News on the latest products from the U.S. Bureau of the Census

1990 Public Use Microdata Samples (PUMS)

"What a find — pick your variables and cross-tabulate them become your own Census Bureau!"

You can design tabulations the way you want with Public Use Microdata Samples (PUMS) available on computer tape and on CD-ROM. Think of these files as a special do-ityourself tabulation kit! In this profile we explain what PUMS are, describe some of the applications, and give you the information you need to order these valuable files from the Census Bureau.

What Are Microdata?

Microdata files from the 1990 census show the full range of responses made on individual questionnaires.

You see, for example, how one household or one household member answered questions on occupation, place of work, and so forth. The files contain records for a sample of all housing units, with information on the characteristics of each unit and the people in it.

All identifying information is removed to ensure confidentiality. The records selected are a sample of those households that received the "long-form" questionnaire.

In the 1990 census, about one in every six housing units (approximately 16-percent of all housing units) received the long-form or sample questionnaire, out of a total of 102 million housing units.

A short-form questionnaire containing only the "100percent questions" went to the balance of housing units. Those who received the long-form questionnaire were asked both the sample questions and the 100percent questions.

The sample questionnaire included questions on income, education, language spoken at home, journey to work, occupation, condominium status, shelter costs, vehicles available, and other subjects, as well as 100percent questions on age, sex, tenure, and so forth.

The Census Bureau has produced 1990 PUMS for the United States and those outlying areas which meet a 100,000 minimum population size threshold.

The standard PUMS products are the 5-percent and 1-percent samples for the United States and Puerto Rico, 10-percent samples for Guam and the U.S. Virgin Islands, and a special 3percent sample dealing specifically with the elderly population. The 5-percent and 1-percent samples are similar in content to the "A" and "B" files made available in 1980. For a further discussion of sample size and content, see page 4.

Figure 1. Subjects in the 1990 Census PUMS

Items in the housing record include:

Bedrooms Condominium status Contract rent (monthly rent) Cost of utilities and fuels Family income in 1989 Family, subfamily, and household relationships Farm status and value Fire, hazard, and flood insurance Fuels used Gross rent House heating fuel Household income in 1989 Household type Kitchen facilities Linguistic isolation* Meals included in rent Mortgage status and selected monthly owner costs Plumbing facilities Presence and age of own children Presence of subfamilies in household Property value Real estate taxes Residence State, PUMA** MSA, Division

Sewage disposal Source of water Telephone in housing unit Tenure Units in structure Vacancy status Vehicles available Year householder moved into unit Year structure built

Items in the person record include:

Ability to speak English Ancestry Children ever born Citizenship Class of worker Disability status Educational attainment Hispanic origin Hours worked Income in 1989 by type Industry Language spoken at home Last week work status Marital status Means of transportation to work

Military status, periods of active duty military service, veteran period of service Mobility status Occupation Personal care limitation Place of birth Place of work State Place of work PUMA**

Migration PUMA**

Migration State

Poverty status in 1989 Relationship School enrollment and type

of school

Time of departure for work Travel time to work Vehicle occupancy Weeks worked in 1989 Work status in 1989 Work limitation status Year of entry

- Households in which no person, age 14 or over, speaks only English or speaks English very well.
- ** PUMA's are public use microdata areas.

product profile

The full range of population and housing information collected in 1990 census is available in the PUMS: Over 500 occupation categories, age by single years up to 90, wages in dollars up to \$140,000 and over, and so forth. For most questions asked on the long-form questionnaire, the response is given in these files - and you design tabulations to aggregate the responses in ways useful to you.

Summary Data and Microdata — What's the Difference?

Summary data are the type of information found in Census Bureau's printed reports, summary tape files (STF's), and most special tabulations. The basic unit of analysis is a specific geographic entity — census tract, county, State, etc. — for which counts of persons, families, households, or housing units in particular categories are provided.

In microdata, the basic unit is an individual housing unit and the people who live in it. As mentioned above, the record shows all the information associated with a specific housing unit (each group quarters person selected in the sample was considered a housing unit) except for names, addresses, or other identifying information.

