

PRASC



**Project for the Regional
Advancement of Statistics
in the Caribbean**

**Projet régional pour
l'avancement de la statistique
dans les Caraïbes**

Funded by the
Government
of Canada

Canada



Stratified and Probability Proportional to Size Sampling

Project for the Regional Advancement of Statistics in the Caribbean (PRASC)

Gavin Thompson
March 30, 2021, Virtual, from Ottawa, Canada



Delivering insight through data for a better Canada



Bad Math II

NASA Pilot Study



National Aviation Operational Monitoring Service

- From 2001-2004, NASA did a survey of US Pilots (all types) on certain questions of flying time, in-air occurrences, etc.
- In all they interviewed approximately 30,000 pilots.
- The total cost of the exercise was \$11M US

National Aviation Operational Monitoring Service

- On December 31st, 2007, they released 16,000 pages of raw survey results.
- Without any kind of analysis being done on the data.
- In PDF.

National Aviation Operational Monitoring Service

■ Results

- The survey approximated 4 times the number of engine failures than were recorded with the FAA.
- The estimate for the number of diversion landings was impossibly high.
- Overall, the data was pretty much meaningless.

National Aviation Operational Monitoring Service

- On the plus side:
 - No one got hurt conducting the survey.
 - NASA claims it was a survey to explore survey methodology, they now know this system doesn't work.
 - There were already 150 different measuring tools in place to capture the data they were examining!



Statistical Methods

Stratified Sampling

and

PPS Sampling



Overarching Processes

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

Stratified Sampling

Is a sampling process consisting of:

- Dividing the frame into homogeneous, mutually exclusive groups called strata
- Selecting independent samples from each of the strata. Any sample selection method can be used to sample within strata (e.g. SRS, Systematic, PPS)

Stratified Sampling

	18-34	35-64	65+
Age			
M			
F			
Sex			

Example - Farm Expenditure

Farms	Acreage	Expenditure (\$)
1	50	26,000
2	1,000	470,000
3	125	63,800
4	300	145,000
5	500	230,000
6	25	12,500
Total		947,300
Mean		157,883

Example (cont'd) - Stratified Sample of n=4 Farms

Step 1 – Divide population into strata (e.g. small/large)

STRATUM 1 - Small (Under 300 Acres)		
Farm	Acreage	Expenditure (\$)
1	50	26,000
3	125	63,800
6	25	12,500

STRATUM 2 – Large (300 Acres or Over)		
Farm	Acreage	Expenditure (\$)
2	1,000	470,000
4	300	145,000
5	500	230,000



Example (cont'd) - Stratified Sample of n=4 Farms

Step 2 - Select 2 farms from each stratum (SRS of size 2 in each stratum)

Possible samples of size 4	Mean of Farm Expenditure (\$)
1,3,2,4	176,200
1,3,2,5	197,450
1,6,2,4	163,375
1,6,2,5	184,625
1,3,4,5	116,200
1,6,4,5	103,375
3,6,2,4	172,825
3,6,2,4	172,825
3,6,4,5	112,825
Average over all Samples	157,883

Comparison of Methods



Possible samples of size 4	Mean of expenditure (\$) (SRS)	Mean of expenditure (\$) (SRS Stratified)
1,2,3,4	176,200	176,200
1,2,3,5	197,450	197,450
1,2,3,6	143,075	
1,2,4,5	217,750	
1,2,4,6	163,375	163,375
1,2,5,6	184,625	184,625
1,3,4,5	116,200	116,200
1,3,4,6	61,825	
1,3,5,6	83,075	
1,4,5,6	103,375	103,375
2,3,4,5	227,200	
2,3,4,6	172,825	172,825
2,3,5,6	194,075	194,075
2,4,5,6	214,375	
3,4,5,6	112,825	112,825
Average over all Samples	157,883	157,883

Stratified Sampling

Advantages

- Can increase precision of overall estimates of population
- Can be operationally or administratively more convenient

Disadvantages

- Stratification variables must be available for all units of the frame
- Requires restructuring of frame prior to sampling
- Stratification might not be good for all variables of interest

Stratified Sampling - Factors that Influence Efficiency

- Choice of stratification variables with respect to survey characteristics of interest
- Number of strata
- Determination of stratum limits
- Allocation of sample to strata - proportionate or disproportionate

Determination of Stratum Limits

STRATUM 1 - Small (Under 300 Acres)		
Farm	Acreage	Expenditure (\$)
1	50	26,000
3	125	63,800
6	25	12,500
STRATUM 2 – Large (300 Acres or Over)		
Farm	Acreage	Expenditure (\$)
2	1,000	470,000
4	300	145,000
5	500	230,000

Number of Strata

- 2017 LFS Survey Redesign

Stratum Type	Count
Total	1,153
Regular	930
Aboriginal	20
High Income	42
Immigrant	55
Large-Cluster	153
One-Stage	138
Remote	555

Stratified Sampling - Allocation of Sample to Strata

Proportional Allocation (also called N-Proportional Allocation)

- Sample allocated proportionally to the total number of units in each stratum

Disproportional Allocation

- Sample allocated disproportional to the total number of units in each stratum

Why Disproportionate?

