Estimating and Mapping Reliability for American Community Survey Data

Written by Barbara M. Parmenter and Janet Lau, revised 2/14/2013

1. OVERVIEW OF THE EXERCISE
2. BRIEF INTRODUCTION TO AMERICAN COMMUNITY SURVEY
3. CALCULATING THE COEFFICIENT OF VARIATION (CV)
4. EXPLORING THE COEFFICIENT OF VARIATION (CV) FOR MEDIAN HOUSEHOLD INCOME IN THE BOSTON AREA IN EXCEL
5. MAPPING ACS ESTIMATES AND CVS IN ARCMAP – HOUSEHOLD MEDIAN INCOME EXAMPLE
6. YOUR TURN TO CALCULATE SE AND CV'S FOR COMMUTE MODE ACS DATA
7. MAPPING THE COEFFICIENT OF VARIATION FOR MEANS OF TRANSPORTATION TO WORK

1. Overview of the exercise

This is an exercise designed to be done in the Tufts GIS Lab and to familiarize students who have been introduced to using American Community Survey data from the Census Bureau website (<u>http://www.census.gov/</u>).

After completing this exercise, you should:

- Know how to calculate the Coefficient of Variation (CV) from the Margin of Error (MOE)
- Know how to visualize and map out ranges of error inherent in your ACS estimates
- Understand the error trends, benefits and limitations of ACS

More information about American Community Survey estimates is available through the U.S. Census Bureau <u>ACS Guidance for Data Users Main</u> web site. On that site, the <u>Handbooks</u> for different types of users are especially helpful.

2. Brief Introduction to American Community Survey

The American Community Survey (ACS) is an ongoing survey that provides data every year. The ACS asks more detailed questions than the decennial Census for the purpose of giving communities current information to plan investments and services. To help communities, state governments, and federal programs, the ACS asks questions about:

- Age (also in decennial census)
- Race (also in decennial census)

• Sex (also in decennial census)

- Household and family and relationships (also in decennial census)
- Income and benefits
- Health insurance
- Education

- Veteran status
- Disabilities
- Where people work and how they get there
- How much people pay for some essentials

3. Calculating the Coefficient of Variation (CV)

Since ACS estimates are based on a sample, data are published with margins of error (MOEs) for every estimate. These MOEs are based on a 90-percent confidence level and give users an idea of how reliable or precise estimates actually are. As the MOE gets larger, relative to the size of an estimate (the smaller the sample, the larger the MOE), the estimate becomes less reliable. A measure called the "coefficient of variation" (CV) is used to discern the level of reliability of an estimate. To calculate the CV, users first calculate a standard error from the MOE, and then divide that standard error by the estimate. The exact steps for the ACS MOE are below.

a. Calculate the standard error: SE = MOE/1.645

b. Calculate the Coefficient of Variation: CV = SE/Estimate * 100

CVs are a standardized indicator of the reliability of an estimate and can help users to quickly gauge the usability of that estimate. <u>The lower the CV, the more reliable the data</u>. There are no hard-and-fast rules for setting acceptable thresholds of reliability. The amount of acceptable error in an estimate depends on the analysis or situation at hand.

What is a good CV?

U.S. Census case studies:

- High reliability: CVs less than 15%
- Medium Reliability: CVs between 15-30% be careful
- Low Reliability: CVs over 30% use with extreme caution

The cut-off depends on the application of the data – estimates with CVs beyond 15 percent may be acceptable for a very general portrait of an area; estimates with lower CVs may be needed if those estimates need to be more precise.

ESRI, the makers of ArcGIS, in its own use of ACS, uses the following reliability thresholds:¹

- High Reliability: Estimates with CVs less than or equal to 12, indicating that the sampling error is small relative to the estimate and the estimate is reasonably reliable
- Medium Reliability: Estimates with CVs between 12 and 40 use with caution
- Low Reliability: Estimates with large CVs over 40, indicating that the sampling error is large relative to the estimate. The estimate here is considered very unreliable.

