Developing the Sample Design for the New Annual Integrated Economic Survey

Katherine Jenny Thompson, Senior Mathematical Statistician

Associate Directorate for Economic Programs

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)

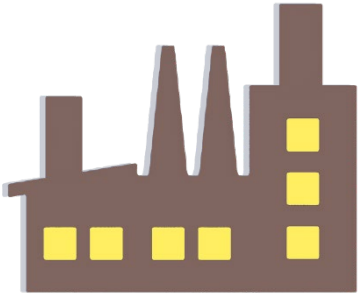
United States® Census Bureau

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)

United States® Census Bureau

More Definitions



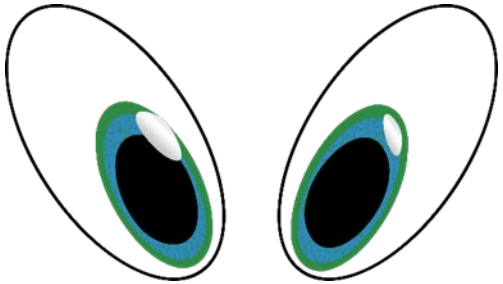
- Establishment – a <u>single</u> business location

- Company – any formal business entity for profit, which may be [illegible]orati[illegible] partnership, association or individual proprietorship (https://[illegible]y.l[illegible]
  - Multi-unit company
  - [illegible]pany

United States® Census Bureau

The Project

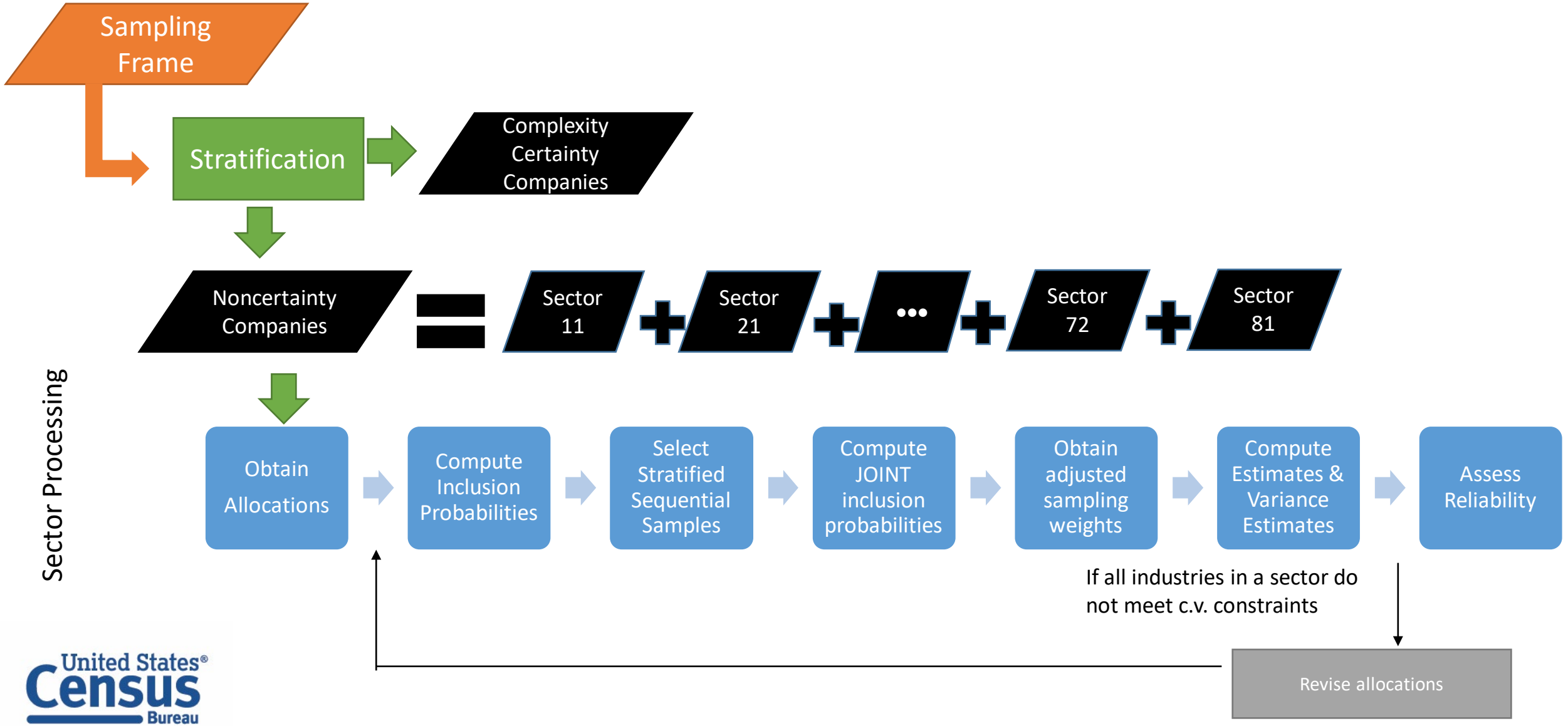
- Replace six annual surveys with one annual survey
  - Recommended by National Academy of Science Panel (2018)
  - Annual Integrated Economic Survey (AIES)



- Requirements for new survey
  - National Industry estimates
    - At least 5 key items (multipurpose)
    - C.V. target ≈ 2%
    - Disaggregated NAICS levels: 4-digit NAICS (NAICS4), 5-digit NAICS (NAICS5), **6-digit NAICS (NAICS6)**
  - Geographic Industry estimates
    - 4 key items
    - C.V. target ≈ 15%
    - Aggregated NAICS levels: **3-digit NAICS (NAICS3)**
      - 23 “Direct use” states
      - 4 “Balance of region” categories

United States® Census Bureau

Overview of AIES Sample Design Process



United States® Census Bureau

Sampling Frame…

Sampling Unit

Components

Classification Information

Geo Infor

Measure of Size (MOS)

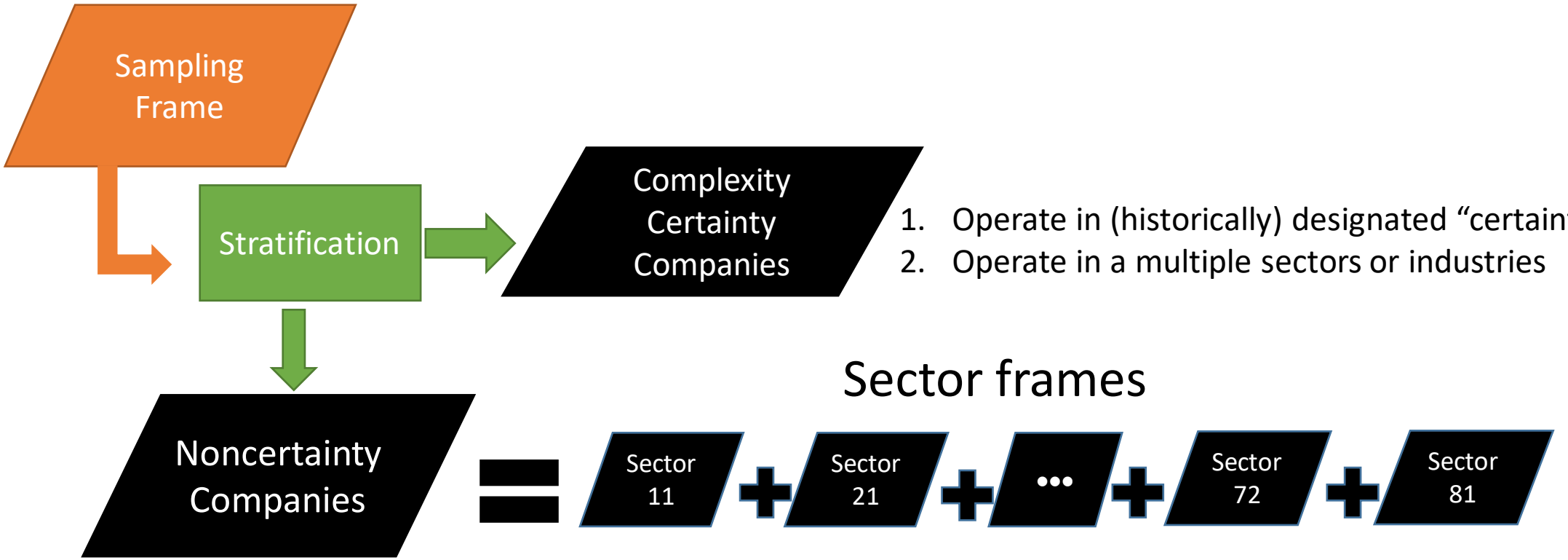
Quantitative oles

Company	Establishment	ector	Industry					
Global Dynamics, Inc.	Mom & Pop Hardware 1	Retail Trade	Hardware Stores	444130	MO	600	20	1,380
	Mom & Pop Hardware 2			444130	MO	800		
	Mom & Po Hardware		Hardware Stores	444130	MO	762	25	1,753
	GD Plant		Ma	332722	TN	9,452		
	Dad's Warehouse	Wholesale Trade	Hardware Merchant Wholesalers	423710	TN	475	16	489
The Shop								

