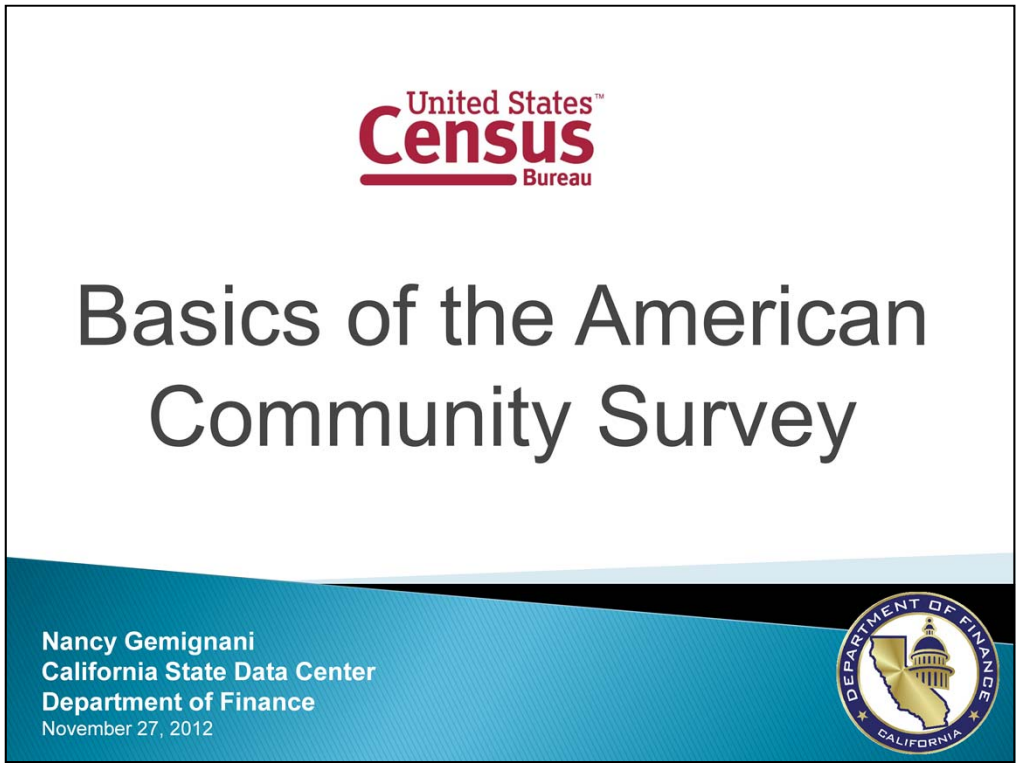


This presentation is approximately one hour.

We are going to look at the data and methodologies of the American Community Survey.

Not the way it used to be…

- 2010 Census focused on counting the U.S. population with basic demographic info – age, sex, race/Hispanic, and owner/renter. There was only one form – no long form.
- The American Community Survey collects those same questions as in the 2010 Census.
- PLUS the census long form is replaced by the American Community Survey. Social and economic data are now collected **only** in the ACS.

ACS & Census Content

Demographic *	Social	Economic	Housing
Sex	Families *	Income	Tenure *
Age	Education	Poverty	Occupancy *
Race	Marital Status	Benefits – Food Stamps/SNAP	Structure Type
Hispanic Origin	Fertility	Employment Status	Housing Value
Household Relationship	Grandparents	Occupation	Taxes & Insurance
Group Quarters	Veterans	Industry	Utilities
	Disability Status	Journey to Work	Mortgage
	Language at Home	Place of Work	Monthly Rent
	Citizenship	Health Insurance	
	Migration		

Items with an "*" and in red were collected in both the 2010 Census and the ACS.

You can see that most of these variables are familiar.

Those with an asterisk "*" and in red are collected in both the ACS and the 2010 Census. If you need only these characteristics, it is better to use the Census right now since we are so close to the census date.

Items in blue are only collected in the ACS

ACS Sample Size – 2011

- In the United States, the Census Bureau expanded the ACS sample in June of 2011 to 3.54 million addresses, about 2.6 percent of the total.
- Usable responses in 2011 to develop the ACS estimates:
  1,917,799 (1.5% of total) in the US
  176,551 (1.3% of total) in California
- Designed to measure **characteristics** and trends. The ACS is not the official population counts.
- A single address (housing unit) will receive the survey more than once every five years. However, if someone moves, they may be surveyed more often.