Situations which indicate the need for disproportional allocation:

- Estimates are required at the stratum level (smaller strata might need to be over-sampled)
- There are differences among the strata in terms of:

Non-response

and/or

The variability among units

and/or

The cost per units

Allocation Affected by Non-Response

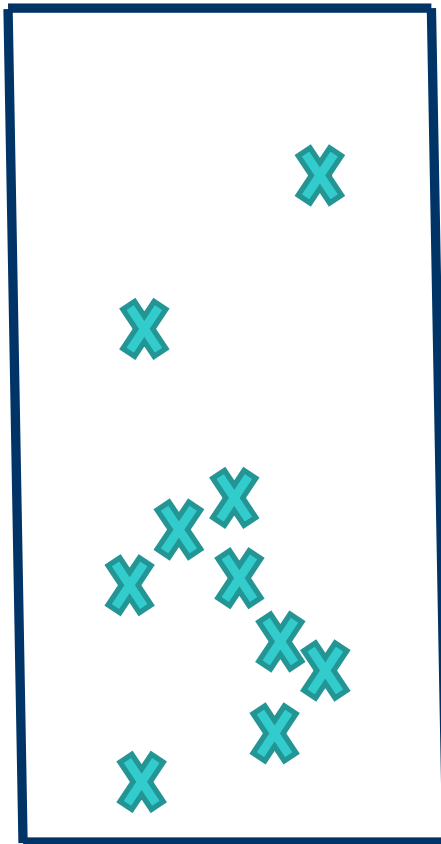
- Suppose we calculate the necessary number of respondents from a stratum is 1,000.
- Historically Stratum A has a response rate of 80% and Stratum B has a response rate of 50%.
- Stratum A would need to have (at a minimum) a sample size of 1,250
- Stratum B would need to have (at a minimum) a sample size of 2,000

Allocation Affected by Variability

- Imagine a survey to estimate household net worth
- A low income neighbourhood will have a fairly stable estimate as all respondents have roughly the same income
- A high income neighbourhood could have very high variability in their estimates

Allocation Affected by Cost

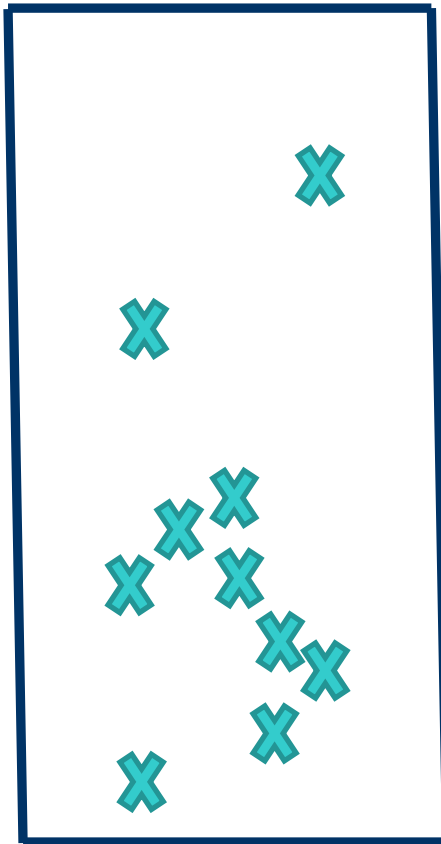
Saskatchewan



- Suppose a survey is being conducted in Sask.
- 2,000 interviews will be conducted face to face
- 1,000 in Regina and Saskatoon and 1,000 elsewhere.
- The cost for an interview is \$50 in the city and \$100 elsewhere...
- The cost of collection will be \$150,000

Allocation Affected by Cost

Saskatchewan



- Suppose analysis shows that 500 non-city interviews will give the necessary results for the survey.
- Two things can happen:
 - Collection cost can be reduced by $500 \times \$100 = \$50,000$
 - Boost # of interviews by 1,000 in cities ($\$50,000 / \50)... Giving 2,500 interviews overall

Disproportional Allocation Methods

SQRT(N)-Proportional Allocation

- Used when interested in having good precision for stratum estimates.

Y-Proportional Allocation

- Allocates sample proportional to the **size** of all units in a stratum (not the number of units). Often used in business and farm surveys as a way to deal with unit to unit variability.

Neyman Allocation

- Assigns more sample units to larger strata, and/or strata with the highest variances.

Optimum Allocation

- Takes cost into account by under-sampling those strata where the costs are higher.

Population of the 10 Provinces, n=10,000

Province	Population (N)	Sample Allocation
		N-Prop
NFLD	519,716	148
PEI	142,907	41
NS	923,598	264
NB	747,101	213
QC	8,164,361	2,330
ON	13,448,494	3,838
MB	1,278,365	365
SK	1,098,352	313
AB	4,067,175	1,161
BC	4,648,055	1,327
Total	35,038,124	10,000



Population of the 10 Provinces, n=10,000

























































Province	Population (N)	SQR(N)	Sample Allocation	
			N-Prop	SQR(N)-Prop
NFLD	519,716	720.9	148	456
PEI	142,907	378.0	41	239
NS	923,598	961.0	264	608
NB	747,101	864.4	213	547
QC	8,164,361	2,857.3	2,330	1,808
ON	13,448,494	3,667.2	3,838	2,321
MB	1,278,365	1,130.6	365	716
SK	1,098,352	1,048.0	313	663
AB	4,067,175	2,016.7	1,161	1,276
BC	4,648,055	2155.9	1,327	1,364
Total	35,038,124	15,800.2	10,000	10,000

Take-All Stratification

Census

SRS

Systematic

Exercise 3: Systematic Stratified Sampling

1. Draw a sample of 10 farms
 - Stratified by farm type
 - Allocate the sample (How many in each stratum?)
 - Allocated proportionally to the number of farms.
 - Calculate k (keep one decimal place)
 - Random start: Stratum 1: 3 Stratum 2: 6
2. Calculate the mean of expenditures *in each stratum* for this sample



