Assessment of an Imputation Process Used in the 2017 Census of Agriculture

Tara Murphy
Habtamu Benecha
Denise A. Abreu
Darcy Miller

The findings and conclusions in this presentation are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

National Agricultural Statistics Service (NASS)

- Conducts over 100 surveys each year, as well as the Census of Agriculture

- Prepares more than 500 reports annually covering every facet of U.S. agriculture

  For example:

  - Production and food supplies
  - Prices paid and received by farmers
  - Farm income and finances
  - Number of farms and land in farms

Census of Agriculture

- Conducted every 5 years (years ending in 2 and 7) using list-based frame
  - Census Mail List (CML)
- Count of all U.S. agricultural operations
  - Any place from which $1,000 or more of agricultural products were produced and sold or normally would have been sold during the year
- Only source of uniform, comprehensive agricultural data for every county or county equivalent in the U.S.
- Leading source of information on characteristics of people operating farms

Background

- In 2012, NASS began to use Dual System Estimation (DSE) to adjust for various sources of error
  - Undercoverage
    - Not all agricultural operations appear on the CML
  - Non-response
    - Not all agricultural operations on the CML respond
  - Misclassification
    - Incorrectly classifying farms as non-farms or non-farms as farms

- Requires two independent surveys
    - Census of Agriculture (COA)
    - June Area Survey (JAS)

June Area Survey (JAS)

- Area-frame based
- Conducted annually via in-person interviews
- Segments of land sampled
- Sampled segments divided into tracts representing unique land operating arrangements
- Measures the incompleteness of the CML

DSE Dataset

- Matched dataset consisting of:
  - Census of Agriculture records overlapping JAS tracts (not all Census records)
  - All JAS tracts

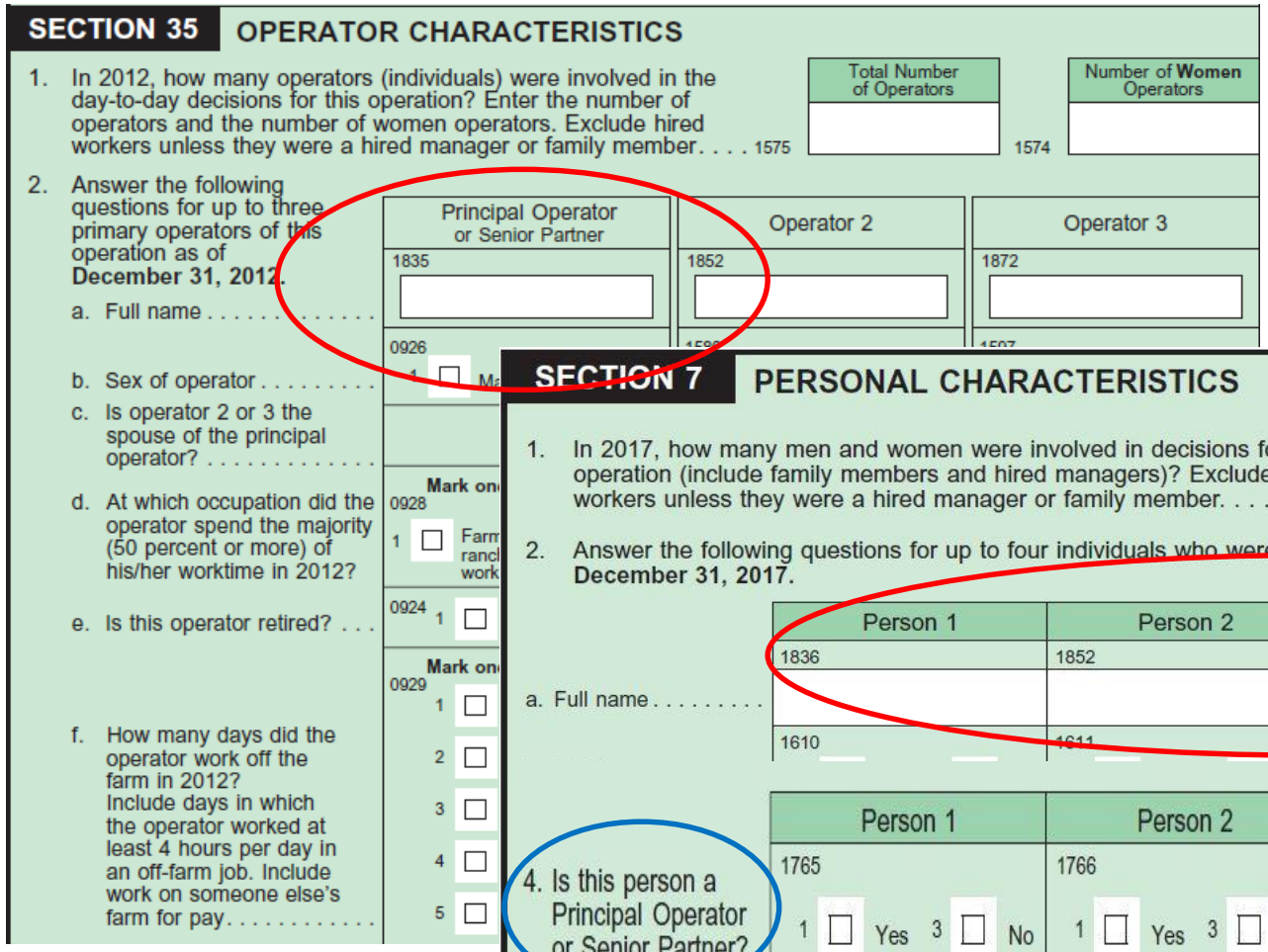
- Records in the 2017 JAS sample (120,000) were matched to the 2017 Census (3 Million) using probabilistic record linkage

