

# Developing the Sample Design for the New Annual Integrated Economic Survey (AIES)

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. (Approval ID: CBDRB-FY22-ESMD006-014).

# Acknowledgements

James Hunt\*

Amy Newman-Smith\*

Laura Bechtel

Nicole Czaplicki

Stephen Kaputa

Adam Smeltz

Justin Z. Smith

Lucas Streng

Thomas Kirk White

Ahntuan Ty

Yeng Xiong

Michael Christen

Carol Caldwell (retired)

James Burton

\*Co-leads, with myself

# Definitions

- Sector: an area of the economy in which businesses share the same or related business activity, product, or service ([www.investopedia.com](http://www.investopedia.com))
- Industry: a group of companies that are related based on their primary business activities or service ([www.investopedia.com](http://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)

# More Definitions

- Establishment – a single business location
- Company – any formal business entity for profit, which may be a partnership, association or individual proprietorship (<https://census.gov>)
  - Multi-unit company



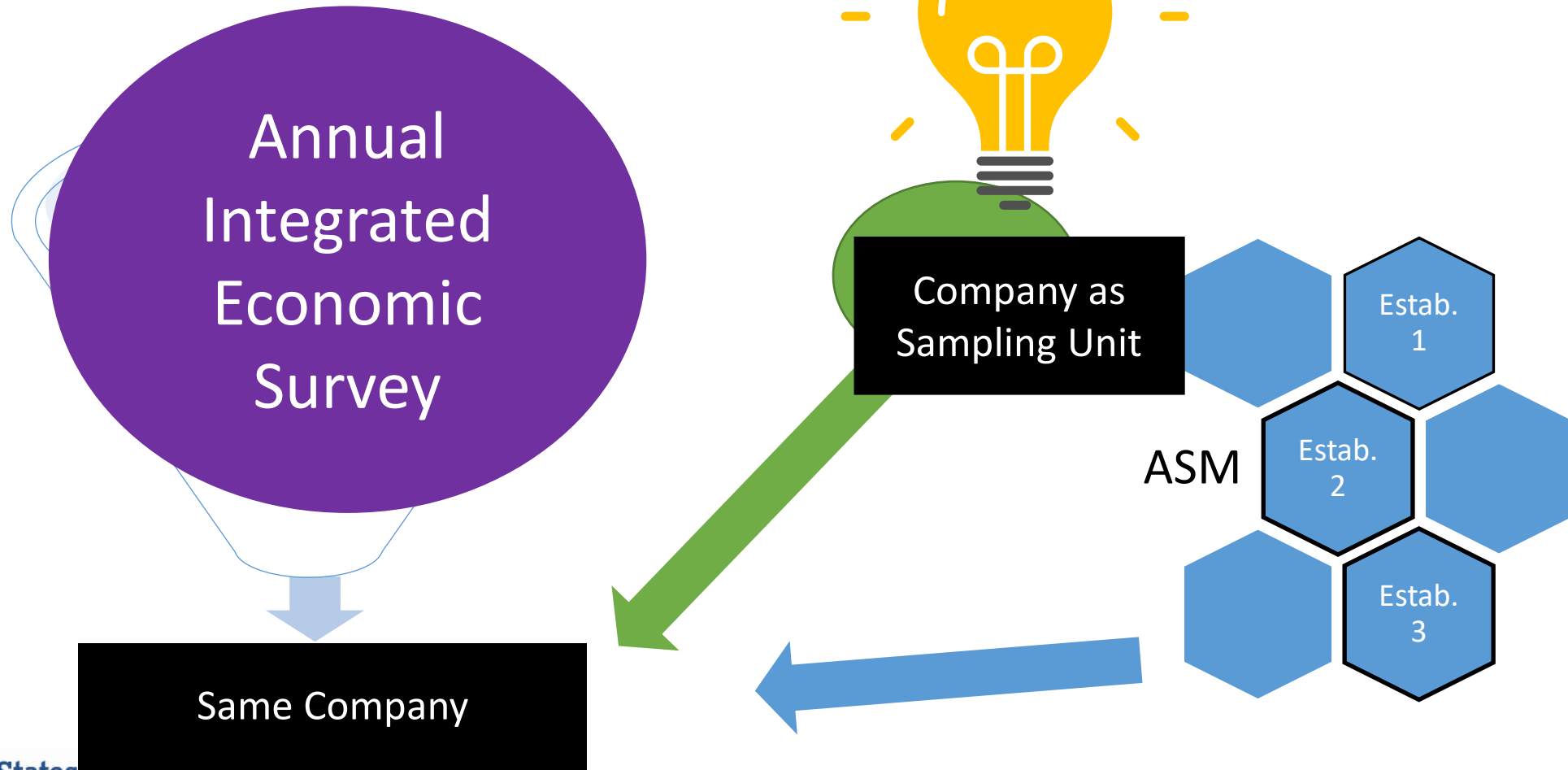
# Current Set of Annual Surveys

Survey	NAICS by
Annual Retail Trade Survey (ARTS)	Operation
Annual Wholesale Trade Survey	2-through-6-digit NAICS by
Service	Tax Exempt Status
Company	National:
Establishment	Natio
	Sub

Reengineering the Census Bureau's  
Annual Economic Surveys  
National Academy of Sciences Panel (2018)