Stratum	Obs	Type	Expenses								
1	1	Cattle	20,000	2	1	Corn	6,000	2	36	Other	3,000
1	2	Cattle	56,000	2	2	Corn	3,800	2	37	Corn	8,700
1	3	Cattle	48,000	2	3	Corn	6,700	2	38	Other	5,000
1	4	Cattle	22,500	2	4	Corn	11,000	2	39	Corn	34,000
1	5	Cattle	19,000	2	5	Corn	4,800	2	40	Corn	7,300
1	6	Cattle	17,500	2	6	Corn	7,600	2	41	Corn	23,000
1	7	Cattle	21,000	2	7	Corn	4,000	2	42	Other	7,000
1	8	Cattle	65,000	2	8	Corn	10,000	2	43	Other	3,800
1	9	Cattle	39,000	2	9	Corn	2,700	2	44	Other	17,000
1	10	Cattle	23,000	2	10	Corn	8,500	2	45	Other	8,000
1	11	Cattle	28,000	2	11	Corn	12,500	2	46	Corn	24,000
1	12	Cattle	18,500	2	12	Corn	12,500	2	47	Corn	13,000
1	13	Cattle	26,000	2	13	Corn	17,500	2	48	Corn	12,500
1	14	Cattle	19,000	2	14	Corn	22,500	2	49	Corn	7,500
1	15	Cattle	22,000	2	15	Corn	8,000	2	50	Corn	24,000
1	16	Cattle	17,000	2	16	Corn	5,000	2	51	Other	4,000
1	17	Cattle	21,000	2	17	Corn	18,500	2	52	Corn	7,500
1	18	Cattle	96,000	2	18	Corn	6,800	2	53	Other	5,000
1	19	Cattle	52,500	2	19	Corn	12,500	2	54	Other	4,000
1	20	Cattle	84,000	2	20	Corn	7,500	2	55	Other	3,500
1	21	Cattle	65,000	2	21	Corn	45,000	2	56	Other	6,000
1	22	Cattle	48,000	2	22	Corn	12,000	2	57	Corn	13,000
1	23	Cattle	27,500	2	23	Corn	4,500	2	58	Corn	24,000
1	24	Cattle	58,000	2	24	Corn	15,000	2	59	Corn	11,500
1	25	Cattle	23,000	2	25	Corn	22,000	2	60	Corn	7,000
1	26	Cattle	62,500	2	26	Corn	31,000	2	61	Other	22,500
1	27	Cattle	120,000	2	27	Corn	17,000	2	62	Other	26,000
1	28	Cattle	19,000	2	28	Corn	5,700	2	63	Corn	8,000
1	29	Cattle	96,000	2	29	Other	4,800	2	64	Corn	23,000
1	30	Cattle	28,000	2	30	Other	3,200	2	65	Corn	12,500
1	31	Cattle	48,000	2	31	Corn	10,000	2	66	Corn	20,000
1	32	Cattle	14,000	2	32	Corn	8,500				
1	33	Cattle	9,600	2	33	Corn	16,500				
1	34	Cattle	31,000	2	34	Other	5,000				
				2	35	Corn	6,200				

Calculate Allocation
N=100



Stratum	Obs	Type	Expenses								
1	1	Cattle	20,000	2	1	Corn	6,000	2	36	Other	3,000
1	2	Cattle	56,000	2	2	Corn	3,800	2	37	Corn	8,700
1	3	Cattle	48,000	2	3	Corn	6,700	2	38	Other	5,000
1	4	Cattle	22,500	2	4	Corn	11,000	2	39	Corn	34,000
1	5	Cattle	19,000	2	5	Corn	4,800	2	40	Corn	7,300
1	6	Cattle	17,500	2	6	Corn	7,600	2	41	Corn	23,000
1	7	Cattle	21,000	2	7	Corn	4,000	2	42	Other	7,000
1	8	Cattle	65,000	2	8	Corn	10,000	2	43	Other	3,800
1	9	Cattle	39,000	2	9	Corn	2,700	2	44	Other	17,000
1	10	Cattle	23,000	2	10	Corn	8,500	2	45	Other	8,000
1	11	Cattle	28,000	2	11	Corn	12,500	2	46	Corn	24,000
1	12	Cattle	18,500	2	12	Corn	12,500	2	47	Corn	13,000
1	13	Cattle	26,000	2	13	Corn	17,500	2	48	Corn	12,500
1	14	Cattle	19,000	2	14	Corn	22,500	2	49	Corn	7,500
1	15	Cattle	22,000	2	15	Corn	8,000	2	50	Corn	24,000
1	16	Cattle	17,000	2	16	Corn	5,000	2	51	Other	4,000
1	17	Cattle	21,000	2	17	Corn	18,500	2	52	Corn	7,500
1	18	Cattle	96,000	2	18	Corn	6,800	2	53	Other	5,000
1	19	Cattle	52,500	2	19	Corn	12,500	2	54	Other	4,000
1	20	Cattle	84,000	2	20	Corn	7,500	2	55	Other	3,500
1	21	Cattle	65,000	2	21	Corn	45,000	2	56	Other	6,000
1	22	Cattle	48,000	2	22	Corn	12,000	2	57	Corn	13,000
1	23	Cattle	27,500	2	23	Corn	4,500	2	58	Corn	24,000
1	24	Cattle	58,000	2	24	Corn	15,000	2	59	Corn	11,500
1	25	Cattle	23,000	2	25	Corn	22,000	2	60	Corn	7,000
1	26	Cattle	62,500	2	26	Corn	31,000	2	61	Other	22,500
1	27	Cattle	120,000	2	27	Corn	17,000	2	62	Other	26,000
1	28	Cattle	19,000	2	28	Corn	5,700	2	63	Corn	8,000
1	29	Cattle	96,000	2	29	Other	4,800	2	64	Corn	23,000
1	30	Cattle	28,000	2	30	Other	3,200	2	65	Corn	12,500
1	31	Cattle	48,000	2	31	Corn	10,000	2	66	Corn	20,000
1	32	Cattle	14,000	2	32	Corn	8,500				
1	33	Cattle	9,600	2	33	Corn	16,500				
1	34	Cattle	31,000	2	34	Other	5,000				
				2	35	Corn	6,200				

