

---

# Public User Microdata Samples (PUMS): Do-It-Yourself Census Data

*by Sylvia Andrews*  
*The Indiana State Data Center*  
*Indiana State Library*

Census data users can now join the ranks of the do-it-yourselfers and create customized demographic tabulations and research data with the Quicktab Software on the PUMS CD-ROM.

If this sounds too good to be true, you have probably used the Summary Tape Files on CD-ROM and are familiar with the constraints of the predefined tables. Although these tables provide a wealth of data from the 1990 census short and long form questionnaires, they do not allow optimum access to the detail and variety of census information available. Users cannot always access data by the characteristics needed. The advantage of PUMS is that users can rely on the huge sample size of the 1990 census to allow them to use smaller amounts of more easily manipulated data and choose their own variables.

What is PUMS? It has nothing to do with plays on words, French apples, English ale houses, or humming a song when you don't know the words. First, it will be useful to define PUMS and give a short history of its development. It stands for Public User Microdata Samples and contains records representing five-percent or one-percent samples of the housing units in the United States and the people in them. There will also be a special three-percent sample dealing specifically with the elderly population. Selected people in group quarters are included. The sample is expandable to the entire population by applying the person and housing unit weights provided.

Geographic coverage is for states and some of their geographic levels. The five-percent sample is basically a county level file and the one-percent sample is a metropolitan area file. It is important to remember when using PUMS that while PUMS five-percent areas are based on counties, they may be groups of counties or parts of counties. The PUMA or Public User Microdata Area is an area of at least 100,000 people but not more than 200,000 people. None of the five-percent PUMA's cross state lines. Maps are available for most states. To see the components of each area and the corresponding PUMS codes, refer to the PUMAS.DOC file on the CD-ROM and the maps in the technical documen-



---

teristics of households using certain products. Because of its versatility, PUMS has applications in a wide variety of settings from libraries to academic research institutions to sales and marketing organizations.

The software included on the PUMS disk is called Quicktab. There are two components needed for Quicktab, a data file and a data dictionary. "Frequency" and "Crosstab" are the Quicktab programs that do frequency counts and cross tabulations. These programs can perform either weighted or unweighted counts.

### **Installing Quicktab and Extracting PUMA Files**

1. What do you need to do? First, copy the software from the PUMA's and Quicktab subdirectories on the CD-ROM to the hard drive. They are located in the appropriately named "tools" directory. Put all of the files in corresponding PUMA's and Quicktab subdirectories in a directory you create on the hard drive which you can call "tools." You should also make another directory on the hard drive at this time for your PUMS area files. (Referred to as PUMSGEO below.)

2. Next, you must use the PUMA2DSK command to extract a PUMA-level PUMS file to your hard drive. You must know the PUMA code for the area you are interested in. These are in PUMSA.DOC in the CD-ROM PUMA subdirectory. To use PUMA2DSK, change directories to the directory with PUMA2DSK on your hard drive. If the directory you created on the hard drive is named "tools," change to C:\TOOLS\PUMAS and then to D: or the drive where your CD-ROM is located. Then at the D: prompt type D:\C:PUMA2DSK IN A 00107 C:\PUMSGEO. This will copy an Indiana PUMS area to the C drive and PUMSGEO directory where it can be used by Quicktab.

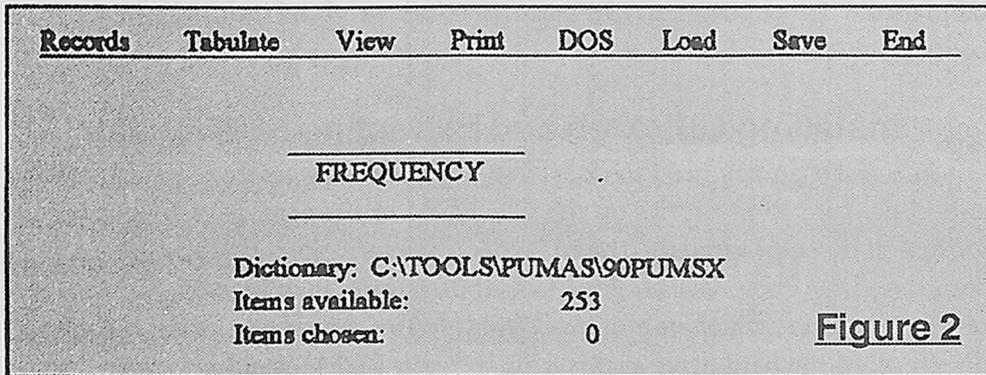
3. Third, download the most current PUMS files from the Census Electronic Bulletin Board (or request on disk from the Census Bureau) including the new data dictionary, 90PUMSX.DD. You can put them in the C:\TOOLS\QUICKTAB subdirectory.

