Agenda for Today

- Announcements
- Using the census to understand difference: Hispanic homeownership (by Rocio Sanchez-Moyano)
- Break
- Interpreting the American Community Survey
- Lab 2: ACS data [Part 1]

Announcements

https://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive/2018.html

Assignment 1

- **My partner dropped! What do I do?**
  - We will assign you another partner TODAY
- **How to download the ACS tables required for the assignment?**
  - We will teach you how to do it on **Labs 2 and 3**
  - Both are already available on bCourses
- **What software/tool to use?**
  - Whichever work best for you (Excel, Jupyter/Python, R, etc.)

Don’t forget to set your WordPress account by Tuesday. You will need it for Wednesday Lab and Assignment 1

Statistics for the American Community Survey

Berkeley
UNIVERSITY OF CALIFORNIA

Introduction to Urban Data Analytics

Manuel Santana Palacios

June 1, 2020

Content

- 2018 ACS Questionnaire
- The ‘Universe’ of a Variable
- Sampling
- Margins of Error (MoE) and Confidence Intervals (CI)

Questionnaire

https://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive/2018.html

Household composition

Sex

Age

Race & Ethnicity

Education

Income

Housing quality

Housing cost

Public utilities

Labor

Transportation

The ‘Universe’

Variables Reported for Different Universes

- **Individuals**

Variables Reported for Different Universes

- Individuals
- **Families** v. **Households**

Variables Reported for Different Universes

- Individuals
- **Families v. Households**
  - A family consists of two or more people (one of whom is the householder) related by birth, marriage, or adoption residing in the same housing unit. A household consists of all people who occupy a housing unit regardless of relationship.

Variables Reported for Different Universes

- Individuals
- Families v. Households
  - A family consists of two or more people (one of whom is the householder) related by birth, marriage, or adoption residing in the same housing unit. A household consists of all people who occupy a housing unit regardless of relationship.
- **Housing Units**

Variables Reported for Different Universes

- Individuals
- Families v. Households
  - A family consists of two or more people (one of whom is the householder) related by birth, marriage, or adoption residing in the same housing unit. A household consists of all people who occupy a housing unit regardless of relationship.
- Housing Units
- **Labor force** (by residence)

Why the Universe Matters

FIPS	County	Population	Population in occupied housing units
06001450102	Alameda	9434	3765
06001422800	Alameda	8368	2934
06075015700	San Francisco	7832	5420
06075033201	San Francisco	3683	1530
06001422700	Alameda	4885	2937
06001004031	Alameda	2238	1253

Why the Universe Matters

FIPS	County	Population	Population in occupied housing units	Diff
06001450102	Alameda	9434	3765	5669
06001422800	Alameda	8368	2934	5434
06075015700	San Francisco	7832	5420	2412
06075033201	San Francisco	3683	1530	2153
06001422700	Alameda	4885	2937	1948
06001004031	Alameda	2238	1253	985

Why the Universe Matters

From population table

From tenure table

FIPS	County	Population	Population in occupied housing units	Diff
06001450102	Alameda	9434	3765	5669
06001422800	Alameda	8368	2934	5434
06075015700	San Francisco	7832	5420	2412
06075033201	San Francisco	3683	1530	2153
06001422700	Alameda	4885	2937	1948
06001004031	Alameda	2238	1253	985

Group Quarters

- A group quarters is a place where people live or stay other than the usual house, apartment, or mobile home.
- Two general types of group quarters are recognized: **institutional** (e.g., nursing homes, mental hospitals or wards, hospitals or wards for chronically ill patients, hospices, and prison wards) and
- **noninstitutional** (for example, college or university dormitories, military barracks, group homes, shelters, missions, and flophouses).
- Group quarters may have housing units on the premises for staff or guests.

Why the Universe Matters

Table 1. Population in Census Tracts – San Francisco Bay Area

FIPS	County	Population	Population in occupied housing units
06001450102	Alameda	9434	3765
06001422800	Alameda	8368	2934
06075015700	San Francisco	7832	5420
06075033201	San Francisco	3683	1530
06001422700	Alameda	4885	2937
06001004031	Alameda	2238	1253

Source: 201X ACS X-Year Estimates. Population estimates in column 3 includes all individuals residing in occupied housing units and group quarters

Rule #1

Always include the universe in the axis title or source information for your table/chart

Sampling

https://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive/2018.html

Draws a Sample from the Entire Population

Sample year 1

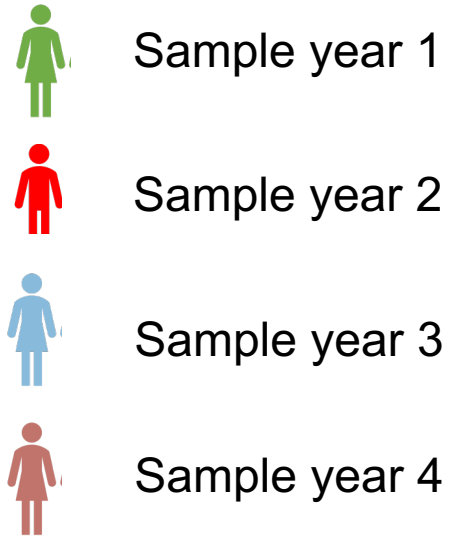
Sample year 2

Sample year 3

Sample year 4

From Sample to Population Estimates

The ACS merges together samples from multiple years to get a population estimate for cities and tracts -e.g., average income for each tract; number of residents from Hispanic origin



ACS Estimates

3-year estimates*	5-year estimates
36 months of collected data *Example:* 2011-2013 ACS 3-year estimates *Date collected between:* January 1, 2011 and December 31, 2013	60 months of collected data *Example:* 2014-2018 ACS 5-year estimates *Date collected between:* January 1, 2014 and December 31, 2018
Data for areas with populations of 20,000+	Data for all areas

https://www.census.gov/programs-surveys/acs/guidance/estimates.html

MoE & CI

https://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive/2018.html

So to Account for Uncertainty in Estimates

Variable	Census Tract 226	
	Estimate	**Margin of Error***
Mean Household Income	178,919	**+/-26,400**