Annual  
Integrated  
Economic  
Survey

# AIES Sampling Unit



# 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)

## Subnational (Geographic) Industry Estimates

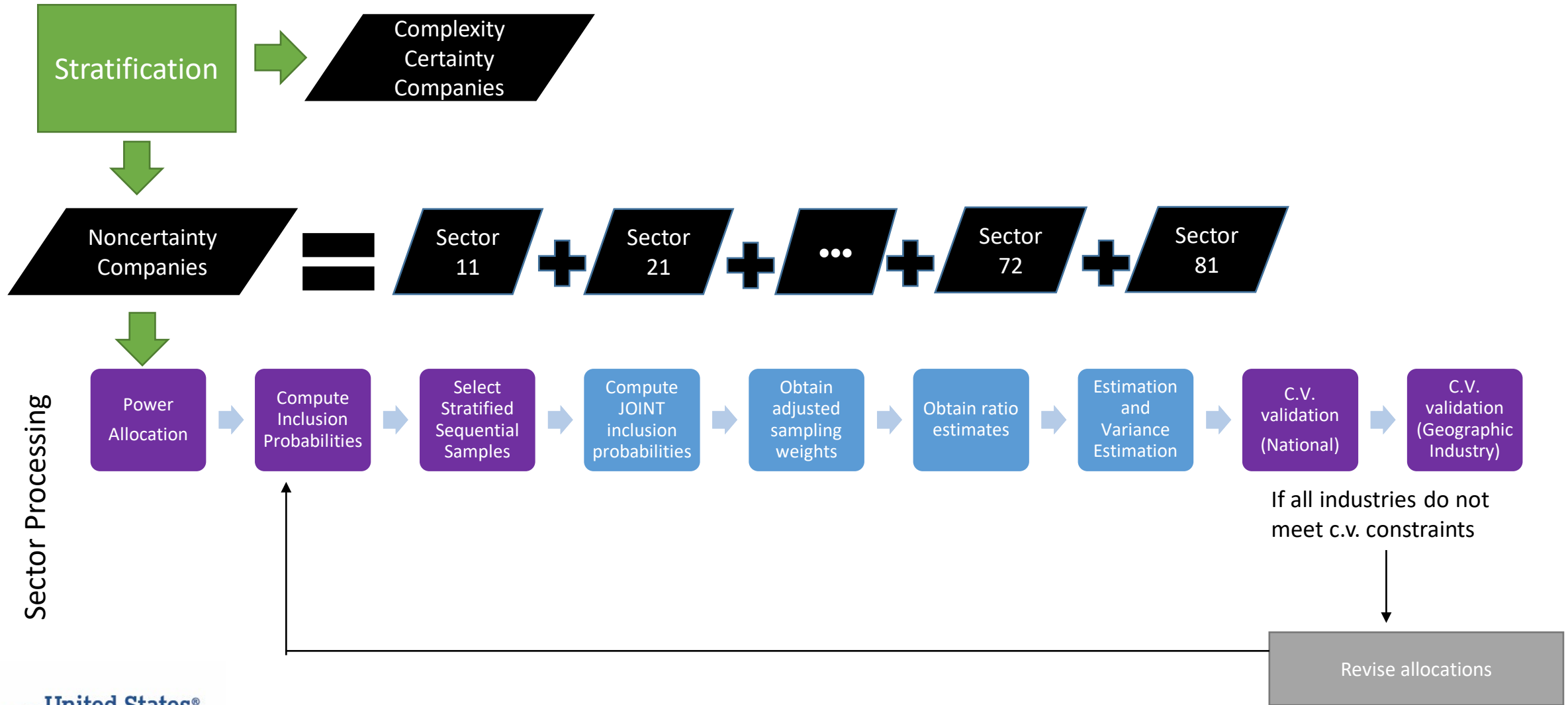
- **Revenue/receipts**, Annual Payroll, 1<sup>st</sup> Quarter Employment, 1<sup>st</sup> Quarter Payroll
- Aggregated NAICS levels (3-digit)
  - 23 “Direct use” states
  - 4 “Balance of region” categories

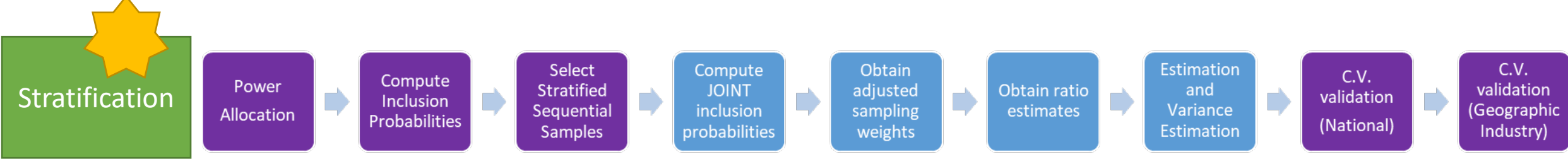
# AIES Frame Variables (Quantitative)

- Annual Payroll (2020 administrative data)
  - Measure of size
- 1<sup>st</sup> 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 Geo