The sample was increased to about 3.54 million addresses in 2011 . Unweighted responses - 1,917,799 to HU - 131,791,065. In California, we end up with unweighted 176,551 (1.3%) which is weighted to 13,682,976.

ACS Sample Size – 2011

- In 2011, the Census Bureau reallocated the sample for census tracts from large census tracts to smaller census tracts.
- While this will slightly reduce the reliability of the data for the larger tracts, there is a clear improvement for less populated tracts.
- Effects will be gradual improvement until 2011-2015 data are published – note the changed coefficient of variances.

Tract Size Category	Average Tract Size	Old Stratification CVs	New Stratification CVs
0 –400	291	66%	41%
401 –1,000	766	41%	30%
1,001 –2,000	1,485	29%	29%
2,000 –4,000	2,636	26%	29%
4,000 –6,000	4,684	19%	29%
6,000 +	8,337	15%	28%

Collecting the ACS

- Data are collected continuously throughout the year.
- Every month questionnaires are sent to a twelfth of the annual sample. If the form is not returned, follow-up is done by phone then a personal visit in the next two months.

ACS sample panel	Month of data collection							
	2005		2006					
	November	December	January	February	March	April	May	June
November 2005	Mail	Phone	Personal visit					
December 2005		Mail	Phone	Personal visit				
January 2006			Mail	Phone	Personal visit			
February 2006				Mail	Phone	Personal visit		
March 2006					Mail	Phone	Personal visit	
April 2006						Mail	Phone	Personal visit
May 2006							Mail	Phone
June 2006								Mail

Data for the American Community Survey are **collected continuously throughout the year**. Interviews conducted between January 1st and December 31st of a given year are aggregated to produce annual estimates for that calendar year. For example, interviews conducted between January 1st and December 31st of 2010 were aggregated to produce estimates for 2010.

Data are released annually.

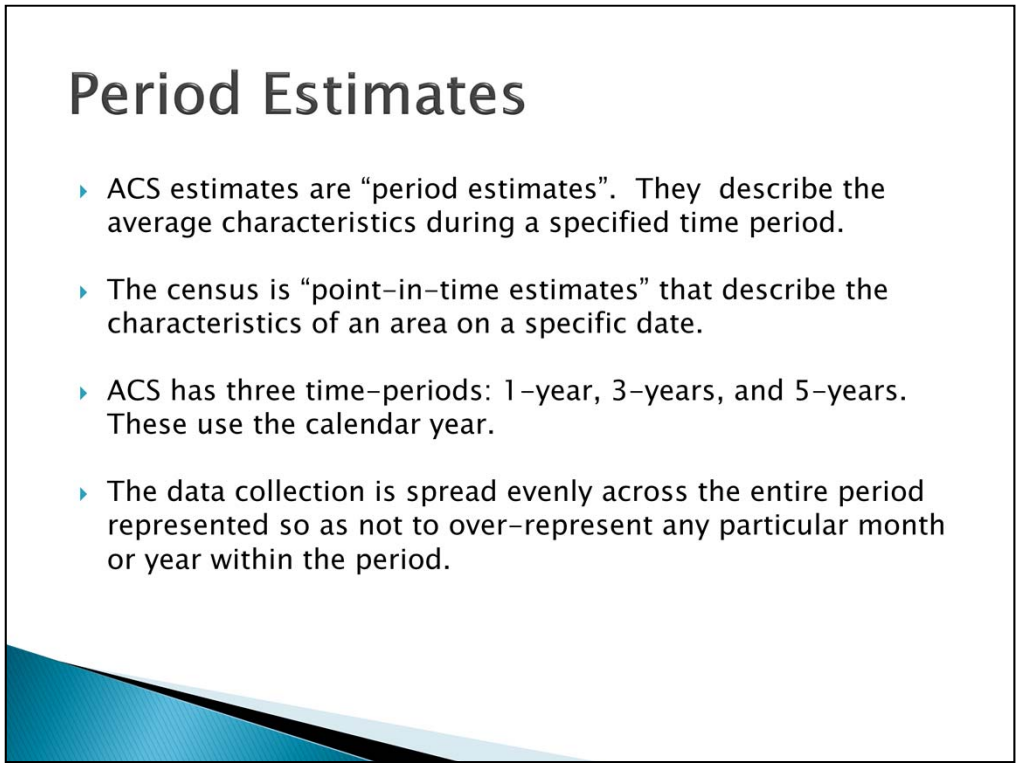
Weighting to the Total

- The Census Bureau's annual population estimates are used as controls to process the ACS responses.
- These population estimates are by sex, age, race, and Hispanic origin, and estimates of total HUs by county.
- For any given tabulation area, a characteristic total is estimated by summing the weights assigned to the people, households, families, or HUs possessing the characteristic.