*ACS MoE associated with its estimates at the 90% Confidence Level

So to Account for Uncertainty in Estimates

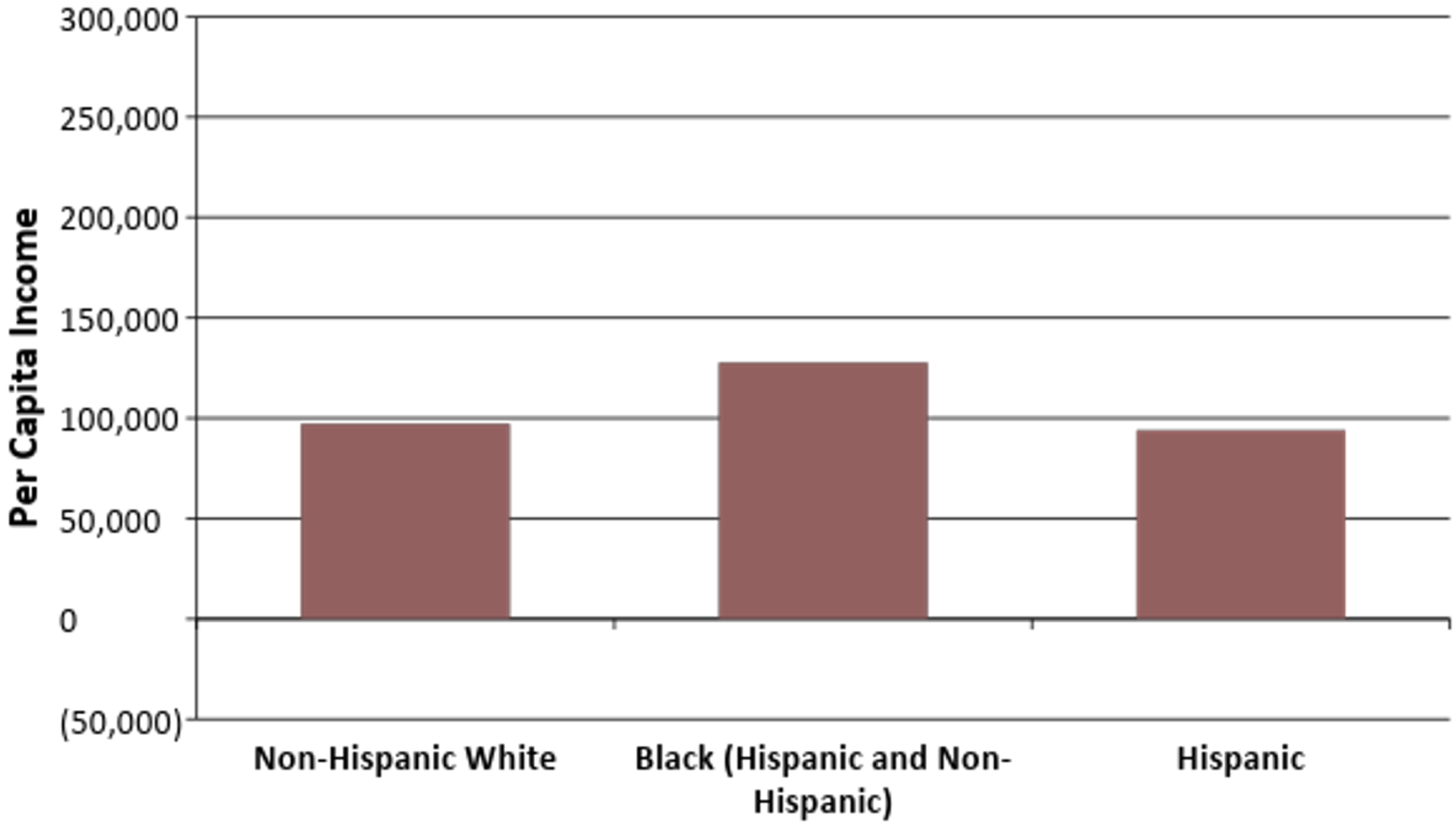
Variable	Census Tract 226 Estimate	Margin of Error	90 % Confidence Interval
Mean Household Income	178,919	**+/-26,400**	**Estimate - MoE ; Estimate + MoE**