> 1 Geo

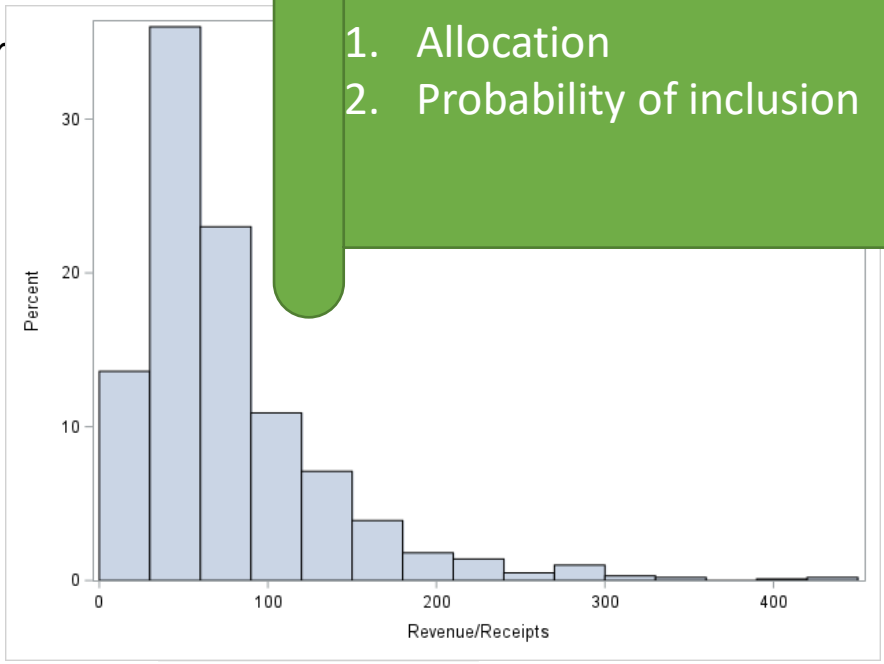
>

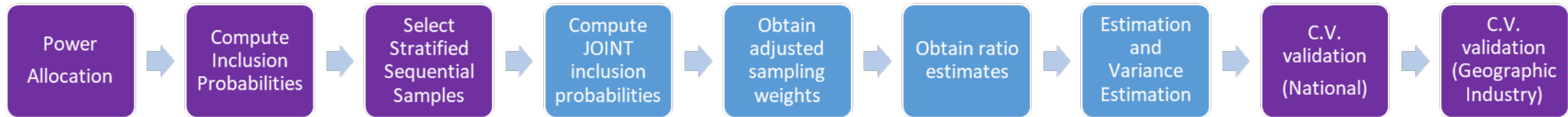
- Historical strata defined by
  - Industry
  - Unit size
- AIES noncertainty strata defined by
  - Industry
  - Geography

Operates establishments in more than one state

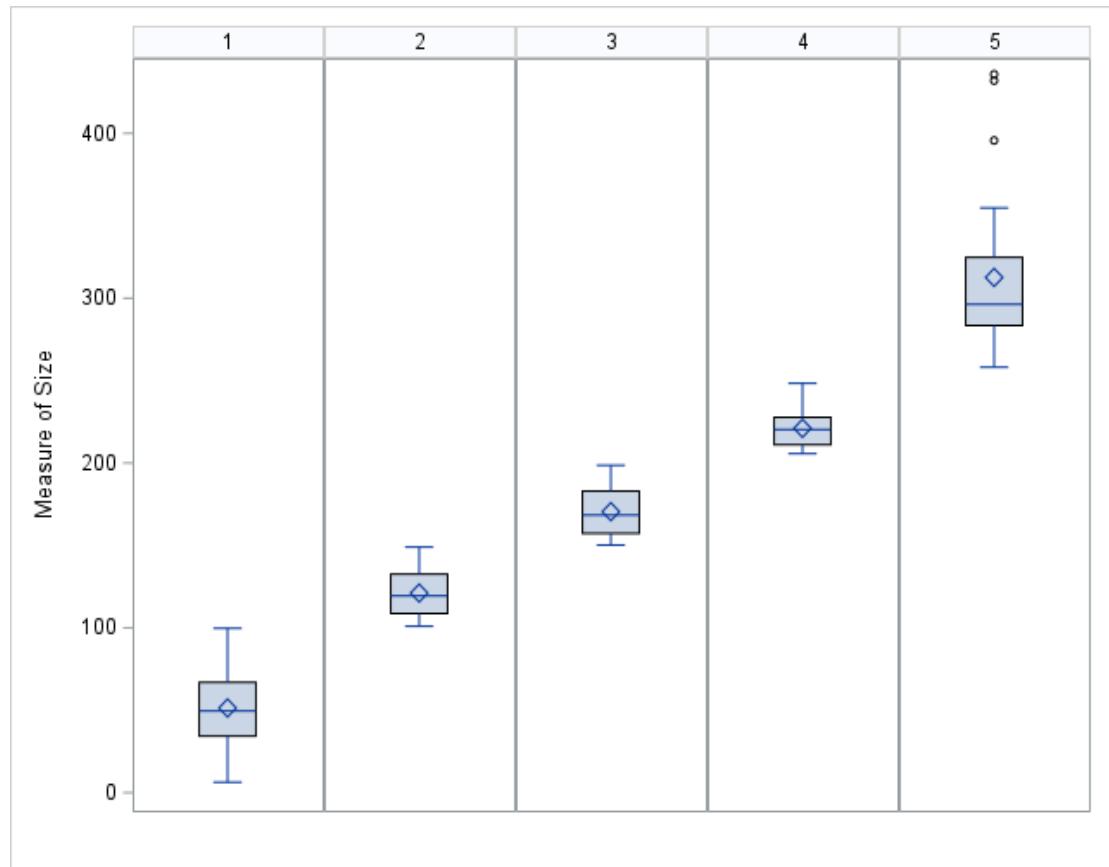
Account for skewed distribution in AIES sample design:

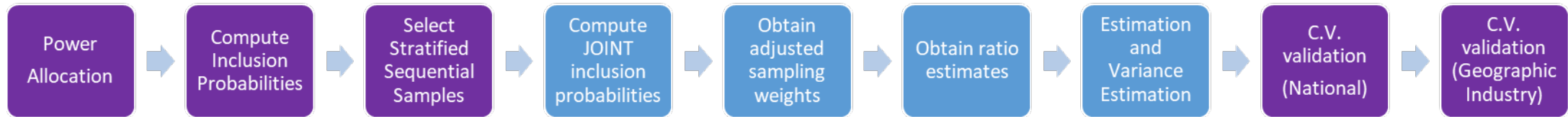
1. Allocation
2. Probability of inclusion