¹ ESRI also warns that users should be aware that these are generalized thresholds

Some estimates may not indicate reliability. In these cases, either the estimate or MOE is missing, or the estimate is zero. When the Census Bureau deems that CVs are too high for tabulation, the entire table provided by ACS may be blanked-out or suppressed.

4. Exploring the Coefficient of Variation (CV) for median household income in the Boston area in Excel

To get you started understanding how the CV can be useful, we have prepared an Excel file that has ACS data by tract for the Boston area showing household median income, the margin of error (MOE) as given by the Census Bureau, the Standard Error (SE), the CV, and the lower and upper bounds of the confidence intervals for each tract based on the MOE.

- 1. Go to S:\classes\UEP_ENV
- 2. Copy whole folder "American Community Survey Error Exploration" to your Desktop
- 3. Make writable right-click on folder \rightarrow properties \rightarrow uncheck read-only
- Using Windows Explorer, go to the following folder: American Community Survey Error Exploration \AFF_data_tables\ Median_HH_Income_tract
- 5. Open the following three Excel files and examine them:
 - a. ACS_10_SF4_B19013_metadata.csv this is the metadata file for the ACS data
 - b. ACS_10_SF4_B19013_Med_HH_Income.xlsx this is the ACS data modified for use in ArcGIS
 - c. ACS_10_SF4_B19013_Med_HH_Income_Error_analysis_worksheet.xlsx this is the ACS data with columns added that calculate the Standard Error (SE), the CV, and the upper and lower bounds of the 90% confidence interval for each tract's household median income estimate
- 6. In the ACS_10_SF4_B19013_Med_HH_Income_Error_analysis_worksheet.xlsx, click in the SE, CV, and 90LowBND and 90UpBND columns to see the calculation we used to derive each of these.
- 7. Write down each of these so you will be able to use them later. The graphic below shows how the SE is calculated as the MOE value in column D (given by the Census Bureau) divided by 1.645:

	E2	- (*	$f_{x} = D$	2/1.645				
	А	В	С	D	E	F	G	Н
1	GEOid	GEOid2	HH_Med_Inc	HH_inc_MOE	SE	CV	90LowBND	90UpBND
2	14000000	25017300100	80143	9661	5872.948328	7.328086	70482	89804
3	14000000	25017301101	68940	7912	4809.726444	6.976685	61028	76852
4	14000000	25017301102	85313	7925	4817.629179	5.647005	77388	93238
5	140000U	25017310100	32190	7570	4601.823708	14.29582	24620	39760
6	140000011	25017310200	65290	9184	5582 978723	8 551047	56106	74474

Note how the CV varies a lot. Smaller CVs mean more reliable estimates. Also take a look at the ranges of household median income between the lower bound and the upper bound. Even with small CVs, these can be large ranges.

Close all the Excel files before the next step! We will be using these in ArcMap now.

5. Mapping ACS estimates and CVs in ArcMap – Household Median Income Example

We will now map this information to see both the estimates for Household Median Income and the CV for each tract.

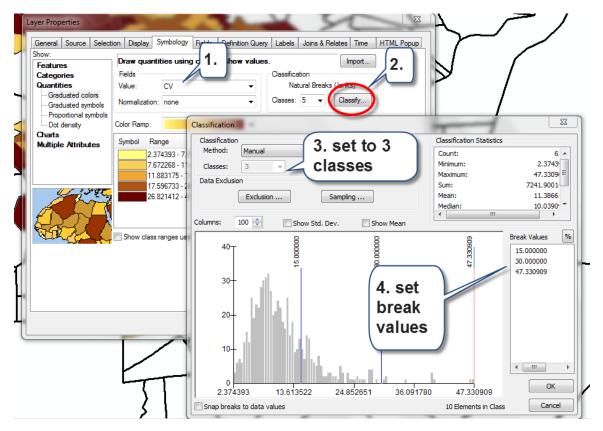
- 1. Double-click on "Exploring Error in the American Community Survey.mxd"
- 2. Join *Median Household Income* table to *Mass Census Tract* shape file as shown below:

Join Data
Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.
What do you want to join to this layer?
Join attributes from a table
1. Choose the field in this layer that the join will be based on:
GEOID 👻
2. Choose the table to join to this layer, or load the table from disk:
🖩 Median_Household_Income\$ 🗾 🖻
Show the attribute tables of layers in this list
3. Choose the field in the table to base the join on:
GEOid2
Join Options
Keep all records
All records in the target table are shown in the resulting table. Unmatched records will contain null values for all fields being appended into the target table from the join table.
Keep only matching records
If a record in the target table doesn't have a match in the join table, that record is removed from the resulting target table.

3. Create a map of Household Median Income – 5 classes by <u>quantiles</u>

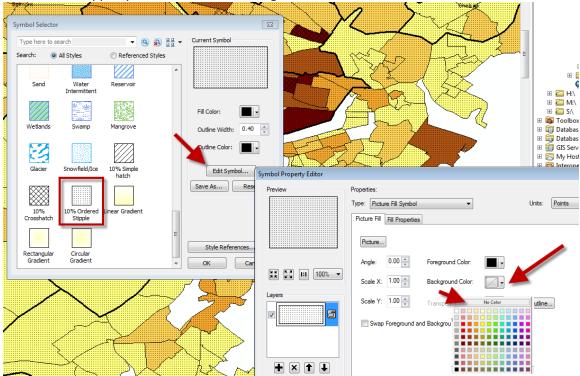
eneral Source Select	ion Displa	ay Symbology Fields Definition (Query Labels Joins & Relates Time	e HT				
now:	Draw o	uantities using color to show va	alues.					
Features Categories	Fields		Classification					
Quantities	Value:	HH_Med_Inc -	Quantile					
Graduated colors Graduated symbols	Normaliz	ation: none 💌	Classes: 5 🗸 Classify					
 Proportional symbols Dot density 	Color Ramp:							
Charts	Symbol	Range	Label					
Multiple Attributes	Symbol							
		12429.000000 - 47371.000000	12429.000000 - 47371.000000					
		47371.000001 - 63425.000000	47371.000001 - 63425.000000					
		63425.000001 - 79241.000000	63425.000001 - 79241.000000					
11 1 2. 9 25		79241.000001 - 99212.000000	79241.000001 - 99212.000000					
		99212.000001 - 218667.000000	99212.000001 - 218667.000000					

- 4. Right-click on Mass Census Tract layer and choose copy
- 5. Right-click on *Layers* and choose **Paste Layers** so that you have a second copy of the tracts
- Use the symbology properties on this *copy* to map the CV 3 classes, with breaks at 15, 30, and max value (remember to set *Normalization* to **NONE**)

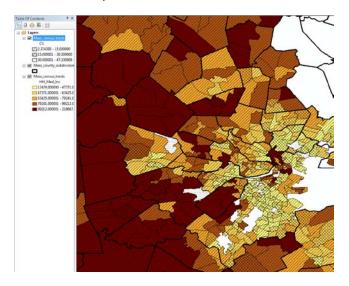


7. Press OK and OK again to see the map of the CV% You can clearly see where the higher and lower CV's are, but you can't see the Household Median Income map this way.

8. Go back to the **Symbology - Properties** your Census Tracts CV layer.You can symbolize the three classes with the styles **Simple Hatch**, **Crosshatch**, and **Ordered Stipple**. Note for Ordered Stipple, you have to set the background color to *No Color*:



9. Now you have estimates overlaid with CV ranges showing how reliable those estimates are. Explore the results.



In this example, only a few tracts have CV's higher than 30%. A good number have CVs in the 15-30% range. Most have CVs less than 15%. We would deem the data for these latter tracts to be

fairly reliable for a general view of household median income in the Boston area, within the 90% confidence intervals.

Exit out of ArcMap – you don't need to save your mapfile.