Multi-unit
Multi-sector
Multi-state

Single-units
Single-sector
Single-state

United States® Census Bureau

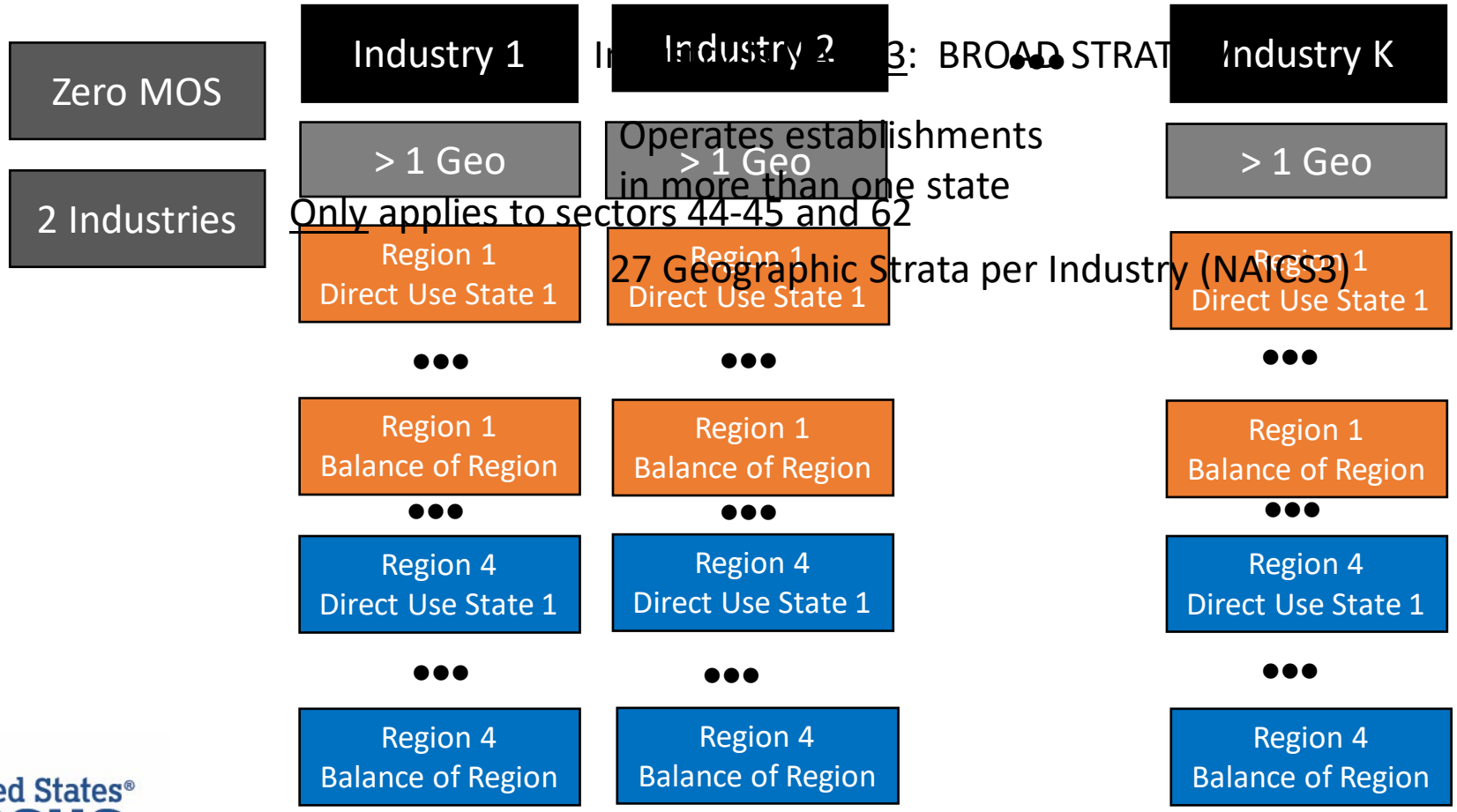


1. Operate in (historically) designated “certainty” industries
2. Operate in a multiple sectors or industries

1. Operate in a single sector
2. Generally operate in one industry (NAICS4)

United States® Census Bureau

Sector Processing:
Stratification for Noncertainty Companies



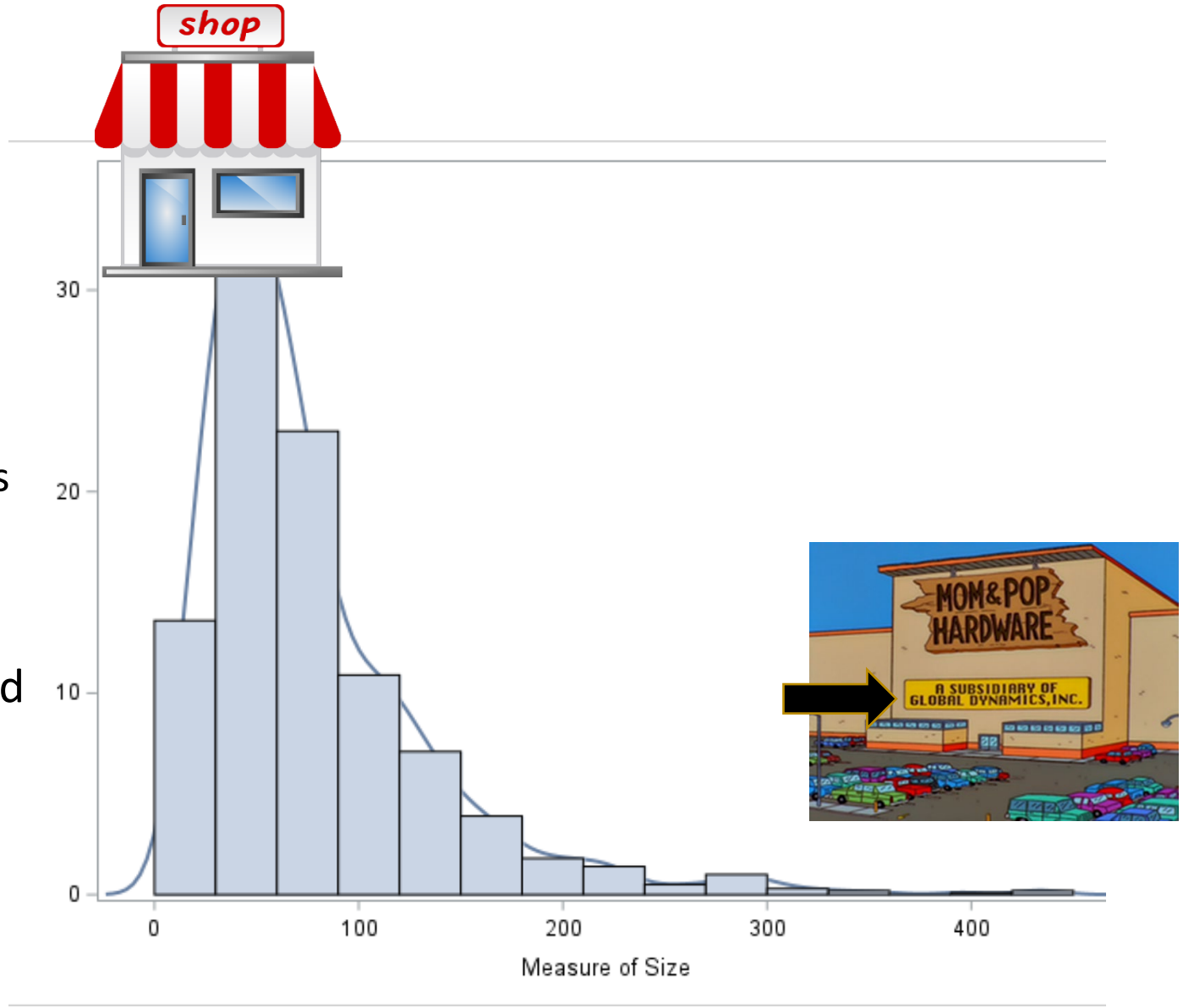
Operates establishments in more than one state

