Addendum to COBRA Subsidy Survey Documentation

Overview

The public-use data file uses categories for the age and pay-amount (i.e., income) variables, whereas the restricted-use data file uses continuous variables. Age was classified into five categories using the standards followed by the U.S. Census[1]. The lowest category for age is "Under 34" because the minimum value for this variable is 21 in the data set. Pay-amount was classified into seven categories to be consistent with the range used in the pay-range-job variable (C15b_1) from the questionnaire. A list of the variables in the restricted-use file and the corresponding transformed categorical variables included in the public-use file is presented in table 1.

Table 1. Names for "Age" and "Pay-Amount" variables in the restricted-use and public-use data files.

Variable Name: Restricted-Use File	Variable Name: Public-Use File
Age (Age at Job Loss)	Age_category (Age at Job Loss Categorical)
C15a_1 (Pay-amount-job 1)	C15a_1_category (Pay-amount-job 1 – categorical)
C15a_2 (Pay-amount-job 2)	C15a_2_category (Pay-amount-job 2 – categorical)
C15a_3 (Pay-amount-job 3)	C15a_3_category (Pay-amount-job 3 – categorical)
C15a_4 (Pay-amount-job 4)	C15a_4_category (Pay-amount-job 4 – categorical)
C15a_5 (Pay-amount-job 5)	C15a_5_category (Pay-amount-job 5 – categorical)
C15a_6 (Pay-amount-job 6)	C15a_6_category (Pay-amount-job 6 – categorical)

Note: Variable labels are provided in the brackets.

Usage Notes

To utilize the SAS data sets for earlier versions (e.g., SAS 9.3) without applying the value formats, include the option "option nofmterr;" at the top of the SAS program.

To apply the value formats to the SAS dataset, include the option "option fmtsearch = (library.formatfilename);"[2].

[1] https://www.census.gov/content/dam/Census/library/publications/2015/demo/p60-252.pdf
[2] See http://support.sas.com/resources/papers/proceedings12/048-2012.pdf for details on SAS formats.

Procedures Used to Construct New Categorical Variables

Age (Age at Job Loss): Age is a continuous variable in the restricted-use file and was transformed into "Age_category" in the public-use file with five categories: "Under 34", "35 to 44", "45 to 54", "55 to 64", and "Over 65".

C15a_1 (Pay-amount-job 1): "C15a_1" is a continuous variable in the restricted-use file and was transformed into a categorical variable "C15a_1_category" in the public-use file using the "C15a_1" and "C15a_Per_1" variables. The details of the construction of this variable are provided below.

- Step 1: Pay frequencies reported in "C15a_Per_1" were transformed to calculate the pay amount per annum following the standard units of 40 hours a week, 52 weeks per year, 12 months per year, and 260 days per year.

C15a_Per_1 (Pay frequency job1)	C15a_Per_1 (Pay frequency job1 transformed units per annum)
1 (Per hour)	2080
2 (Per week)	52
3 (Once every two weeks)	26
4 (Twice a month)	24
5 (Per month)	12
6 (Per year)	1
7 (Per day)	260

- Step 2: Multiply "C15a_1" with transformed C15a_Per_1 (Pay frequency job1 transformed units per annum) to get the annual pay amount.
- Step 3: From the annual pay amount, a new categorical variable "C15a_1_category" was constructed with 7 categories: "Under $10,000", "$10,000 to $19,999", "$20,000 to $39,999", "$40,000 to $49,999", "$50,000 to $74,999", "$75,000 to $99,999", and "$100,000 and over".

Similarly, following the above steps new categorical variables were constructed for "C15a_2" to "C15a_6". Please note that "C15a_1" to "C15a_6" and "C15a_Per_1" to "C15a_Per_6" are not available in the public-use file and instead "C15a_1_category" to "C15a_6_category" variables are available in the public-use file[3].

[3] The total number of variables in the dataset is 1189. The following variables were dropped from the data set: "Age", "C15a_1", "C15a_Per_1", "C15a_2", "C15a_Per_2", "C15a_3", "C15a_Per_3", "C15a_4", "C15a_Per_4", "C15a_5", "C15a_Per_5", "C15a_6", "and C15a_Per_6". The new categorical variables added to the data set are as follows: "Age_category", "C15a_1_category", "C15a_2_category", "C15a_3_category", "C15a_4_category", "C15a_5_category", and "C15a_6_category".