6. Your Turn to Calculate SE and CV's for Commute Mode ACS Data

The household median income data appeared to be fairly reliable – that's because most every household sampled has an income to report, so the sample base was relatively large.

We have also data detailing how people who work get to work. Officially this is *Table B08301* – *Means of Transportation to Work* from the ACS for 2006-2010 as well. The universe for this data is **workers 16 and over**.

- Using Windows Explorer, go to the following folder: American Community Survey Error Exploration \AFF_data_tables\ commute_means_tract
- 2. **Open** the metadata file for an explanation of all the fields ACS_10_SF4_B08301_metadata.csv
- 3. Now open ACS_10_SF4_B08301_commute_means_modified.xlsx You see that we have modified the original downloaded table to just show the following:
 - Tot the estimate of the universe for this table (here, the total number of workers 16 and over)
 - TotMOE the margin of error for Tot
 - CarTot the estimate of workers who drove a car, van, or truck to work (plus CarMOE)
 - PubTot the estimate of workers who took public transit to work (plus PubMOE)
 - BikeTot the estimate of workers who biked to work (plus BikeMOE)
 - WalkTot the estimate of workers who walked to work (plus WalkMOE)

Insert additional columns in this Excel file to calculate the SE and CV for all 5 of the main estimates (total, car total, public transit total, bike total, and walk total).

Note the Insert Sheet Columns button is on the Home tab, over to the right side:

	Calculation Note	* * *	Inse		te Form		Σ Au J Fil 2 Cl		ZU Sort &	Find & Select *	
				Insert (Cells Sheet <u>R</u> o	ws		Ed	liting		
Q	R	S	÷	Insert S	iheet <u>C</u> o iheet	lumns	5	/	W	X	

Here are the formulas again:

a. First calculate the standard error: SE = MOE/1.645 – e.g., for BikeSE, this would be BikeMOE/1.645.

b. Then calculate the Coefficient of Variation: CV = SE/Estimate * 100, where "estimate" is the data value for which you are calculating the CV - e.g., for BikeCV, this would be (BikeSE/BikeTot)*100

If you get stuck, we have an Excel file with the error columns calculated - ACS_10_SF4_B08301_commute_means_error_filled out.xlsx

Note that for the PubCV, BikeCV, and WalkCV, there are errors when you calculate the formula (#DIV/0) – this is because the estimate is 0 in some of these tracts, and Excel cannot divide by 0. This will create problems in ArcMap, so it is important to convert these #DIV/0 errors to some number like 9999 before proceeding. We have already done this in the Excel file,

ACS_10_SF4_B08301_commute_means_error_filled out_fix_div_zero.xlsx- so use this one if you need to.

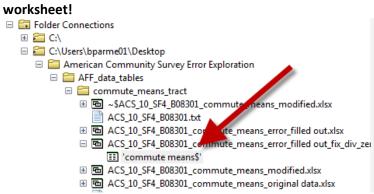
We fixed this in Excel by clicking on the **DATA** tab, then doing a **SORT** on the entire worksheet, based on the CV column we wanted to fix, and sorting *Largest to Smallest* – this got all the #DIV/0 values in one area, and we replaced these with 9999's.