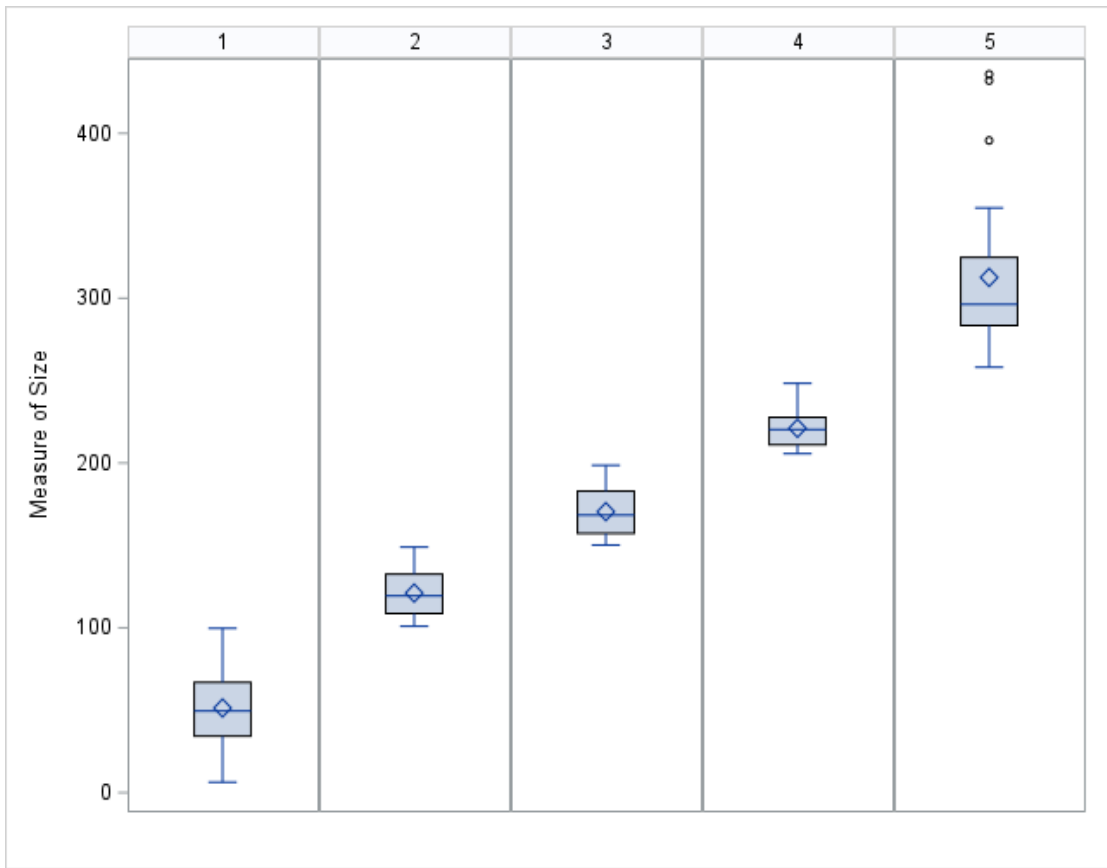
	1	2	3	4	5
Number of Units ( $N_h$ )	760	146	53	20	21
Total Measure of Size ( $MOS_h$ )	39023	17675	9035	4421	6564
Average Size Per Unit	51.4	121.1	170.5	221.1	312.6



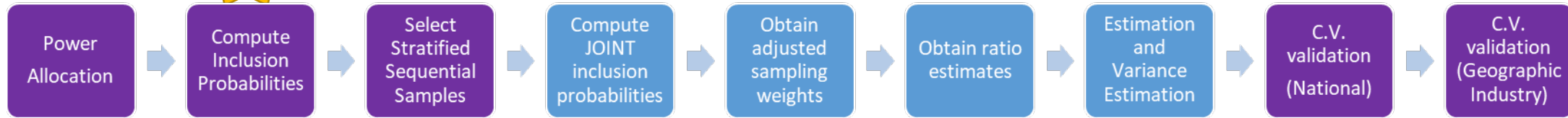


Allocation for sample size ( $n$ ) of 50 from Population ( $N$ ) of 1000

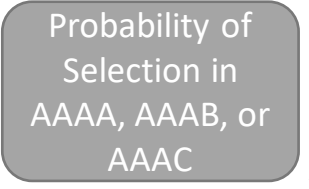
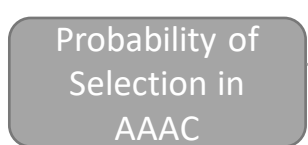
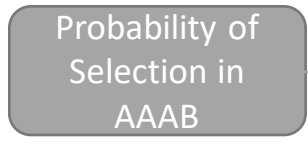
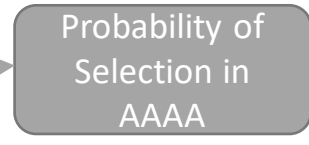
	1	2	3	4	5
Number of Units ( $N_h$ )	760	146	53	20	21
Total Measure of Size ( $MOS_h$ )	39023	17675	9035	4421	6564
Average Size Per Unit	51.4	121.1	170.5	221.1	312.6
Proportional Allocation (units)	38	7	3	1	1
Proportional Allocation (MOS)	25	12	6	3	4
Power Allocation $\frac{n}{\sqrt{MOS_h}}$	17	12	8	6	7



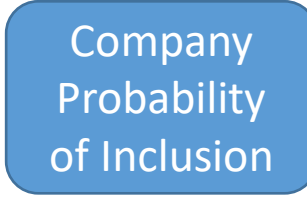
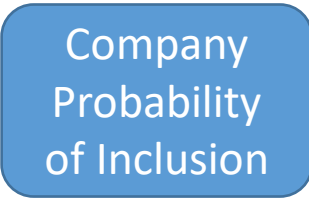
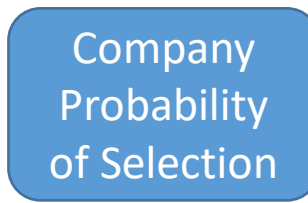
$$m_h \approx (n) \left( \frac{N_h MOS_h}{\sum_h MOS_h} \right)$$