It is you, the user, who determines the structure of the tabulation and the characteristics to be tabulated. Only large geographic areas are identified on microdata records — large cities, counties, and metro areas. You will find parts of these areas (actually groupings of census tracts) when they have a population of 100,000 people or more.

Figure 2.

What's the Difference Between Summary Data and Microdata?

Summary Data

- Basic unit is a statistical table for an identified geographic area (e.g., counts of race groups in one county) or a set of areas (e.g., housing values for metropolitan areas).
- Files contain population and housing statistics similar to those found in printed reports, though in more detail.
- Available for large and small geographic areas (census blocks, block groups, tracts counties, States etc.).
- Table structure generally limits number of data items included.

		Occupied	Persons	Renter	Gross Rent		
Place	Total population	housing units	per unit	occupied units	\$100	\$100- \$149	\$150- \$199
Weston City	110,938	49,426	2.2	31,447	158	1,967	6,282
Smithville	21,970	7,261	3.1	2,492	17	90	766
Junction	17,152	5,494	2.7	822	11	29	238

Public Use Microdata

- Basic unit is a record for a housing unit, a household, or a person.
- Shows all of the responses made to questions on the questionnaire.
- Contains geographic codes identifying only very large areas called PUMAs (public
 use microdata areas) that contain at least 100,000 people, such as metropolitan
 areas, county groups, or large groups of census tracts to further ensure confidentiality.
- Allows study of relationships among characteristics of one household (e.g., earnings
 of plumbers under age 30 living in metro areas).

	Housing unit no.	Relation- ship	Sex	Age	Race	Place of birth	Occu- pation	Earnings
Person a	1	House- holder	M	37	W	Kansas	Plumber	\$28,100
Person b	1	Spouse	F	35	W	Virginia	NA	None
Person c	1	Child	M	6	W	Virginia	NA	None

The Census Bureau uses the minimum population threshold to help avoid disclosure of information about any household or individual.

To further protect confidentiality, there is limited detail on items such as place of residence, place of work, high incomes, and other items.

See figure 1, on page 1, for a listing of some of the items in the housing and person records. Also, see figure 2 on page 2 for more on how microdata records differ from the summary data found in STF's.

Why Use PUMS?

For many data users, the STF's and the reports will suffice. Microdata are for those users who want to create do-it-yourself tabulations, to be able to further draw on the richness of detail recorded in the census. (See figure 3 for a comparison of STF detail and PUMS.)

Microdata users frequently want to look at relationships among variables not shown in the standard products offered by the Census Bureau. For example, what are the characteristics of unemployed homeowners? Or what characteristics do families with four or more children have in common? Or what are the characteristics of husbands and wives crosstabulated by educational attainment and income? Or what ethnic group makes the highest income in specified occupations by level of education?

The advantage of PUMS is that users can rely on the vast size of the 1990 census sample to design reliable tabulations for any desired set of variables.

In short, PUMS allow users to tabulate data according to the characteristics they need to know about.

Who Can Use PUMS?

Carmen Campbell, whose staff coordinated the release of these files, notes that more users than ever will find it easier to use PUMS. "In the past, PUMS were used mostly by academic researchers adept at using the appropriate software or at programming for the tabulations they wanted."

"Now, with PUMS on CD-ROM's, microcomputer users will have access to these files. We've put data retrieval software on the discs. And the commercial software is much easier to use than in the past. We envision a whole new generation of users for this product."

PUMS are useful for research that does not require the identification of specific small geographic areas (like a single block or a single tract) or detailed cross-tabulations for small populations.

Users interested in special population tallies can select, for example, those households with one or more disabled persons, households with more boys than girls, or particular income groups who send their children to private schools.

For those who wish to choose their own combination of variables, endless possibilities exist. For example, you can arrange poverty areas by geographic (metro areas) or social (occupation) delimiters or determine what kinds of Hispanic families in a particular geographic area rent or own their own homes.

Figure 3.