So to Account for Uncertainty in Estimates

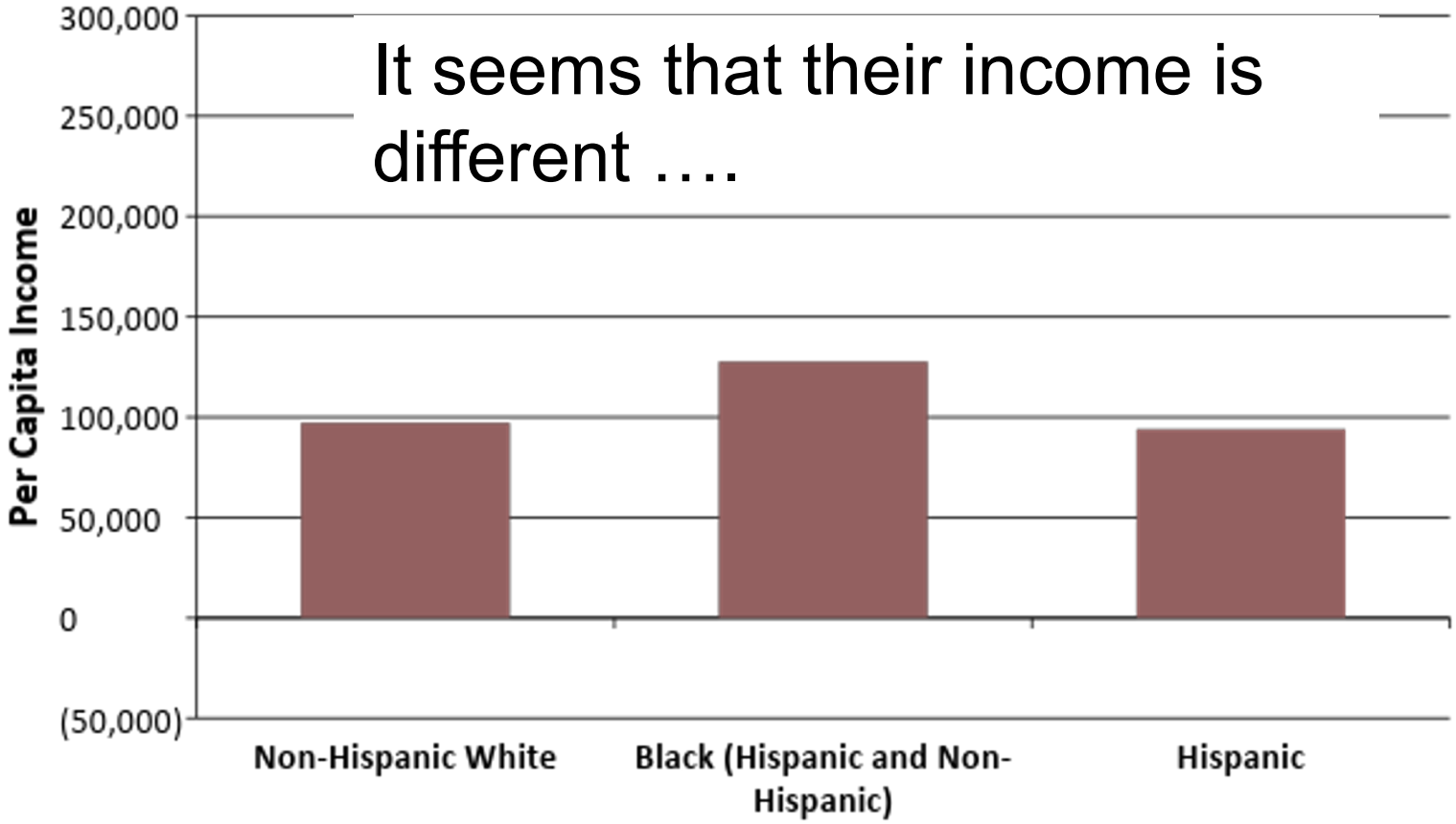
Census Tract 226			
Variable	Estimate	**Margin of Error**	**90 % Confidence Interval**
Mean Household Income	178,919	**+/-26,400**	**152,519 ; 205,319**

We can be 90% confident that the TRUE mean household income for CT226 falls between $152,519 and $205,319.

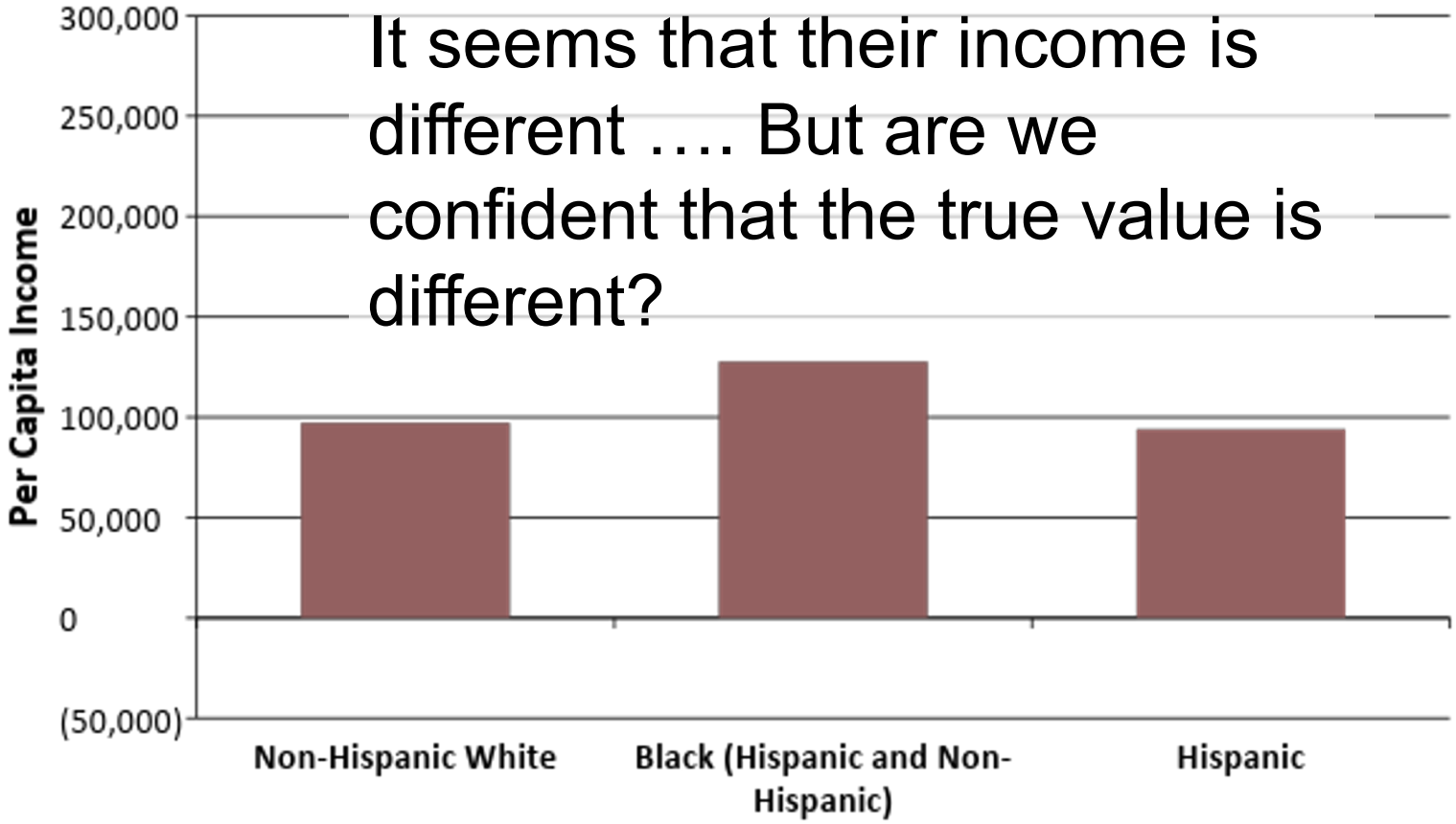
Why does accounting for uncertainty matter