- 2010 Census results were used to create the intercensal estimates for 2000-2009
- Also were used to create simple extrapolated estimates for 2010
- Both of the above were used in processing the 2010 data products last year.
- 2011 will use the first post-censals based on 2010

Missing Data a.k.a. Item Nonresponse

- It occurs when a respondent fails to provide an answer to a required item or that answer is invalid.
- To fill in the blanks, the Census Bureau uses two imputation methods – "assignment" and "allocation."
- **Assignment** uses rules to determine acceptable answers. It involves logical imputation where a response to one question implies the value for a missing response to another question. For example, first name can often be used to assign a value to sex.
- **Allocation** uses answers from similar housing units or people who provided the item information. It involves using statistical procedures, such as within-household or nearest neighbor matrices populated by "donors" (the response source), to impute for missing values.



American Community Survey period estimates describe the average characteristics of the population or housing over a specified period of time. In the case of American Community Survey one-year estimates, the period is the calendar year.

For example, the 2011 American Community Survey data describe the population and characteristics of an area from January 1, 2011 through December 31, 2011, not for any specific day within the year.

The American Community Survey collects survey information continuously nearly every day of the year and then aggregates the results over a specific time period – one year, three years, or five years.

The data collection is spread evenly across the entire period represented so as not to over-represent any particular month or year within the period.

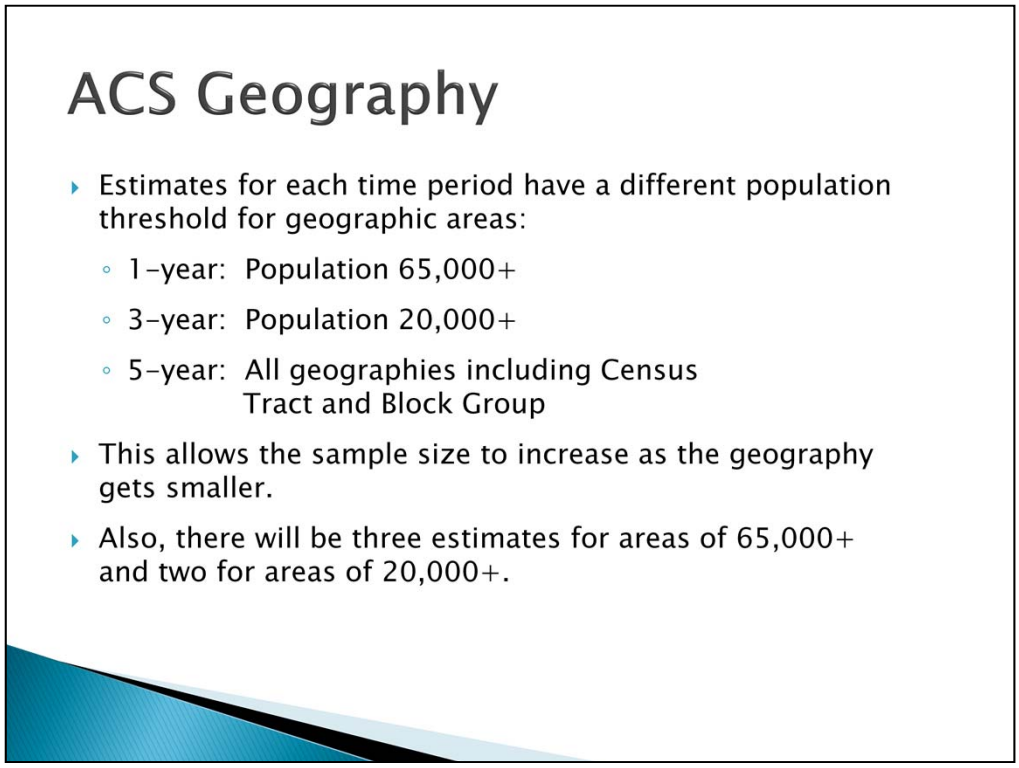
Time Periods

Data File Releases	Estimates are averages of surveys from	Survey Years				
1-year	2011					2011
3-year	2009-2011			2009	2010	2011
5-year	2007-2011	2007	2008	2009	2010	2011

American Community Survey period estimates describe the average characteristics of the population or housing over a specified period of time. In the case of American Community Survey one-year estimates, the period is the calendar year.

