Developing the Sample Design for the New Annual Integrated Economic Survey

Katherine Jenny Thompson, Senior Mathematical Statistician
Associate Directorate for Economic Programs

Any opinions and conclusions expressed herein are those of the author(s) and do not reflect the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied. (CBDRB-FY22-ESMD006-014, CBDRB-FY22-ESMD010-011, and CBDRB-FY23-ESMD010-003).
Acknowledgements

James Hunt*
Amy Newman-Smith*
Laura Bechtel
Nicole Czaplicki
Stephen Kaputa
Adam Smeltz
Justin Z. Smith
Lucas Streng
Thomas Kirk White
Robert Miller
Joseph Haskell
Shelby Plude
Michael McElhatten

Valerie Mastalski
Christina Gouvatsos
Ahntuan Ty
Suiyun Chen
Gerald A McGarvey
Yeng Xiong
Michael Christen
Quatracia (Traci) Lucky
Carol Caldwell (retired)
James Burton

*Co-leads, with myself
(katherine.j.Thompson@census.gov)
Preliminaries
Definitions

• **Sector**: an area of the economy in which businesses share the same or related business activity, product, or service (www.investopedia.com)

• **Industry**: a group of companies that are related based on their primary business activities or service (www.investopedia.com)

• **Industrial classification**: industry code assigned to an individual business, usually based on the business’ largest source(s) of revenue

• **NAICS**: North American Industry Classification System
  • Digits indicate level of detail used for classification (more digits = more criteria)
Simple NAICS Example from the Retail Trade Sector (44_45)
More Definitions

• Establishment – a single business location

• Company – any formal business entity for profit, which may be a corporation, partnership, association or individual proprietorship (https://dictionary.law.com/)

• Multi-unit company

• Single-unit company
### Annual Economic Surveys (2022)

<table>
<thead>
<tr>
<th>Survey</th>
<th>Sampling Unit</th>
<th>Estimation Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Retail Trade Survey (ARTS)</td>
<td>Company and EIN</td>
<td>National: 2- through-6-digit NAICS by Type of Operation</td>
</tr>
<tr>
<td>Annual Wholesale Trade Survey</td>
<td>Company and EIN</td>
<td>National: 2- through-5-digit NAICS by Type of Operation</td>
</tr>
<tr>
<td>Services Annual Survey (SAS)</td>
<td>Company and EIN</td>
<td>National: 2- through-6-digit NAICS by Tax Exempt Status</td>
</tr>
<tr>
<td>Annual Capital Expenditures Survey (ACES)</td>
<td>Company</td>
<td>National: 3- or 4-digit NAICS</td>
</tr>
<tr>
<td>M3 Unfilled Orders (M3UFO)</td>
<td>Company</td>
<td>National: M3 industry category</td>
</tr>
<tr>
<td>Annual Survey of Manufactures (ASM)</td>
<td>Establishment</td>
<td>National: 2- through-6-digit NAICS</td>
</tr>
<tr>
<td></td>
<td>Subnational: 2- through-4-digit NAICS (50 states plus DC)</td>
<td></td>
</tr>
</tbody>
</table>
AIES Sampling Unit

Annual Integrated Economic Survey

Same Company

Company as Sampling Unit

ASM

Estab. 1

Estab. 2

Estab. 3
Sampling Design Requirements (Totals)
Multipurpose Survey Design

National Industry Estimates

- Revenue/receipts, Annual Payroll, Inventories, Capital Expenditures, Operating Expenses...
- Disaggregated industry (NAICS) levels
  - 4-digit NAICS (NAICS4)
  - 5-digit NAICS (NAICS5)
  - 6-digit NAICS (NAICS6)
- C.V. target ≈ 2%