1990 Educational Attainment Data STF 3A, STF4A, and PUMS

Summary Tape File 3A: Educational Attainment Persons 25 years and over

Less than 9th grade
9th to 12th grade,
no diploma
High school graduate
(includes equivalency)
Some college, no degree
Associate degree
Bachelor's degree
Graduate or professional
degree

Suppose you're studying the educational attainment of women. You'll get little help in STF 3A. The attainment tabulation is limited here. Also there is no cross-tabulation by sex. Nevertheless, there is great geographic flexibility (small area data) available on the very popular STF 3A.

Summary Tape File 4A: Educational Attainment Persons 25 years and over

No school or less than 1st grade 1st to 4th grade 5th to 8th grade 9th grade 10th grade 11th grade 12th grade, no diploma High school graduate (includes equivalency) Some college, no degree Associate degree in college (occupational program) Associate degree in college (academic program) Bachelor's degree Master's degree Professional school degree Doctorate degree

STF 4 shows educational attainment by sex and repeats tables for race and Hispanic groups, another useful feature. But many users may find this file too large to manage. (Over 10,000 cells/items of sample population and housing characteristics for each geographic area.) Also it comes only on tape.

PUMS: Educational Attainment

No school completed Nursery school Kindergarten 1st, 2nd, 3rd, or 4th grade 5th, 7th, or 8th grade 9th grade 10th grade 11th grade 12th grade, no diploma High school graduate, diploma or GED Some college, but no degree Associate degree in college (occupational program) Associate degree in college (academic program) Bachelor's degree Master's degree Professional degree

Doctorate degree

The PUMS file from the 1990 Census provides the greatest detail on educational attainment. Here are all the attainment levels recorded in the census. What makes the files even more useful is that you can cross-tabulate this information by other variables — sex, income, occupation, language spoken at home, and so forth. PUMS is really for the user who needs access to the full range of detail gathered in the census. It comes on both tape and disc.

Figure 4.

PUMA boundary maps for the 5-percent and 1-percent files come with the technical documentation. Here is an example of a 5-percent PUMA State map.

Alabama 5% PUMA Boundaries



Jefferson County PUMA's: 01801-01806

A five-digit number, unique within the State, identifies each PUMA. The first three digits is the PUMA code and the last two are the sub-PUMA. The sub-PUMA is used when counties or metropolitan areas are sub-divided by groupings of census tracts.

Alabama 5% PUMA Boundaries

Counties

PUMA

PUMS allow the user to do work with correlation or regression analyses leading to projections and predictions. Users in academic life, such as scientists, economists, psychologists, and sociologists have found this PUMS application of great value and interest. For example, users can use modelling to determine how new tax policies will affect certain types of families or households.

PUMS appeals to marketers looking for particular characteristics of the populations using their products, such as incomes of families with adolescent girls in a given geographical area.

PUMS files on CD-ROM or extracts from PUMS files are perfect for those who are looking for greater accessibility to inexpensive data (see order form on page 8 for prices) such as students or small business owners looking for specially targeted markets.

Those in the public sector concerned with public policy issues can use PUMS to determine the environmental or social impact on a particular population in certain occupations.

Three PUMS Samples Are Available

The Census Bureau has prepared three PUMS files from the 1990 census for the U.S. — a 1-percent, a 3-percent, and a 5-percent sample. None of the three samples includes any of the same housing units. Each record in these files also contains a variable that permits selection of scientifically designated subsamples or extracts (for example, a 1-in-1,000 subsample).

The 1-percent PUMS for a given State are drawn from

the full census sample for that State. For example, if the full census sample for a State was 20 percent, then the census sample was divided into 20 subsamples of equal size. One of the 20 subsamples was randomly designated as the 1-percent sample.

From the remaining 19 subsamples, 5 1-percent subsamples were chosen at random and merged to produce the 5-percent PUMS for that State.

The 3-percent sample, referred to as the "elderly sample," is sponsored by other Federal agencies and focuses on the older population. It includes only housing units with at least one person age 60 or older (all other persons in the housing unit are included as well).

It was produced in the same way as the 5-percent, but required an extra step. The 3-percent subsamples were merged and the elderly household and person records, households with at least one person age 60 years or more, or group quarters persons 60 years old and older, were selected and designated as the elderly PUMS file.

