

Population and Householders by Age Before and After Differential Privacy

Ken Hodges

Rachel Cortes



The Claritas “Use Case”

- Build information products
 - For many businesses
 - With many use cases
- Annual demographic estimates
 - All block groups nationwide
 - Build from census data
- Concerned with impact of Differential Privacy (**DP**)

Previous Demonstration Data

- Claritas presented at the 2019 Workshop
- Noted inconsistent Population and Household data
 - For Claritas: The most troubling feature of DP
 - Impossible values of average household size
 - Areas with households but no household population
 - Areas with household population but no households
- Made our peace with PL-94 data
 - Still inconsistencies
 - But improvement since 2019
 - We “corrected” inconsistent 2020 totals
- Now reviewing **DHC Demonstration Data**

DHC Demonstration Data

- Tabular data from the IPUMS National Historical Geographic Information System
 - 2010 Census data with Differential Privacy
 - 2010 Census data as published in SF1

DHC Demonstration Data

- **Age/Sex:** the most important DHC variable for Claritas
- Input to estimates
 - Population by age/sex for block groups
- Consistency with SF1 much improved
 - Index of dissimilarity
 - 2019 Demonstration data = 35.4
 - DHC Demonstration data = 8.9
- But we also estimate Householders by Age
 - Businesses want HH Income by Age of Householder
 - Start with Population by Age
 - Convert to Householders by Age
 - Using age-specific Householder/Population rates

DHC Demonstration Data

- Need Population by Age and Householders by Age
 - Percent of persons that are householders
- How good are householder rates after DP?
 - Reason to worry
 - Pop vs. HH inconsistencies in previous Demonstration Data
- What do the DHC Demonstration Data show?
- Calculated householder rates for 9 age categories
 - 15-24 60-64
 - 35-34 65-74
 - 35-44 75-84
 - 45-54 85 and older
 - 55-59

DHC Demonstration Data

Block Group Mean Percent of Population that is Householder

	DP	SF1
Age	Pct HHr	Pct HHr
15-24	12.7	11.9
25-34	44.2	42.3
35-44	53.7	52.3
45-54	57.0	56.1
55-59	61.2	58.4
60-64	65.1	60.4
65-74	66.1	62.9
75-84	73.3	66.9
85+	87.7	67.3

DP shows higher householder rates especially for older age categories

DHC Demonstration Data

- Inconsistencies for individual block groups
 - Householder rates greater than 100 percent
 - Example:
 - HHr = 500
 - Pop = 400
 - Rate = $500/400 * 100 = 125.0$
 - 125 percent of persons are householders
- 217,740 total block groups
 - 74,857 (34.4 pct) at least one age with more Householders than Pop
 - 99,379 total rates over 100 (some BGs have more than one)
- In contrast
 - SF1: Householders never exceeded Population

DHC Demonstration Data

BGs by Number of Householder/Pop Rates Greater Than 100

N	DP		SF1	
0	142,883	65.6	217,740	0.0
1	56,427	25.9	0	0.0
2	13,738	6.3	0	0.0
3	3,595	1.7	0	0.0
4	852	0.4	0	0.0
5	190	0.1	0	0.0
6	52	0.0	0	0.0
7	3	0.0	0	0.0
8	0	0.0	0	0.0
9	0	0.0	0	0.0
Total	217,740	100.0	217,740	0.0

DHC Demonstration Data

- How different are the DP and SF1 ratios?
- Measure: (DP HHr rate – SF1 HHr rate)
 - Example: DP = 51.9
 - SF1 = 53.3
 - Difference = - 1.4
 - Absolute value so positive and negative differences don't offset
- Look at mean error for each age category
- Mean across age categories for a one-number measure

DHC Demonstration Data

Mean Absolute Difference: DP Ratio vs. SF1 Ratio

Age	N	Mean
15-24	216,614	3.47
25-34	216,616	6.69
35-44	216,553	7.00
45-54	216,536	6.48
55-59	216,252	11.40
60-64	216,108	14.77
65-74	216,055	12.12
75-84	215,375	18.20
85+	210,631	39.24

The ACS Alternative?

