Title: Extracting Custom Cross-tabs for Travel Demand Modeling Data with the PUMS 2000 DVD-ROM Beyond 20/20 Software

Mike Jaffe
Mid-Willamette Valley Council of Governments
105 High Street, S.E. Salem 97301
(503)-588-6177

Introduction: Travel demand modeling requires that input of cross-classifications of socio-demographic variables, such as household size, income, and age of head of household (HIA). This data is needed at a regional level in order to estimate it at smaller geographic levels, such as census tracts or transportation analysis zones (TAZs) used in travel demand models.

The Census can provide distributions for each variable, but does not provide customized three-way joint distributions. However, the PUMS 1-percent or 5-percent data can be manipulated to construct custom cross-tabulations using the software Beyond 20/20, which is included on the 2000 PUMS DVD. This tutorial shows the steps for creating a three-way joint distribution for PUMAs in Oregon, including recoding existing fields to new categories (aka "bins").

Steps

1. Opening the PUMS file.

   Start the DVD Census 2000 PUMS software.

   Click "Select File Set".

   Pick 5-Percent Sample.

   Pick the Person Unit Records – click OK.

   Pick "Oregon" from scrolling list.

   The Beyond 20/20 Browser software will start. The browser basically has a cross-tab template which you drag and drop fields, then view the data. The fields are on the right side of the screen, in alphabetical order, first the household fields followed by the person fields.

2. To create the HIA (household size, income, and age of head of household) crosstabs using the custom categories needed, we can't just drag the fields into the template. Instead, we have to recode the data into specific bins, shown later.

3. Find the field (aka "dimension") in the list called "p_relate" and drag the field to the top row of the cross-tab template (from here on, just called template). It will turn yellow and say "p_relate: To…"

4. Find the field h_persons and right click with mouse. From the pop-up, choose Define Recode… A dialog box will warn you that you must save a working copy of the extract before you can modify the extract. Click Ok. Then save a copy to the hard drive with whatever name you like (e.g. 2000PUMScopy.IVX). After the file is saved, a dialog box will pop up called "Define Recode". Give the New Source Field Name a name you want (e.g. hhsize). Next, give the first bin a name in the New Code box (e.g. "Vacant"), select the available code of "00 Vacant Unit" in the lower box and click "Add-->" Type the next code of 1-person in the New Code box, pick the Available Code "01 Householder living alone" and click "Add-->" (Don't be concerned that as you add a new code, the previous code no longer appears; it is saved.) Continue to recode the hhsize field as shown below:

New Code	**Available Codes**
Vacant	00 Vacant
1-person	01 Householder living alone
2-person	02
3-person	03
4+ person	04 to 17

Once all the available codes have been re-coded, you can click OK to the Define Recode dialog box. Click OK when it asks if you want to create a new field.

5. Drag the new field "hhsize" (which will be at the bottom of the "h" fields and before the "p" fields) to the second row of the template. The recoded hhsize field will appear as columns with the headings "Total" 1-person 2-person 3-person 4+ person.

6. The first row heading will be the 5-percent PUMA areas. Fine the field "h_puma5" to the column under hhsize. It will look like this table below, with row and column headings but no data yet

P_relate: **To…**						
hhsize H_puma5	**Total**	1-person	2-person	3-person	4+ persons	Vacant
Total						
4100100						
4100200						
Etc.						

At this point, you may want to hide some of the PUMA area. If so, right click a PUMA and click hide (if you make a mistake, right click and pick Show All)

7. Next you have to recode the household income field (h_hinc) to the bins you want. Here's how I did it:

New Source Field Name: **h_hinc_1**

New Code	Available Codes
< $15K	01 – 11
$15-30K	12 – 14
$30-60K	15 --19
>$60K	20 -- 31

8. This next part is a little tricky. You want to drag the h_hinc_1 field just to the right of the PUMS field. You know you've hit the right spot when the vertical line turns black. Once you've dropped it in the right spot, it will nest within the PUMAs, like this

P_relate: **To…**							
	hhsize	**Total**	1-person	2-person	3-person	4+ persons	Vacant
H_puma5	H_hinc_1						
Total	Total						
	< $15K						
	$15-30K						
	$30-60K						
	>$60K						
4100100	Total						
	< $15K						
	$15-30K						
	$30-60K						
	>$60K						

(Note: If you happen to drag the wrong field to the template, you can drag it back to the list of fields to remove it)