Geographic Coverage

• Five-Percent Sample.
Besides the obvious
difference in file size, the
5-percent and the 1-percent
files differ in the geography
around which the files are
constructed. Each sample
has its own public use microdata area (PUMA) as the
lowest level of geography
identified on the file.

The 5-percent is basically a county-level file; that is, the PUMA can be a single county (county equivalent), a group of counties, a place, or county/place parts if the total county or place has more than 200,000 persons. The 5-percent sample records carry a code identifying the PUMA where the housing unit is located. None of the PUMA's in the 5-percent sample crosses State lines. See figure 4 on page 4.

- One-Percent Sample.
 The 1-percent sample is basically a metropolitan-area file. For this, the PUMA is a metropolitan area (MA), groups of MA's, parts of MA's when the metro area is larger than 100,000 persons, or nonmetropolitan areas. In the 1-percent sample, each household record includes codes associating it with either a MA or a nonmetropolitan area.
- Three-Percent Sample. Records in the 3-percent sample include an area designed to service the needs of the elderly called a planning service area (PSA). A three-digit code is designated whenever the PSA has the same boundaries as the 1-percent and 5-percent PUMA already identified in the record. The PUMA has the same geographic boundaries as the 5-percent sample. See figure 4 on page 4.

Further geographic information is provided in the PUMS technical documentation (which comes with the files). It includes maps identifying all PUMA's for each State.

Selecting the Right PUMS

Reliability increases with sample size, so the choice of sample size must represent a balance between the level of precision desired and the resources available for working with microdata files. Interest in county or metro areas will also be a factor. The file you select should be determined by what you plan to do with it. Using tables in PUMS technical documentation, you can estimate how much sampling error will affect any specific number prepared from a microdata file of a particular sample size.

In many instances, you will find the 1-percent sample, or an extract from it, large enough to provide sufficiently reliable estimates. On the other hand, if you contemplate extremely detailed tabulations, or if you are concerned with small segments of the population, like men 65 years and over who are of Polish ancestry, the 5percent sample would be a more reliable choice because of its larger size. You may even combine all the PUMS for a larger sample size, if geography is not an issue.

PUMS Records

There are two basic record types of 231 characters each: the housing unit-record and the person record. Each has a unique identifier. Each of the records contains a serial number that links the persons in the housing unit to the proper housing-unit record. The file is sorted to maintain the relationship between both record types.

The Census Bureau releases the PUMS in this format because of the tremendous amount of data contained in one record. Although these records are extremely large, they can be handled by most statistical or report-writing software. Each record has an individual weight which allows users to produce population estimates close to those in other products showing sample data.

Users Speak

In using PUMS, most analysts pick two or more variables and cross-tabulate them. Academic researchers are routine users. The constraints of summary statistics don't apply. Using these files, you become your own Census Bureau!

In fact, Census Bureau researchers in conjunction with the Russell Sage Foundation are doing just that - producing cross-tabulations of PUMS social and economic data for racial groups. For example, Roderick Harrison of the Bureau's Population Division is working with occupation by race by gender. His concern is with the distribution of occupation, income, and geographic categories as they affect racial diversity.

Rebecca Clark of the Urban Institute is working with the Census Bureau on a joint study using 1980 PUMS files to determine whether the neighborhood is a factor in dropout rates of high school age boys. She is using some tract-level data from 1980 STF files to get neighborhood detail. The study is using the 1980 PUMS files to look at the characteristics of young men, 10 years older than the boys in the major grouping, who returned to school, and is trying to determine whether their income has been affected. Clark says that the PUMS files provide a sufficiently comprehensive number of cases to allow a researcher to study enough boys of a particular ethnic group nationwide.

Jack Beresford, formerly of the Census Bureau, used 1980 PUMS data to estimate the number and characteristics of entrepreneurs in the United States for a study

commissioned by the Small **Business Administration** (SBA). The SBA saw self-employment as an important and growing trend in society and needed data on entrepreneurs to set policy and strategies. The results provided the first information about the number and characteristics of home-based businesses. Beresford adds, "We kept the cost low by using an extract of the census microdata file for tabulations of data about entrepreneurs and by using printed results from the census to get data for total population."