Only applies to sectors 44-45 and 62

27 Geographic Strata per Industry (NAICS3)

Digression #1: Establishment (Business) Sample Survey Designs

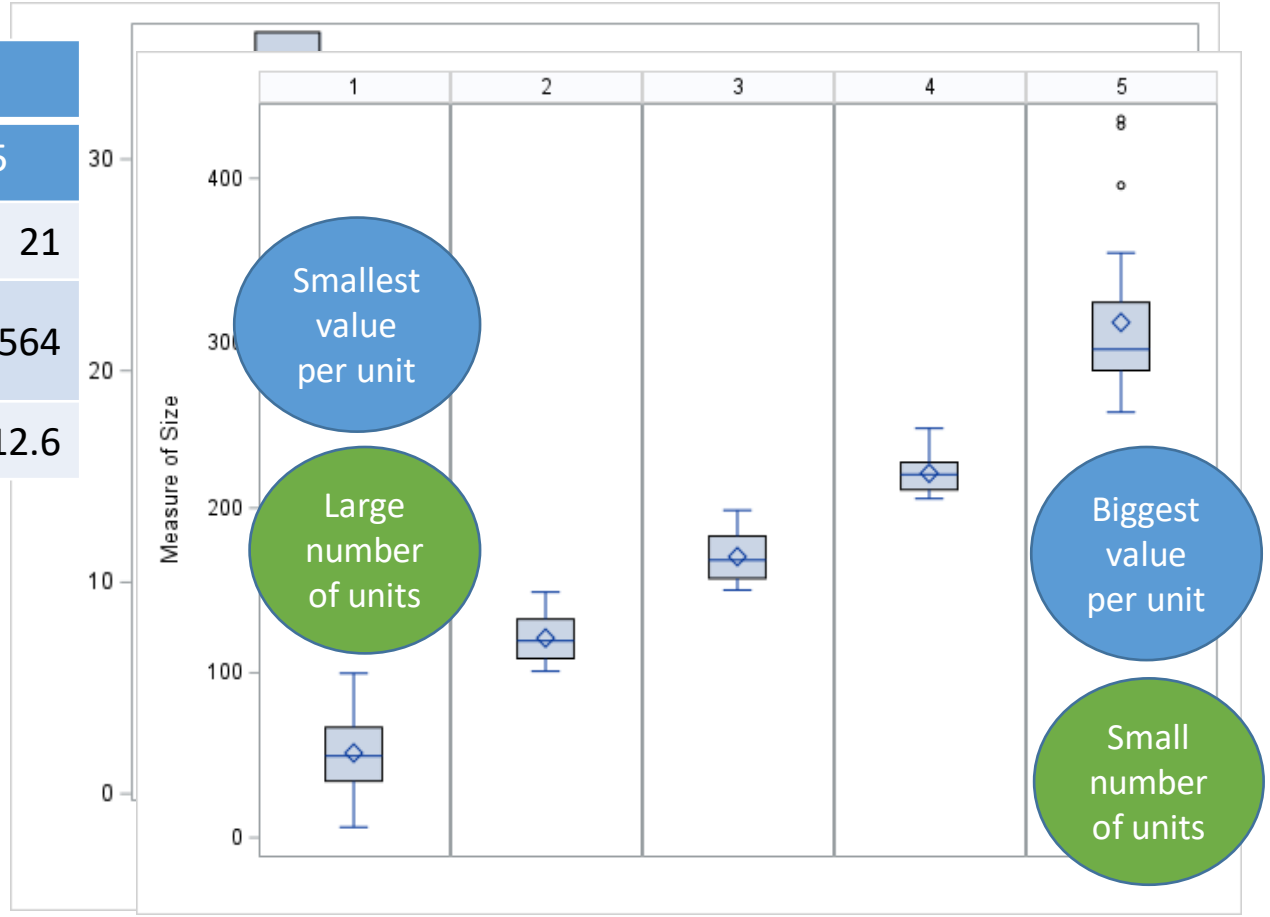
- Populations are skewed!
  - Small number of large companies
  - Majority small companies
- Publish TOTALS (and Ratios)
- Sample design accounts for skewed distribution in
  - Stratification (size)
  - Allocation (# sampled units)
  - Unit inclusion probability



United States® Census Bureau

Digression #2: Allocation for a Multipurpose Survey Sampled from a Skewed Populatoin

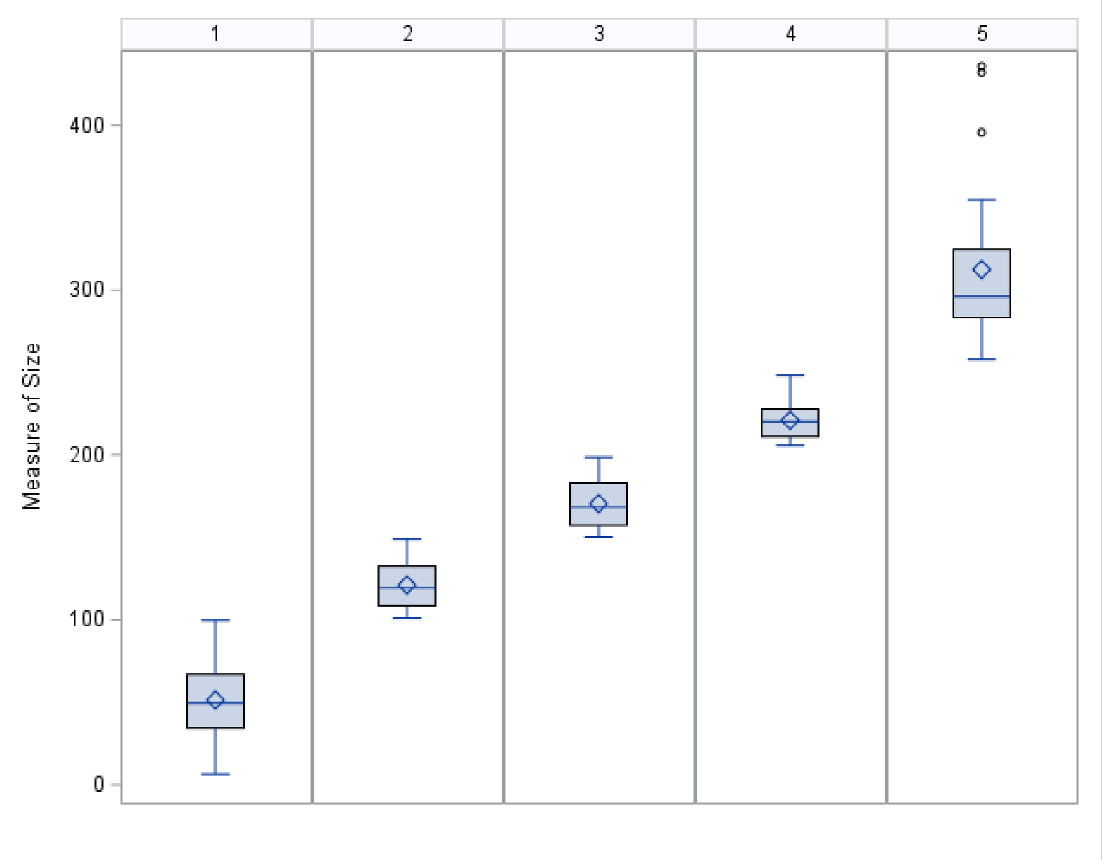
	Sampling Strata				
	1	2	3	4	5
Number of Units (N_h)	760	146	53	20	21
Total Measure of Size (MOS_h)	39023	17675	9035	4421	6564
Average Size Per Unit	51.4	121.1	170.5	221.1	312.6



United States® Census Bureau

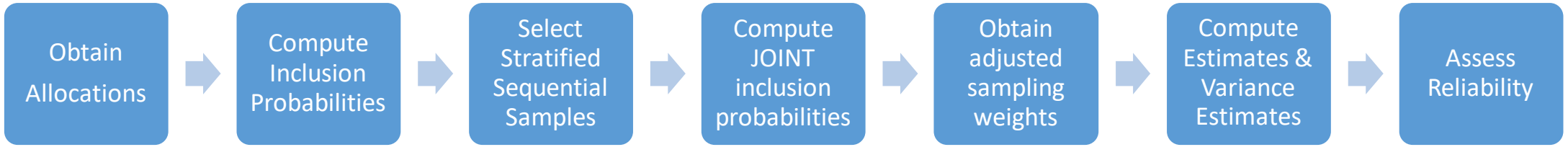
Allocation for sample size (*n*) of 50 from Population (*N*) of 1000