PO	mulas	Data R	wiew view	Add-ens	ACTO	pat						
ons	Refree All y	Connection Properties Edit Links nections	ns 24 27 24 Sort	J. A	Clear Reapply Advance		Remove Data Duplicates Validati Data	Consolidate	What-If Analysis *	Group L		Show Detail Hide Detail
02/	1.6											
	H	1		к		L	M	N		o	Р	Q
DE	CarTot	CarMOE	PubTot	PubMOE	P	ubSE	PubCV	Bike_Tot	Bike_MC	DE	BikeSE	BikeCV
325	1337	324		0	127	77.20364742	9999		0	127	76.96969697	#DIV/01
154	713	136		0	127	77.20364742	9999		D	127	76.96969697	#DIV/01
123	1674	128		0	127	77.20364742	9999		D	127	76.96969697	#DIV/0!
313	3128	323		0	127	77.20364742	9999		0	127	76.96969697	#DIV/01
200	1778	191	Sort		-	-	-	Contraction of the		7 ×	76.96969697	#DIV/01
142	1714	147	Son	-	-	the summaries	-				76.96969697	#DIV/0!
224	1379	229	PI Add Lev	el X Delei	te Level	Copy Level		tions	V My data	has beader	s 76.96969697	#DIV/01
105	2942	397								-	76.96969697	#DIV/0!
175	2281	169				Sort On		Order			10.90909091	99.173553
201	2235	195	Sorton	ueCV	-	Values		Largest to !	smallest		76.96969697	#DIV/01
272	2699	266								-	76.96969697	WDIV/01
160	3417	387	-								76.96969697	#DIV/0!
216	2488	227									76.96969697	#DIV/01
183	2296	188									76.96969697	WDIV/01
293	2586	287									76.96969697	#DIV/0!
156	1765	149									76.96969697	#DIV/0!
97	718	96							ж	Cancel	76.96969697	WDIV/0!
353	3241	347		-	-		-			-	76.96969697	#DIV/01
148	659	152	2	0	29	17.62917933	88.14589666		0	127	76.96969697	#DIV/01
237	1094	217	4	10	55	33.43465046	83.58662614		D	127	76.96969697	#DIV/0!

When you are done, save and exit out of Excel.

7. Mapping the Coefficient of Variation for Means of Transportation to Work

- Start up ArcMap again and add the Mass Census Tracts shape file from your American Community Survey Error Exploration \geography\ folder
- Now add to the ArcMap session either your own modified Excel file or our modified one (ACS_10_SF4_B08301_commute_means_error_filled out_fix_div_zero.xlsx) – remember you have to click on the + sign by the Excel file to see and drag in the



3. **Right-click** on the tracts again, and **join** the *Commute Means* table to Tracts. See the Join dialog box below:

Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.	
What do you want to join to this layer?	
Join attributes from a table	•
1. Choose the field in this layer that the join will be based on:	
GEOID	
2. Choose the table to join to this layer, or load the table from the	
💷 'commute means\$'	
Show the attribute tables of layers in this list	
3. Choose the field in the table to base the join on:	
GEOid2 🗸	
Join Options	
○ Keep all records	
All records in the target table are shown in the resulting table. Unmatched records will contain null values for all fields being appended into the target table from the join table.	

4. Try mapping the percent of bike commuters out of total workers (hint: you need to normalize *Bike_tot* by *Tot*):

General Source Select	ion Display	Symbology Fields	Definition Query	Labels Joins &	Relates Time	HTML Popup
now: Features	Draw quant	ities using color	to show value	assification	Import	
Categories Quantities Graduated colors Graduated symbols Proportional symbols Dot density	Value: Normalization	Bike_Tot		Natural Brea Classes: 5 -	ks (Jenks) Classify	
Charts Multiple Attributes	0.0 0.0 0.0 0.0 0.0	nge 0000000 - 0.00787 07874017 - 0.0254 25477708 - 0.0493 49380208 - 0.0924 92476490 - 0.1720 ranges using featu	77707 0.00 30207 0.02 76489 0.04 59825 0.09	el 0 - 0.007874 7875 - 0.02548 549 - 0.04938 939 - 0.09248 249 - 0.1721	Advanced	Y

- 5. What do you think? Does this look reasonable?
- 6. Create a copy of the Mass Census Tracts in the Table of Contents.
- 7. Try mapping the CV (remember to set *Normalization* back to **NONE**)
- 8. Try setting the number of classes to 0 and the class breaks to 15, 30 and the max value. What happens? You see the minimum value for the CV% is 23, so you can't set the first break lower than that! Leave it and set the second break at 30. Leave the max alone.
- 9. Make a map of this over your Census Tracts showing bike commuters.