- The challenge: How to estimate Householders by Age?
 - While maintaining consistency with Population by Age.
 - One-third of BGs have at least one impossible HHr rate
 - Many others likely have large errors (*but which ones?*)
- Would ACS data be better?
 - Even for block groups?
 - Let's look
- ACS 5-year estimates 2008-2012 (centered on 2010)
 - Compared with 2010 Census SF1
 - Block groups nationwide

DHC Demonstration Data

Block Group Mean Percent of Population that is Householder

	DP	SF1	ACS
Age	Pct HHr	Pct HHr	Pct HHr
15-24	12.7	11.9	11.3
25-34	44.2	42.3	42.9
35-44	53.7	52.3	52.9
45-54	57.0	56.1	56.3
55-59	61.2	58.4	58.3
60-64	65.1	60.4	60.0
65-74	66.1	62.9	62.5
75-84	73.3	66.9	65.9
85+	87.7	67.3	64.8

Note: DP mean rates inflated by those over 100

The ACS Alternative?

BGs by Number of Householder/Pop Rates Greater Than 100

N	DP		SF1		ACS	
0	142,883	65.6	217,740	0.0	217,739	0.0
1	56,427	25.9	0	0.0	0	0.0
2	13,738	6.3	0	0.0	0	0.0
3	3,595	1.7	0	0.0	0	0.0
4	852	0.4	0	0.0	0	0.0
5	190	0.1	0	0.0	0	0.0
6	52	0.0	0	0.0	0	0.0
7	3	0.0	0	0.0	0	0.0
8	0	0.0	0	0.0	0	0.0
9	0	0.0	0	0.0	0	0.0
Total	217,740	100.0	217,740		217,739	

The ACS Alternative?

Mean Absolute Difference: DP Ratio vs. SF1 Ratio

	DP vs SF1		ACS vs SF1	
Age	N	Mean	N	Mean
15-24	216,614	3.47	214,378	8.48
25-34	216,616	6.69	213,831	13.31
35-44	216,553	7.00	214,703	12.48
45-54	216,536	6.48	215,473	11.19
55-59	216,252	11.40	211,619	17.43
60-64	216,108	14.77	209,312	18.38
65-74	216,055	12.12	210,686	14.86
75-84	215,375	18.20	199,214	19.15
85+	210,631	39.24	157,509	26.74
	210,196	12.77	143,301	15.78

The ACS Alternative?

- ACS has no Householder / Pop inconsistencies
- By that measure
 - ACS is vastly better than DP
 - DP worse than error-prone, sample-based, 5-year period estimates
- But DP Householder / Pop closer to SF1 67 percent of the time
 - DP usually better than error-prone, sample-based, 5-year period estimates
 - Faint praise
- Unfortunate reality (for small areas)
 - DHC will provide many householder rates that are unusable
 - ACS avoids this glaring flaw (At least until DP is applied to ACS)
 - But ACS is not a simple substitute

Race and Hispanic Ethnicity

- **But wait !!!!** Age is not the only problem!
- We also estimate Householders by race/ethnicity
 - 7 races by Hispanic/not-Hispanic (14 categories)
 - Start with population by race/ethnicity
 - Use householder rates by race/ethnicity
- Work in progress but race/ethnicity looks even worse

Early Findings

- 153,667 block groups
 - At least one race/ethnicity with more householders than pop
 - 60.7 pct of block groups
- 296,604 total rates over 100 percent
 - 9.6 pct of all householder rates
- SF1 had no inconsistencies

Race and Hispanic Ethnicity

BGs by Number of Householder/Pop Rates Greater Than 100

N	DP		SF1	
0	66,667	30.3	220,334	0.0
1	71,493	32.4	0	0.0
2	45,111	20.5	0	0.0
3	21,897	9.9	0	0.0
4	9,470	4.3	0	0.0
5	3,710	1.7	0	0.0
6	1,349	0.6	0	0.0
7	463	0.2	0	0.0
8	140	0.1	0	0.0
9	27	0.0	0	0.0
10	7	0.0	0	0.0
11-14	0	0.0	0	0.0
Total	220,334	100.0	220,334	0.0

Conclusion

- DHC variables combining Pop and HHs are important to our work
- We look forward to receiving the DHC product
- But are bracing for what we will find
 - And what we will need to deal with it
- Still assessing the Demonstration Data
- And look forward to engaging with the Census Bureau and the data user community

Thank You