Evaluating Publication Rules for the County Agricultural Production Survey Using Simulation

Andrew J. Dau, Nathan B. Cruze, Joe L. Parsons, and Linda J. Young
USDA/NASS
Disclaimer

• The findings and conclusions in this presentation are those of the author and should not be construed to represent any official USDA or U.S. Government determination or policy.
County Agricultural Production Survey

• Annual Survey
  – Small Grains
  – Row Crops

• Publication of county level estimates of planted and harvested acreage – yield and production

• 2019 and earlier publication rule:
  – 30 reports
    OR
  – 3+ reports AND >= 25% coverage.
Goals/Objectives

• Publish accurate estimates

• Publish as much as possible
Administrative Data

• USDA program

• Excellent Coverage

• For this simulation, treating as the population.
Monte Carlo Simulation

US Level Dataset -> County Level Data

- n = 3, 1000x
- n = 5, 1000x
- n = 10, 1000x
- n = 20, 1000x
- n = 30, 1000x
Simulation Design

• Probability Proportionate to Size (PPS)

• Weighted Yield (Ratio Production/Acres)

• Variance Estimated

• Coefficient of Variation Calculated
Pennsylvania County Example
Coefficient of Variation (CV)
Pennsylvania County Example

Yield

![Box plot showing yield variation across different sample sizes for single Pennsylvania County](image-url)
Low CV & Far From Truth

• What is a “low” CV?
  – For this example, we will use 20%

• What is “far” from population parameter?
  – For this example, we will use 5 bushels/acre
Pennsylvania County Example

n=3
Pennsylvania County Example

difference, coverage, cv, and sample size
Pennsylvania County Example

Old Publication Rule
Expanding

• For visualization purposes
  – 20 random counties
  – 10 different states
20 Counties
Difference, Coverage, CV, and Sample size
20 Counties
Old Publication Standard
20 Counties
Old Publication Standard – Not Published
20 Counties
old publication standard
20 Counties
Old Publication Standard = “Yes”
A new standard

- 30 reports
  OR
- 3 reports and 25% coverage
  OR
- 10 reports and 10% coverage ***new***
20 Counties
New Publication Standard
Old vs New

Old Publication Standard

New Publication Standard
Old vs New

Old Publication Standard

New Publication Standard
New Publication Standard
Simulation estimates published under new standard, but not old
## Summary

<table>
<thead>
<tr>
<th>Pub Rule</th>
<th>Total Simulations</th>
<th>Number Published</th>
<th>Number Published With Greater than 5 Bushels/Acre from Population</th>
<th>Good Pub Rate (% published within 5 bushels/acre of Population)</th>
<th>Absolute Average Difference from Population bushels/acre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Rule</td>
<td>100,000</td>
<td>30,869</td>
<td>9,458</td>
<td>0.69</td>
<td>4.09</td>
</tr>
<tr>
<td>Proposed Rule</td>
<td>100,000</td>
<td>46,583</td>
<td>16,453</td>
<td>0.65</td>
<td>4.57</td>
</tr>
<tr>
<td>CV &lt; 20%</td>
<td>100,000</td>
<td>98,268</td>
<td>50,313</td>
<td>0.49</td>
<td>12.02</td>
</tr>
</tbody>
</table>
Additional Factors

• County Agricultural Production Survey modeled indications
  – Cruze et al (Statistical Science 2019 vol 34)
  – Strength added to yield indication from National Commodity Crop Productivity Index (NCCPI) and location

• Additional References
  – Bell and Barboza (2012) ICES proceedings
  – Cruze et al (2018) JSM proceedings
  – Cruze and Young (2019) JSM proceedings
Conclusion

• One size fits all is difficult

• Simulation allows for a variety of different rules to be explored.