For example, the 2011 American Community Survey data describe the population and characteristics of an area from January 1, 2011 through December 31, 2011, not for any specific day within the year.

The American Community Survey collects survey information continuously nearly every day of the year and then aggregates the results over a specific time period – one year, three years, or five years. There is no "base year" the estimates represent the entire time period.



Each time period has a different population threshold. Larger areas will be available in each survey period – areas of 65,000+ will also be available in the 3- and 5-year files; areas of 20,000+ will be available in the 5-year file.

Note that at these thresholds, we will get all counties in CA only in the 5-year file.

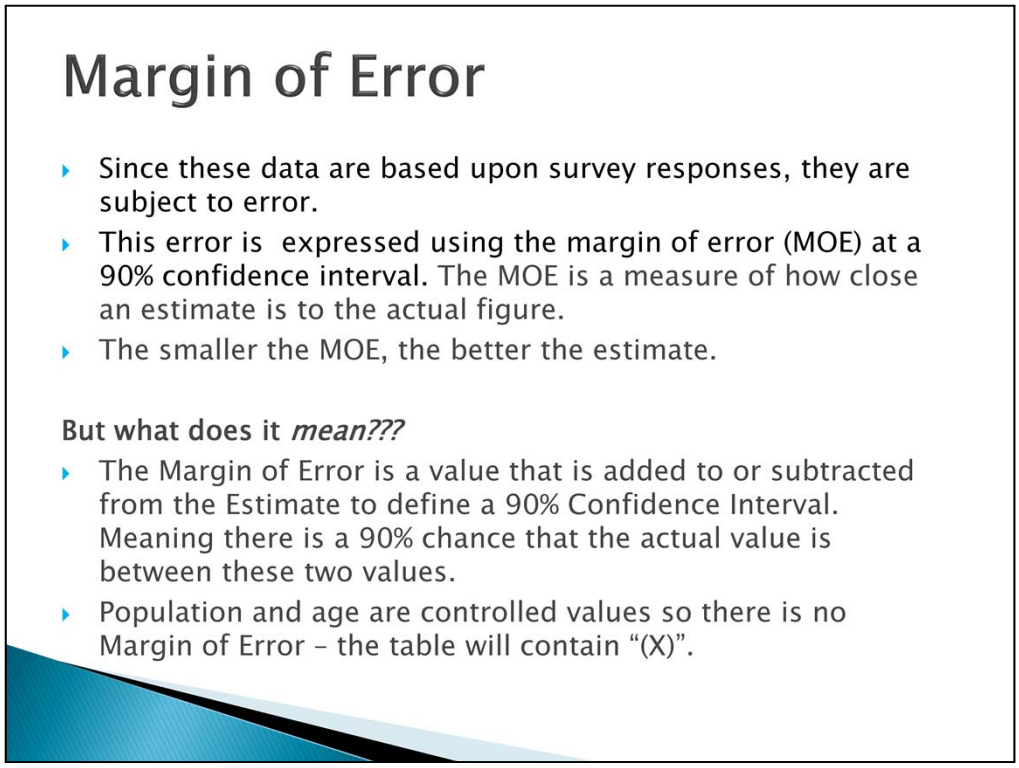
Census will have data down to the block level.

ACS Special 5-Year Files

- In addition to the standard 5-year Summary Files, special tabulations will also be generated for the 5-year time period including:
  - Selected Population Tables (detailed tables for race, Hispanic, and ancestry group to the census tract) – Similar to the 2000 Census, Summary File 4
  - American Indian and Alaska Native Tables (detailed tables by American Indian tribe)
  - Census Transportation Planning Product (detailed tables for transportation planning)
  - EEO Special File (detailed occupation by race/Hispanic and gender for areas of 50,000 or more population)
- Due to the detailed tabulations in these files, the geography available will be larger than the standard 5-year file.

EEO special file
http://www.census.gov/people/eeotabulation/data/eeoupcoming.html

CTPP: http://ctpp.transportation.org/



When the ACS data are released, they will be published with a margin of error. The margin of error is for a 90% confidence interval – 1.65 times the standard error. The MOE is a measure of how close an estimate is to the actual figure. The smaller the MOE, the better the estimate.