Subnational (Geographic) Industry Estimates

- Revenue/receipts, Annual Payroll, 1st Quarter Employment, 1st Quarter Payroll
- Aggregated NAICS levels (3-digit)
  - 23 “Direct use” states
  - 4 “Balance of region” categories
- C.V. target ≈ 15%
<table>
<thead>
<tr>
<th>Company</th>
<th>Establishment</th>
<th>Sector</th>
<th>Industry</th>
<th>NAICS6</th>
<th>State</th>
<th>2020 Payroll</th>
<th>2020 Emp</th>
<th>2017 Receipts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Dynamics, Inc.</td>
<td>Mom &amp; Pop Hardware 1</td>
<td>Retail Trade</td>
<td>Hardware Stores</td>
<td>444130</td>
<td>MO</td>
<td>600</td>
<td>20</td>
<td>1,380</td>
</tr>
<tr>
<td>Global Dynamics, Inc.</td>
<td>Mom &amp; Pop Hardware 2</td>
<td>Retail Trade</td>
<td>Hardware Stores</td>
<td>444130</td>
<td>MO</td>
<td>800</td>
<td>27</td>
<td>1,840</td>
</tr>
<tr>
<td>Global Dynamics, Inc.</td>
<td>Mom &amp; Pop Hardware 3</td>
<td>Retail Trade</td>
<td>Hardware Stores</td>
<td>444130</td>
<td>MO</td>
<td>762</td>
<td>25</td>
<td>1,753</td>
</tr>
<tr>
<td>GD Plant</td>
<td>GD Plant</td>
<td>Manufacturing</td>
<td>Bolt, Nut, Screw, Rivet, &amp; Washer</td>
<td>332722</td>
<td>TN</td>
<td>9,452</td>
<td>189</td>
<td>16,068</td>
</tr>
<tr>
<td>Dad’s Warehouse</td>
<td>Dad’s Warehouse</td>
<td>Wholesale Trade</td>
<td>Hardware Merchant Wholesalers</td>
<td>423710</td>
<td>TN</td>
<td>475</td>
<td>16</td>
<td>489</td>
</tr>
<tr>
<td>The Shop</td>
<td>The Shop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
AIES Frame Variables (Quantitative)

• Annual Payroll (2020 administrative data)
  • Measure of size

• 1st Quarter Employment (2020 administrative data)
  • Analysis (evaluation) variable (EMP)

• Revenue/Receipts (2017 Economic Census)
  • Analysis (evaluation) variable (RCPT)
Overview of AIES Sample Design Process

1. **Sampling Frame**
   - Stratification
   - Complexity
   - Certainty Companies
   - Noncertainty Companies
   - Sector Processing
     - Sector 11 + Sector 21 + ... + Sector 72 + Sector 81

2. Obtain Allocations
3. Compute Inclusion Probabilities
4. Select Stratified Sequential Samples
5. Compute JOINT inclusion probabilities
6. Obtain adjusted sampling weights
7. Compute Estimates & Variance Estimates
8. Assess Reliability

If all industries in a sector do not meet c.v. constraints, Revise allocations.
Stratification
1. Operate in (historically) designated “certainty” industries
2. Operate in a multiple sectors or industries

1. Operate in a single sector
2. Generally, operate in one industry (NAICS4)
Sector Processing:
Stratification for Noncertainty Companies

Industry is NAICS3: BROAD STRATUM

- Zero MOS
- 2 Industries

**Industry 1**
- Region 1 Direct Use State 1
- Region 1 Balance of Region
- Region 4 Direct Use State 1
- Region 4 Balance of Region

**Industry 2**
- Region 1 Direct Use State 1
- Region 1 Balance of Region
- Region 4 Direct Use State 1
- Region 4 Balance of Region

**Industry K**
- Region 4 Direct Use State 1
- Region 4 Balance of Region

> 1 Geo

Only applies to sectors 44-45 and 62

Operates establishments in more than one state

27 Geographic Strata per Industry (NAICS3)
Digression #1:
Establishment (Business)
Sample Survey Designs

- Populations are skewed!
  - Small number of large companies
  - Majority small companies
- Publish TOTALS (and Ratios)
- Sample design accounts for skewed distribution in
  - Stratification (size)
  - Allocation (# sampled units)
  - Unit inclusion probability

United States Census Bureau
Allocation

Obtain Allocations → Compute Inclusion Probabilities → Select Stratified Sequential Samples → Compute JOINT inclusion probabilities → Obtain adjusted sampling weights → Compute Estimates & Variance Estimates → Assess Reliability
### Digression #2: Allocation for a Multipurpose Survey Sampled from a Skewed Population

