

Example 7 - Calculating medians (or any percentile) on-the-fly

Other features illustrated: Column formulas and IF conditions

Background: The DataFerrett spreadsheet allows you to calculate medians (or any percentile) on-the-fly by creating your own bins (microdata) or utilizing income distribution counts (aggregate data), and using standard formulas to calculate the linear percentile you have defined within a given distribution.

Goal: Create a table that calculates the median income for age groups (for the population 15 and over) using the March CPS supplement (Annual Social and Economic Supplement).

Steps:

1. Open the Ferrett Tabulation File named [cps_income_percentile+mean.ftf](#).
2. "Go Get Data" to see the tabulation.
3. This tabulation calculates the median (the 50th percentile) which is defined in column 2 (C2) as 0.5. This value can be changed to any percentile you want (number greater than 0 and less than 1, e.g 0.25 for first quartile). To change the value, click in the gray C2 column header. This will show the formula in the formula bar. Edit the formula.
4. This table uses the 2006 March file (which reports income for 2005). To calculate for a different year, leave the spreadsheet window open and go to Step 1 in the DataFerrett window to select a different year. Then go back to the spreadsheet window and "Go Get Data".
5. You can also calculate for different ages or other variables. To do so, you either select new variables or create new recodes (while the spreadsheet window is open). New variables or recodes will become available in the spreadsheet window. Then, go to "Edit - Clear > All rows" to clear the existing age variable. Then drop your variable(s) into column 1 and "Go Get Data".
6. This table's universe eliminates income values of zero through the Advanced SQL Option in the Step 2 tab of DataFerrett. If you wanted to include these values, or if you wanted to determine the medians for only positive income (current table includes negative income), you can do so in the Advanced SQL Option. The existing SQL universe is
(AGE1 in (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17) and ((PTOTVAL between -999999 and -1) or (PTOTVAL between 1 and 999999)))
If you wanted to include the zero PTOTVAL values, you could simply remove everything related to PTOTVAL like:
(AGE1 in (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17))
Or if you wanted to use only positive income values, change the universe to: (AGE1 in (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17) and (PTOTVAL between 1 and 999999))

NOTE: These percentile calculations are not statistically accurate and will not match published numbers exactly, although they should be within the confidence interval.