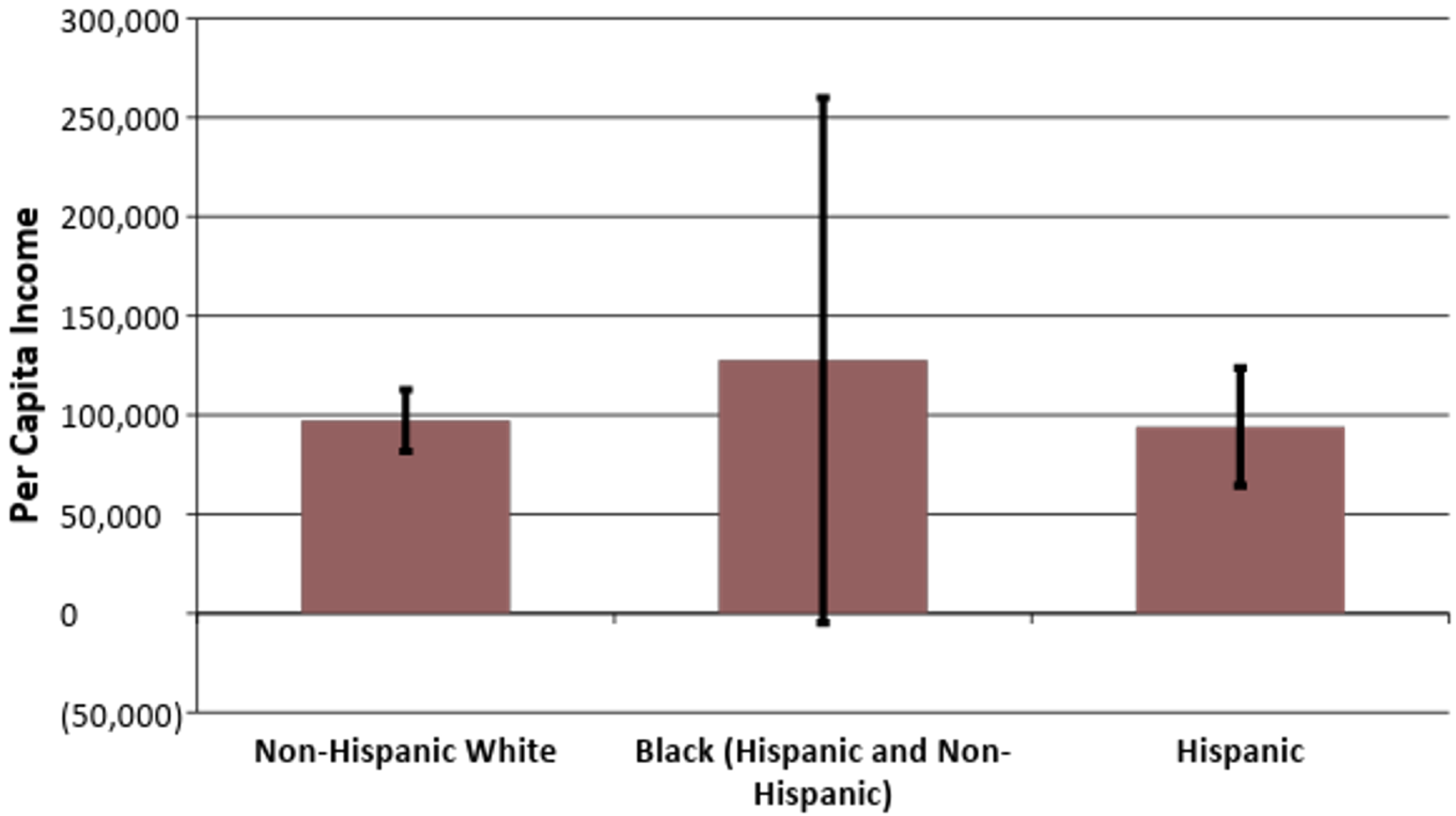
Why does accounting for uncertainty matter



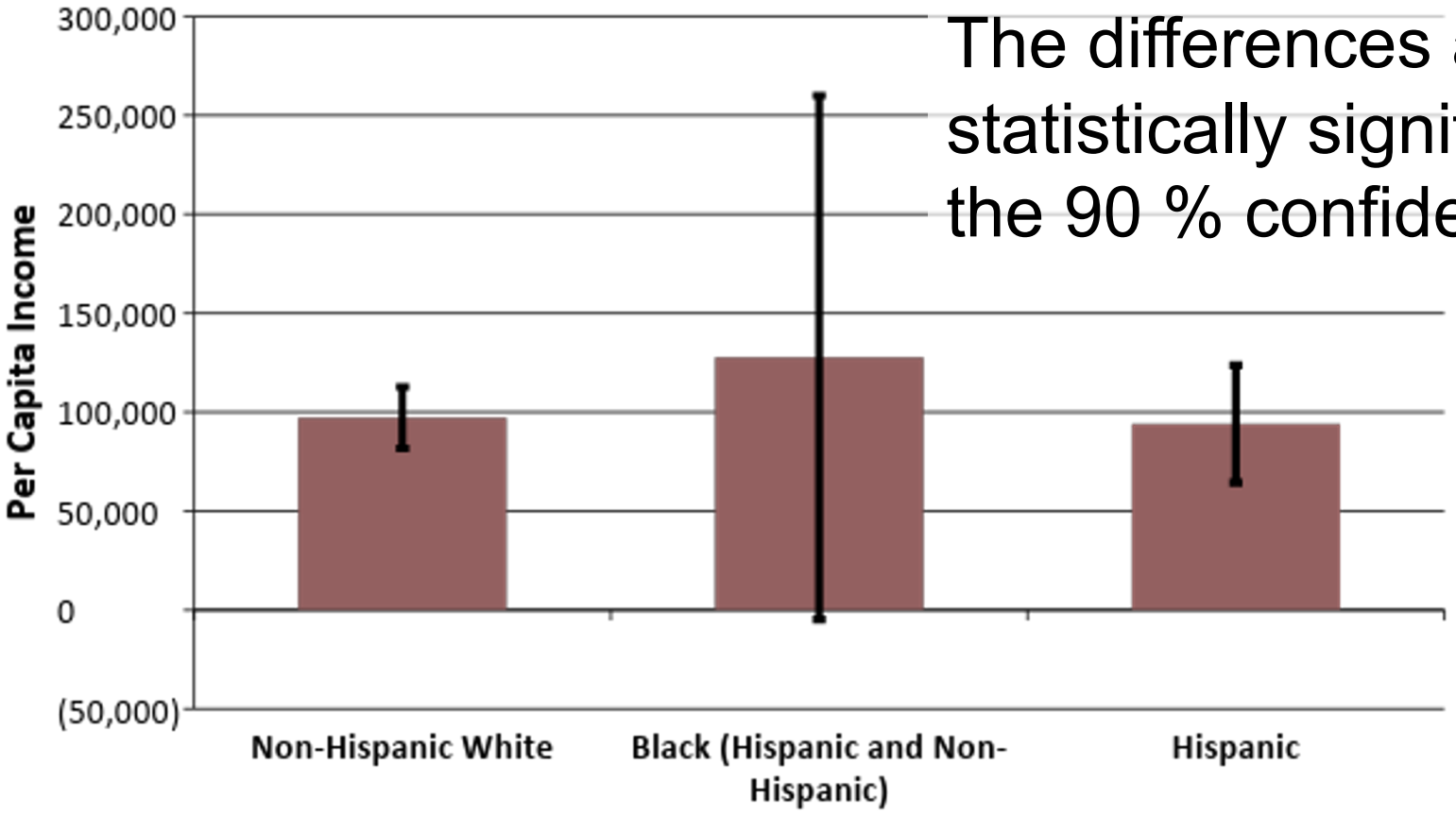
Why does accounting for uncertainty matter



