food at home*. The latter will increase the value of the mean ($1289.73) expenditure by a small amount and may be more effective at describing the purchasing patterns of those consumer units who actually made an expenditure on the item. The former is the mean which is descriptive of the entire sample ($1272.82).

Project 2: Calculating mean food at home expenditures with the MTBI file

The MTBI files in the Interview Survey contain expenditures and other information at the Universal Classification Code (UCC) level. Only the variables of interest for this exercise (NEWID, UCC, and COST) are selected from the relevant data file MTBI141x. The "141" identifies the year (2014) and quarter (1) in which the data were collected. The 'x' denotes that the file is produced using the top coding and other methods from the year 2014. There exists a similar file which does not contain an 'x,' this file functions as the fifth quarter of the previous year which is necessary for producing calendar year estimates.

The MTBI files include one record for every UCC reported. Therefore, consumer units reporting multiple expenditures will appear multiple times in the file. To find the total expenditure on *food at home* for a consumer unit, it is necessary to produce one record that shows expenditures on this item for the consumer unit. Restrict the dataset to only the necessary UCCs ("790240" and "190904"). The expenditure records (COST) are to be collapsed by NEWID so that each record in the final data set will include all expenditures of interest for each individual consumer unit reporting at least one expenditure of interest. From here, directly compute the mean ($1289.73) which will match the mean computed with the summary variables with zeroes restricted in the previous project. This is because the MTBI files do not contain records for consumer units which did not make an expenditure on a particular UCC.

Project 3: Calculating mean food at home expenditures with the MTBD file

The EXPD files in the Diary Survey are similar to the MTAB files in the Interview Survey. That is, they contain expenditures and other information at the UCC level. Only the variables of interest for this exercise (NEWID, UCC, and COST) are selected from the relevant data file: EXPD141. The "141" identifies the year (2014) and quarter (1) in which the data were collected.

As with the MTAB files, the EXPD files include one record for every UCC reported. Therefore, a consumer unit reporting expenditures only for oranges and other citrus will appear twice in the EXPD file. To find the total expenditure on citrus fruit for a consumer unit, it is necessary to produce one record that shows expenditures for both items for this consumer unit. The first step in completing this task is to generate new variables to identify the amount of the expenditure for each item. In this case, the new variable "ORANGES" equals "COST" for each record for which UCC="110310", and "OTHER_CITRUS" equals COST for each record for which UCC="110510." The expenditure records are to be collapsed by NEWID so that each record in the final data set will include all expenditures of interest for each individual diary reporting at least one expenditure of interest.

(NOTE: The last character of NEWID identifies the week (first or second) for which the consumer unit completes the diary. Therefore, a consumer unit reporting expenditures for oranges in each week will appear twice in EXPD_SUM. However, for this demonstration, each diary is treated independently in computing average expenditures and related statistics. The result is that all statistics computed are for weekly expenditures. For example, if the same consumer unit spent $4 for oranges in week 1 and $6 in week 2, the sample includes two diaries, each with average weekly expenditures of $5.)

The data set now includes a single record for each diary reporting weekly expenditures for at least one of the three items of interest: apples; oranges; or citrus fruits excluding oranges. The next step is to compute the amount each consumer unit spent on combinations of fruits: either all citrus fruit, or apples and oranges.

(Example: A consumer unit reporting $5 for ORANGES and $10 for OTHER_CITRUS will have a value of $15 for ALL_CITRUS. If this consumer unit also reports $1 for apples, then APPLES_AND_ORANGES will have a value of $6.)

There should be 1198 total records describing the purchase of at least one item of interest. However, only 606 purchases are apples, 454 purchases are oranges, and 676 are citrus other than oranges. Because the total number of purchasers (1,736=606+454+676) exceeds the number of consumer units reporting purchase of at least one type of fruit (1,198), some consumer units purchased at least two types of fruit. Related to this, note that the mean, standard error, minimum, and maximum values computed for each type of fruit are for purchasers only, not all of the records in the sample. This is because the MTBD files, like the MTBI files, do not contain records for consumer units which did not make an expenditure on a particular UCC.