Note that sample data from the 2000 Census also has a margin of error but the Census Bureau never published it.

Margin of Error (MOE) Since these data are based upon survey responses, they are subject to error. This error is expressed using the margin of error (MOE) at a 90% confidence interval. This MOE is added and subtracted to the estimate to represent a range within which we are 90% sure the actual value occurs. The MOE relative to the estimate may act as a measure of the quality of the estimate. An estimate with a MOE which is 50% or more of that estimate should be considered unreliable.

Using the Margin of Error

So for the Employed,

- The estimate is 16,243,172 and the MOE is "plus or minus" 45,906.
- We are 90% sure that the actual estimate is between 16,289,078 and 16,197,266

California, 2010 ACS	Estimate	Estimate Margin of Error	Percent	Percent Margin of Error
EMPLOYMENT STATUS				
Population 16 years and over	29,159,527	+/-10,611	29,159,527	(X)
In labor force	18,759,274	+/-41,728	64.30%	+/-0.1
Civilian labor force	18,625,515	+/-41,714	63.90%	+/-0.1
Employed	16,243,172	+/-45,906	55.70%	+/-0.2
Unemployed	2,382,343	+/-26,929	8.20%	+/-0.1
Armed Forces	133,759	+/-4,527	0.50%	+/-0.1
Not in labor force	10,400,253	+/-43,179	35.70%	+/-0.1

Note the "(X)" for the MOE of the universe. Also the percent does not show 100% but the actual count. This tells you that the data below are the percent of that number.

Guidelines for Using Estimates/MOE

- Since the MOE measures the error in an estimate, we use it to determine if an estimate has too much error to be reliable.
- We calculate the MOE relative to the estimate by dividing the MOE by the estimate and multiplying it by 100. We call this the "relative MOE".
- There is not a standard for the ranges of an acceptable "relative MOE". Ultimately is up to the user to access the quality of each ACS estimate and decide whether a particular estimate is suitable to his/her needs.
- Our recommendation is below 33% is "good", 33% to 49% is "use with caution", and 50% or more is "unreliable".

California, 2010 ACS	Estimate	Estimate Margin of Error	Relative MOE
EMPLOYMENT STATUS			
Population 16 years and over	3,244	278	9%
In labor force	2,356	293	12%
Civilian labor force	2,344	293	13%
Employed	2,083	248	12%
Unemployed	261	121	*46%*
Armed Forces	12	18	*150%*
Not in labor force	888	213	24%

Looking at these data for a single census tract – Tract 12 in Sacramento – we see that the data for unemployed is borderline and the data for Armed Forces are unreliable.

This comparison between the MOE and the estimate is NOT done on AFF – this is something the user must do in their analysis of the reliability of the data.

Compare 1-year to 5-year MOE

California, 2010 ACS	Estimate	Estimate Margin of Error	Relative MOE
EMPLOYMENT STATUS			
Population 16 years and over	29,159,527	+/-10,611	0.0%
In labor force	18,759,274	+/-41,728	0.2%
Civilian labor force	18,625,515	+/-41,714	0.2%
Employed	16,243,172	+/-45,906	0.3%
Unemployed	2,382,343	+/-26,929	1.1%
Armed Forces	133,759	+/-4,527	3.4%
Not in labor force	10,400,253	+/-43,179	0.4%

California, 2006-2010 ACS	Estimate	Estimate Margin of Error	Relative MOE
EMPLOYMENT STATUS			
Population 16 years and over	28,445,585	+/-4,159	0.0%
In labor force	18,418,306	+/-17,253	0.1%
Civilian labor force	18,274,871	+/-17,360	0.1%
Employed	16,632,466	+/-19,046	0.1%
Unemployed	1,642,405	+/-9,158	0.6%
Armed Forces	143,435	+/-2,902	2.0%
Not in labor force	10,027,279	+/-16,830	0.2%

ACS Balancing Act: Currency vs. Reliability

- Using multiyear estimates is more than simply a choice between using the one-year or the multiyear estimates. Many areas will be in all files > one-, three- and five-year.
- Consider the trade-off between currency and reliability. The one-year estimates for an area reflect the most current data but they tend to have higher margins of error than the three- and five-year estimates.
- The main advantage of using multiyear estimates is the increased statistical reliability for smaller geographic areas and small population groups.
- There are no hard-and-fast rules on choosing between one-, three-, and five-year data, but the margins of error provided with ACS data can help data users decide on the trade-off between currency and reliability.