Calculate Allocation
N=100

Stratum 1: 34/100
=3.4
=3

Stratum 2: 66/100
=6.6
=7

K=N/n



Stratum	Obs	Type	Expenses								
1	1	Cattle	20,000	2	1	Corn	6,000	2	36	Other	3,000
1	2	Cattle	56,000	2	2	Corn	3,800	2	37	Corn	8,700
1	3	Cattle	48,000	2	3	Corn	6,700	2	38	Other	5,000
1	4	Cattle	22,500	2	4	Corn	11,000	2	39	Corn	34,000
1	5	Cattle	19,000	2	5	Corn	4,800	2	40	Corn	7,300
1	6	Cattle	17,500	2	6	Corn	7,600	2	41	Corn	23,000
1	7	Cattle	21,000	2	7	Corn	4,000	2	42	Other	7,000
1	8	Cattle	65,000	2	8	Corn	10,000	2	43	Other	3,800
1	9	Cattle	39,000	2	9	Corn	2,700	2	44	Other	17,000
1	10	Cattle	23,000	2	10	Corn	8,500	2	45	Other	8,000
1	11	Cattle	28,000	2	11	Corn	12,500	2	46	Corn	24,000
1	12	Cattle	18,500	2	12	Corn	12,500	2	47	Corn	13,000
1	13	Cattle	26,000	2	13	Corn	17,500	2	48	Corn	12,500
1	14	Cattle	19,000	2	14	Corn	22,500	2	49	Corn	7,500
1	15	Cattle	22,000	2	15	Corn	8,000	2	50	Corn	24,000
1	16	Cattle	17,000	2	16	Corn	5,000	2	51	Other	4,000
1	17	Cattle	21,000	2	17	Corn	18,500	2	52	Corn	7,500
1	18	Cattle	96,000	2	18	Corn	6,800	2	53	Other	5,000
1	19	Cattle	52,500	2	19	Corn	12,500	2	54	Other	4,000
1	20	Cattle	84,000	2	20	Corn	7,500	2	55	Other	3,500
1	21	Cattle	65,000	2	21	Corn	45,000	2	56	Other	6,000
1	22	Cattle	48,000	2	22	Corn	12,000	2	57	Corn	13,000
1	23	Cattle	27,500	2	23	Corn	4,500	2	58	Corn	24,000
1	24	Cattle	58,000	2	24	Corn	15,000	2	59	Corn	11,500
1	25	Cattle	23,000	2	25	Corn	22,000	2	60	Corn	7,000
1	26	Cattle	62,500	2	26	Corn	31,000	2	61	Other	22,500
1	27	Cattle	120,000	2	27	Corn	17,000	2	62	Other	26,000
1	28	Cattle	19,000	2	28	Corn	5,700	2	63	Corn	8,000
1	29	Cattle	96,000	2	29	Other	4,800	2	64	Corn	23,000
1	30	Cattle	28,000	2	30	Other	3,200	2	65	Corn	12,500
1	31	Cattle	48,000	2	31	Corn	10,000	2	66	Corn	20,000
1	32	Cattle	14,000	2	32	Corn	8,500				
1	33	Cattle	9,600	2	33	Corn	16,500				
1	34	Cattle	31,000	2	34	Other	5,000				
				2	35	Corn	6,200				

Calculate Allocation
N=100

Stratum 1: 34/100
=3.4
=3

Stratum 2: 66/100
=6.6
=7

$k = N/n$
 $k_1 = 34/3 = 11.3$
 $k_2 = 66/7 = 9.4$



Stratum	Obs	Type	Expenses								
1	1	Cattle	20,000	2	1	Corn	6,000	2	36	Other	3,000
1	2	Cattle	56,000	2	2	Corn	3,800	2	37	Corn	8,700
1	3	Cattle	48,000	2	3	Corn	6,700	2	38	Other	5,000
1	4	Cattle	22,500	2	4	Corn	11,000	2	39	Corn	34,000
1	5	Cattle	19,000	2	5	Corn	4,800	2	40	Corn	7,300
1	6	Cattle	17,500	2	6	Corn	7,600	2	41	Corn	23,000
1	7	Cattle	21,000	2	7	Corn	4,000	2	42	Other	7,000
1	8	Cattle	65,000	2	8	Corn	10,000	2	43	Other	3,800
1	9	Cattle	39,000	2	9	Corn	2,700	2	44	Other	17,000
1	10	Cattle	23,000	2	10	Corn	8,500	2	45	Other	8,000
1	11	Cattle	28,000	2	11	Corn	12,500	2	46	Corn	24,000
1	12	Cattle	18,500	2	12	Corn	12,500	2	47	Corn	13,000
1	13	Cattle	26,000	2	13	Corn	17,500	2	48	Corn	12,500
1	14	Cattle	19,000	2	14	Corn	22,500	2	49	Corn	7,500
1	15	Cattle	22,000	2	15	Corn	8,000	2	50	Corn	24,000
1	16	Cattle	17,000	2	16	Corn	5,000	2	51	Other	4,000
1	17	Cattle	21,000	2	17	Corn	18,500	2	52	Corn	7,500
1	18	Cattle	96,000	2	18	Corn	6,800	2	53	Other	5,000
1	19	Cattle	52,500	2	19	Corn	12,500	2	54	Other	4,000
1	20	Cattle	84,000	2	20	Corn	7,500	2	55	Other	3,500
1	21	Cattle	65,000	2	21	Corn	45,000	2	56	Other	6,000
1	22	Cattle	48,000	2	22	Corn	12,000	2	57	Corn	13,000
1	23	Cattle	27,500	2	23	Corn	4,500	2	58	Corn	24,000
1	24	Cattle	58,000	2	24	Corn	15,000	2	59	Corn	11,500
1	25	Cattle	23,000	2	25	Corn	22,000	2	60	Corn	7,000
1	26	Cattle	62,500	2	26	Corn	31,000	2	61	Other	22,500
1	27	Cattle	120,000	2	27	Corn	17,000	2	62	Other	26,000
1	28	Cattle	19,000	2	28	Corn	5,700	2	63	Corn	8,000
1	29	Cattle	96,000	2	29	Other	4,800	2	64	Corn	23,000
1	30	Cattle	28,000	2	30	Other	3,200	2	65	Corn	12,500
1	31	Cattle	48,000	2	31	Corn	10,000	2	66	Corn	20,000
1	32	Cattle	14,000	2	32	Corn	8,500				
1	33	Cattle	9,600	2	33	Corn	16,500				
1	34	Cattle	31,000	2	34	Other	5,000				
				2	35	Corn	6,200				