- This dataset is the foundation for modeling probabilities of coverage, response, and misclassification of farm/non-farms

Problem

- The 2017 demographics section of the Census of Agriculture questionnaire was redesigned to allow up to **four principal producers** per farm, whereas the 2017 JAS collected demographic information on only **one principal operator**

2012 and 2017 Census of Agriculture Demographic Sections



SECTION 35 OPERATOR CHARACTERISTICS

1. In 2012, how many operators (individuals) were involved in the day-to-day decisions for this operation? Enter the number of operators and the number of women operators. Exclude hired workers unless they were a hired manager or family member. . . . 1575

Total Number of Operators		Number of **Women** Operators
	1574	

2. Answer the following questions for up to three primary operators of this operation as of **December 31, 2012.**

	Principal Operator or Senior Partner	Operator 2	Operator 3
a. Full name	1835	1852	1872
b. Sex of operator	0926 1 ☐ Ma		
c. Is operator 2 or 3 the spouse of the principal operator?			
d. At which occupation did the operator spend the majority (50 percent or more) of his/her worktime in 2012?	Mark on 0928 1 ☐ Farm ranch work		
e. Is this operator retired?	0924 1 ☐		
f. How many days did the operator work off the farm in 2012? Include days in which the operator worked at least 4 hours per day in an off-farm job. Include work on someone else's farm for pay.	Mark on 0929 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐		

SECTION 7 PERSONAL CHARACTERISTICS

1. In 2017, how many men and women were involved in decisions for this operation (include family members and hired managers)? Exclude hired workers unless they were a hired manager or family member. 1571

Men		Women
	1574	

2. Answer the following questions for up to four individuals who were involved in the decisions for this operation as of **December 31, 2017.**

	Person 1	Person 2	Person 3	Person 4
a. Full name	1836	1852	1872	1873
	1610	1611	1612	1613
4. Is this person a Principal Operator or Senior Partner?	1765 1 ☐ Yes 3 ☐ No	1766 1 ☐ Yes 3 ☐ No	1767 1 ☐ Yes 3 ☐ No	1768 1 ☐ Yes 3 ☐ No