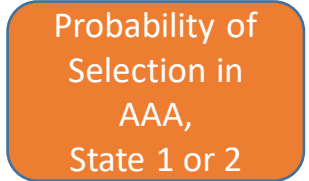
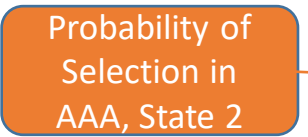
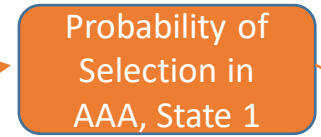
National



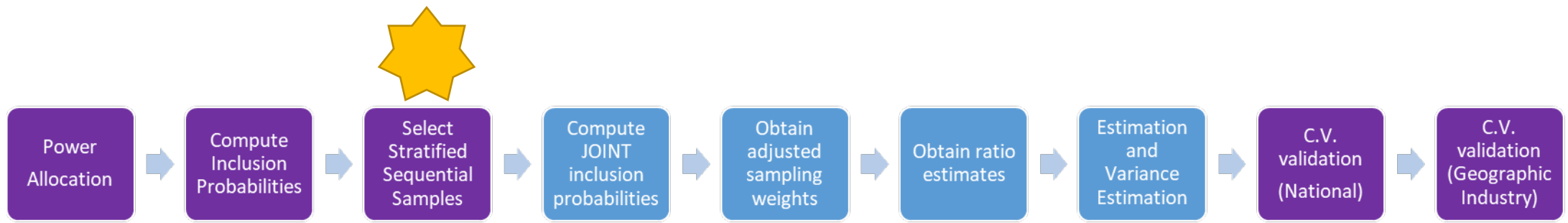
Select MAXIMUM



Geographic



Without "stratification" property, selection probabilities to allocation



- AIES Sampling Requirements

- Fixed sample size
- PPES sampling (unequal probabilities)
- Representation in sampled units for NAICS4, NAICS5, NAICS6 within NAICS3

- Stratified Sequential (Chromy) Sampling

- List sequential

# Why Use Stratified Sequential Sampling?

Unit	Industry	MOS	
1	AAA1	61	0.1755
2	AAA1	117	
6	AAA1	60	0.1720
7			
3			
4			
5			
9			
8	AAA3	79	0.2272
10	AAA3	308	0.8859

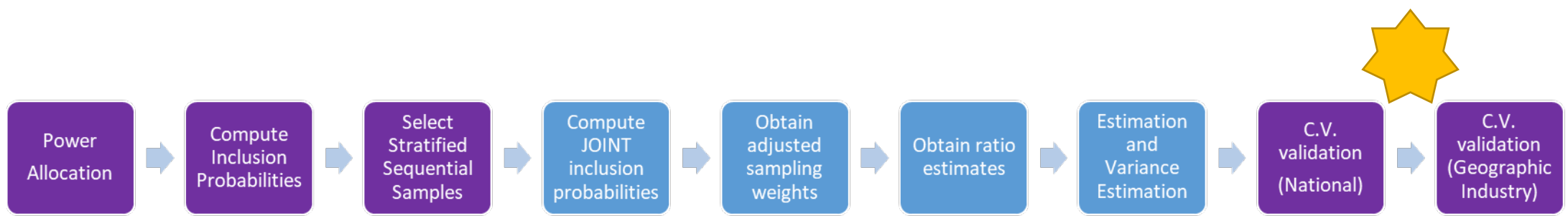
All companies in the same NAICS3 stratum

Each company associated with one NAICS4 code

Slightly worse (over repeated samples) for industry representation

Allows ACCURATE variance estimation

PPS Samples PROC SURVEYSELECT (Sequential)			
	AAA1	AAA2	AAA3
Sample 1	1	1	1
Sample 2	0	1	2
Sample 3	1	1	1
Sample 4	1	1	1
Sample 5	1	1	1



- National Estimates: NAICS4, NAICS5, NAICS6
  - $0 \leq cv(RCPT) \leq 0.02$  and  $0 \leq cv(EMP) \leq 0.02$
- Geographic Industry Estimates: NAICS3 by geography category
  - $0 \leq cv(RCPT) \leq 0.15$  and  $0 \leq cv(EMP) \leq 0.15$
- Any C.V. above 0.30 is considered to have “serious issues” per the Census Quality Standards