Summary Statistics for the New Categorical Variables

Frequency and percentage tabulations for the newly constructed variables are provided below.

Variable name: Age_category

Variable label: Age at Job Loss Categorical
Variable type: Numeric (categorical)

Number of distinct values: 5
Total number of cases: 3,476
Cases with missing values / total number of cases (percent): 0 / 3,476 (0.0%)

Value	Label	Frequency	Percentage
1	Under 34	668	19.2%
2	35 to 44	728	20.9%
3	45 to 54	972	28.0%
4	55 to 64	861	24.8%
5	Over 65	247	7.1%

Variable name: C15a_1_category

Variable label: Pay-amount-job 1 - categorical
Variable type: Numeric (categorical)

Number of distinct values: 7
Total number of cases: 3,476
Cases with missing values / total number of cases (percent): 121 / 3,476 (3.5%)

Value	Label	Frequency	Percentage
1	Under $10,000	20	0.6%
2	$10,000 to $19,999	291	8.4%
3	$20,000 to $39,999	1608	46.3%
4	$40,000 to $49,999	463	13.3%
5	$50,000 to $74,999	588	16.9%
6	$75,000 to $99,999	205	5.9%
7	$100,000 and over	180	5.2%

Variable name: C15a_2_category

Variable label: Pay-amount-job 2 - categorical
Variable type: Numeric (categorical)

Number of distinct values: 7
Total number of cases: 3,476
Cases with missing values / total number of cases (percent): 866 / 3,476 (24.9%)

Value	Label	Frequency	Percentage
1	Under $10,000	115	3.3%
2	$10,000 to $19,999	460	13.2%
3	$20,000 to $39,999	1130	32.5%
4	$40,000 to $49,999	257	7.4%
5	$50,000 to $74,999	368	10.6%
6	$75,000 to $99,999	135	3.9%
7	$100,000 and over	145	4.2%

Variable name: C15a_3_category

Variable label: Pay-amount-job 3 - categorical
Variable type: Numeric (categorical)

Number of distinct values: 7
Total number of cases: 3,476
Cases with missing values / total number of cases (percent): 2,039 / 3,476 (58.7%)

Value	Label	Frequency	Percentage
1	Under $10,000	55	1.6%
2	$10,000 to $19,999	262	7.5%
3	$20,000 to $39,999	590	17.0%
4	$40,000 to $49,999	145	4.2%
5	$50,000 to $74,999	216	6.2%
6	$75,000 to $99,999	79	2.3%
7	$100,000 and over	90	2.6%

Variable name: C15a_4_category

Variable label: Pay-amount-job 4 - categorical
Variable type: Numeric (categorical)

Number of distinct values: 7
Total number of cases: 3,476
Cases with missing values / total number of cases (percent): 2,797 / 3,476 (80.5%)

Value	Label	Frequency	Percentage
1	Under $10,000	33	0.9%
2	$10,000 to $19,999	121	3.5%
3	$20,000 to $39,999	276	7.9%
4	$40,000 to $49,999	73	2.1%
5	$50,000 to $74,999	93	2.7%
6	$75,000 to $99,999	37	1.1%
7	$100,000 and over	46	1.3%

Variable name: C15a_5_category

Variable label: Pay-amount-job 5 - categorical
Variable type: Numeric (categorical)

Number of distinct values: 7
Total number of cases: 3,476
Cases with missing values / total number of cases (percent): 3,178 / 3,476 (91.4%)

Value	Label	Frequency	Percentage
1	Under $10,000	14	0.4%
2	$10,000 to $19,999	49	1.4%
3	$20,000 to $39,999	106	3.0%
4	$40,000 to $49,999	38	1.1%
5	$50,000 to $74,999	57	1.6%
6	$75,000 to $99,999	21	0.6%
7	$100,000 and over	13	0.4%

Variable name: C15a_6_category

Variable label: Pay-amount-job 6 - categorical
Variable type: Numeric (categorical)

Number of distinct values:	7
Total number of cases:	3,476
Cases with missing values / total number of cases (percent):	3,345 / 3,476 (96.2%)

Value	Label	Frequency	Percentage
1	Under $10,000	1	0.0%
2	$10,000 to $19,999	18	0.5%
3	$20,000 to $39,999	50	1.4%
4	$40,000 to $49,999	20	0.6%
5	$50,000 to $74,999	23	0.7%
6	$75,000 to $99,999	13	0.4%
7	$100,000 and over	6	0.2%