2017 JAS Demographic Section

SECTION P - OPERATOR CHARACTERISTICS

1. **Age of operator as of December 31, 2016?**
[*Check (√) age of operator and enter code.*]

		Code
☐ Less than 25 years	= 1	821
☐ 25 - 34 years	= 2	
☐ 35 - 44 years	= 3	
☐ 45 - 54 years	= 4	
☐ 55 - 64 years	= 5	
☐ 65 years and over	= 6	

2. **Ethnicity of operator?** [*Check (√) one and enter code.*]

☐ Hispanic or Latino	= 1	822
☐ Not Hispanic or Latino	= 3	

3. **Race of operator?** [Check (√) one or more and enter code.]

☐ White	= 1	823
☐ Black or African American	= 2	
☐ American Indian or Alaska Native (*Specify tribe*:)	= 3	
☐ Asian	= 4	
☐ Native Hawaiian or Other Pacific Islander	= 5	

4. **Sex of operator?** [Complete from observation and enter code.]

☐ Male	= 1	824
☐ Female	= 2	

	Year
5. **In what year did the operator begin to operate any part of this operation?**	896

Problem

- JAS records are a crucial element for modeling coverage of the CML

- Because COA publications include demographic estimates at the county level, it is essential for the demographic variables to be included in the model

- There are missing demographic variables, associated with producers 2, 3, and 4, in the matched dataset for the JAS records

Imputation Strategy for DSE

- Impute producers 2, 3, and 4 on the JAS, using hot deck method
  - Form groups within regions based on demographic characteristics of
    - JAS records (recipients): Only person reported on the JAS record
    - COA records (donors): Person listed in the first column of the COA (most often primary producer)
  - For each JAS record in a group, producer information for producers 2 – 4 were drawn from a single COA record in the same group to impute information for producers 2 – 4 on the JAS record
    - Distribution of number of producers was preserved
      - e.g. imputed values could all be 0, preserving the distribution of single producer farms in the JAS

Imputation Use in DSE

- Once demographic variables for potential producers 2, 3, and 4 on the JAS were imputed, a full matched dataset was formed

- This dataset was used in DSE for the published estimates

Study

- Perform DSE modeling, with variable selection, using data where producers 2, 3, and 4 were not imputed for the JAS

- Compare study DSE estimates to published DSE estimates for demographic variables

Expectation

- Based on research used to redesign the 2017 COA demographics section, there is an expectation to capture more young and female producers (Ridolfo, 2016)

Findings

- Both the number of young producers and the number of female producers increased with the imputation efforts

- Note: Due to confidentiality, we are not able to show exact estimates; findings are shown in percent differences