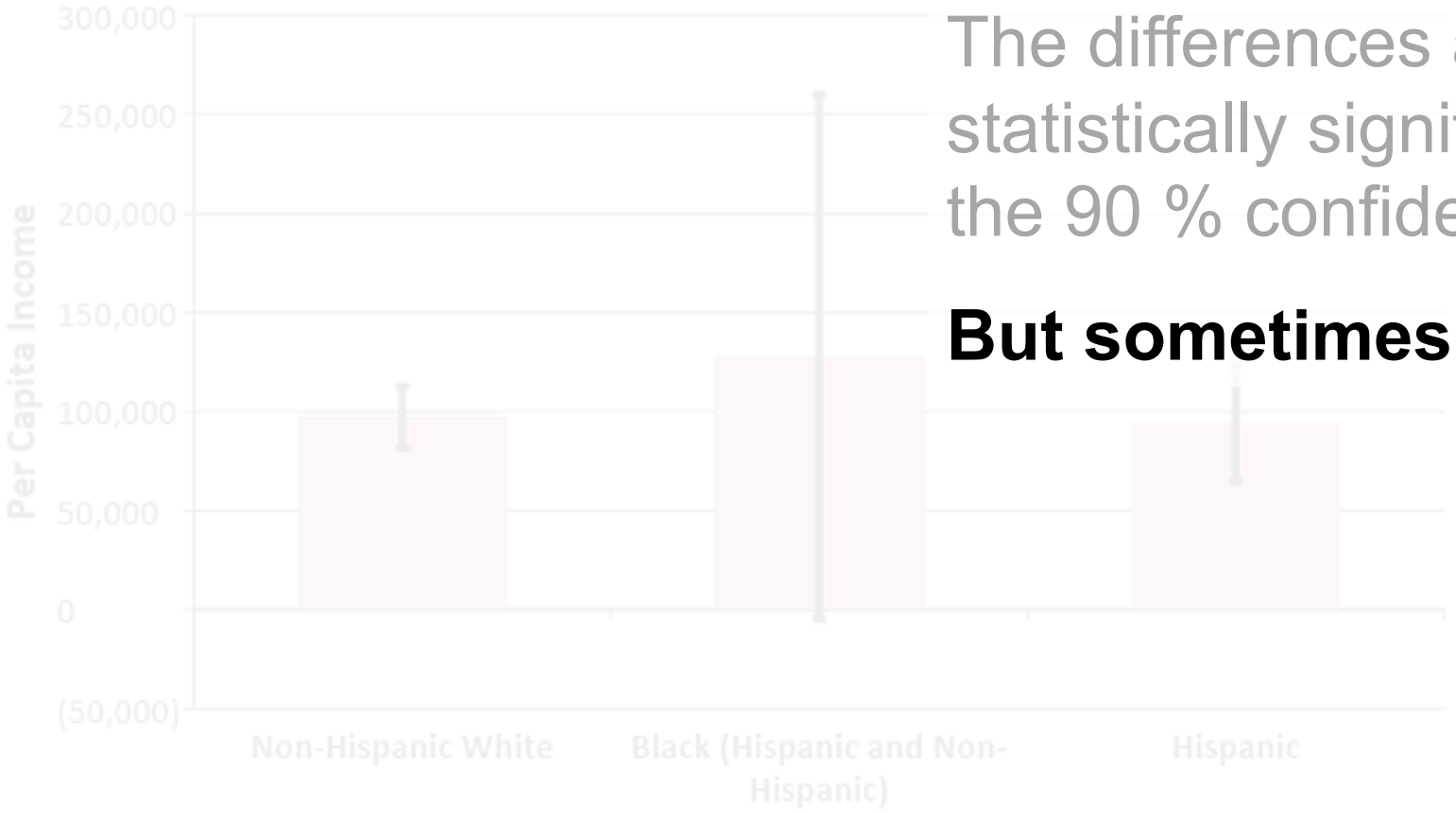
Why does accounting for uncertainty matter