<table>
<thead>
<tr>
<th>Sampling Strata</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Units ($N_h$)</td>
<td>760</td>
<td>146</td>
<td>53</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Total Measure of Size ($MOS_h$)</td>
<td>39023</td>
<td>17675</td>
<td>9035</td>
<td>4421</td>
<td>6564</td>
</tr>
<tr>
<td>Average Size Per Unit</td>
<td>51.4</td>
<td>121.1</td>
<td>170.5</td>
<td>221.1</td>
<td>312.6</td>
</tr>
</tbody>
</table>

The box plots illustrate the distribution of the measure of size per unit across different sampling strata. The largest value per unit is observed in stratum 5, while the smallest value is found in stratum 2. The average sizes per unit range from 51.4 in stratum 1 to 312.6 in stratum 5. The total measure of size ($MOS_h$) across all strata is divided among the units, with stratum 5 having the highest measure of $6564$. The box plots show the variability within each stratum, with stratum 5 having the highest spread.
Allocation for sample size ($n$) of 50 from Population ($N$) of 1000

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Units ($N_h$)</td>
<td>760</td>
<td>146</td>
<td>53</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Total Measure of Size ($MOS_h$)</td>
<td>39023</td>
<td>17675</td>
<td>9035</td>
<td>4421</td>
<td>6564</td>
</tr>
<tr>
<td>Average Size Per Unit</td>
<td>51.4</td>
<td>121.1</td>
<td>170.5</td>
<td>221.1</td>
<td>312.6</td>
</tr>
<tr>
<td>Proportional Allocation (units)</td>
<td>38</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Proportional Allocation (MOS)</td>
<td>25</td>
<td>12</td>
<td>6</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

\[ n_h \approx (n) \left( \frac{N_h}{\sum N_h} \right) \left( \frac{MOS_h}{\sum MOS_h} \right)^{-1} \text{sample size of selected stratum} \]

in stratum 1
Zero MOS

Small allocated sample size

Industry K

Designate any stratum containing 15 or fewer units as certainty

- Obtain Allocations
- Compute Inclusion Probabilities
- Select Stratified Sequential Samples
- Compute JOINT inclusion probabilities
- Obtain adjusted sampling weights
- Compute Estimates & Variance Estimates
- Assess Reliability

- Region 1
  - Direct Use State 1
  - Balance of Region
- Region 4
  - Direct Use State 1
  - Balance of Region

2 Industries

Obtain Allocations

Select Stratified Sequential Samples

Compute JOINT inclusion probabilities

Obtain adjusted sampling weights

Compute Estimates & Variance Estimates

Assess Reliability

Small allocated sample size

Designate any stratum containing 15 or fewer units as certainty
2 Industries

Zero MOS

Industry 1: 
- Region 1 Direct Use State 1
- Region 1 Balance of Region
- Region 4 Direct Use State 1
- Region 4 Balance of Region

Industry 2: 
- Region 1 Direct Use State 1
- Region 1 Balance of Region
- Region 4 Direct Use State 1
- Region 4 Balance of Region

Industry K: 
- Region 1 Direct Use State 1
- Region 1 Balance of Region
- Region 4 Direct Use State 1
- Region 4 Balance of Region

Unit size is NOT accounted for in stratification

Number of companies can vary greatly by stratum
Step 1: AIES Allocation Procedure

Proportional Allocation to Broad Stratum (NAICS3)

N = 21,000
n = 1,000
f = 1,000/21,000 ≈ 0.05

Total MOS for Stratum

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Total MOS for Stratum</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>(1,000) x 0.05 ≈ 48</td>
</tr>
<tr>
<td>AAB</td>
<td>(20,000) x 0.05 ≈ 952</td>
</tr>
</tbody>
</table>
## Step 2: AIES Allocation Procedure

**Power Allocation Within Broad Stratum (NAICS3)***

<table>
<thead>
<tr>
<th></th>
<th>AAA</th>
<th>AAB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AAA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Companies ($N_h$)</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>Allocation ($n_h$)</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>Sampling fraction ($f_h$)</td>
<td>0.29</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>AAB</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Companies ($N_h$)</td>
<td>245</td>
<td>3,330</td>
</tr>
<tr>
<td>Allocation ($n_h$)</td>
<td>10</td>
<td>202</td>
</tr>
<tr>
<td>Sampling fraction ($f_h$)</td>
<td>0.04</td>
<td>0.06</td>
</tr>
</tbody>
</table>

*Total $\sqrt{MOS}$ for Stratum*
Step 3: AIES Allocation Procedure

Revise allocations if necessary

Increase allocation to sector sampling fraction (0.05).