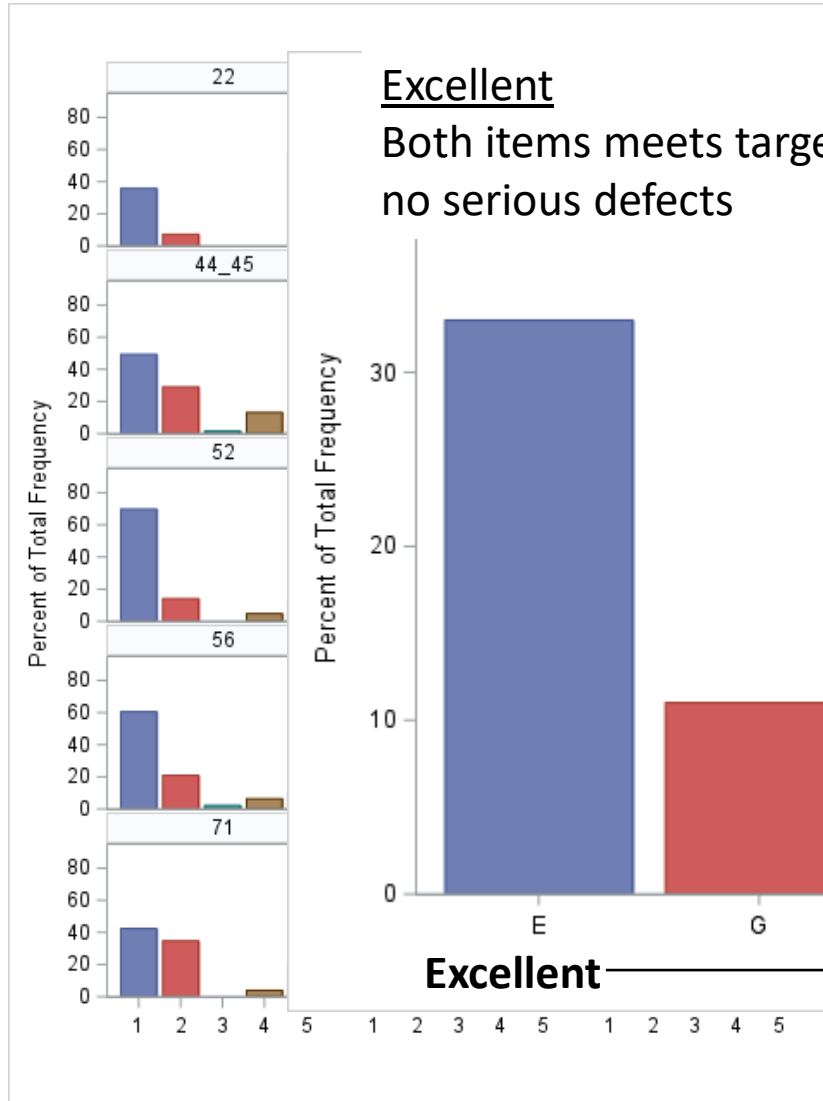


# End-to-end test #1

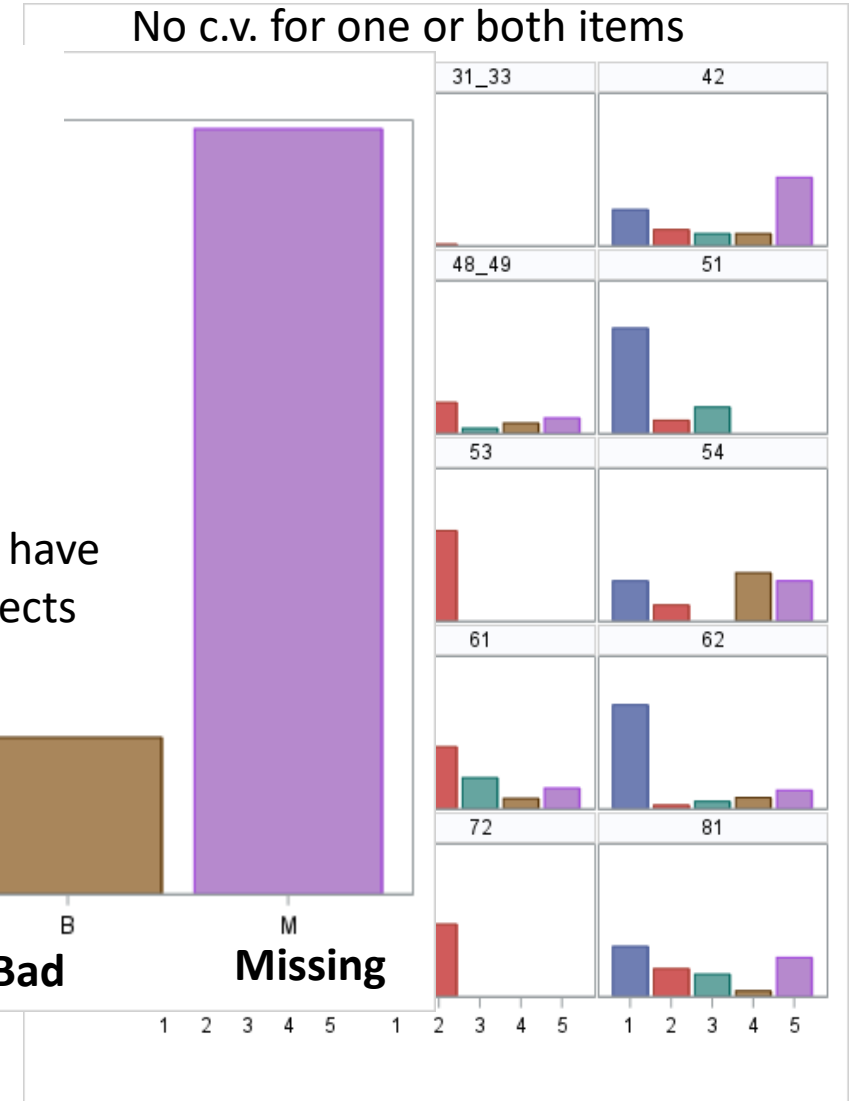
- Overall sample size: 407,000 companies
  - Complexity Certainty Companies  $\approx$  52,000 companies
  - Initial Noncertainty Sample  $\approx$  355,000 companies
- Fully automated
  - Default allocations, inclusion probabilities, and strata definitions
  - Within strata sort
    - Unique NAICS designation (company)
    - State code
    - Measure of size (company)