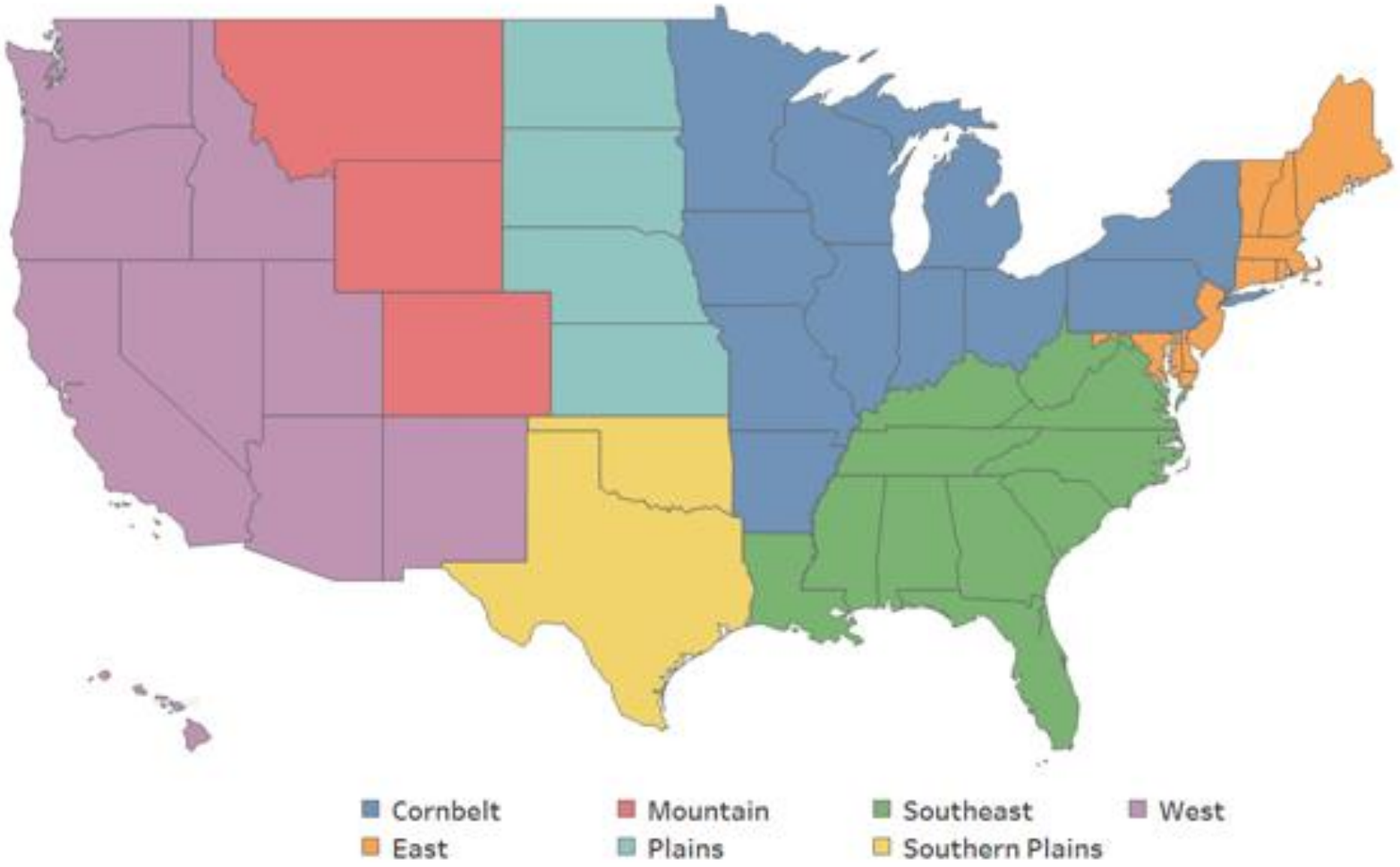
Young Producers (age < 25)

- Difference in farms with at least one *principal* producer aged less than 25 was found to be significant at the $p=0.11$ level nationally