Your use of the ACS will be a balancing act. The choices posed for using multiyear estimates is more than simply a choice between using the one-year or the multiyear estimates, because **for many areas there will also be the choice of which multiyear estimate to use, three- or five-year.**

For small areas, only five-year estimates are released, but for larger areas, each annual release will provide one-, three-, and five-year estimates. For example, in 2010, there will be three sets of commuting data for San Diego County – one-year estimates for 2009, three-year estimates reflecting 2007-2009, and five-year estimates for the period of 2005-2009. Users need to decide which is the most appropriate for their needs including the requirements of the grant application.

In making this choice, you need to consider the tradeoff between currency and reliability. The three-year and five-year estimates for an area have larger samples and smaller margins of error than the one-year estimates, but they are less current because the larger samples include data that were collected in earlier years. The main advantage of using multiyear estimates is the increased statistical reliability for smaller geographic areas and small population groups.

There are no hard-and-fast rules on choosing between one-, three-, and five-year data, but the margins of error provided with ACS data can help data users decide on the tradeoff between currency and reliability.

Comparing ACS Data

The Census Bureau recommends that you:

- **DO** compare similar period lengths, for example, 3-year to 3-year.
- **DON'T** compare estimates from different period lengths, for example, 1-year to 3-year.
- **DO** compare estimates from non-overlapping periods, for example, compare a 2005-2007 ACS 3-year estimate to a 2008-2010 ACS 3-year estimate.
- **DON'T** compare overlapping periods, for example, the 2005-2007 ACS 3-year estimates to the 2006-2008 ACS 3-year estimates.

Source: www.census.gov/acs/www/guidance_for_data_users/comparing_data/

18

To aggregate an estimate, simply sum the estimates together.

Comparing ACS to Census Data

- ACS 1-year, 3-year, or 5-year estimates can be compared **with** Census 2000 data.
- However, there may be differences in the **universe**, **question wording**, **residence rules**, **reference periods**, and **the way in which the data are tabulated** which can impact comparability with Census 2000.
- Use the "Table Comparison Lookup Tool" to check for table-specific comparability at www.census.gov/acs/www/guidance_for_data_users/table_comparisons/

19

To aggregate an estimate, simply sum the estimates together.

Getting the Estimates You Need

- Sometimes you may need to aggregate estimates because:
  - You need an estimate for a less specific group than the detail it is published.
  - You want a more robust estimate
  - You are creating custom geographies based on adding/subtracting published areas.
- The "new" estimate is easy – it's the sum of the published estimates.
- BUT you cannot simply add the MOEs together.

20

To aggregate an estimate, simply sum the estimates together.

Need a less specific estimate?

- FIRST – Verify that it is not available in another table on American FactFinder, such as:
  - Profile tables (DP02 – DP05)
  - Subject tables (Sxxx)
  - Collapsed Detailed Tables (Cxxxxx)
- Example data need: Unemployment rate 16+ population
  - Table B23001: Sex by Age by Employment Status Requires aggregating over 26 cells to estimate unemployed and labor force then calculate the rate.
  - This will require a lot of work calculating and the MOE will be less accurate
  - However, S2301: Employment Status has it calculated already down to tracts. In the 2011 ACS, Table B23025 provides the high-level data.

21

NOTE: Tables that start with B or C are available on the Summary File. But "S" tables and "DP" Data Profiles are on American FactFinder only. So the S and DP tables are NOT available for block groups.

Need a less specific estimate?

- THEN – if you determine you can't find what you need elsewhere, try to find a table with fewest cells that need to be combined
- Example: Suppose interested in unemployment rate for 16-24 males
  - Table B23001 has ages 16-19, 20-21, 22-24
  - Table C23001 has age ranges 16-19, 20-24
  - Using the collapsed table, C23001, will be easier and your calculated MOE generally will be more accurate.

Need a different geography?

- Try to cover the geography with a fewer number of larger geographies than several smaller geographies – for example, use one place rather than 10 census tracts.
- Or subtract an area from a larger one – for example, subtract one block group from a tract rather than adding up four block groups.
- Not only is it less work, your MOEs will be more accurate.