Calculate Allocation
 $N=100$

Stratum 1: $34/100$
 $=3.4$
 $=3$

Stratum 2: $66/100$
 $=6.6$
 $=7$

$k=N/n$
 $k_1=34/3=11.3$
 $k_2=66/7=9.4$
 $rs_1=3$
 $rs_2=6$



Stratum	Obs	Type	Expenses								
1	1	Cattle	20,000	2	1	Corn	6,000	2	36	Other	3,000
1	2	Cattle	56,000	2	2	Corn	3,800	2	37	Corn	8,700
1	3	Cattle	48,000	2	3	Corn	6,700	2	38	Other	5,000
1	4	Cattle	22,500	2	4	Corn	11,000	2	39	Corn	34,000
1	5	Cattle	19,000	2	5	Corn	4,800	2	40	Corn	7,300
1	6	Cattle	17,500	2	6	Corn	7,600	2	41	Corn	23,000
1	7	Cattle	21,000	2	7	Corn	4,000	2	42	Other	7,000
1	8	Cattle	65,000	2	8	Corn	10,000	2	43	Other	3,800
1	9	Cattle	39,000	2	9	Corn	2,700	2	44	Other	17,000
1	10	Cattle	23,000	2	10	Corn	8,500	2	45	Other	8,000
1	11	Cattle	28,000	2	11	Corn	12,500	2	46	Corn	24,000
1	12	Cattle	18,500	2	12	Corn	12,500	2	47	Corn	13,000
1	13	Cattle	26,000	2	13	Corn	17,500	2	48	Corn	12,500
1	14	Cattle	19,000	2	14	Corn	22,500	2	49	Corn	7,500
1	15	Cattle	22,000	2	15	Corn	8,000	2	50	Corn	24,000
1	16	Cattle	17,000	2	16	Corn	5,000	2	51	Other	4,000
1	17	Cattle	21,000	2	17	Corn	18,500	2	52	Corn	7,500
1	18	Cattle	96,000	2	18	Corn	6,800	2	53	Other	5,000
1	19	Cattle	52,500	2	19	Corn	12,500	2	54	Other	4,000
1	20	Cattle	84,000	2	20	Corn	7,500	2	55	Other	3,500
1	21	Cattle	65,000	2	21	Corn	45,000	2	56	Other	6,000
1	22	Cattle	48,000	2	22	Corn	12,000	2	57	Corn	13,000
1	23	Cattle	27,500	2	23	Corn	4,500	2	58	Corn	24,000
1	24	Cattle	58,000	2	24	Corn	15,000	2	59	Corn	11,500
1	25	Cattle	23,000	2	25	Corn	22,000	2	60	Corn	7,000
1	26	Cattle	62,500	2	26	Corn	31,000	2	61	Other	22,500
1	27	Cattle	120,000	2	27	Corn	17,000	2	62	Other	26,000
1	28	Cattle	19,000	2	28	Corn	5,700	2	63	Corn	8,000
1	29	Cattle	96,000	2	29	Other	4,800	2	64	Corn	23,000
1	30	Cattle	28,000	2	30	Other	3,200	2	65	Corn	12,500
1	31	Cattle	48,000	2	31	Corn	10,000	2	66	Corn	20,000
1	32	Cattle	14,000	2	32	Corn	8,500				
1	33	Cattle	9,600	2	33	Corn	16,500				
1	34	Cattle	31,000	2	34	Other	5,000				
				2	35	Corn	6,200				

Calculate Allocation
N=100

Stratum 1: $34/100$
 $=3.4$
 $=3$

Stratum 2: $66/100$
 $=6.6$
 $=7$

$k=N/n$
 $k_1=34/3=11.3$
 $k_2=66/7=9.4$
 $rs_1=3$
 $rs_2=6$



Stratum	Obs	Type	Expenses								
1	1	Cattle	20,000	2	1	Corn	6,000	2	36	Other	3,000
1	2	Cattle	56,000	2	2	Corn	3,800	2	37	Corn	8,700
1	3	Cattle	48,000	2	3	Corn	6,700	2	38	Other	5,000
1	4	Cattle	22,500	2	4	Corn	11,000	2	39	Corn	34,000
1	5	Cattle	19,000	2	5	Corn	4,800	2	40	Corn	7,300
1	6	Cattle	17,500	2	6	Corn	7,600	2	41	Corn	23,000
1	7	Cattle	21,000	2	7	Corn	4,000	2	42	Other	7,000
1	8	Cattle	65,000	2	8	Corn	10,000	2	43	Other	3,800
1	9	Cattle	39,000	2	9	Corn	2,700	2	44	Other	17,000
1	10	Cattle	23,000	2	10	Corn	8,500	2	45	Other	8,000
1	11	Cattle	28,000	2	11	Corn	12,500	2	46	Corn	24,000
1	12	Cattle	18,500	2	12	Corn	12,500	2	47	Corn	13,000
1	13	Cattle	26,000	2	13	Corn	17,500	2	48	Corn	12,500
1	14	Cattle	19,000	2	14	Corn	22,500	2	49	Corn	7,500
1	15	Cattle	22,000	2	15	Corn	8,000	2	50	Corn	24,000
1	16	Cattle	17,000	2	16	Corn	5,000	2	51	Other	4,000
1	17	Cattle	21,000	2	17	Corn	18,500	2	52	Corn	7,500
1	18	Cattle	96,000	2	18	Corn	6,800	2	53	Other	5,000
1	19	Cattle	52,500	2	19	Corn	12,500	2	54	Other	4,000
1	20	Cattle	84,000	2	20	Corn	7,500	2	55	Other	3,500
1	21	Cattle	65,000	2	21	Corn	45,000	2	56	Other	6,000
1	22	Cattle	48,000	2	22	Corn	12,000	2	57	Corn	13,000
1	23	Cattle	27,500	2	23	Corn	4,500	2	58	Corn	24,000
1	24	Cattle	58,000	2	24	Corn	15,000	2	59	Corn	11,500
1	25	Cattle	23,000	2	25	Corn	22,000	2	60	Corn	7,000
1	26	Cattle	62,500	2	26	Corn	31,000	2	61	Other	22,500
1	27	Cattle	120,000	2	27	Corn	17,000	2	62	Other	26,000
1	28	Cattle	19,000	2	28	Corn	5,700	2	63	Corn	8,000
1	29	Cattle	96,000	2	29	Other	4,800	2	64	Corn	23,000
1	30	Cattle	28,000	2	30	Other	3,200	2	65	Corn	12,500
1	31	Cattle	48,000	2	31	Corn	10,000	2	66	Corn	20,000
1	32	Cattle	14,000	2	32	Corn	8,500				
1	33	Cattle	9,600	2	33	Corn	16,500				
1	34	Cattle	31,000	2	34	Other	5,000				
				2	35	Corn	6,200				