	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 $\sqrt{MOS_h}$	17	12	8	6	7

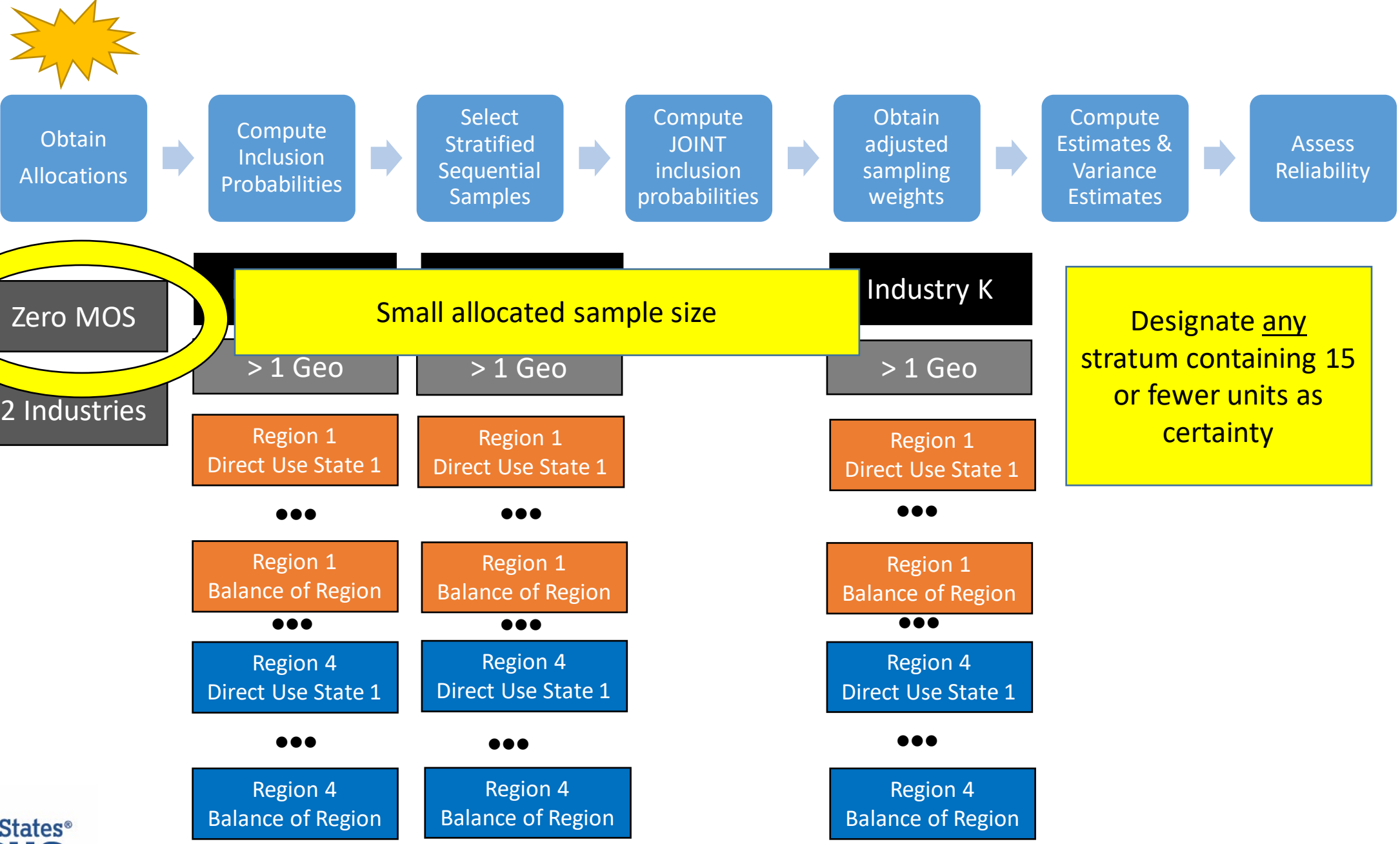


$$n_h \approx (n)\left(\frac{N_h}{\sum_h N_h}\right) \quad n_h \approx (n)\left(\frac{MOS_h}{\sum_h MOS_h}\right) \quad n_h \approx (n)\left(\frac{\sqrt{MOS_h}}{\sum_h \sqrt{MOS_h}}\right)$$

in stratum 1



United States® Census Bureau


Obtain Allocations
Compute Inclusion Probabilities
Select Stratified Sequential Samples
Compute JOINT inclusion probabilities
Obtain adjusted sampling weights
Compute Estimates & Variance Estimates
Assess Reliability
Zero MOS
2 Industries
Small allocated sample size
Industry K
> 1 Geo
> 1 Geo
> 1 Geo
Region 1 Direct Use State 1
Region 1 Direct Use State 1
Region 1 Direct Use State 1
Region 1 Balance of Region
Region 1 Balance of Region
Region 1 Balance of Region
Region 4 Direct Use State 1
Region 4 Direct Use State 1
Region 4 Direct Use State 1
Region 4 Balance of Region
Region 4 Balance of Region
Region 4 Balance of Region
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

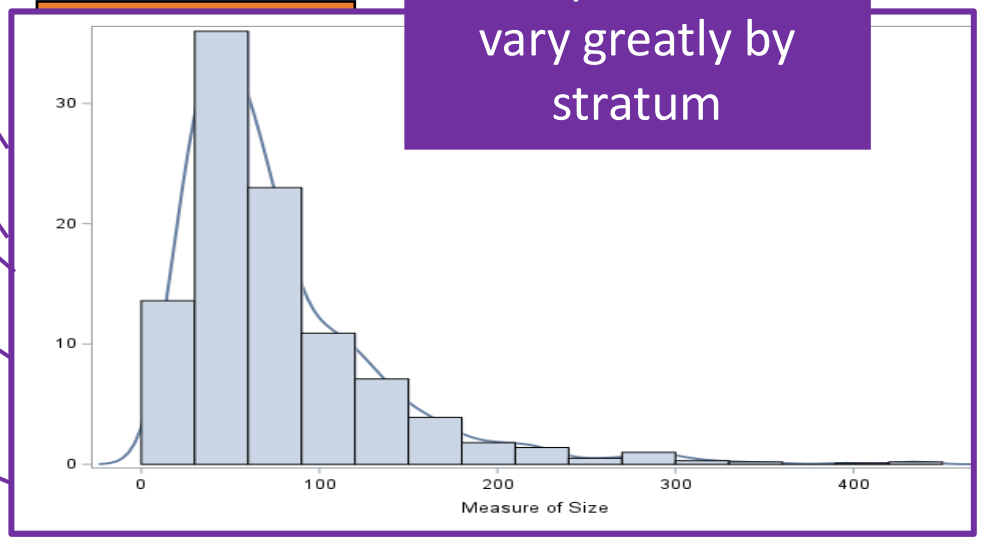
Zero MOS

2 Industries

Industry 1	Industry 2	•••	Industry K
> 1 Geo	> 1 Geo		> 1 Geo
Region 1 Direct Use State 1	Region 1 Direct Use State 1		Region 1 Direct Use State 1
•••	•••		•••
Region 1 Balance of Region	Region 1 Balance of Region		
•••	•••		
Region 4 Direct Use State 1	Region 4 Direct Use State 1		
•••	•••		
Region 4 Balance of Region	Region 4 Balance of Region		

Unit size is NOT accounted for in stratification

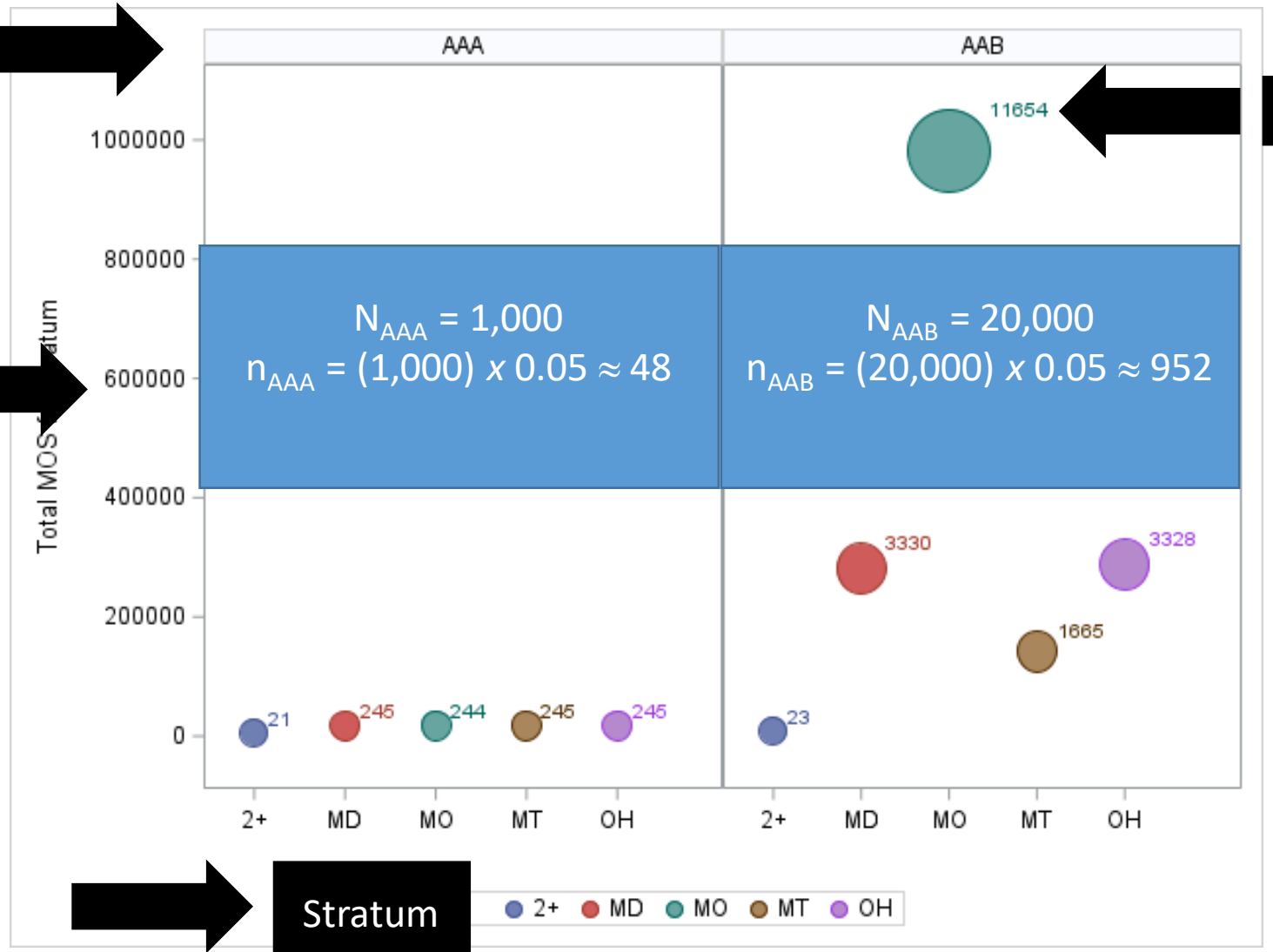
Number of companies can vary greatly by stratum