Reduce other noncertainty allocations assigned to industry AAB.
Compute Inclusion Probabilities for Each Company

1. Obtain Allocations
2. Compute Inclusion Probabilities
3. Select Stratified Sequential Samples
4. Compute JOINT inclusion probabilities
5. Obtain adjusted sampling weights
6. Compute Estimates & Variance Estimates
7. Assess Reliability
• For each company, need to account for
  • Contribution to National estimates
  • Contribution to Geographic estimates

• Not all units are equally important
  • Small unit at the national level could be “large” at the subnational level
  • “Medium sized” company could operate in several states
# Create Sampling Units from Frame

## Sampling Units

<table>
<thead>
<tr>
<th>Company</th>
<th>Sampling Stratum</th>
<th>MOS&lt;sub&gt;c&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&gt; 1 State</td>
<td>181</td>
</tr>
<tr>
<td>2</td>
<td>&gt; 1 State</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>AAA_MO</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>AAA_MO</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Company</th>
<th>Establish-</th>
<th>Industry</th>
<th>State</th>
<th>MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>AAAAAAA</td>
<td>MO</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>AAAAAAA</td>
<td>MO</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>AAAAAAA</td>
<td>TN</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>AAAAAAB</td>
<td>MO</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>AAAAAAB</td>
<td>MO</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>AAAAAAB</td>
<td>TN</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>AAAAAAA</td>
<td>MO</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>AAAAAAA</td>
<td>TN</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>AAAAAAA</td>
<td>MO</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>AAAAAAB</td>
<td>MO</td>
<td>100</td>
</tr>
</tbody>
</table>
### Company National Industry \( \pi \)

<table>
<thead>
<tr>
<th>“Sampling Stratum”</th>
<th>Company</th>
<th>Pseudo-Unit</th>
<th>MOS</th>
<th>( \pi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAAAA</td>
<td>1</td>
<td>1_AAAAAA</td>
<td>104</td>
<td>0.4685</td>
</tr>
<tr>
<td>AAAAAA</td>
<td>2</td>
<td>2_AAAAAA</td>
<td>64</td>
<td>0.2883</td>
</tr>
<tr>
<td>AAAAAA</td>
<td>3</td>
<td>3_AAAAAA</td>
<td>54</td>
<td>0.2432</td>
</tr>
<tr>
<td>AAAAAAB</td>
<td>1</td>
<td>1_AAAAAAB</td>
<td>77</td>
<td>0.4350</td>
</tr>
<tr>
<td>AAAAAAB</td>
<td>4</td>
<td>4_AAAAAAB</td>
<td>100</td>
<td>0.5650</td>
</tr>
</tbody>
</table>

\[
\text{Probability that Company 1 is NOT sampled in AAAAAA} = 1 - ((1 - 0.4685)(1 - 0.4350)) = 0.6997
\]

### National Industry Pseudo-Units

<table>
<thead>
<tr>
<th>Company</th>
<th>MOS</th>
<th>( \pi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MO</td>
<td>37</td>
</tr>
<tr>
<td>2</td>
<td>TN</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>MO</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>MO</td>
<td>100</td>
</tr>
</tbody>
</table>

### Probability that Company 1 is NOT sampled in AAAAAA

- 1
  - 1
  - 0.6997

### Probability that Company 1 is NOT sampled in AAAAAAB

- 2
  - 2
  - 0.2883

- 3
  - 3
  - 0.2432

- 4
  - 4
  - 0.5650
Within stratum, rake all company selection probabilities to allocation.

Select Stratified Sequential Random Samples

1. Obtain Allocations
2. Compute Inclusion Probabilities
3. Select Stratified Sequential Samples
4. Compute JOINT inclusion probabilities
5. Obtain adjusted sampling weights
6. Compute Estimates & Variance Estimates
7. Assess Reliability
Obtain Allocations $\rightarrow$ Compute Inclusion Probabilities $\rightarrow$ Select Stratified Sequential Samples $\rightarrow$ Compute JOINT inclusion probabilities $\rightarrow$ Obtain adjusted sampling weights $\rightarrow$ Compute Estimates & Variance Estimates $\rightarrow$ Assess Reliability