Calculate Allocation
N=100

Stratum 1: $34/100$
 $=3.4$
 $=3$

Stratum 2: $66/100$
 $=6.6$
 $=7$

$k=N/n$
 $k_1=34/3=11.3$
 $k_2=66/7=9.4$
 $rs_1=3$
 $rs_2=6$



Stratum	Obs	Type	Expenses								
1	1	Cattle	20,000	2	1	Corn	6,000	2	36	Other	3,000
1	2	Cattle	56,000	2	2	Corn	3,800	2	37	Corn	8,700
1	3	Cattle	48,000	2	3	Corn	6,700	2	38	Other	5,000
1	4	Cattle	22,500	2	4	Corn	11,000	2	39	Corn	34,000
1	5	Cattle	19,000	2	5	Corn	4,800	2	40	Corn	7,300
1	6	Cattle	17,500	2	6	Corn	7,600	2	41	Corn	23,000
1	7	Cattle	21,000	2	7	Corn	4,000	2	42	Other	7,000
1	8	Cattle	65,000	2	8	Corn	10,000	2	43	Other	3,800
1	9	Cattle	39,000	2	9	Corn	2,700	2	44	Other	17,000
1	10	Cattle	23,000	2	10	Corn	8,500	2	45	Other	8,000
1	11	Cattle	28,000	2	11	Corn	12,500	2	46	Corn	24,000
1	12	Cattle	18,500	2	12	Corn	12,500	2	47	Corn	13,000
1	13	Cattle	26,000	2	13	Corn	17,500	2	48	Corn	12,500
1	14	Cattle	19,000	2	14	Corn	22,500	2	49	Corn	7,500
1	15	Cattle	22,000	2	15	Corn	8,000	2	50	Corn	24,000
1	16	Cattle	17,000	2	16	Corn	5,000	2	51	Other	4,000
1	17	Cattle	21,000	2	17	Corn	18,500	2	52	Corn	7,500
1	18	Cattle	96,000	2	18	Corn	6,800	2	53	Other	5,000
1	19	Cattle	52,500	2	19	Corn	12,500	2	54	Other	4,000
1	20	Cattle	84,000	2	20	Corn	7,500	2	55	Other	3,500
1	21	Cattle	65,000	2	21	Corn	45,000	2	56	Other	6,000
1	22	Cattle	48,000	2	22	Corn	12,000	2	57	Corn	13,000
1	23	Cattle	27,500	2	23	Corn	4,500	2	58	Corn	24,000
1	24	Cattle	58,000	2	24	Corn	15,000	2	59	Corn	11,500
1	25	Cattle	23,000	2	25	Corn	22,000	2	60	Corn	7,000
1	26	Cattle	62,500	2	26	Corn	31,000	2	61	Other	22,500
1	27	Cattle	120,000	2	27	Corn	17,000	2	62	Other	26,000
1	28	Cattle	19,000	2	28	Corn	5,700	2	63	Corn	8,000
1	29	Cattle	96,000	2	29	Other	4,800	2	64	Corn	23,000
1	30	Cattle	28,000	2	30	Other	3,200	2	65	Corn	12,500
1	31	Cattle	48,000	2	31	Corn	10,000	2	66	Corn	20,000
1	32	Cattle	14,000	2	32	Corn	8,500				
1	33	Cattle	9,600	2	33	Corn	16,500				
1	34	Cattle	31,000	2	34	Other	5,000				
				2	35	Corn	6,200				