United States® Census Bureau

AIES Allocation Procedure: Step 1
Proportional Allocation to Broad Stratum (NAICS3)

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



Number of Companies

Total MOS for Stratum

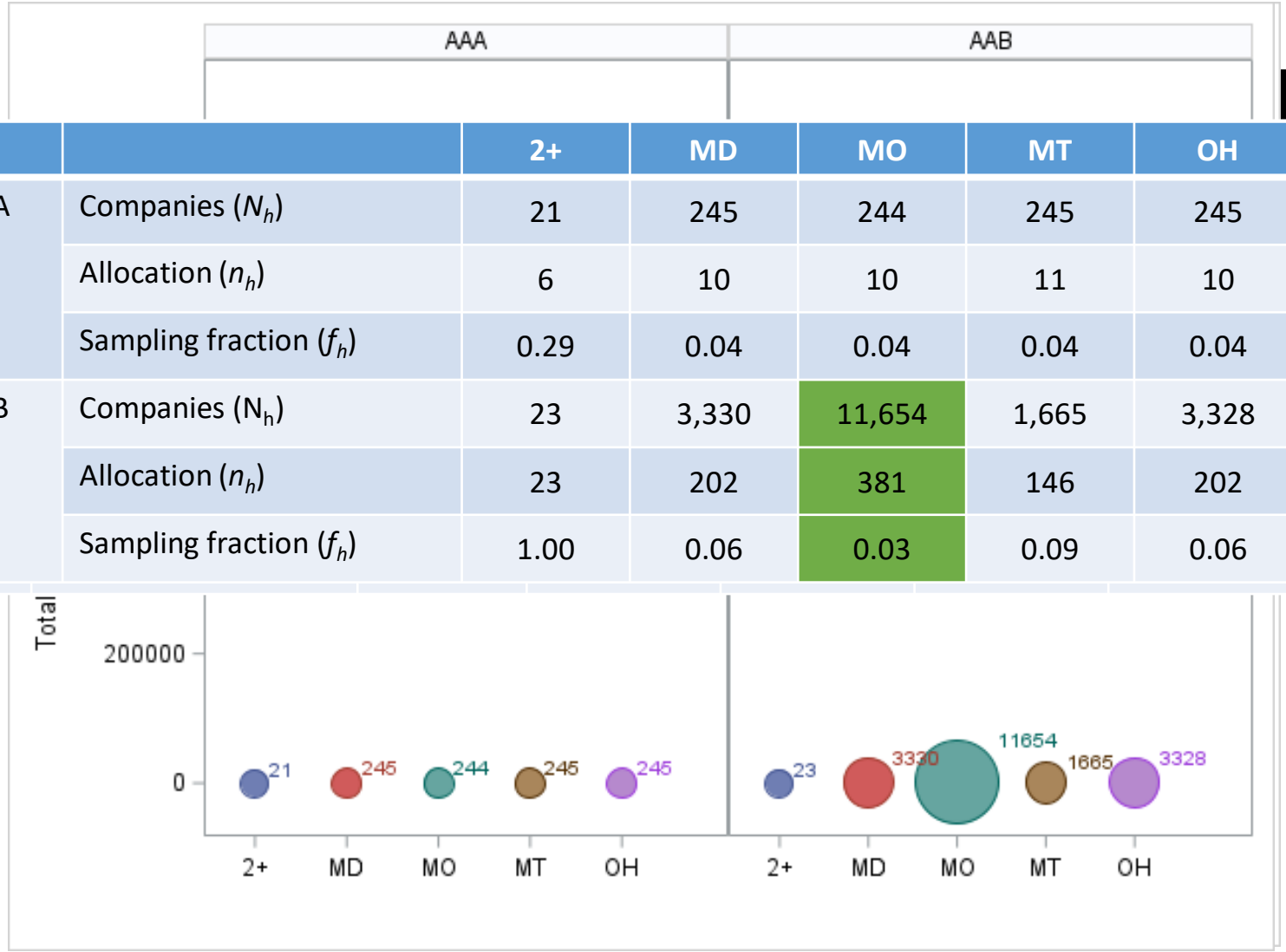
United States® Census Bureau

AIES Allocation Procedure: Step 2
Power Allocation Within Broad Stratum (NAICS3)

Number of Companies

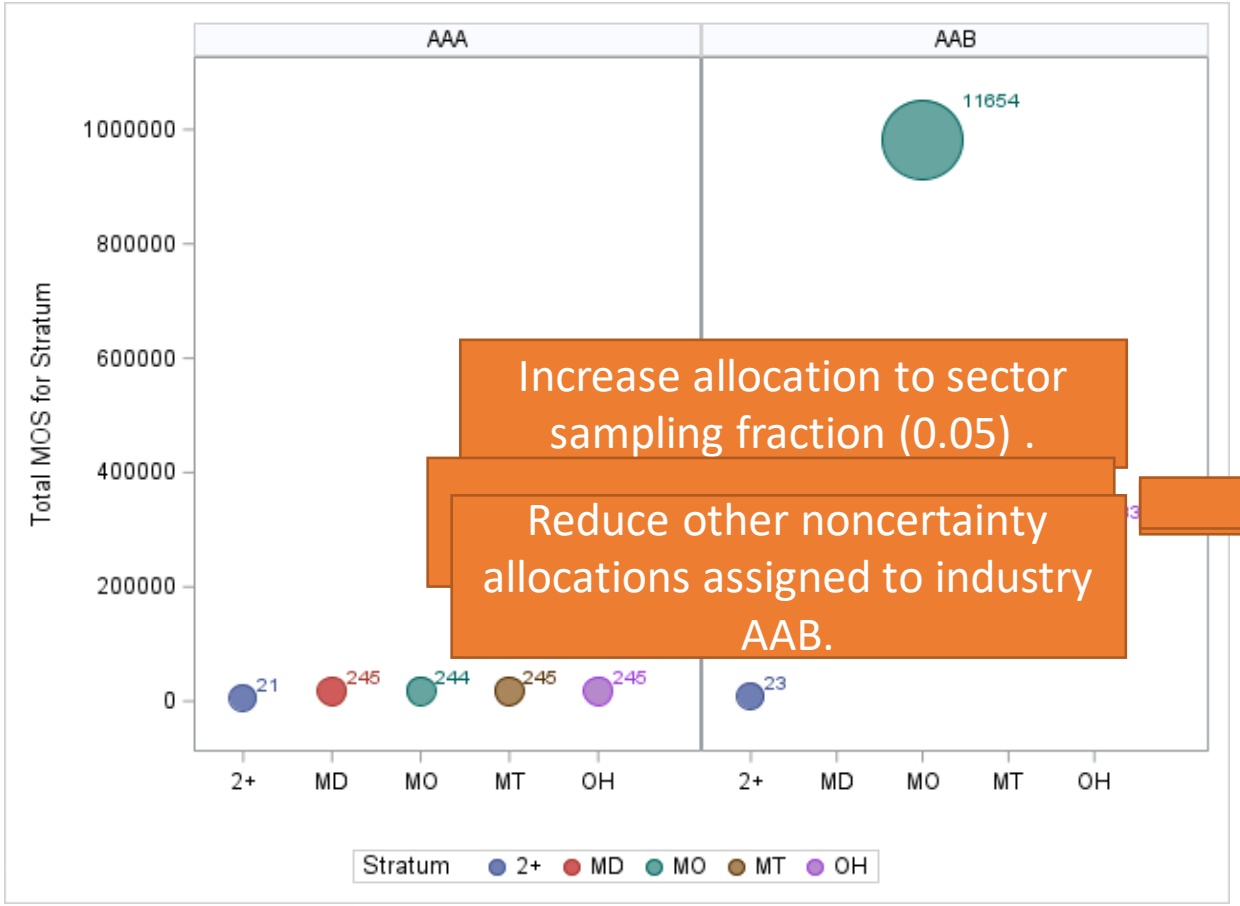
Total $\sqrt{MOS}$ for Stratum

		2+	MD	MO	MT	OH
AAA	Companies (N_h)	21	245	244	245	245
	Allocation (n_h)	6	10	10	11	10
	Sampling fraction (f_h)	0.29	0.04	0.04	0.04	0.04
AAB	Companies (N_h)	23	3,330	11,654	1,665	3,328
	Allocation (n_h)	23	202	381	146	202
	Sampling fraction (f_h)	1.00	0.06	0.03	0.09	0.06