4. These three steps will enable you to use Quicktab, but the last and most important step is to read the technical documentation. PUMS users should be aware of changes in definitions, techniques of error estimation, and the topcodes and should look over the appendixes which include notes on selected data items and maps. Pay particular attention to the information on verifying data, accuracy of microdata sample estimates, and comparing esti-

mates from PUMS to other census data.

## Using Quicktab

Running Quicktab is a fairly simple matter of following the menu, choosing the appropriate files, and adding the housing weight or person weight. (See Sample Menu: Figure 2)



1. To do a frequency distribution report, start by typing "freq" at the DOS prompt in the quicktab subdirectory (e.g.: C:\TOOLS\QUICKTAB >FREQ). Enter the path and name of the new data dictionary, C:\TOOLS\QUICKTAB\90PUMSX.DD (or enter the path C:\TOOLS\QUICKTAB and hit F2 to choose from file list). Choose "records" from the menu. PUMS has two record types, housing and person. Select H-record and press "enter." Highlight the variable you are interested in such as Units1 (units in structure) and press "enter." Next select a second variable such as tenure. Several more variables may be chosen. Descriptions of variables are in the technical documentation. When finished choosing variables hit "escape."
2. Choose P-items the same way.
3. To generate the report choose the "tabulate" option from the menu. Enter the path and data file name or just enter the path and hit function key F2 to get a list of the files. (e.g., X:\, F2 where X is the CD-ROM drive where the PUMS disk is located.) If you have used PUMS2DSK to transfer the area files to another directory on the hard drive use that path. (e.g.: C:\PUMSGEO, F2)
4. Enter the report name or use the default. After you enter the report name, the tables are executed and appear on the screen. A row is

generated for each of the defined values and an additional row for undefined values.

5. For weighted frequency counts move the cursor to the fifth column under "weight" on the housing or person record window and press "enter." From the list choose "houswgt" which is the weight variable for H-Record.
6. Repeat the process for the p-record choosing "pwgt1." Press "escape" to return to the frequencies menu.
7. Tabulate the data, and you will notice that all of the frequencies have increased. These are the corresponding "weighted" frequencies to the counts produced in the previous report.
8. To define the universe or the characteristics of the report population, go to the fourth column and press "y". A window appears which allows you to choose conditions and link them with logical connectors. "And"

**Figure 3a**

Data file: C:\PUMSGEO\IN00101A.TXT

Record: HOUSING-RECORD

Universe: BEDROOMS = THREE

Weight: HOUSWGT

Item: UNITS1

Values	Frequency				Cumulative	
	Total	Percent	% Def.	% Valid	Total	Percent
MOBILE-TRAILER.....	16	.1	.1	.1	16	.1
DETACHED.....	16980	81.3	81.3	81.3	16996	81.4
ATTACHED.....	1697	8.1	8.1	8.1	18693	89.5
APARTMENT-2.....	123	.6	.6	.6	18816	90.1
APARTMENT-3-4.....	413	2.0	2.0	2.0	19229	92.0
APARTMENT-5-9.....	1179	5.6	5.6	5.6	20408	97.7
APARTMENT-10-19.....	311	1.5	1.5	1.5	20719	99.2
APARTMENT-20PLUS.....	54	.3	.3	.3	20773	99.4
BOAT.....	-	-	-	-	20773	99.4
OTHER.....	117	.6	.6	.6	20890	100.0
NA.....	-	-	-	-	20890	100.0
Undefined.....	-	-	-	-	20890	100.0

Item: TENURE

Values	Frequency				Cumulative	
	Total	Percent	% Def.	% Valid	Total	Percent
MORTGAGE-OWNED.....	13060	62.5	62.5	64.9	13060	62.5
CLEAR-OWNED.....	4285	20.5	20.5	21.3	17345	83.0
RENTED-FOR-CASH.....	2679	12.8	12.8	13.3	20024	95.9
RENTED-NO-CASH.....	98	.5	.5	.5	20122	96.3
NA.....	768	3.7	3.7	-	20890	100.0
Undefined.....	-	-	-	-	20890	100.0

**Figure 3b**

Data file: C:\PUMSGEO\IN00101A.TXT

Record: PERSON-RECORD

Universe: MARITAL = DIVORCED

Weight: PWGT1

Item: MARITAL

Values	Frequency			Cumulative	
	Total	Percent	% Valid	Total	Percent
NOW-MARRIED.....	-	-	-	0	.0
WIDOWED.....	-	-	-	0	.0
DIVORCED.....	11417	100.0	100.0	11417	100.0
SEPARATED.....	-	-	-	11417	100.0
NEVER-MARRIED.....	-	-	-	11417	100.0
Undefined.....	-	-	-	11417	100.0