Mean stratum 1:
= $(48,000+19,000+23,000)/3$
= \$30,000

Mean stratum 2:
= \$10,057



Stratum	Obs	Type	Expenses								
1	1	Cattle	20,000	2	1	Corn	6,000	2	36	Other	3,000
1	2	Cattle	56,000	2	2	Corn	3,800	2	37	Corn	8,700
1	3	Cattle	48,000	2	3	Corn	6,700	2	38	Other	5,000
1	4	Cattle	22,500	2	4	Corn	11,000	2	39	Corn	34,000
1	5	Cattle	19,000	2	5	Corn	4,800	2	40	Corn	7,300
1	6	Cattle	17,500	2	6	Corn	7,600	2	41	Corn	23,000
1	7	Cattle	21,000	2	7	Corn	4,000	2	42	Other	7,000
1	8	Cattle	65,000	2	8	Corn	10,000	2	43	Other	3,800
1	9	Cattle	39,000	2	9	Corn	2,700	2	44	Other	17,000
1	10	Cattle	23,000	2	10	Corn	8,500	2	45	Other	8,000
1	11	Cattle	28,000	2	11	Corn	12,500	2	46	Corn	24,000
1	12	Cattle	18,500	2	12	Corn	12,500	2	47	Corn	13,000
1	13	Cattle	26,000	2	13	Corn	17,500	2	48	Corn	12,500
1	14	Cattle	19,000	2	14	Corn	22,500	2	49	Corn	7,500
1	15	Cattle	22,000	2	15	Corn	8,000	2	50	Corn	24,000
1	16	Cattle	17,000	2	16	Corn	5,000	2	51	Other	4,000
1	17	Cattle	21,000	2	17	Corn	18,500	2	52	Corn	7,500
1	18	Cattle	96,000	2	18	Corn	6,800	2	53	Other	5,000
1	19	Cattle	52,500	2	19	Corn	12,500	2	54	Other	4,000
1	20	Cattle	84,000	2	20	Corn	7,500	2	55	Other	3,500
1	21	Cattle	65,000	2	21	Corn	45,000	2	56	Other	6,000
1	22	Cattle	48,000	2	22	Corn	12,000	2	57	Corn	13,000
1	23	Cattle	27,500	2	23	Corn	4,500	2	58	Corn	24,000
1	24	Cattle	58,000	2	24	Corn	15,000	2	59	Corn	11,500
1	25	Cattle	23,000	2	25	Corn	22,000	2	60	Corn	7,000
1	26	Cattle	62,500	2	26	Corn	31,000	2	61	Other	22,500
1	27	Cattle	120,000	2	27	Corn	17,000	2	62	Other	26,000
1	28	Cattle	19,000	2	28	Corn	5,700	2	63	Corn	8,000
1	29	Cattle	96,000	2	29	Other	4,800	2	64	Corn	23,000
1	30	Cattle	28,000	2	30	Other	3,200	2	65	Corn	12,500
1	31	Cattle	48,000	2	31	Corn	10,000	2	66	Corn	20,000
1	32	Cattle	14,000	2	32	Corn	8,500				
1	33	Cattle	9,600	2	33	Corn	16,500				
1	34	Cattle	31,000	2	34	Other	5,000				
				2	35	Corn	6,200				

Mean stratum 1:
= $(48,000+19,000+23,000)/3$
= \$30,000

Mean stratum 2:
= \$10,057

Overall Mean:
 $(30,000 \times 34 + 10,057 \times 66) / 100$
= \$16,837

Probability Proportional to Size (PPS) Sampling

- PPS Sampling is a technique that uses auxiliary data to select sample. If population units vary in size and these sizes are known, such information can be used during sampling.
- The probabilities of selection are unequal. Units with larger sizes have more chance to be selected.

When PPS?

- Measures of size are available for sampling units and these measures are accurate.
- A strong correlation exists between the size measures of units and characteristics of interest.

PPS shifts from “count” to “size”

Possible samples of size n=1 (Farm Selected)	Acreage	Expenditure (\$)
1	50	26,000
2	1,000	470,000
3	125	63,800
4	300	145,000
5	500	230,000
6	25	12,500

PPS shifts from “count” to “size”

Possible samples of size n=1 (Farm Selected)	Acreage	Expenditure (\$)
1	50	26,000
2	1,000	470,000
3	125	63,800
4	300	145,000
5	500	230,000
6	25	12,500

PPS Sampling – Simple Random Selection Method

PPS Sampling - Systematic Selection Method



Example: Possible PPS Samples of Size $n=1$

Possible samples of size $n=1$ (Farm Selected)	Acreage	Probability of selection	Expenditure (\$)
1	50	$50/2,000$	26,000
2	1,000	$1,000/2,000$	470,000
3	125	$125/2,000$	63,800
4	300	$300/2,000$	145,000
5	500	$500/2,000$	230,000
6	25	$25/2,000$	12,500

Acreage is a size measure of unit - Acreage is correlated with expenditure

PPS Sampling - Random method (n=1)

- 1) For each unit, cumulate size measures and determine range of numbers representing the size
- 2) Select a random number between 1 and the total cumulative size
- 3) Select the unit whose range contains the random number

Example:

-n=1

-Random Nb = 70

-Selection:

*70 Farm 2
(51 ≤ 70 ≤ 1,050)*

→

Farm	Acreage	Cumulative Size	Range	Farm Expenditure (\$)
1	50	50	1-50	26,000
2	1,000	1,050	51-1,050	470,000
3	125	1,175	1,051-1,175	63,000
4	300	1,475	1,176-1,475	145,000
5	500	1,975	1,476-1,975	230,000
6	25	2,000	1,976-2,000	12,500

PPS Sampling – Systematic Method

- 1) For each unit, calculate cumulative size measures and determine range of numbers representing the size
- 2) Determine Sampling Interval:
$$k = \text{Cumulative Size} / \text{Sample Size}$$
- 3) Determine a Random start, r , between 1 and k
- 4) Select the units whose range contains the random numbers $r, r+k, r+2k, \dots, r+(n-1)k$

Example:

- $n=2$

- $k=2,000/2=1,000$

- $r=725$ (random start)

-Selection:

725 \rightarrow Farm 2

1,725 \rightarrow Farm 5

Farm	Acreage	Cumulative Size	Range	Farm Expenditure (\$)
1	50	50	1-50	26,000
2	1,000	1,050	51-1,050	470,000
3	125	1,175	1,051-1,175	63,000
4	300	1,475	1,176-1,475	145,000
5	500	1,975	1,476-1,975	230,000
6	25	2,000	1,976-2,000	12,500

PPS Sampling

Advantage

- It can improve the statistical efficiency of the sampling strategy by using auxiliary information

Disadvantage

- Requires a frame that contains good quality, up-to-date auxiliary information for all units on the frame that can be used as size measures
- It can result in a sampling strategy that is less statistically efficient than SRS for survey variables that are not positively correlated with the size variables