Available on Tapes and CD-ROM

PUMS files are released on 9-track computer tape, cartridge, and CD-ROM. The 1- and 3-percent tapes are sold by State, geographic division, or as a compete set. See page 7 for ordering information. The 1990 PUMS are available in EBCIDIC or ASCII format accompanied by electronic data dictionaries which describe the file format. The data dictionaries can be used with software such as SAS and SPSS to locate data fields within each record.

By far, the most inexpensive option is compact disc. The files for all States fit on seven discs costing \$700. For information on CD-ROM contact Customer Services (301-763-4100).

The technical documentation includes a section on "how to use this file," where software concerns are addressed. Users may also write their own programs enabling them to perform custom tabulations on their system of choice. See page 7 for technical and ordering information.

Other Sources of Assistance

For More Information

- If you would like information concerning the PUMS files, contact— Carmen Campbell Data User Services Division (301-763-2005)
- For ordering information, contact —
 Customer Services
 Census Bureau
 (301-763-4100)
- For information about the Product Profiles series, contact — Molly Abramowitz Data User Services Division (301-763-1584)

State Data Center (SDC) Program

If you need further information on PUMS, or other Census Bureau data products, the State Data Centers (SDC's) may provide just the service you need.

These organizations are not part of the Census Bureau, but they are vital partners in helping us disseminate data. They are generally State or local government agencies, libraries, and academic centers.

Begun in 1978, this program now numbers over 1,400 organizations. SDC's are found in every State, the District of Columbia, Puerto Rico, Guam, and the U.S. Virgin Islands. SDC's receive Census
Bureau data for their States
and surrounding areas and
make the data available to
the public, often at or below
cost. Many provide special
services, some of which are
not available from the
Census Bureau.

The Oregon, New Jersey, and Michigan SDC's, for example, have produced a variety of housing profiles.

The California and Massachusetts SDC's have been heavy users of PUMS. They distribute information requested by the governor, local governments, and academic users.

Another program to check is the Business and Industry Data Center Program (BIDC), an outgrowth of the State Data Center program. The BIDC was set up to better serve the information needs of the business community and is similar in design to the SDC program. The BIDC's help new and existing businesses.

National Clearinghouse for Census Data Services

The Census Bureau maintains a list of businesses that release products and perform services involving Census Bureau data. The list also includes a few government and academic organizations. These groups are enrolled in our National Clearinghouse for Census Data Services. They do

marketing studies, geocoding, and computer mapping; they download tapes, conduct surveys, and serve clients in numerous other ways.

Particularly if you need data accessible only in machinereadable form or if you need a customized tabulation, you may find it useful to consult them.

The Census Bureau releases some software for accessing data on compact disc, but for more sophisticated applications, you can develop your own or consult a vendor such as those listed in the clearinghouse.

Many of these organizations can provide such software.

Call Customer Services for a List

For a list of State Data Centers, the Business and Industry Data Centers, and clearinghouse organizations, contact Customer Services (301-763-4100), and ask for free copies.

Census Regional Offices

If you have questions about any of the Census Bureau's products and services, a regional office is the first place to turn.

Information specialists in the 12 regional offices answer more than 100,000 inquires a year.

These information specialists answer phone inquires, conduct seminars, speak at conferences, and staff exhibit booths

They know the local scene well and may be able to tell you where to turn for other sources of assistance. Call and put these experts to work for you!

404-347-2274
617-565-7078
704-371-6144
312-353-0980
214-767-7105
303-969-7750
313-354-4654
913-236-3711
818-904-6339
212-264-4730
215-597-8313
206-728-5314

1990 Census Public Use Microdata Samples—CD-ROM and Tape Options

(Please indicate choice)

CD-ROM Options		Tape Format Options	Product Pricing		
PUMS 5-Percent \$	3150 per disc	Computer tape reels are industry standard 9 track, 2,400 foot reels.	Priority shipping and handling service is available at additional cost.		
A1 Alabama-CaliforniaA2 Colorado-IdahoA3 Illinois-MarylandA4 Massachusetts-New HampshireA5 New Jersey-North DakotaA6 Ohio-TennesseeA7 Texas-Wyoming(package price) \$700		Density: Translation coding: 6250 BPIASCII1600 BPIEBCDIC Labeling:Standard IBM LabelUnlabeled Or you may select:	Upon receipt, an acknowledgment letter will be sent noting an approximate shipping date. See chart below for ordering tapes.		
B1 Alabama-Montana B2 Nebraska-Wyomir		IBM 3480 Tape Cartridge A machine-readable data dictionary is included without charge.			