Item: RAGECHLD

Values	Frequency				Cumulative	
	Total	Percent	% Def.	% Valid	Total	Percent
UNDER6.....	186	1.6	1.6	2.5	186	1.6
SIX-17.....	1600	14.0	14.0	21.8	1786	15.6
UNDER6-AND-6-17.....	222	1.9	1.9	3.0	2008	17.6
NONE.....	5334	46.7	46.7	72.7	7342	64.3
NA.....	4075	35.7	35.7		11417	100.0
Undefined.....	-	-	-		11417	100.0

connectors take precedence over "or" connectors. Conditions with inequalities and ranges of values can also be chosen. Escape and tabulate as previously.

9. You can save the selections you have made to the disk by using the "save" option from the menu. To use the selections later highlight the "load" option on the menu.

Crosstab uses essentially the same steps to produce tables rather than frequencies. Type "crosstab" at the prompt. (e.g., C:\TOOLS\QUICKTAB>CROSSTAB). There can be a maximum of 120 rows and 20 columns. Print the tables using "print" on the menu. (See Figures 3a-3c for examples of frequency and crosstab tables showing h-records and p-records, using weights, and limiting the universe.) The housing-record tables show units in the structure and tenure with universe = three bedroom units, and weight = housing weight. The person-record tables show marital status and presence and age of own children with universe = divorced, and weight = person weight.

Some interesting questions have been answered using Quicktab on the

**Figure 3c**

Data file: C:\PUMSGEO\IN00102A2.TXT

Record: HOUSING-RECORD

Universe: BEDROOMS = THREE

Weight: HOUSWGT

Count: UNITS1 by TENURE

	Total:	MORTGAGE:	CLEAR-OW:	RENTED-F:	RENTED-N:	NA:Undef
Total.....	20890:	13060	4285	2679	98	768
MOBILE-TRAILER.....	16:	16	-	-	-	-
DETACHED.....	16980:	12072	3969	335	35	569
ATTACHED.....	1697:	799	244	548	25	81
APARTMENT-2.....	123:	7	-	100	-	16
APARTMENT-3-4.....	413:	39	6	355	-	13
APARTMENT-5-9.....	1179:	102	-	1004	38	35
APARTMENT-10-19.....	311:	-	19	263	-	29
APARTMENT-20PLUS.....	54:	-	-	29	-	25
BOAT.....	-:	-	-	-	-	-
OTHER.....	117:	25	47	45	-	-
NA.....	-:	-	-	-	-	-
Undefined.....	-:	-	-	-	-	-

Page 2 07/02/94 09:00:06

IMPS Version 3.0

Data file: C:\PUMSGEO\IN00102A2.TXT

Record: PERSON-RECORD

Universe: MARITAL = DIVORCED

Weight: PWGT1

Count: MARITAL by RAGECHLD

	Total:	UNDER6:	SIX-17:	UNDER6-A:	NONE:	NA:Undef
Total.....	11417:	186	1600	222	5334	4075
NOW-MARRIED.....	-:	-	-	-	-	-
WIDOWED.....	-:	-	-	-	-	-
DIVORCED.....	11417:	186	1600	222	5334	4075
SEPARATED.....	-:	-	-	-	-	-
NEVER-MARRIED.....	-:	-	-	-	-	-
Undefined.....	-:	-	-	-	-	-

PUMS CD-ROM at the Indiana State Data Center including:

- ✓ What is the Russian population over age 16 in Indianapolis, and what language do they speak at home?
- ✓ What are the labor force characteristics such as employment and unemployment of people in Elkhart by level of educational attainment?
- ✓ How many female heads of household are there in Marion County and at what income levels?

The PUMS CD-ROMs are available from the Census Bureau for \$150 per disk or \$700 for the entire set (seven disks) which includes PUMS five-percent samples for all the states. PUMS one-percent disks are also \$150 per disk or \$200 package price (two disks). For ordering information, contact Census

---

Bureau Customer Services at (301)763-4100.

In summary, the Quicktab software on the PUMS CD-ROM is very useful in a reference or data center context for questions relating to areas of 100,000 or more when other census CD-ROM products do not provide the data in the required combinations. Questions can be answered in a timely manner with a very reasonable coefficient of variance (4.4 percent) between the PUMS estimate and the full census estimate. Additional software is not required when loaning disks to users, and Quicktab is very user friendly. Users have been very happy with the additional flexibility. So good luck in plumbing the data depths, no puns intended.

### **Sources**

1. Bureau of the Census, *Census of Population and Housing, 1990: Public User Microdata Samples: Technical Documentation*. Washington, D.C.,