United States® Census Bureau

AIES Allocation Procedure: Step 3
Revise allocations if necessary

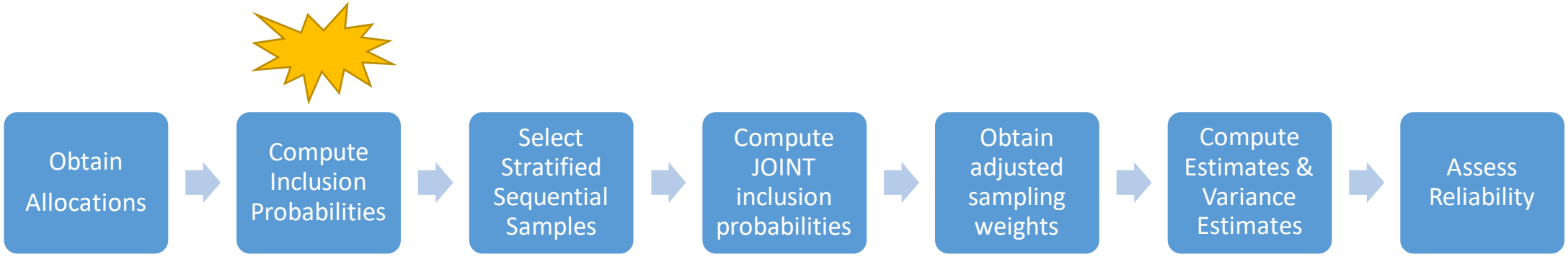


Increase allocation to sector sampling fraction (0.05) .

Reduce other noncertainty allocations assigned to industry AAB.

		2+	MD	MO	MT	OH
AAA	Companies (N_h)	21	245	244	245	245
	Allocation (n_h)	6	10	10	11	10
	Sampling fraction (f_h)	0.29	0.04	0.04	0.04	0.04
AAB	Companies (N_h)	23	3,330	11,654	1,665	3,328
	Allocation (n_h)	23	145	583	103	146
	Sampling fraction (f_h)	1.00	0.04	0.05	0.06	0.04

United States® Census Bureau



- 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

United States® Census Bureau

Create Sampling Units from Frame

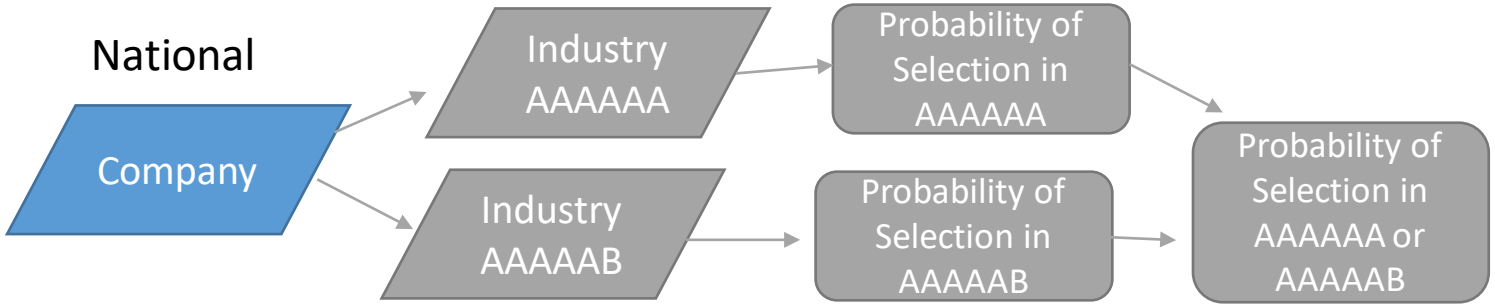
FRAME

Company	Establish-ment	Industry	State	MOS
1	1	AAAAAA	MO	56
	2	AAAAAA	MO	9
	3	AAAAAA	TN	39
	4	AAAAAB	MO	33
	5	AAAAAB	MO	20
	6	AAAAAB	TN	24
2	1	AAAAAA	MO	37
	2	AAAAAA	TN	27
3	1	AAAAAA	MO	54
4	1	AAAAAB	MO	100

Sampling Units

Company	Sampling Stratum	MOS_c
1	> 1 State	181
2	> 1 State	64
3	AAA_MO	54
4	AAA_MO	100

United States® Census Bureau



FRAME

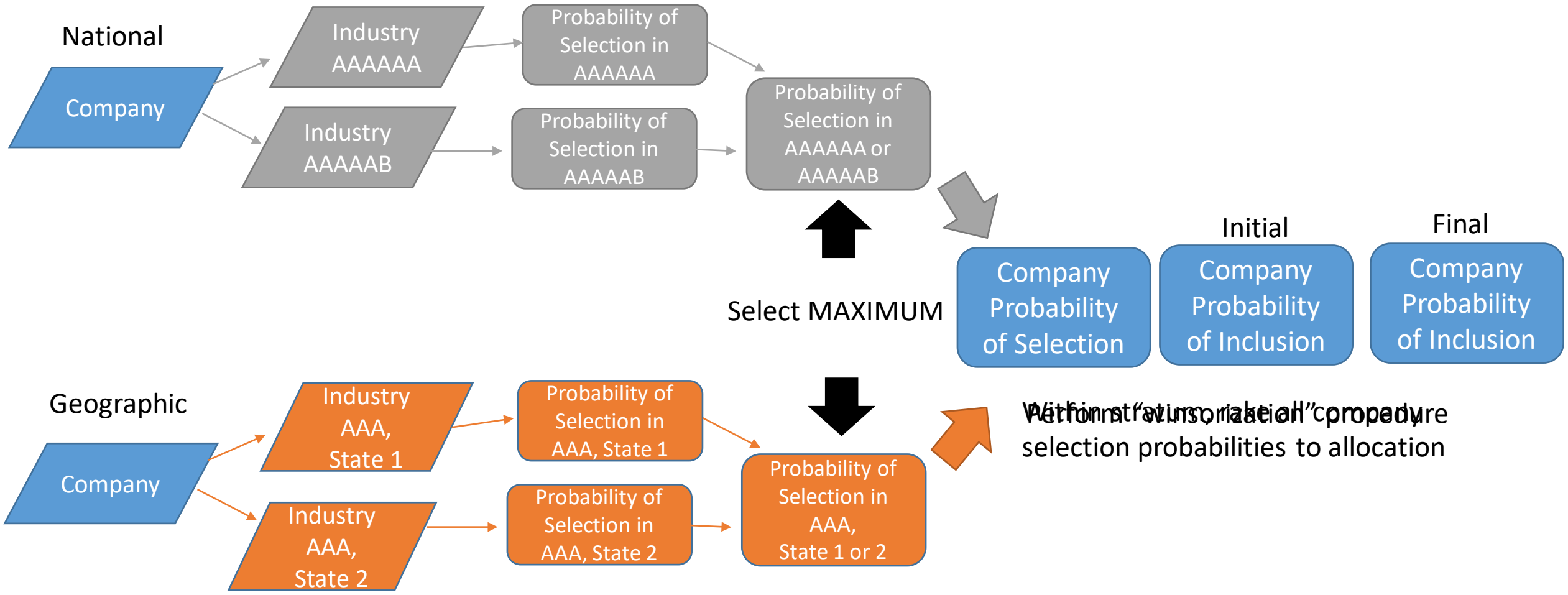
"Sampling Stratum"	Company	Pseudo-Unit	MOS	π
AAAAAA	1	1_AAAAAA	104	0.4685
AAAAAA	2	2_AAAAAA	64	0.2883
AAAAAA	3	3_AAAAAA	54	0.2432
AAAAAB	1	1_AAAAAB	77	0.4350
AAAAAB	4	4_AAAAAB	100	0.5650
	6	AAAAAB	TN	24
2	1	AAAAAA	MO	37
	2	AAAAAA	TN	27
3	1	AAAAAA	MO	54
4	1	AAAAAB	MO	100

Company National Industry π

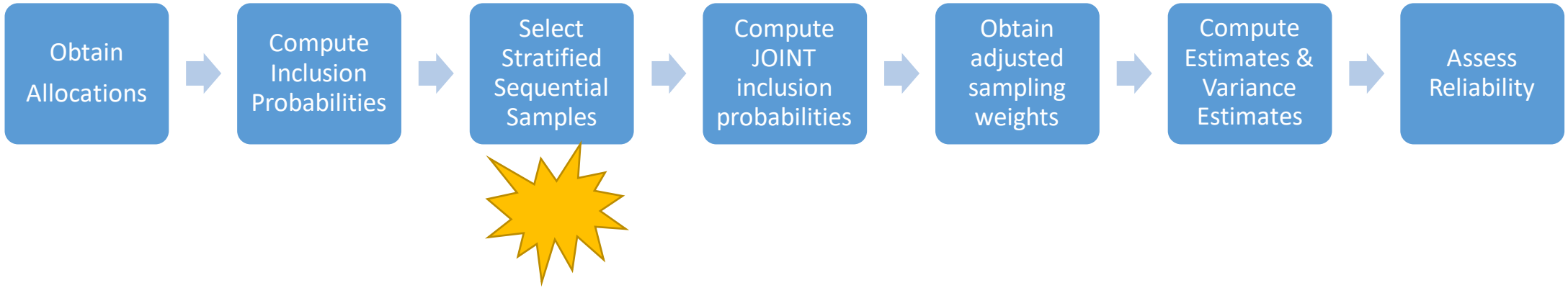
Company	π
1	1 – ((1-0.4685)(1-0.4350)) = 0.6997
2_AAAAAA	2	64
3_AAAAAA	3	54
4_AAAAAB	4	100