Sampling Conditions
1. Fixed sample size
2. Unequal probability sampling
3. Stratified sample

Want
- Sampled companies from each NAICS4, NAICS5, & NAICS6 within NAICS3
- Sampled companies from each state (for “balance of region”)
- Variety of company sizes

Use list sequential sampling procedure

### Why Use Stratified Sequential Random Sampling?

#### Unit

<table>
<thead>
<tr>
<th>Unit</th>
<th>Industry</th>
<th>MOS</th>
<th>Probability of Inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AAA1</td>
<td>61</td>
<td>0.1755</td>
</tr>
<tr>
<td>2</td>
<td>AAA1</td>
<td>117</td>
<td>0.3365</td>
</tr>
<tr>
<td>3</td>
<td>AAA2</td>
<td>67</td>
<td>0.1927</td>
</tr>
<tr>
<td>4</td>
<td>AAA2</td>
<td>76</td>
<td>0.2186</td>
</tr>
<tr>
<td>5</td>
<td>AAA2</td>
<td>101</td>
<td>0.2905</td>
</tr>
<tr>
<td>6</td>
<td>AAA1</td>
<td>60</td>
<td>0.1726</td>
</tr>
<tr>
<td>7</td>
<td>AAA1</td>
<td>74</td>
<td>0.2128</td>
</tr>
<tr>
<td>8</td>
<td>AAA3</td>
<td>79</td>
<td>0.2272</td>
</tr>
<tr>
<td>9</td>
<td>AAA2</td>
<td>100</td>
<td>0.2876</td>
</tr>
<tr>
<td>10</td>
<td>AAA3</td>
<td>308</td>
<td>0.8859</td>
</tr>
</tbody>
</table>

#### Industry

- **AAA1**: 4 companies, 40%
- **AAA2**: 4 companies, 40%
- **AAA3**: 2 companies, 20%

#### Stratification

- All companies in the same NAICS3 stratum
- Each company associated with one NAICS4 code

#### PPS Samples

**PROC SURVEYSELECT (Systematic)**

<table>
<thead>
<tr>
<th>Sample</th>
<th>AAA1</th>
<th>AAA2</th>
<th>AAA3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sample 2</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Sample 3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sample 4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sample 5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**PROC SURVEYSELECT (Sequential)**

<table>
<thead>
<tr>
<th>Sample</th>
<th>AAA1</th>
<th>AAA2</th>
<th>AAA3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sample 2</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Sample 3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sample 4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sample 5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

- Slightly worse (over repeated samples) for industry representation
- Allows ACCURATE variance estimation

---

**GREAT in terms of industry representation** (implicit stratification works!)

**TERRIBLE for variance estimation**
Compute JOINT inclusion probabilities (given sample)
.obtain allocations → compute inclusion probabilities → select stratified sequential samples → compute joint inclusion probabilities → obtain adjusted sampling weights → compute estimates & variance estimates → assess reliability

\[
\hat{v}(\hat{y}_k) = \sum_{h=1}^{H} \frac{1}{2} \sum_{i=1}^{n_h} \left( \frac{\pi_{hi}}{\pi_h,\pi_{h,l}} \right) \left( \frac{y^k_{hi}}{\pi_{hi}} - \frac{y^k_{hl}}{\pi_{hl}} \right)^2
\]

KNOWN inclusion probability for company \(i\) in stratum \(h\)

Joint inclusion probability of companies \(i\) and \(l\) in stratum \(h\)
- \(= 0\) if companies are in different strata
- Dependent on full sample design (LOW entropy design)

Obtain adjusted sampling weights
Multi-State Company Stratum (AAA)

Company 1

Company 2

Company 3

Company 4

Company 5

Company 6

Company 7

<table>
<thead>
<tr>
<th>Industry (National)</th>
<th>Frame MOS</th>
<th>Contributing Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAAAAAA</td>
<td>496</td>
<td>1, 2, 3, 4, 5</td>
</tr>
<tr>
<td>AAAAAAB</td>
<td>285</td>
<td>1, 4, 6, 7</td>
</tr>
</tbody>
</table>
Multi-State Company Stratum (AAA)

Company 1: \(\pi = 0.62\)
- Industry (National): AAAA
- Frame MOS: 56
- Contributing Companies: 1, 2, 3, 4, 5
- Sampled Companies: 1, 3