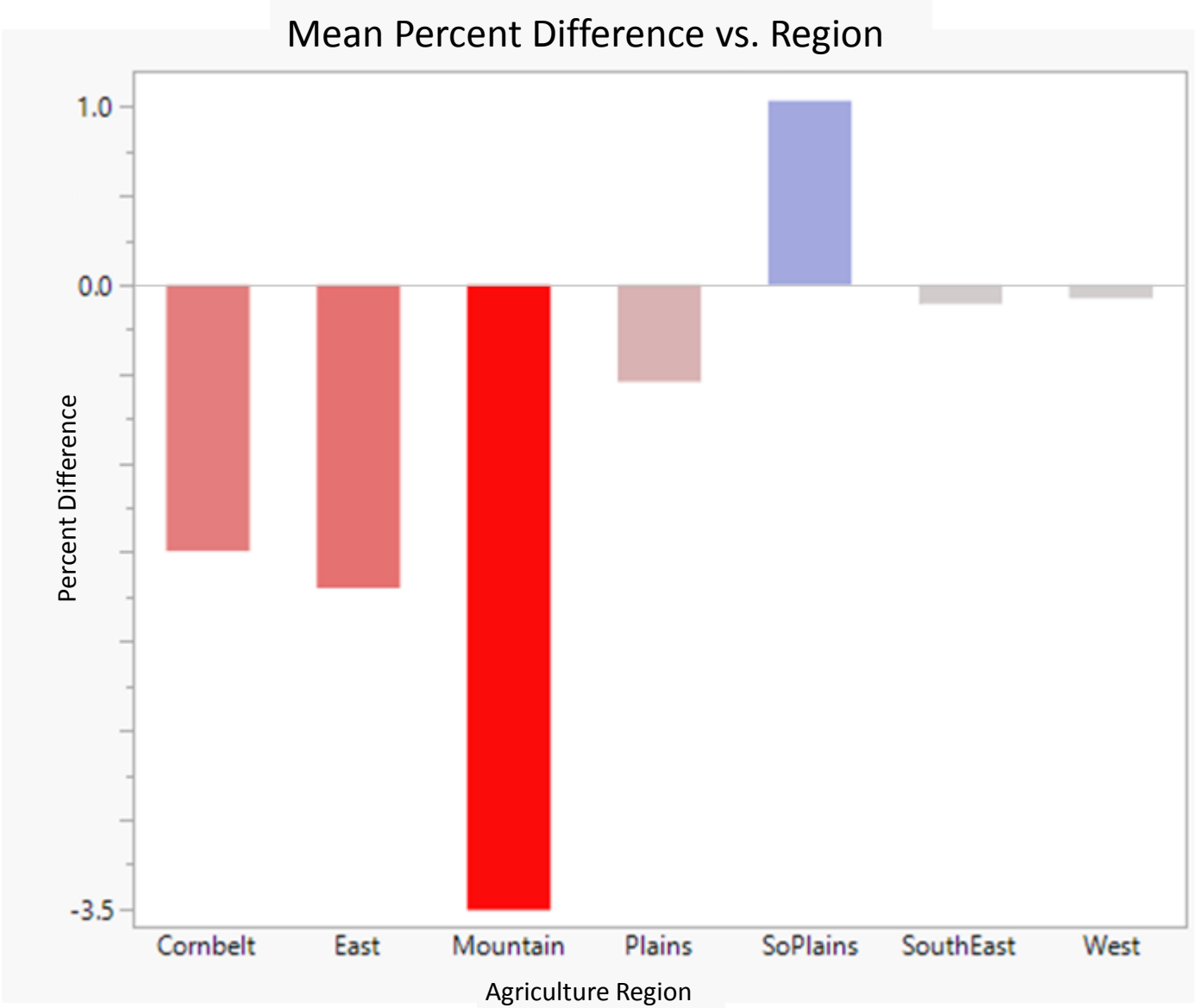
- Difference in farms with *any* producer aged less than 25 was found to be significant at the $p<0.01$ level nationally

- When national level estimates were found to be significant, regional analysis occurred