# Results for National Estimates

## NAICS6

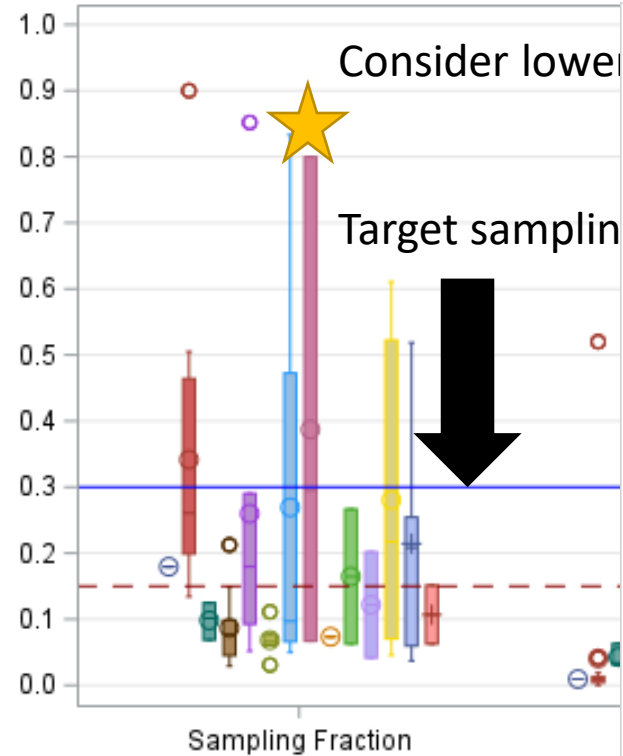
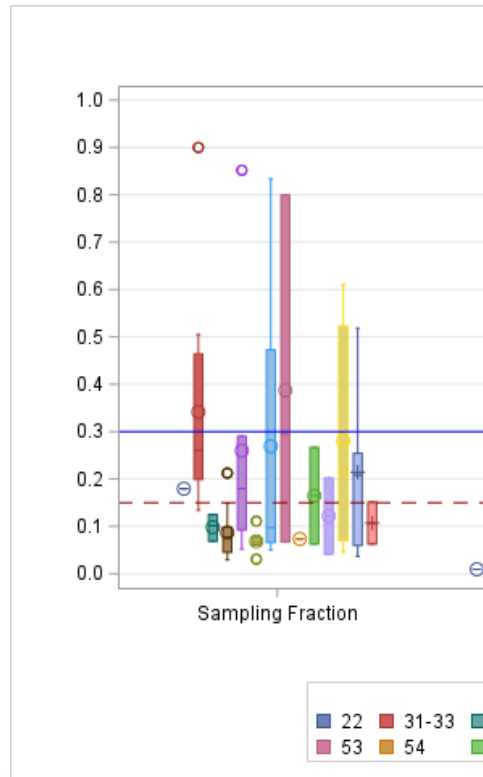