Why does accounting for uncertainty matter



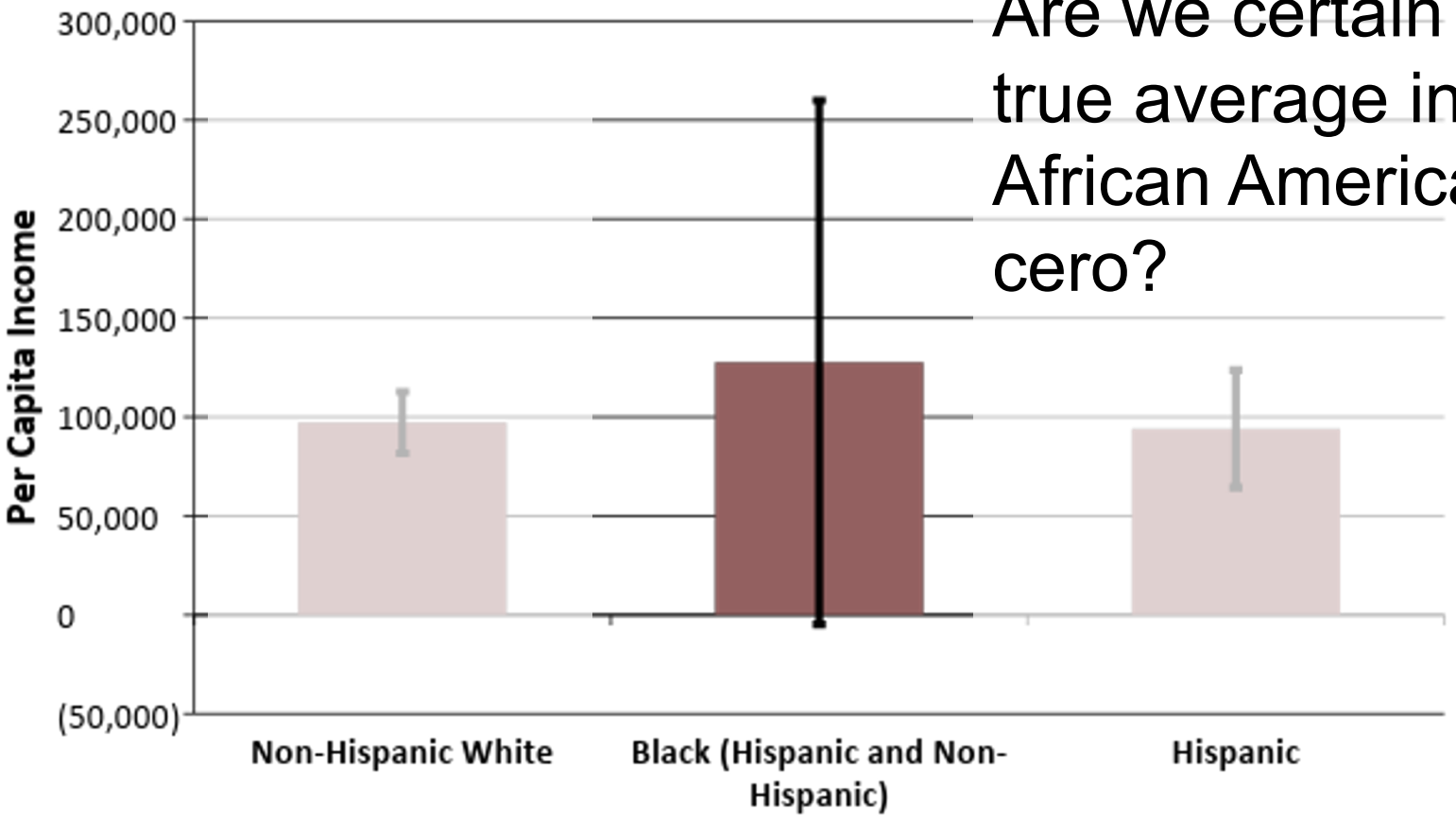
The differences are not statistically significant at the 90 % confident level