Company 2: 37

Company 3: \(\pi = 0.59\)
- Industry (National): AAAA
- Frame MOS: 96
- Contributing Companies: 1, 2, 3
- Sampled Companies: 1, 3

Company 4: 38

Company 5: 43

Company 6: 100

Multi-State Company Stratum (AAA)

Company 7: \(\pi = 0.13\)
- Industry (National): AAAA
- Frame MOS: 1
- Contributing Companies: 1, 2
- Sampled Companies: 1

Missouri Stratum (AAA)

Company 2: 38

Company 5: 54

Company 6: 100

Company 7: \(\pi = 0.13\)
- Industry (National): AAAA
- Frame MOS: 1
- Contributing Companies: 1, 2
- Sampled Companies: 1

Company 5: 54

Company 6: 100

Company 7: \(\pi = 0.13\)
- Industry (National): AAAA
- Frame MOS: 1
- Contributing Companies: 1, 2
- Sampled Companies: 1

Company 5: 54

Company 6: 100

Company 7: \(\pi = 0.13\)
- Industry (National): AAAA
- Frame MOS: 1
- Contributing Companies: 1, 2
- Sampled Companies: 1

Company 5: 54

Company 6: 100

Company 7: \(\pi = 0.13\)
- Industry (National): AAAA
- Frame MOS: 1
- Contributing Companies: 1, 2
- Sampled Companies: 1

Multi-State Company Stratum (AAA)

Company 1

- Company 1
  - Multi-State Company Stratum (AAA)
  - Missouri Stratum (AAA)

Company 3

- Company 3
  - Multi-State Company Stratum (AAA)
  - Missouri Stratum (AAA)

Company 6

- Company 6
  - Multi-State Company Stratum (AAA)
  - Missouri Stratum (AAA)

### Industry (National)

<table>
<thead>
<tr>
<th>Industry</th>
<th>Frame MOS</th>
<th>Estimated MOS</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAAAA</td>
<td>496</td>
<td>505</td>
<td>0.98</td>
</tr>
<tr>
<td>AAAAAAB</td>
<td>285</td>
<td>300</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Obtain Allocations → Compute Inclusion Probabilities → Select Stratified Sequential Samples → Compute JOINT inclusion probabilities → Obtain adjusted sampling weights → Compute Estimates & Variance Estimates → Assess Reliability

Multi-State Company Stratum (AAA)

- Company 1
  - $w_{1,AAAAB} = 1.6129$
  - $\hat{w}_{1,AAAAB} = 1.5323$

- Company 3
  - $w_{3,AAAAAA} = 1.6949$
  - $\hat{w}_{3,AAAAAA} = 1.6610$

Missouri Stratum (AAA)

- Company 6
  - $w_{6,AAAAB} = 1.7544$
  - $\hat{w}_{6,AAAAB} = 1.6667$
\[ \hat{v}(\tilde{Y}^k) = \left( \hat{R}_s^k \right) \sum_{h=1}^{H} \frac{1}{2} \sum_{i=1}^{n_h} \left( \frac{\pi_{hi} \pi_{hl} - \pi_{h,il}}{\pi_{h,il}} \right) \left( \hat{e}_{hi}^k - \hat{e}_{hi}^k \right)^2 \]

Ratio adjustment factor (noncertainty units)

Linearized ratio estimate
Assess Reliability

1. Obtain Allocations
2. Compute Inclusion Probabilities
3. Select Stratified Sequential Samples
4. Compute JOINT inclusion probabilities
5. Obtain adjusted sampling weights
6. Compute Estimates & Variance Estimates
7. Assess Reliability
\[ \hat{Z} = \text{Ratio estimate from sampled units} + \text{Total from certainty units} \]

\[ v(\hat{Z}) = \text{variance estimate for ratio estimate (sampled units)} \]

\[ c.v.(\hat{Z}) = \frac{\sqrt{v(\hat{Z})}}{\hat{Z}} \]
Different c.v. requirements for national and geographic estimates

\[ \omega_c = \sqrt{cv(\hat{Z}_{1c})cv(\hat{Z}_{2c})} \]