Division/State	5%	1%	3% (Elderly)	Division/State	5%	1%	3% (Elderly)
New England Div.	\$275	\$175	\$175	North Carolina	175		
Connecticut	175		5K	South Carolina	175		
Maine	175			Florida	275		
Massachusetts	175			Georgia	175		
New Hampshire	175						
Rhode Island	175			East S. Central Div.	\$300	\$175	\$175
Vermont	175			Alabama	175		
				Kentucky	175		
Middle Atlantic Div.	\$750	\$175	\$450	Mississippi	175		
New York	350			Tennessee	175		
New Jersey	175					A	4225
Pennsylvania	225			West S. Central Div.	\$550	\$175	\$325
	12121212			Arkansas	175		
East N. Central Div.	\$825	\$175	\$500	Louisiana	175		
Indiana	175			Oklahoma	175		
Ohio	225			Texas	350		
Illinois	225			Manual Dis	6375	¢175	6175
Michigan	175			Mountain Div.	\$275	\$175	\$175
Wisconsin	175			Arizona	175		
				Colorado	175		
West N. Central Div.	\$350	\$175	\$225	Idaho	175		
Minnesota	175			Montana	175		
Missouri	175			Nevada	175		
South Dakota	175			New Mexico	175		
Iowa	175			Utah	175		
Kansas	175			Wyoming	175		
Nebraska	175			D 16 D1	4750	4175	6450
North Dakota	175			Pacific Div.	\$750	\$175	\$450
	****	A485	4550	Alaska	175		
South Atlantic Div.	\$925	\$175	\$550	Hawaii	175		
Delaware	175			Oregon	175		
District of Columbia	175			Washington	175		
Maryland	175			California	575	4005	** ***
Virginia	175			All States/divisions	\$4,000	\$800	\$2,400
West Virginia	175			(package price)			

product profile

Customer Typology

To enable us to learn more about our customers, please mark (X):	College/University- Student Other Academic	Trade and Professional OrganizationsOther	Is your organization affiliated with the Census Bureau in any way?		
Private Citizen	Private CitizenResearch Organizations		Please mark (X):		
Governmental Sector Congressional/ Legislative Federal GovernmentState GovernmentLocal/County/Regional GovernmentTribal Organizations Academic/Research SectorCollege/University- Faculty	Business Sector Media (radio,	Public Service SectorSocial Service and Religious OrganizationsCivic OrganizationsNational Minority OrganizationsPolitical Organizations International SectorForeign GovernmentOther	State Data Center/ Business and Industry Data Center/Affiliate National Services Information Center Affiliate Clearinghouse Organization None Do you work in your organization's library? Yes No		
Order Form					
Customer address: (please prin	it legibly or type):	Shipping address: (if different than customer address)			
Company or personal name:		Company or personal name:			
Additional address/attention li	ne:	Additional address/attention line:			
Street Address (No P.O. box numb ping address. The shipping service we		Street Address (No P.O. box numbers please, if this is also the shipping address. The shipping service we use cannot deliver to a P.O. box.)			
City, State, ZIP Code:		City, State, ZIP Code:			
Telephone: Voice F	ax	Telephone: Voice Fax			
To be completed by customer: Method of Payment (Please mark (X): Check payable to Commerce-Census Census deposit account:					
Total this order:	VISA or MasterCard	account:			
All orders must be prepaid. Customer Services					
All orders must be prepaid. Customer Services Bureau of the Census Washington, DC 20233-8300 PUMS6.93 Customer Services Bureau of the Census On card: 301-763-4100					