| IPUMS.ORG

IPUMS USA

HOME | SELECT DATA | MY DATA | FAQ | HELP

DATA
BROWSE AND SELECT DATA
DOWNLOAD OR REVISE EXTRACTS
ANALYZE DATA ONLINE
IPUMS ABACUS *(BETA)*
IPUMS REGISTRATION

DOCUMENTATION
VARIABLES
SAMPLES
USER'S GUIDE
GEOGRAPHIC TOOLS
FAQ

RESOURCES
ENUMERATION FORMS
PUBLISHED CENSUS VOLUMES
ERRATA AND REVISIONS

RESEARCH
CITATION AND USE
BIBLIOGRAPHY
RELATED SITES

CONTACT US
HELP
USER FORUM
IPUMS STAFF
HOW TO HELP
DONATE TO IPUMS

SAMPLE DESIGN AND ESTIMATION IN THE AMERICAN COMMUNITY SURVEY (ACS) AND THE PUERTO RICO COMMUNITY SURVEY (PRCS) [1]

[Go Back to Sampling Procedures Index](#)

This section discusses the sample selection and estimation procedure for the Public Use Microdata Samples (PUMS) extracted from the annual American Community Survey (ACS) and Puerto Rico Community Survey (PRCS). The sample design of the ACS/PRCS (hereafter referred to only as the ACS) approximates the Census 2000 long form sample design and oversamples areas with smaller populations. Each month a systematic sample is drawn to represent each U.S. county or county equivalent. The selected monthly sample is mailed the ACS survey at the beginning of the month. Nonrespondents are contacted via telephone for a computer assisted telephone interview (CATI) one month later. One third of the nonrespondents to the mail or telephone survey are contacted in person for a computer assisted personal interview (CAPI) one month following the CATI attempt. In addition to this national sample (referred to as the "supplementary" sample), ACS collects data at 31 selected test sites that represent different combinations of county population sizes and difficult to enumerate areas. The test sites were selected to provide detailed comparisons with the Census 2000 long form data at the county and smaller geographic levels. Further information about the ACS sample is discussed in the [ACS documention](#)   .

SAMPLE DESIGN

Housing unit records for the ACS PUMS consist of all records from the supplementary sample and a sample of records from the 31 test sites. Persons in the selected occupied housing units constitute the ACS PUMS person sample. The process of selecting housing units from the test sites was performed independently for each state that has test sites (26 states).

The housing units in the test sites were classified into three types: vacant, occupied mail/CATI, and occupied CAPI. Sampling rates were then determined separately for each of these types. Sampling rates were determined based on the size of test site housing unit weights compared to supplementary sample housing unit weights in the same state. Weights of test site housing units are generally smaller than those in the supplementary sample, so it would be easy to identify a test site case in some instances. This is a disclosure risk that we eliminate by choosing sampling rates so that the distribution of PUMS weights for the selected test site housing units is similar to that of supplementary sample housing units so test site housing units with very small weights do not stand out. In most cases, this need necessitated the stratification of the housing units based on their weight. housing units with a weight less than a certain value would be sampled at one rate with the remaining housing units sampled at a higher rate. The table below shows two hypothetical examples of sampling rates for each type of housing unit. In state 1 for example, mail/CATI cases with weight less than 16 are sampled at

a rate of 1 in 10, and the remaining mail/CATI cases are sampled at a rate of 1 in 5. In state 2, all mail/CATI cases are sampled at a rate of 1 in 5, and CAPI occupied housing units are sampled at a rate of 1 in 3.

State		Mail/CATI		CAPI		Vacant	
State 1	Weight	<16	>=16	<25	>=25	<23	>=23
	Sampling Interval	10	5	7	3	9	4
State 2	Weight	All		All		<41	>=41
	Sampling Interval	5		3		10	3

These combinations of housing unit type and value of weight determined the cells that the test site housing units were stratified into. In the table above, state 1 has six stratification cells and state 2 has five cells. Sampling was done independently in each cell. After stratification, the housing units in each cell were sorted. The cells for vacant housing units were sorted by reason for vacancy, census tract, and weight. The cells for occupied housing units were sorted by tenure, race of householder, census tract, and weight. The categories for vacancy, tenure, and race are:

> *Reason for Vacancy: For sale, For rent*
>
> *Tenure: Owner, Renter*
>
> *Race of Householder: White Non-Hispanic, Black Non-Hispanic, American Indian/Alaska Native Non-Hispanic, Asian Non-Hispanic, Native Hawaiian/Other Pacific Islander Non-Hispanic, Hispanic*

The householder is, in most cases, the person or one of the people in whose name the home is owned, being bought, or rented and who is listed on line one of the survey questionnaire. If there is no such person in the household, any adult household member 15 years old and over could be designated as the householder.

After stratification and sorting, sampling is done in each cell as follows. A random integer between 1 and the sampling interval is generated to select the first record. After the first record is selected, every kth subsequent record is chosen, where k is the sampling interval. The PUMS housing unit weight is calculated by multiplying the original housing unit weight by the sampling interval.

The PUMS person sample is obtained by selecting all persons that are in the selected housing units. The PUMS person weight is calculated by multiplying a person factor by the PUMS weight for the person's housing unit. The person factor is defined as:

> Person Factor = PWGT/WGT
> Where PWGT = The ACS person weight
> WGT = The ACS housing unit weight of the person's housing unit

Under this method for calculating the PUMS person weight, the ratio of the person weight to the weight of the person's housing unit is preserved in the PUMS.

PRODUCTION OF ESTIMATES

The ACS PUMS sample is not self - weighted. To produce estimates or tabulations of characteristics from the ACS PUMS simply add the weights of all persons or housing units that possess the characteristic of interest. For instance, if the characteristic of interest is "total number of black teachers", simply determine the race and occupation of all persons and

cumulate the weights of those who match the characteristics of interest. To get estimates of proportions simply divide the weighted estimate of persons or housing units with a given characteristic by the weighted estimate of the base. For example, the proportion of "black teachers" is obtained by dividing the weighted estimate of black teachers by the PUMS estimate of teachers.

ENDNOTES:

1. Originally published as "Sample Design" and "Production of Estimates" sections, *PUMS Accuracy of Data (2003)*, U.S. Department of Commerce, Bureau of the Census, Washington, DC, 2004, pp. 10-12.

Go Back to Sampling Procedures Index

⏶ Back to Top

SUPPORTED BY

HOME | DATA | SAMPLES | HELP | SEARCH