

Example 2 - Multi-variable recode for microdata

Other features illustrated: Using percent buttons, sorting on a column, creating a time series graph from a table cell(s), creating a table that averages across time, creating a table that shows data over time, turning off/on Automatic Totals, cross-variable explosion, nesting vs. not nesting

Background: In microdata datasets, you can create new variables with values based on the combinations of multiple variables' values.

Goals: 1) Determine the percentage of people with and without health care coverage by age groups. 2) Examine health insurance coverage numbers over time.

Steps:

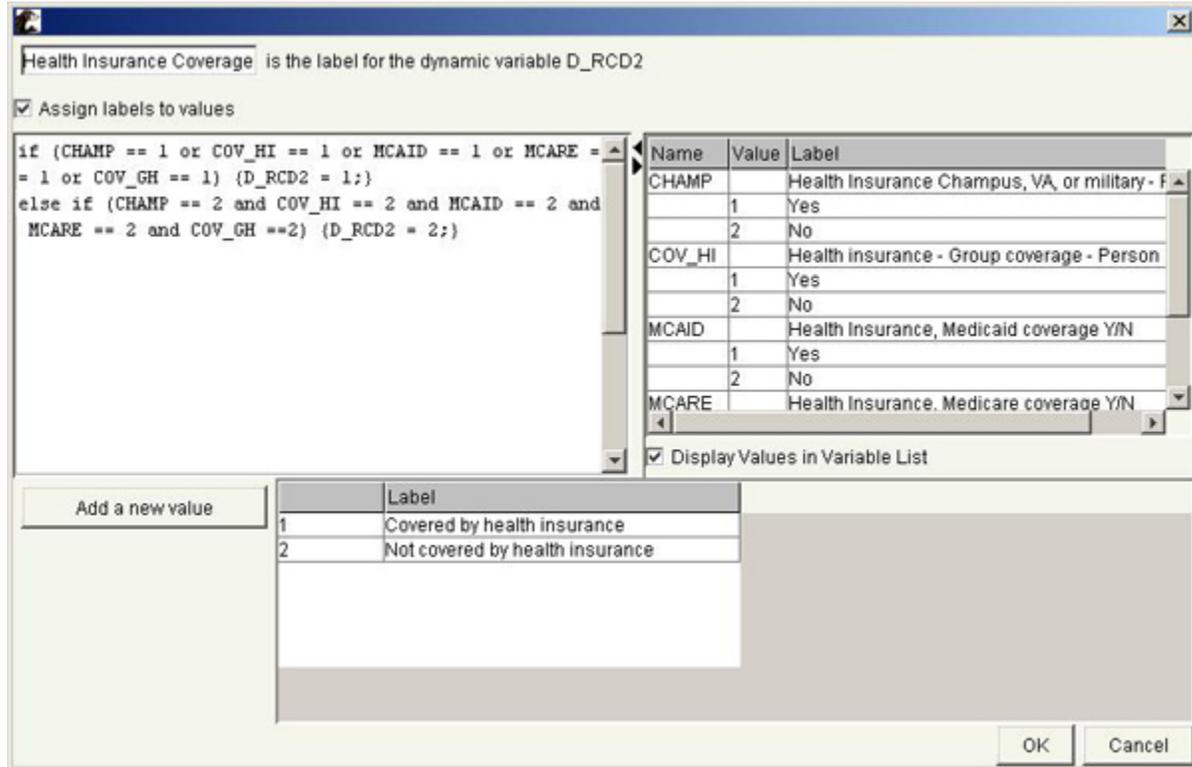
Goal 1 Determine the percentage of people with and without health care coverage by age groups.

1. Open the Ferrett Session File (or saved data basket) named [cps_mar_healthins_multi_databasket.fsf](#).
2. Go to the Step 2 tab to view the variables in the data basket. There are five health insurance variables that are required to determine if someone has some sort of health insurance coverage:
 - o CHAMP - Health Insurance Champus, VA, or military - Person
 - o COV_HI - Health insurance - Group coverage - Person
 - o COV_GH - Health insurance group coverage, including depend
 - o MCAID - Health Insurance, Medicaid coverage Y/N
 - o MCARE - Health Insurance, Medicare coverage Y/N

For a person to have any coverage, the answer to any one of these must be "Yes" and for a person to NOT have coverage, the answer to ALL of these must be "No." We can create a multi-variable recode that does this for us so that we have one new variable that shows whether a person has coverage or not.