## NAICS4

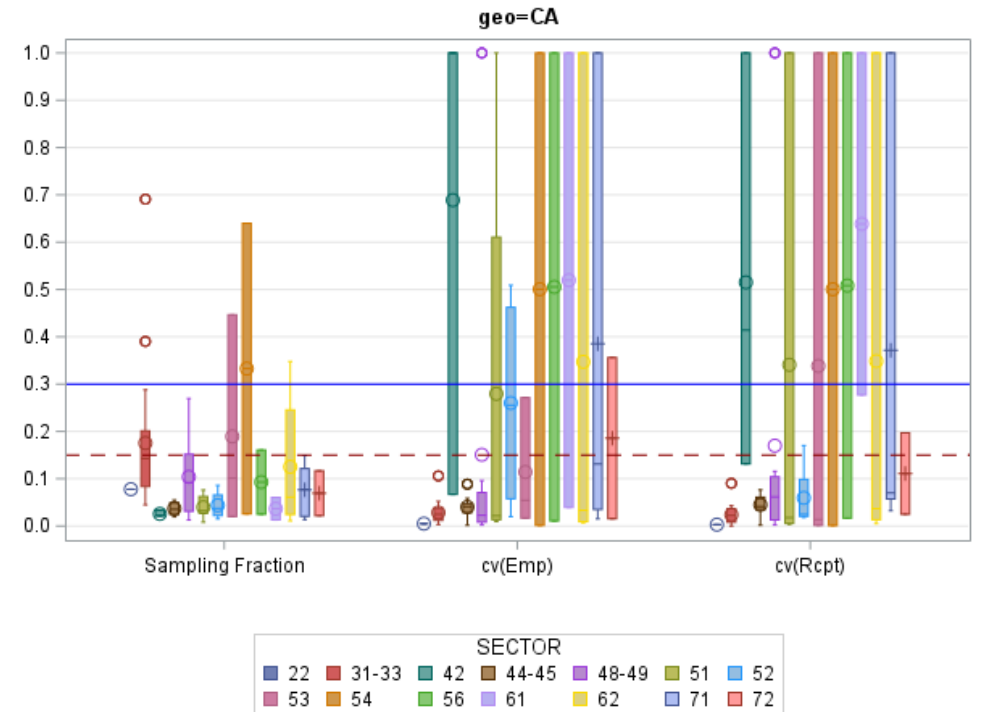


# Results for Geographic Estimates (Excluding Sector 81)

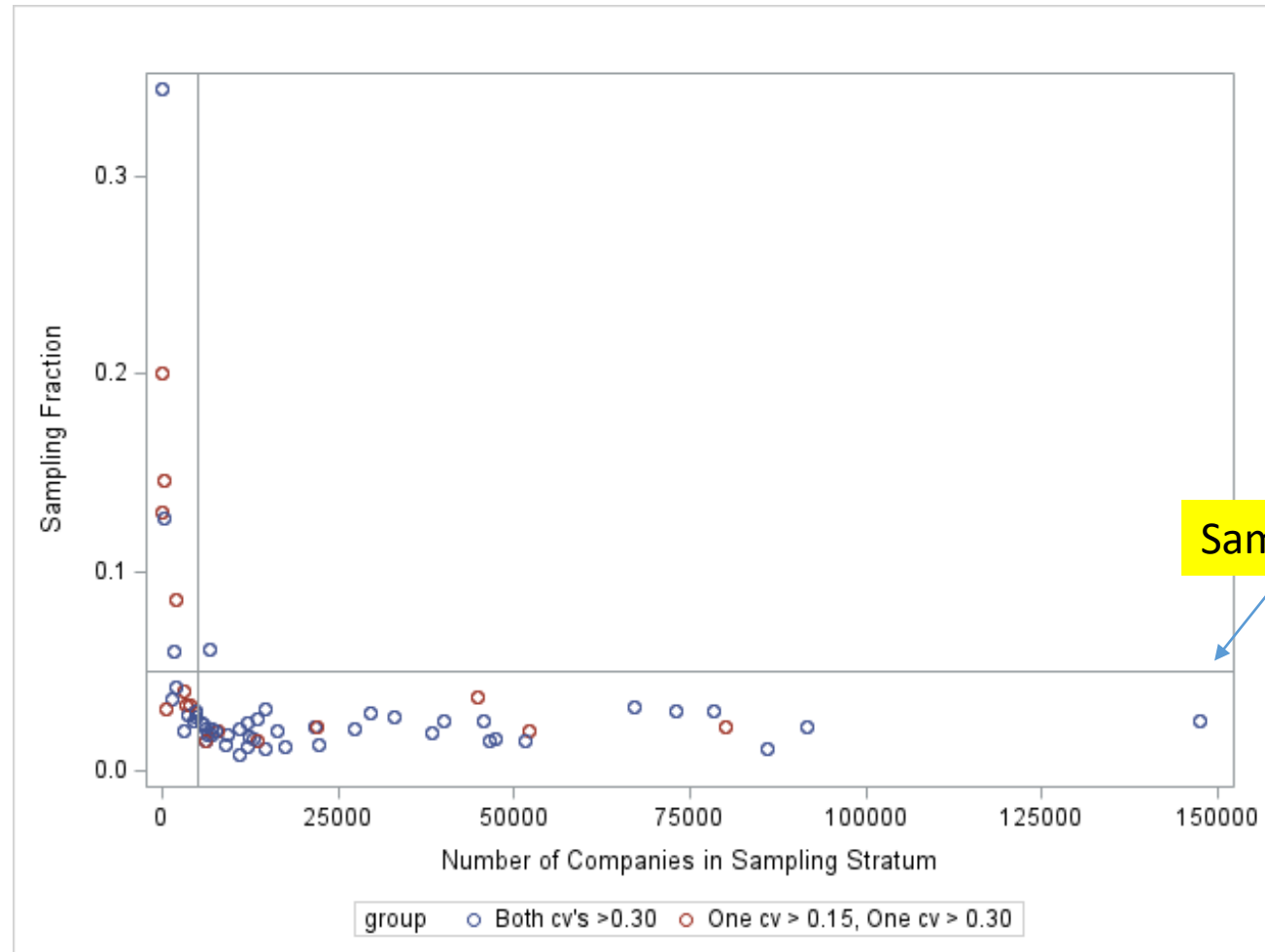
## Majority of States



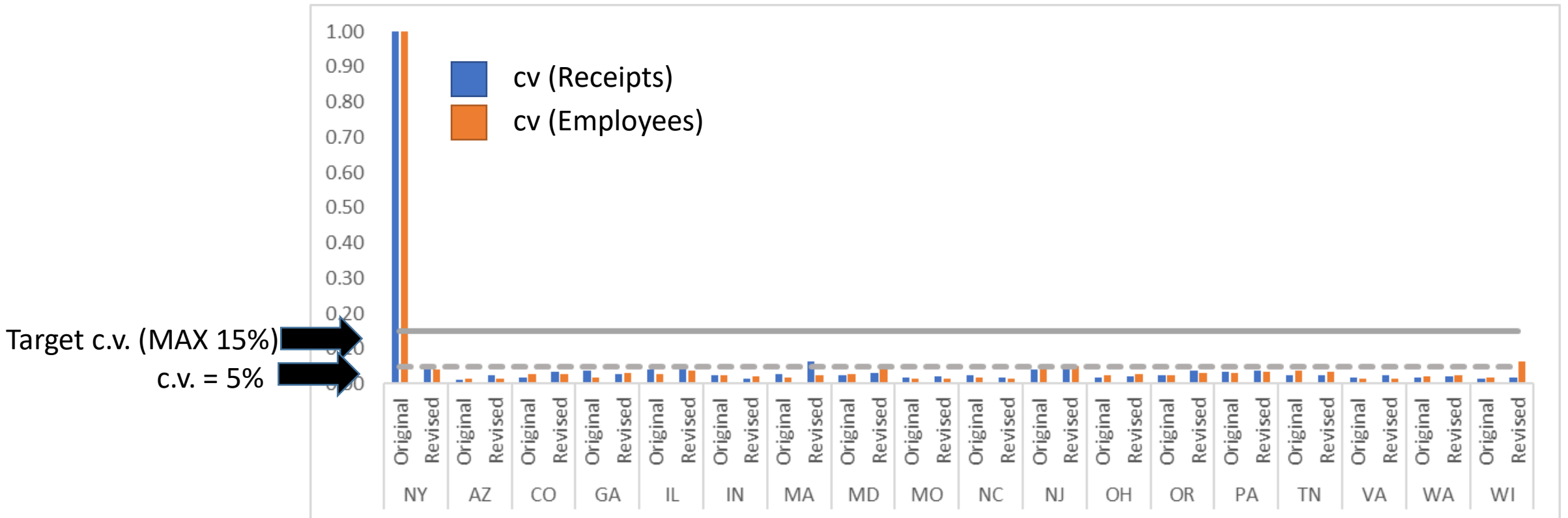
geo=TN



# Further Exploration “Bad” Geographic Results



# Preliminary Results (Revised Allocation) Food and Beverage Stores



Added sample (increased sampling fraction)      Removed sample (decreased sampling fraction)

# Remaining Decisions

- Sample selection frequency
  - Annual, quinquennial, other?
- Economic indicator design(s)
  - Monthly and quarterly estimates
- Estimation, imputation, variance estimation...

Questions, suggestions,  
comments?

[katherine.j.thompson@census.gov](mailto:katherine.j.thompson@census.gov)