9. Next recode the p_age field to the bins you want.

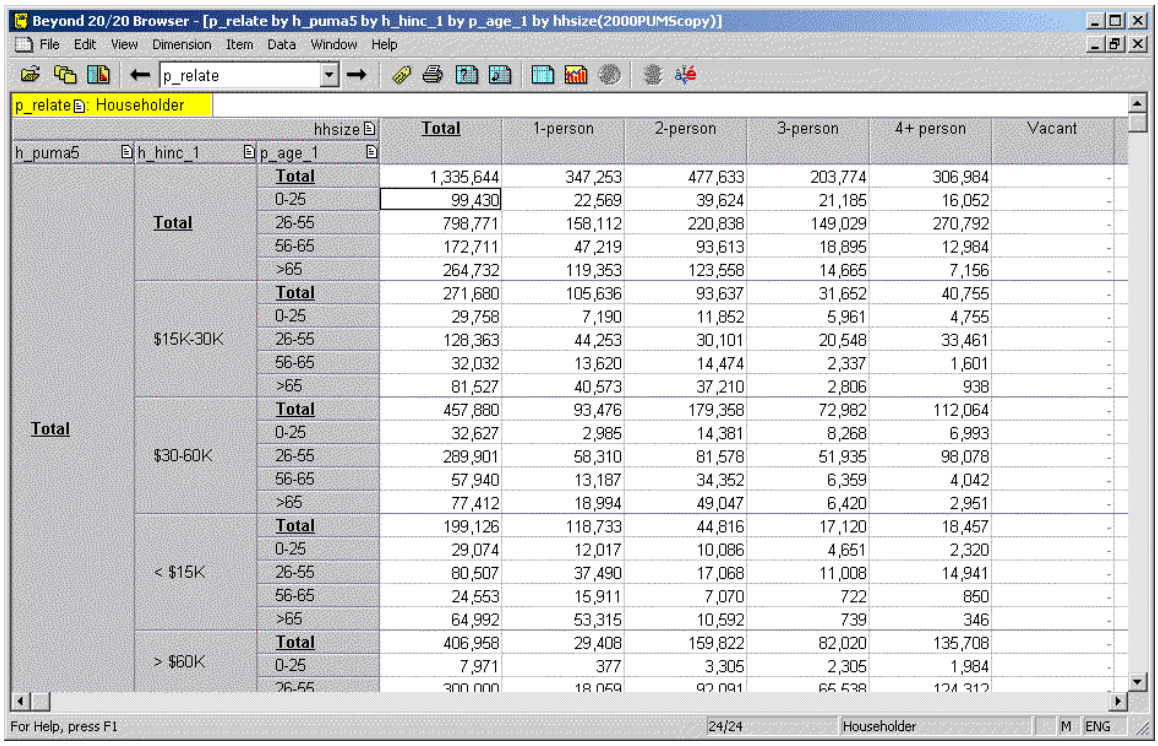
New Source Field Name: p_age_1

New Code	Available Codes
0 – 25	01 – 25
26 – 55	26 – 55
56 – 65	56 - 65
> 65	66 - [last code]

10. Like you did the income, drag the p_age_1 field to the left of the h_hinc_1 column. This completes the cross-tab template.

11. To compute the data, click on the icon that looks like a traffic signal; it's on the toolbar. You'll see some status boxes as it calculates the data.

12. Now, for the amazing part that will let you see data only for the householder, and therefore the age category will be "age of head of household". On the toolbar is a dropdown box with arrows left and right. It contains the names of the fields in your template. Choose the field "p_relate", then click the right arrow once to select only Householders. This will change all the data to only the age of the householder. For the entire state (i.e. PUMA of total), the upper-right cell of the data will show 1,335,644 households, which is a close match to the total households for the state of Oregon. **You now have an HIA cross-tab of household size, income, and age of head of household, for all or any subset of PUMAs, as shown in the screen shot below.**



h_puma5	h_hinc_1	p_age_1	hhsize: Total	1-person	2-person	3-person	4+ person	Vacant
Total	Total	Total	1,335,644	347,253	477,633	203,774	306,984	-
		0-25	99,430	22,569	39,624	21,185	16,052	-
		26-55	798,771	158,112	220,838	149,029	270,792	-
		56-65	172,711	47,219	93,613	18,895	12,984	-
		>65	264,732	119,353	123,558	14,665	7,156	-
	$15K-30K	Total	271,680	105,636	93,637	31,652	40,755	-
		0-25	29,758	7,190	11,852	5,961	4,755	-
		26-55	128,363	44,253	30,101	20,548	33,461	-
		56-65	32,032	13,620	14,474	2,337	1,601	-
		>65	81,527	40,573	37,210	2,806	938	-
	$30-60K	Total	457,880	93,476	179,358	72,982	112,064	-
		0-25	32,627	2,985	14,381	8,268	6,993	-
		26-55	289,901	58,310	81,578	51,935	98,078	-
		56-65	57,940	13,187	34,352	6,359	4,042	-
		>65	77,412	18,994	49,047	6,420	2,951	-
	< $15K	Total	199,126	118,733	44,816	17,120	18,457	-
		0-25	29,074	12,017	10,086	4,651	2,320	-
		26-55	80,507	37,490	17,068	11,008	14,941	-
		56-65	24,553	15,911	7,070	722	850	-
		>65	64,992	53,315	10,592	739	346	-
	> $60K	Total	406,958	29,408	159,822	82,020	135,708	-
		0-25	7,971	377	3,305	2,305	1,984	-
		26-55	300,000	18,059	92,091	65,538	124,312	-

13. To save the data, use File Save As… If you save the table (as an .ivt file), you can reopen the file with all your re-codes and data. You can also choose other formats for exporting data to other programs (Excel is what I used).

Acknowledgement: Thanks to Cam McIntosh, Information Specialist at the Seattle offices of U.S. Census Bureau, for providing me the steps to create the following tutorial

More info:

1. Link to the PUMS TechTalk is http://censtats.census.gov/techdoc/techtalk.html
2. For a method that uses the R-scrpting language, go to http://www.odot.state.or.us/tddtpau/R.html and look for a file name 2000 PUMS data for Oregon Travel Demand Models