MOEs of Aggregated Estimates

- The actual formula is

$$MOE(sum) = \sqrt{[MOE(\hat{X})]^2 + [MOE(\hat{Y})]^2 ... + [MOE(n)]^2}$$

- As much as possible – try to minimize the number of estimates you sum up.
- See Appendix 3 of the ACS User Handbooks for more information – www.census.gov/acs/www/guidance_for_data_users/handbooks/

24

To aggregate an estimate, simply sum the estimates together.

However, you cannot simply sum the MOEs together. Here's the actual formula.

We square the published MOEs for the each estimate, add them together, and take the square root.

Covariances are not published and not easy for users to approximate on their own. Therefore, the approximation given at the bottom of the slide is what is given in all the documentation. Basically the covariance term is ignored.

You should try to minimize the number of estimates that you add up. For example, rather than adding up 8 census tracts, perhaps this is a CDP plus one tract.

Example: Poverty for Persons < 65

Subject	California					
	Total		Below poverty level		Percent below poverty level	
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error
Population for whom poverty status is determined	36,593,372	+/-8,519	5,783,043	+/-74,336	15.8%	+/-0.2
AGE						
Under 18 years	9,157,681	+/-9,562	2,012,585	+/-38,677	22.0%	+/-0.4
18 to 64 years	23,258,720	+/-6,873	3,363,215	+/-39,911	14.5%	+/-0.2
65 years and over	4,176,971	+/-6,408	407,243	+/-10,151	9.7%	+/-0.2
Under 65 years	32,416,401	+/-11,776	5,375,800	+/-55,577		

The formula in Excel is:

$$= sqrt(38677\text{^}2 + 39911\text{^}2) = 55{,}577$$

or

$$= \mathrm{SQRT}(\mathrm{F8\text{^}2} + \mathrm{F9\text{^}2})$$

Remember to remove the "+/-" from the cell values.

Example: % of Persons < 65 in Poverty

Subject	California					
	Total		Below poverty level		Percent below poverty level	
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error
Under 65 years	32,416,401	+/-11,776	5,375,800	+/-55,577	16.6%	0.2

The formula to calculate the MOE of a percentage:

$$MOE(sum) = +/- \frac{\sqrt{[MOE(num)]^2 - [proportion]^2 * [MOE(den)]^2}}{[den]}$$

The formula in Excel is (remember to remove the "+/-"):

$$= \text{ABS(SQRT(55577^2 - (0.166^2*11776^2)))/32416401} = 0.002$$

Note that the MOE for a percent is not shown with a percent sign – because this is not a percent but the percentage points by which the MOE of the percentage should be changed.

MOEs of "Zero" cells

- For a given geographic level, the Census Bureau assigns the same MOE for zero estimates – most of the time it will be "132" in California.
- The Census Bureau recommends that you DO NOT include the MOE for the "0" cells in your MOE calculations.
- For more information see: www.census.gov/acs/www/Downloads/data_documentation/Accuracy/MultiyearACSAccuracyofData2011.pdf

27

.

ACS Data for Block Groups

- ACS data for Block Groups are not available on American FactFinder.
- Not every table in the 5-year file will be available by block group.
- For a complete list of 2006-2010 tables and their geography, see the documentation at www2.census.gov/acs2010_5yr/summaryfile/ACS_2006-2010_SF_Tech_Doc.pdf
- There is an extraction tool that will provide block group data from the 2006-2010 ACS.
  The User Guide is at www2.census.gov/acs2010_5yr/summaryfile/UserTools/SFRetrievalToolUsersGuide.pdf

Other Useful Resources

- ACS Main Web Page – www.census.gov/acs/
- Comparing ACS years and ACS to Census 2000 – www.census.gov/acs/www/guidance_for_data_users/comparing_data/
- Handbooks for Data Users on the American Community Survey – www.census.gov/acs/www/guidance_for_data_users/handbooks/
- ACS Detailed Methodology – www.census.gov/acs/www/methodology/methodology_main/
- California State Data Center – www.dof.ca.gov/research/demographic/ Selected data extracts from Census 2010 and ACS for state, county, places. Census 2010 for census tracts, ZIP Code areas, and school districts.

Stay Up-to-Date

- On the Demographic Research Unit web page, sign up for the DRU e-mail announcements list. You will get announcements of data releases and workshops.
- Subscribe at mailinglists.dof.ca.gov/mailman/listinfo/demographicresearch/

30

Questions? Call Us.

California State Data Center
www.dof.ca.gov/research/demographic/

ficalpop@dof.ca.gov
(916) 323-4086

Thank you!

31