Table: Analysis Variables

<table>
<thead>
<tr>
<th>Company</th>
<th>Establishment</th>
<th>Sector</th>
<th>Industry</th>
<th>NAICS6</th>
<th>State</th>
<th>2020 Payroll</th>
<th>2020 Emp</th>
<th>2017 Receipts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Dynamics, Inc.</td>
<td>Mom &amp; Pop Hardware 1</td>
<td>Retail Trade</td>
<td>Hardware Stores</td>
<td>444130</td>
<td>MO</td>
<td>600</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mom &amp; Pop Hardware 2</td>
<td>Retail Trade</td>
<td>Hardware Stores</td>
<td>444130</td>
<td>MO</td>
<td>300</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mom &amp; Pop Hardware 3</td>
<td>Retail Trade</td>
<td>Hardware Stores</td>
<td>444130</td>
<td>MO</td>
<td>762</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>別の業種</td>
<td>別の業種</td>
<td>別の業種</td>
<td>332721</td>
<td>TN</td>
<td>457</td>
<td></td>
<td>16,068</td>
</tr>
<tr>
<td></td>
<td>別の業種</td>
<td>別の業種</td>
<td>別の業種</td>
<td>489</td>
<td></td>
<td></td>
<td></td>
<td>237</td>
</tr>
</tbody>
</table>
### Evaluation Criteria: National Industry Estimates

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(xcellent)</td>
<td>$\omega_c \leq 0.02$</td>
<td></td>
</tr>
<tr>
<td>G(ood)</td>
<td>$0.02 &lt; \omega_c \leq 0.05$</td>
<td></td>
</tr>
<tr>
<td>A(dequate)</td>
<td>$0.05 &lt; \omega_c \leq 0.10$</td>
<td></td>
</tr>
<tr>
<td>B(ad)</td>
<td>$0.10 &lt; \omega_c \leq 0.30$</td>
<td></td>
</tr>
<tr>
<td>U(nacceptable)</td>
<td>$0.30 &lt; \omega_c$</td>
<td></td>
</tr>
</tbody>
</table>

![Graph showing CVI vs CV2 with different evaluation categories represented by colors]
Evaluation Criteria: Unacceptable/Bad Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(xcellent)</td>
<td>$\omega_c \leq 0.02$</td>
<td></td>
</tr>
<tr>
<td>G(ood)</td>
<td>$0.02 &lt; \omega_c \leq 0.05$</td>
<td></td>
</tr>
<tr>
<td>A(dequate)</td>
<td>$0.05 &lt; \omega_c \leq 0.10$</td>
<td></td>
</tr>
<tr>
<td>B(ad)</td>
<td>$0.10 &lt; \omega_c \leq 0.30$</td>
<td></td>
</tr>
<tr>
<td>U(nacceptable)</td>
<td>$0.30 &lt; \omega_c$</td>
<td></td>
</tr>
</tbody>
</table>
Evaluation Criteria
Excellent/Good/Adequate Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(xcellent)</td>
<td>$\omega_c \leq 0.02$</td>
<td>Blue</td>
</tr>
<tr>
<td>G(good)</td>
<td>$0.02 &lt; \omega_c \leq 0.05$</td>
<td>Red</td>
</tr>
<tr>
<td>A(adequate)</td>
<td>$0.05 &lt; \omega_c \leq 0.10$</td>
<td>Green</td>
</tr>
<tr>
<td>B(ad)</td>
<td>$0.10 &lt; \omega_c \leq 0.30$</td>
<td>Orange</td>
</tr>
<tr>
<td>U(acceptable)</td>
<td>$0.30 &lt; \omega_c$</td>
<td>Purple</td>
</tr>
</tbody>
</table>

Reference lines at cv1=0.02 and cv2=0.02
Results: July 2022 AIES End-to-End Test

National Industry Tabulations

Geographic Industry Tabulations

Source: AIES test sample frame created in May 2022 from U.S. Census Business Register
Overview of AIES Sample Design Process

Sampling Frame

Stratification

Complexity
Certainty
Companies

Noncertainty
Companies

Sector Processing

Obtain Allocations

Compute Inclusion Probabilities

Select Stratified Sequential Samples

Compute JOINT inclusion probabilities

Obtain adjusted sampling weights

Compute Estimates & Variance Estimates

Assess Reliability

If all industries in a sector do not meet c.v. constraints

Revise allocations

Sector 11 + Sector 21 + ... + Sector 72 + Sector 81
Questions, suggestions, comments?

katherine.j.thompson@census.gov