The differences are not statistically significant at the 90 % confident level

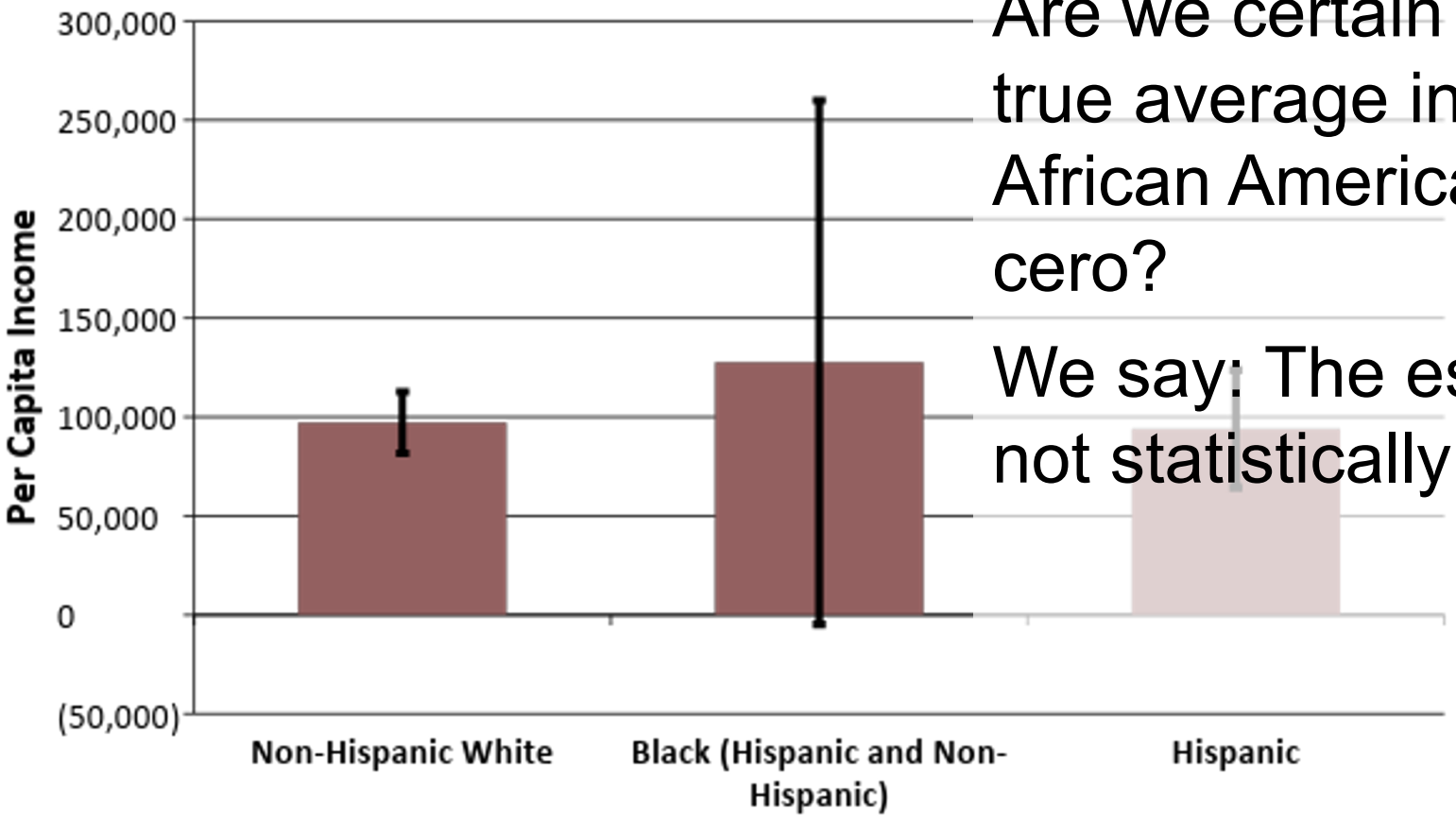
But sometimes they are

Another question



Are we certain that the true average income for African Americans is not cero?

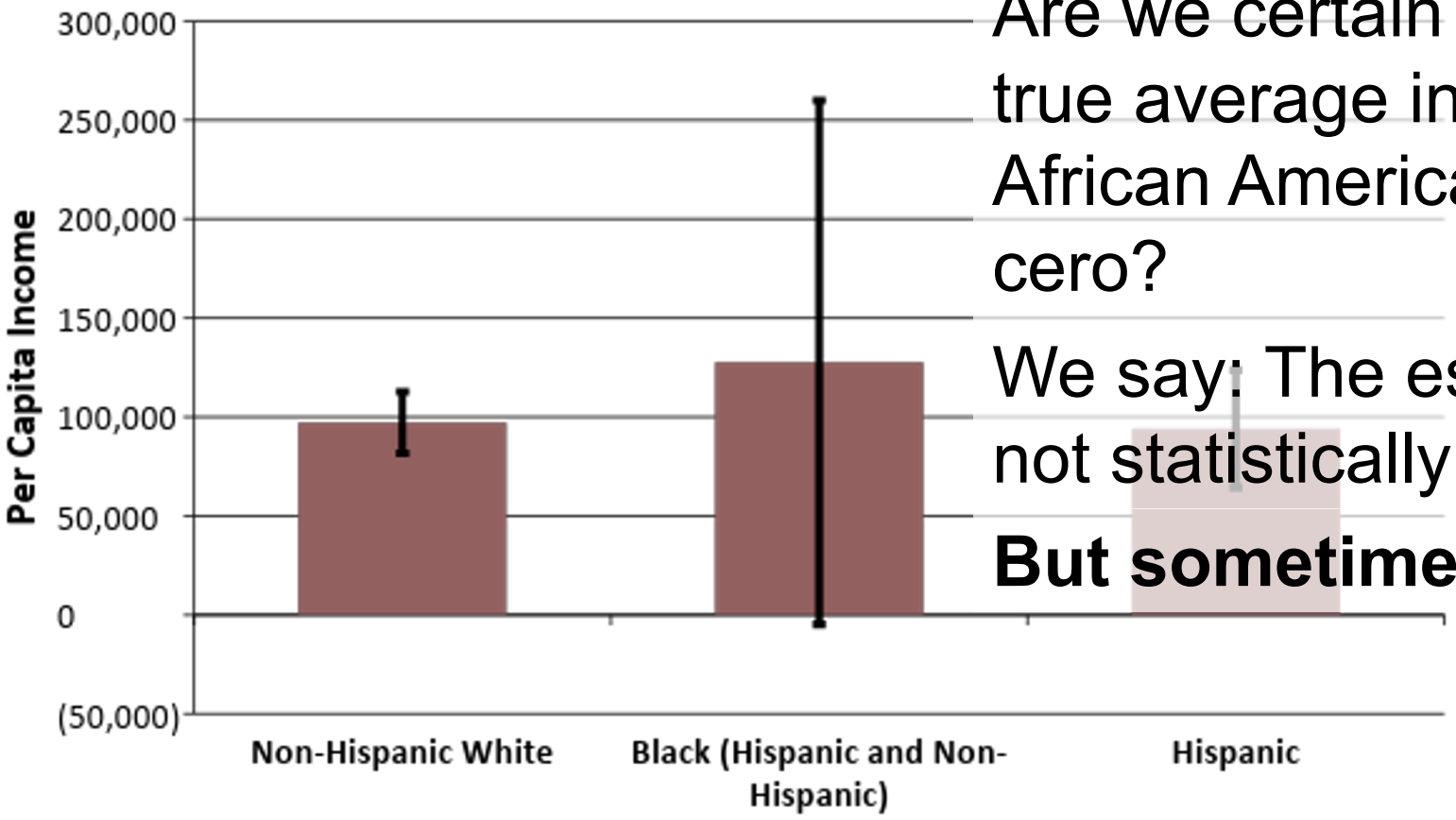
Another question



Are we certain that the true average income for African Americans is not cero?

We say: The estimate is not statistically significant

Another question



Are we certain that the true average income for African Americans is not cero?