3. Highlight the "D_RCD2 Health Insurance Coverage" variable and click on the "Modify" button on the right-hand side. This brings up the multi-variable recode window and shows us the definition of this variable. See the image below:

Screenshot of multi-variable recode for microdata variables:



4. Close the multi-variable recode window by pressing the "Cancel" button at the bottom right.
5. We will now create a new multi-variable recode defined the same way.
 - o Highlight the five health insurance variables listed above (ctrl-mouse click combination), then press the "Create Multi-variable Data Step" button on the right-hand side. This brings up the multi-variable recode dialog window.
 - o Give your new variable a label in the text field at the upper left, something like "Health insurance coverage - Y/N".
 - o Our new variable will have two values - 1=Yes, covered; and 2=No, not covered - and these are defined by the "if" and "else if" statements in the formulas.
 - o To view the values of each variable, check the "Display Values in Variable List" box on the right side below the list of variables.
 - o Enter the formula in the text area exactly as shown in the image above, except you MUST change the D_RCD2 to the name of your new variable shown at the top where you gave it a label.
 - o To assign value labels to your new variable's values, check the "Assign labels to values" box directly above the text area. This will then reveal an area at the bottom for defining value labels.
 - o Click in the "label" area for value 1 and enter "Covered by health insurance".
 - o Click the "Add a new value" button to get a line for value 2.
 - o Click in the "label" area for value 2 and enter "NOT covered by health insurance".
 - o Click the "Ok" button. You now have a new variable in your data basket.
6. Press the "Make a Table" button. This brings up the DataFerrett spreadsheet window.
7. Drop the "RECODE3 Age groups" variable into column 1 to define the rows.
8. Drop our new dynamic recode into C2,R1 to define the columns and press the "Go Get Data" button.

9. Show the percentages of each age group covered and not covered by clicking on the "Show % of first data column" button in the toolbar.
10. Sort the results by column 4 - not covered by health insurance - to see which age groups have the highest and lowest percentages.
11. Turn off sorting and turn off the percentages.

Goal 2 Examine health insurance coverage numbers over time.

1. Create a time series graph - highlight four cells in column 4 - not covered, 4 age groups - then press the Time series Graph button in the toolbar. This automatically creates a time series graph of the same measure going back in time as far as possible, up to 12 time periods.
2. Show the same table, but show the average over 5 years, and add the SEX variable to the table.
 - o DO NOT CLOSE the spreadsheet window, but go to your main DataFerrett window using your Windows taskbar.
 - o Go to the Step 1 tab and notice that there are several years of the March supplement available and currently just Mar 2007 is highlighted. Highlight all available years using the ctrl-click or shift-click combinations.
 - o Now go back to the tabulation window from your taskbar.
 - o Drag the SEX variable into column 1 to break out each age group by sex.
 - o Press the "Go Get Data" button. By looking at the universe and dataset information in the spreadsheet (blue text in the middle of the right-hand side), you can see how many years are listed for the "Dataset(s) selected".
 - o Go to the "Options" menu and you will see that "Average across time" is selected. You could change it to "Cumulative" if you wanted to see the cumulative totals over all years.
3. Create a table that shows the health insurance coverage for EACH of the available years.
 - o Close the current spreadsheet window.
 - o From the Step 2 tab, press the "Make a table" button.
 - o Notice that there is now a variable named "Instances" at the top of your variable list in the spreadsheet. This variable appears when you have more than one instance highlighted in the dataset list and you open a spreadsheet window.
 - o Drag your multi-variable recode into column 1 to define the rows.
 - o Turn off the automatic totals - go to "Options" and click on the "Automatic Totals" to de-select that option.
 - o Drag the Instances variable into C2,R1 to show the years in the columns. There is no total for all years because we turned the automatic totals off before dropping this variable.
 - o "Go Get Data".
 - o Highlight the data cells in row 4, not covered by health insurance, and DO NOT include the label cell (C1,R4).
 - o Press the "Graph" button in the toolbar (looks like a bar graph). The initial graph window that opens will show a bar graph by default.
 - o Using the drop-down list at the bottom left, select "Line Graph" to essentially create a different time series graph.
 - o Since the most recent time is on the left side, you can change it to the opposite by selecting the "Edit" menu and selecting "Reverse Axis".

Additional Table Manipulation

- Close the spreadsheet window.
- In the Step 1 tab, select only Mar 2007.

- Step 2, "Make a Table"
- Health insurance recode in column1.
- Age recode in C2,R1 to define columns.
- Nest SEX in the rows on top of health insurance.
- Turn off "Automatic Totals".
- Drop the race variable in the rows, but not nested - drop two lines below current row definitions.
- Turn on "Automatic Totals".
- Turn off "Cross variable explosion" and nest the Hispanic variable on the race - wpress only category.