Documentation for the 2010 Census PSID Geocode Match File

TABLE OF CONTENTS

The 2010 Census geocode file currently contain 1968 through 2019 waves (except 1969) of PSID respondent geocodes. Census geocodes on this file are from the 2010 Census.

This first release of the PSID 2010 Census geocode file contains: a) 2019 data, and b) geocoding of previous years using the SAS proc geocode process. As a result, the FIPS county identifiers are current as of the time of release.

Address Data Used for the 2010 Geocode Match File

The list of addresses that are geocoded came from the PSID Address files, a confidential set of addresses of PSID respondents kept separately from the main PSID data. As a basis for geocoding the residential addresses of respondents, the address data on some files are problematic:

a) In earlier waves (1968-1985) the physical address at the time of interviewing was not preserved, or multiple versions of the file existed. For these waves, we geocoded the address file closest to the time of interviewing. For more information, see "*Address Data Used for 1970 and 1980 Geocode Match Files*" on page 1 of the document file *70-90CensusDoc.pdf* (enclosed in the package). Later waves were geocoded using a preserved physical address at time of interviewing.

b) Respondents often list a non-physical address as their residence (i.e., PO Box).

c) Some address information reported by respondents is incorrect, inaccurate, or not in a format that the geocoder can recognize.

For various reasons, the number of addresses archived and geocoded was different than the total number of families on the public release files. *Table 1* illustrates.

Table 1: PSID Address Files Used for 2010 Geocode Match File

Interview year	Number of addresses	Number of Families
1968	4,798	4,802
1969	Missing	4,460
1970	4,637	4,645
1971	4,831	4,840
1972	5,047	5,060
1973	5,265	5,285
1974	5,501	5,517
1975	5,709	5,725
1976	5,840	5,862
1977	5,982	6,007

Table 1: PSID Address Files Used for 2010 Geocode Match File

Interview year	Number of addresses	Number of Families
1978	6,133	6,154
1979	6,349	6,373
1980	6,508	6,533
1981	6,595	6,620
1982	6,716	6,742
1983	6,824	6,852
1984	6,895	6,918
1985	7,004	7,032
1986	6,995	7,018
1987	7,040	7,061
1988	7,086	7,114
1989	7,081	7,114
1990	9,341	9,371
1991	9,315	9,363
1992	9,787	9,829
1993	9,932	9,977
1994	10,712	10,765
1995	10,401	10,401
1996	8,511	8,511
1997	6,747	6,747
1999	6,997	6,997
2001	7,406	7,406
2003	7,822	7,822
2005	8,002	8,002
2007	8,289	8,289
2009	8,690	8,690
2011	8,907	8,907
2013	9,063	9,063
2015	9,048	9,048
2017	9,607	9,607
2019	9,569	9,569

2010 Geocoded Addresses

The geocoding for the 2010 Census match file is done differently than in the previous Census match files. For 2010, we did some minimal cleaning of the address data and used SAS 9.4 proc geocode process. We imported the latest TIGER/Line shape files for all states from the Census Bureau. These files can be found here: https://www.census.gov/geo/maps-data/data/tiger-line.html. When an exact match is not found, SAS will use other values to get the closest match. The accuracy of that match is coded in the MATCH10 variable, variable on the 2010 Census match file.   Another accuracy variable "SCORE10" gives a numeric representation of how good the geocode match is. The score is calculated based on the "NOTES10" variable and the tokens listed per geocode. Each token within "NOTES10" has a numeric value that is used to calculate "SCORE10". For example if "NOTES10" contains "AD ZC NM" then that means the street name, zip code, and house number matched.   The "SCORE10" value equals the sum of 20 for AD, 15 for ZC and 10 for NM for a total score of 45.