Probability that Company 1 is NOT sampled in AAAAAA

Probability that Company 1 is NOT sampled in AAAAAB



United States® Census Bureau

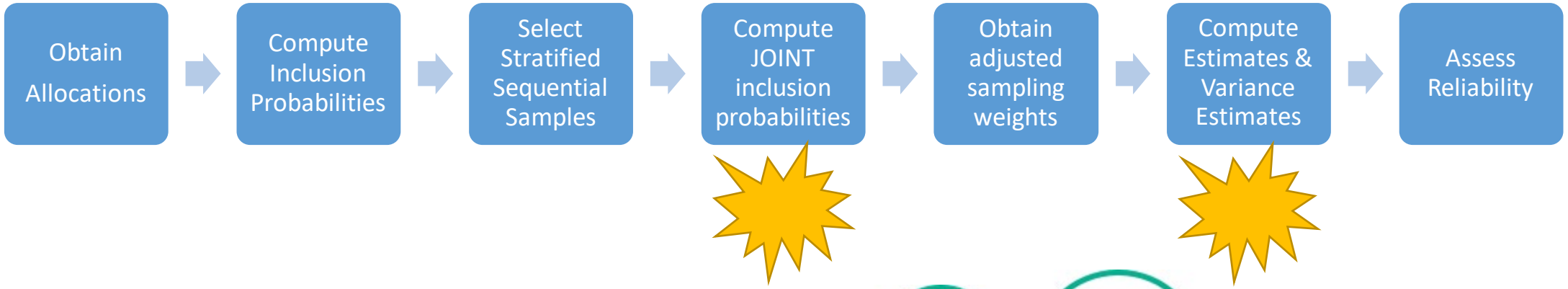


Chromy, J. R. (1979). Sequential Sample Selection Methods. In *Proceedings of the Survey Research Methods Section*, pp. 401–406. Washington, DC: American Statistical Association.

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



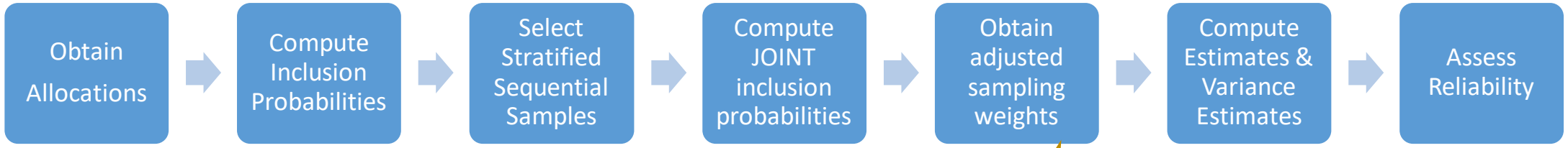
$$\hat{v}\left(\hat{Y}^k\right) = \sum_{h=1}^{H} \frac{1}{2} \sum_{i=1}^{n_h} \left(\frac{\pi_{hi}}{\pi_{h,il}} - \frac{\pi_{h,il}}{\pi_{hi}} \frac{y_{hi}^k}{\pi_{hi}} - \frac{y_{hl}^k}{\pi_{hl}} \right)^2$$

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)

Chauvet, G. (2021). A Note on Chromy's Sampling Procedure, *Journal of Survey Statistics and Methodology*, 9, pp. 1050–1061, https://doi.org/10.1093/jssam/smaa025.

United States® Census Bureau



Frame

Stratum	Company	Estab.	Industry	State	MOS	PrInc
Stratum	Company	Estab.	Industry	State	MOS	PrInc
AAA_2+	1	1	AAAAAA	MO	56	0.62
AAA_2+	1	2	AAAAAA	MO	9	
AAA_2+	1	3	AAAAAA	TN	39	
AAA_2+	1	4	AAAAAB	MO	33	
AAA_2+	1	5	AAAAAB	MO	20	
AAA_2+	1	6	AAAAAB	TN	24	
AAA_2+	2	1	AAAAAA	MO	37	0.19
AAA_2+	2	2	AAAAAA	TN	27	
AAA_2+	3	1	AAAAAA	MO	38	0.59
AAA_2+	3	2	AAAAAA	MO	22	
AAA_2+	3	3	AAAAAA	MO	96	
AAA_2+	3	4	AAAAAA	TN	43	
AAA_2+	4	1	AAAAAB	MO	4	0.60
AAA_2+	4	2	AAAAAB	TN	82	
AAA_2+	4	3	AAAAAA	TN	75	
AAA_MO	5	1	AAAAAA	MO	54	0.31
AAA_MO	6	1	AAAAAB	MO	100	0.57
AAA_MO	7	1	AAAAAB	MO	22	0.13

Industry (National)	Frame MOS	Contributing Companies
AAAAAA	496	1, 2, 3, 4, 5
AAAAAB	285	1, 4, 6, 7

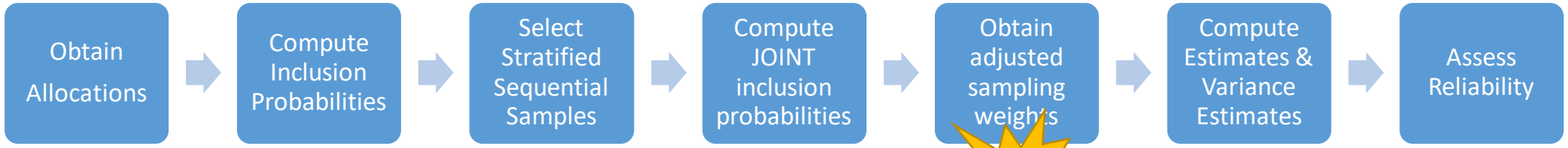
$N_{AAA\_2+} = 4$

$n_{AAA\_2+} = 2$

$N_{AAA\_MO} = 3$

$n_{AAA\_MO} = 1$

United States Census Bureau

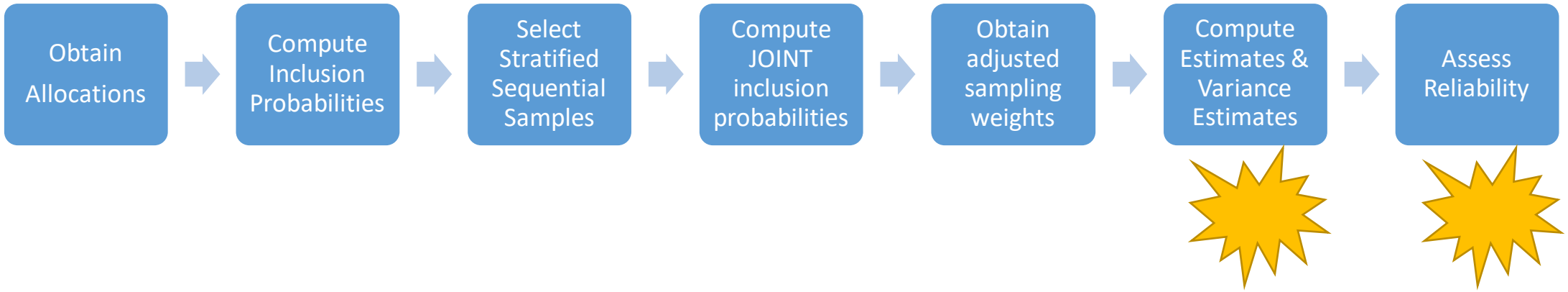


National Industry Sampling Units (Domain Estimates)

Company	Unit	MOS	Princ	Sampling Weight	Adjusted Weight
1	1_AAAAAA	104	0.62	1.6129	1.5841
1	1_AAAAAB	77	0.62	1.6129	1.5841
2	2_AAAAAA	64	0.19	0	0
3	3_AAAAAA	199	0.59	1.6949	1.6646
4	4_AAAAAB	86	0.60	0	0
4	4_AAAAAA	75	0.60	0	0
5	5_AAAAAA	54	0.31	0	0
6	6_AAAAAB	100	0.57	1.7544	1.7230
7	7_AAAAAB	22	0.13	0	0