We say: The estimate is not statistically significant

But sometimes they are

Rule #2:

to meaningfully interpret ACS data (or any sample data), we *must* look at both the **estimate** *and* the **confidence interval** around that estimate

Navigating Statistical Uncertainty: How Urban and Regional Planners Understand and Work with the American Community Survey Data for Guiding Policy

Reading Response Questions

Loren Greene

Jurjevich et al., 2018. Navigating Statistical Uncertainty

In this article Jurjevich talked about the ACS and how it is miss used and interpreted throughout city planners. Jurjevich, interviewed a handful of city planners asking them if they had training in statistics and found that more than half of them did not have training. Furthermore, many of the planners interviewed either didn't know how to read it properly or just completely disregard the MOE which is vital to understand the data's reliability especially when considering a smaller region or subset of the population. It also touched upon cases when clients explicitly "instruct[ed] planners not to report MOE" page 115. In cases like this do you think it is lawful for them to not publish their complete findings? What are the different reasons that an employer would instruct you not to report the MOE? It seems like this is would create discrepancy's in perceptions and create distrust between the public and planners? If a planner were to reference or us data from the ACS would it make it more reliable and accurate if there was a law that they either had to acknowledge the MOE or state it in their analysis? Would that make analysis more transparent?

Reading Response Questions

Loren Greene

Jurjevich et al., 2018. Navigating Statistical Uncertainty

In this article Jurjevich talked about the ACS and how it is miss used and interpreted throughout city planners. Jurjevich, interviewed a handful of city planners asking them if they had training in statistics and found that more than half of them did not have training. Furthermore, many of the planners interviewed either didn't know how to read it properly or just completely disregard the MOE which is vital to understand the data's reliability especially when considering a smaller region or subset of the population. It also touched upon cases when clients explicitly "instruct[ed] planners not to report MOE" page 115. In cases like this do you think it is lawful for them to not publish their complete findings? What are the different reasons that an employer would instruct you not to report the MOE? It seems like this is would create discrepancy's in perceptions and create distrust between the public and planners? If a planner were to reference or us data from the ACS would it make it more reliable and accurate if there was a law that they either had to acknowledge the MOE or state it in their analysis? Would that make analysis more transparent?

Reading Response Questions

Amy Huang

Jurjevich et al., 2018. Navigating Statistical Uncertainty

Through this reading, Jurjevich highlights the stark lack of understanding of important statistical uncertainty by urban and regional planners when working with ACS data, thus calling into question the validity of our past and present public policy decisions. With this understanding, how can we hold our past accountable and our current policies are not disproportionately hurting members of our community? how does this relate to the misconstrued understanding of urban planning as a profession?

How to aggregate categories

Example: How many children, adults, and seniors live in Census Tract 4004?

Subject	Census Tract 4004, Alameda County,	
	Total	
	Estimate	Margin of Error
Total civilian noninstitutionalized population	3,986	+/-250
AGE		
Under 5 years	257	+/-101
5 to 17 years	421	+/-125
18 to 34 years	1,241	+/-204
35 to 64 years	1,569	+/-140
65 to 74 years	329	+/-77
75 years and over	169	+/-59

Source: 2013-2017 ACS 5-Year Estimates, Table S1810.

How to aggregate categories with MoE

Example: How many children, adults, and seniors live in Census Tract 4004?

Subject	Census Tract 4004, Alameda County,	
	Total	
	Estimate	Margin of Error
Total civilian noninstitutionalized population	3,986	+/-250
AGE		
Under 5 years	257	+/-101
5 to 17 years	421	+/-125
18 to 34 years	1,241	+/-204
35 to 64 years	1,569	+/-140
65 to 74 years	329	+/-77
75 years and over	169	+/-59

Source: 2013-2017 ACS 5-Year Estimates, Table S1810.

How to aggregate categories with MoE

Calculating MOEs for Aggregated Count Data

To calculate the MOE for aggregated count data:

1) Obtain the MOE of each component estimate.
2) Square the MOE of each component estimate.
3) Sum the squared MOEs.
4) Take the square root of the sum of the squared MOEs.

The result is the MOE for the aggregated count. Algebraically, the MOE for the aggregated count is calculated as:

$$MOE_{agg} = \pm\sqrt{\sum_{c} MOE_{c}^{2}}$$

where MOE_c is the MOE of the c^{th} component estimate.

From: A Compass for Understanding and Using American Community Survey Data
What General Data Users Need to Know
https://www.census.gov/content/dam/Census/library/publications/2008/acs/ACSGeneralHandbook.pdf
Page A-14, Appendix 3

How to aggregate categories with MoE

Census Tract 4004, Alameda County

Step 1-2

	Estimate	MoE		Estimate	MoE^2
Total population	3,986	+/- 250			
AGE			**GROUP**		
Under 5 years	257	+-101	Children	678	10,201
5 to 17 years	421	+-125			15,625
18 to 34 years	1,241	+-204	Adults	2,810	41,616
35 to 64 years	1,569	+-140			19,600
65 to 74 years	329	+-77	Seniors	498	5,929
75 years and over	169	+-59			3,481

How to aggregate categories with MoE

Census Tract 4004, Alameda County

Step 3

	Estimate	MoE		Estimate	MoE^2	$MoE_a^2 + MoE_b^2$
Total population	3,986	+/- 250				
AGE			**GROUP**			
Under 5 years	257	+-101	Children	678	10,201	25,826
5 to 17 years	421	+-125			15,625	
18 to 34 years	1,241	+-204	Adults	2,810	41,616	61,216
35 to 64 years	1,569	+-140			19,600	
65 to 74 years	329	+-77	Seniors	498	5,929	9,410
75 years and over	169	+-59			3,481	

How to aggregate categories with MoE

Census Tract 4004, Alameda County

Step 4

	Estimate	MoE		Estimate	MoE^2	$MoE_a^2 + MoE_b^2$	$MoE = (MoE_a^2 + MoE_b^2)^{1/2}$
Total population	3,986	+/- 250					
AGE			**GROUP**				
Under 5 years	257	+-101	Children	678	10,201	25,826	+/- 160.7
5 to 17 years	421	+-125			15,625		
18 to 34 years	1,241	+-204	Adults	2,810	41,616	61,216	+/- 247.4
35 to 64 years	1,569	+-140			19,600		
65 to 74 years	329	+-77	Seniors	498	5,929	9,410	+/- 97.0
75 years and over	169	+-59			3,481		

Lab 2 – ACS Census Data [Part 1]