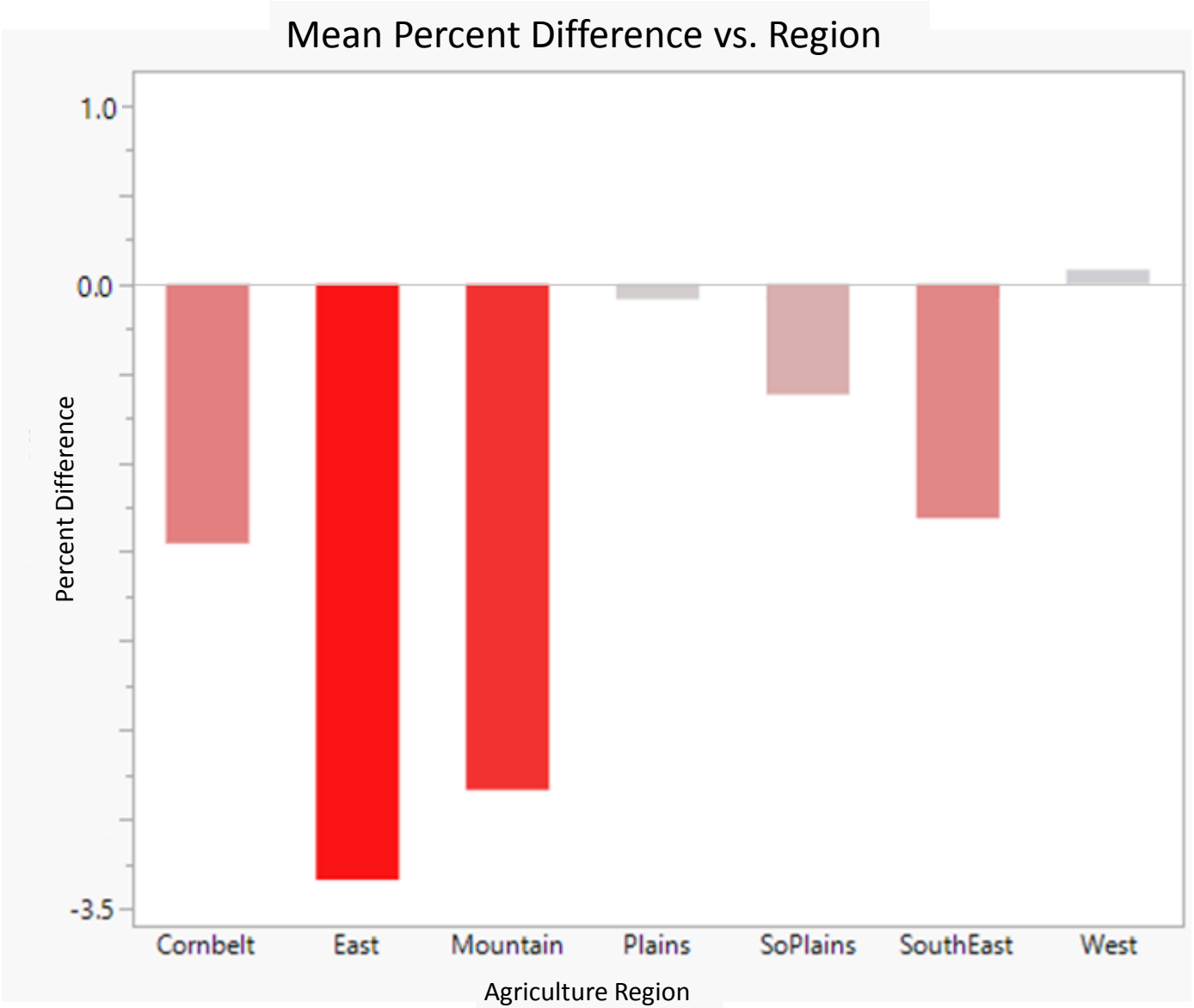
Agriculture Regions



Farms with at least one *Principal* Producer Aged < 25



Farms with *any* Producer Aged < 25

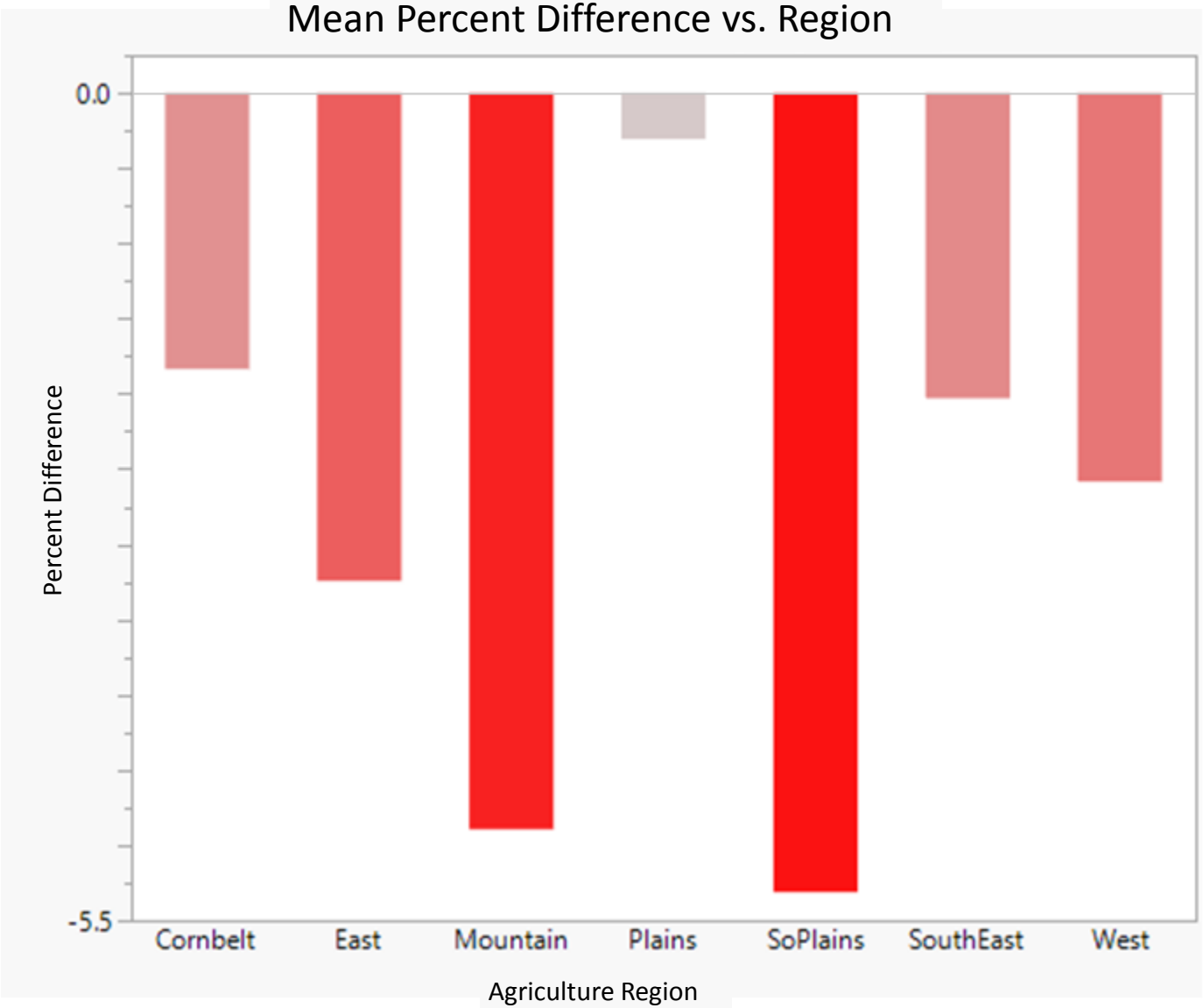


Female Producers

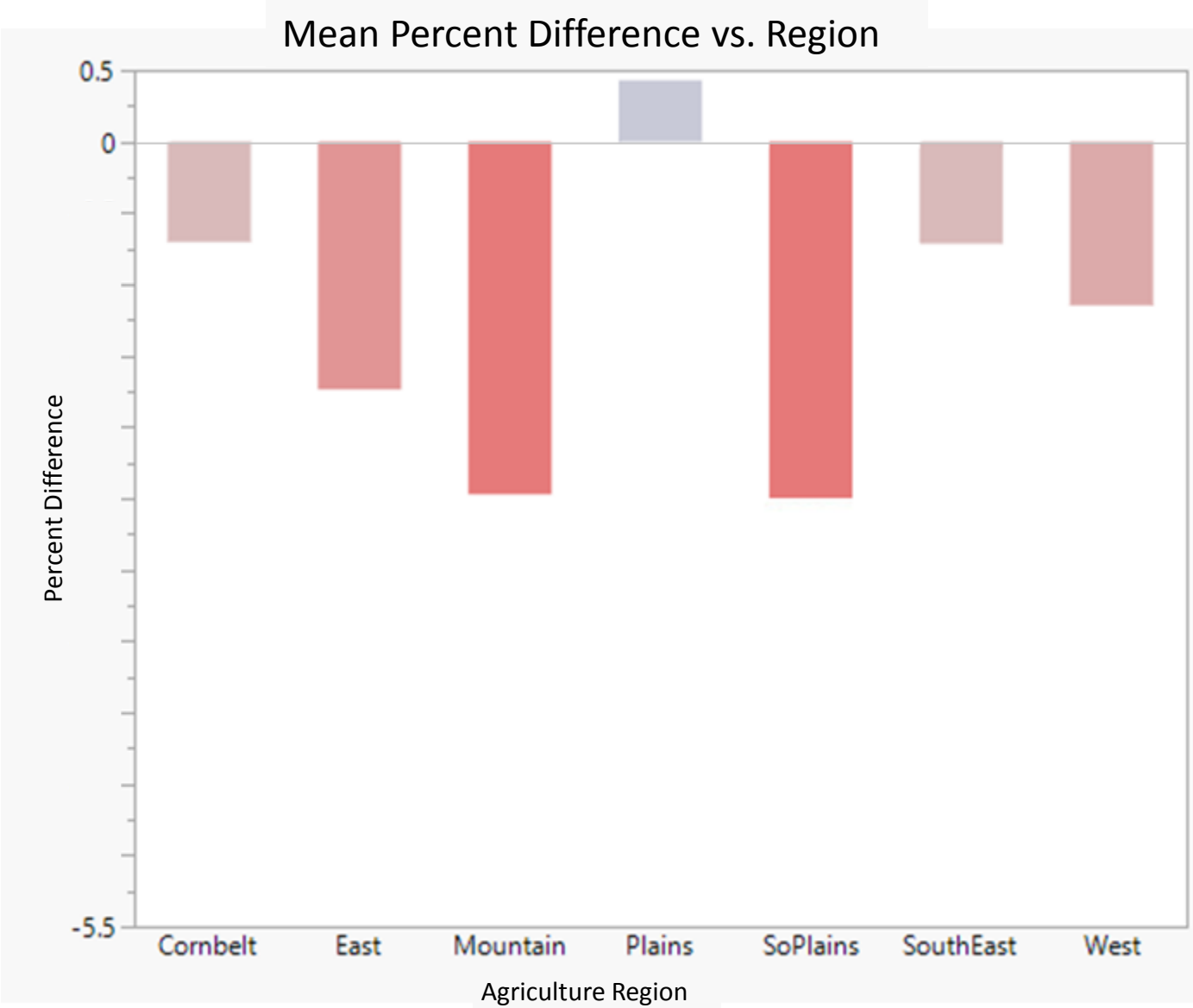
- Difference in farms with at least one female *principal* producer was found to be significant at the $p<0.01$ level nationally

- Difference in farms with *any* female producer was found to be significant at the $p<0.01$ level nationally

Farms with at least one Female *Principal* Producer



Farms with *any* Female Producers



Further Findings

- Variables statistically significant at $p<0.01$ nationally:
  - farms with any *principal* producer between the ages of 35 and 44
  - farms with *any* producer between the ages of 35 and 44
  - farms with any *principal* producer between the ages of 45 and 54
  - farms with *any* producer between the ages of 45 and 54
  - farms with any male *principal* producer
  - farms with *any* male producer

- Number increased with the imputation efforts

Future Work

- Determine whether the imputation of the JAS or the redesign of the COA captured more producers
  - Simulation studies

- Decide if the demographics section of the JAS should be redesigned to allow up to four producers

Thank you!

Tara Murphy

Tara.Murphy@usda.gov