)	Frame MOS	Estimated MOS	Ratio
	496	505	0.98
	285	300	0.95

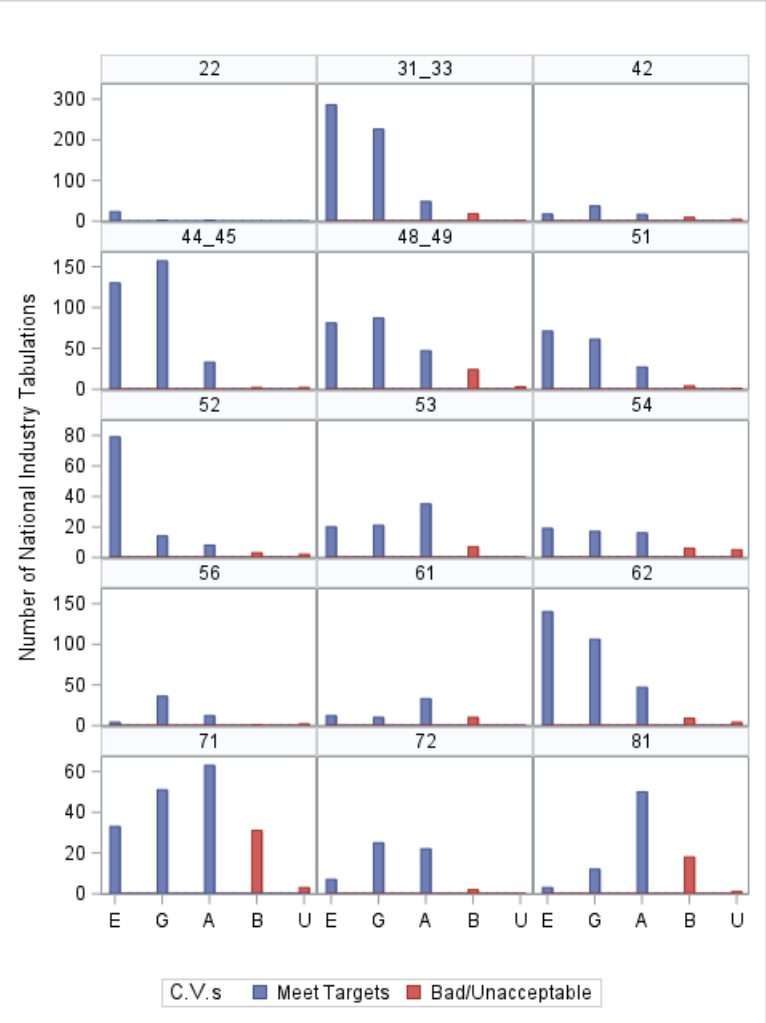
United States® Census Bureau



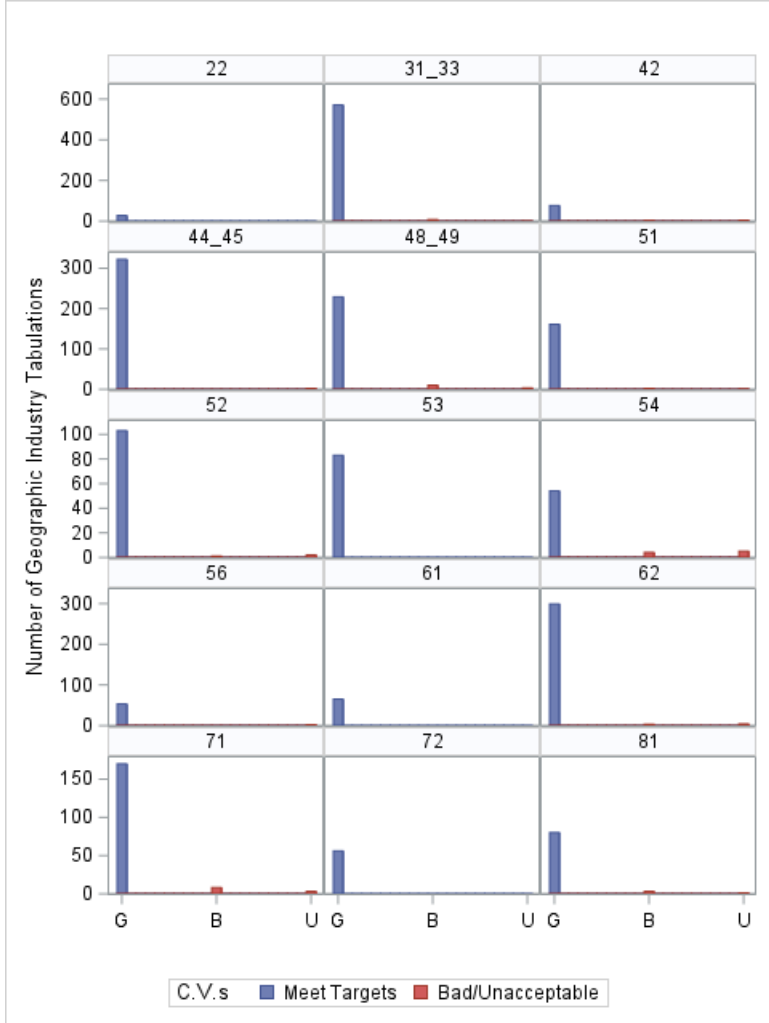
United States® Census Bureau

Results: July 2022 AIES End-to-End Test

National Industry Tabulations



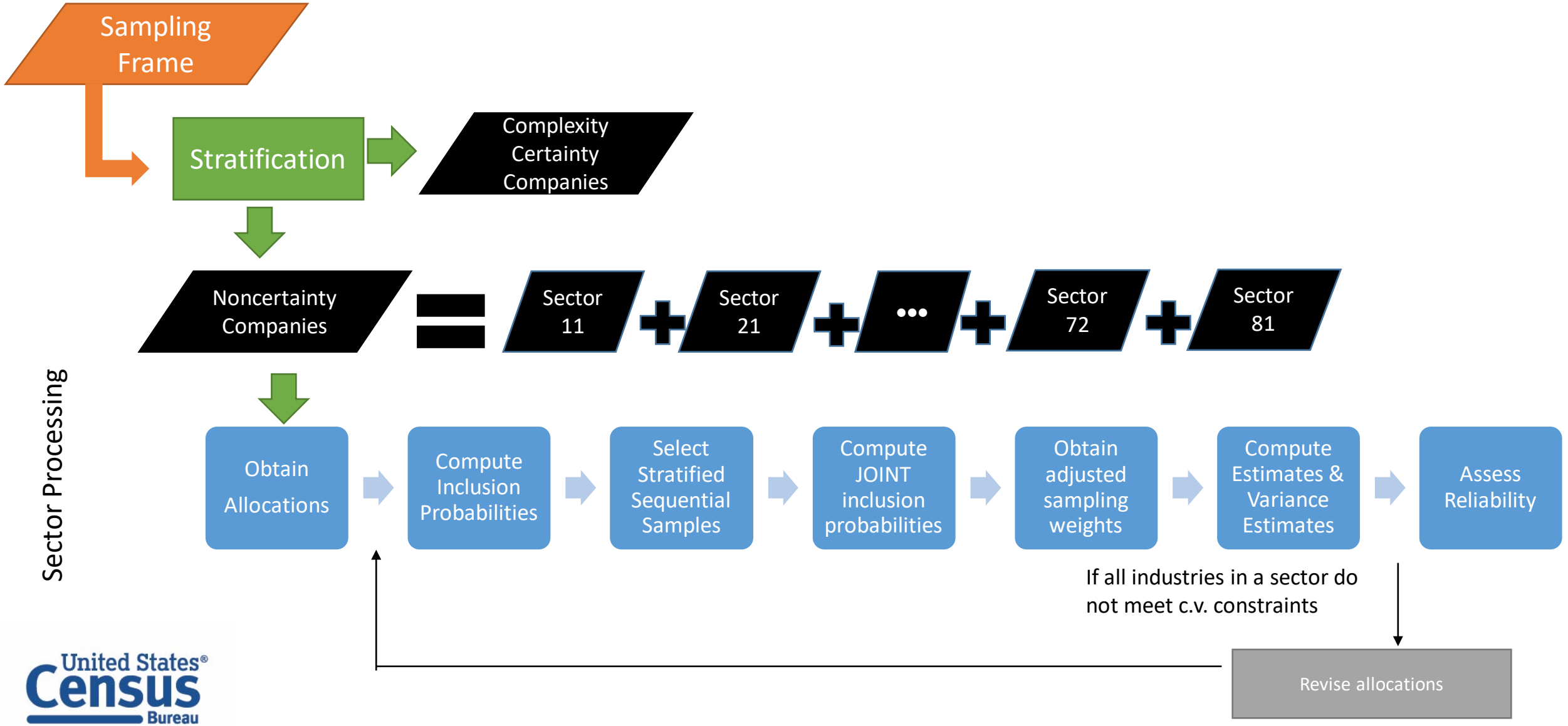
Geographic Industry Tabulations



United States® Census Bureau

Source: AIES test sample frame created in May 2022 from U.S. Census Business Register

Overview of AIES Sample Design Process



United States® Census Bureau

Questions, suggestions, comments?

katherine.j.thompson@census.gov

United States® Census Bureau