Exercise 4: PPS - Systematic Method

- Draw a sample of $n=10$ farms.
- Proportional to number of acres
 - Calculate your Sampling Interval K (keep one decimal!)
 - Random Start (r):
 - Start at 0623



Obs	Acres	Expenses	Cumulative Acres
1	100	6,000	100
2	50	3,800	150
3	120	6,700	270
4	200	11,000	470
5	80	4,800	550
6	130	7,600	680
7	60	4,000	740
8	220	10,000	960
9	40	2,700	1,000
10	140	8,500	1,140
11	240	12,500	1,380
12	300	12,500	1,680
13	400	17,500	2,080
14	600	22,500	2,680
15	150	8,000	2,830
16	100	5,000	2,930
17	500	18,500	3,430
18	100	6,800	3,530
19	220	12,500	3,750
20	200	20,000	3,950
21	800	56,000	4,750
22	600	48,000	5,350
23	220	22,500	5,570
24	180	19,000	5,750
25	150	17,500	5,900
26	250	21,000	6,150
27	900	65,000	7,050
28	400	39,000	7,450
29	200	23,000	7,650
30	250	28,000	7,900
31	150	7,500	8,050
32	100	45,000	8,150
33	300	12,000	8,450
34	175	18,500	8,625

35	240	26,000	8,865
36	75	4,500	8,940
37	280	15,000	9,220
38	420	22,000	9,640
39	620	31,000	10,260
40	320	17,000	10,580
41	120	5,700	10,700
42	360	19,000	11,060
43	250	22,000	11,310
44	160	17,000	11,470
45	260	21,000	11,730
46	1,200	96,000	12,930
47	750	52,500	13,680
48	80	4,800	13,760
49	50	3,200	13,810
50	200	10,000	14,010
51	175	8,500	14,185
52	400	16,500	14,585
53	100	5,000	14,685
54	150	6,200	14,835
55	50	3,000	14,885
56	200	8,700	15,085
57	80	5,000	15,165
58	400	34,000	15,565
59	90	7,300	15,655
60	500	23,000	16,155
61	140	7,000	16,295
62	75	3,800	16,370
63	300	17,000	16,670
64	160	8,000	16,830
65	460	24,000	17,290
66	300	13,000	17,590
67	200	12,500	17,790
68	125	7,500	17,915

69	500	24,000	18,415
70	1,000	84,000	19,415
71	800	65,000	20,215
72	500	48,000	20,715
73	300	27,500	21,015
74	600	58,000	21,615
75	200	23,000	21,815
76	800	62,500	22,615
77	1,500	120,000	24,115
78	200	19,000	24,315
79	1,200	96,000	25,515
80	400	28,000	25,915
81	600	48,000	26,515
82	200	14,000	26,715
83	120	9,600	26,835
84	60	4,000	26,895
85	150	7,500	27,045
86	100	5,000	27,145
87	80	4,000	27,225
88	50	3,500	27,275
89	400	31,000	27,675
90	120	6,000	27,795
91	260	13,000	28,055
92	600	24,000	28,655
93	325	11,500	28,980
94	150	7,000	29,130
95	400	22,500	29,530
96	500	26,000	30,030
97	200	8,000	30,230
98	600	23,000	30,830
99	300	12,500	31,130
100	525	20,000	31,655

$K=N/n$
 $K=31,655/10$
 $K=3,165.5$

 $RS=623$





Obs	Acres	Expenses	Cumulative Acres
1	100	6,000	100
2	50	3,800	150
3	120	6,700	270
4	200	11,000	470
5	80	4,800	550
6	130	7,600	680
7	60	4,000	740
8	220	10,000	960
9	40	2,700	1,000
10	140	8,500	1,140
11	240	12,500	1,380
12	300	12,500	1,680
13	400	17,500	2,080
14	600	22,500	2,680
15	150	8,000	2,830
16	100	5,000	2,930
17	500	18,500	3,430
18	100	6,800	3,530
19	220	12,500	3,750
20	200	20,000	3,950
21	800	56,000	4,750
22	600	48,000	5,350
23	220	22,500	5,570
24	180	19,000	5,750
25	150	17,500	5,900
26	250	21,000	6,150
27	900	65,000	7,050
28	400	39,000	7,450
29	200	23,000	7,650
30	250	28,000	7,900
31	150	7,500	8,050
32	100	45,000	8,150
33	300	12,000	8,450
34	175	18,500	8,625

35	240	26,000	8,865
36	75	4,500	8,940
37	280	15,000	9,220
38	420	22,000	9,640
39	620	31,000	10,260
40	320	17,000	10,580
41	120	5,700	10,700
42	360	19,000	11,060
43	250	22,000	11,310
44	160	17,000	11,470
45	260	21,000	11,730
46	1,200	96,000	12,930
47	750	52,500	13,680
48	80	4,800	13,760
49	50	3,200	13,810
50	200	10,000	14,010
51	175	8,500	14,185
52	400	16,500	14,585
53	100	5,000	14,685
54	150	6,200	14,835
55	50	3,000	14,885
56	200	8,700	15,085
57	80	5,000	15,165
58	400	34,000	15,565
59	90	7,300	15,655
60	500	23,000	16,155
61	140	7,000	16,295
62	75	3,800	16,370
63	300	17,000	16,670
64	160	8,000	16,830
65	460	24,000	17,290
66	300	13,000	17,590
67	200	12,500	17,790
68	125	7,500	17,915

69	500	24,000	18,415
70	1,000	84,000	19,415
71	800	65,000	20,215
72	500	48,000	20,715
73	300	27,500	21,015
74	600	58,000	21,615
75	200	23,000	21,815
76	800	62,500	22,615
77	1,500	120,000	24,115
78	200	19,000	24,315
79	1,200	96,000	25,515
80	400	28,000	25,915
81	600	48,000	26,515
82	200	14,000	26,715
83	120	9,600	26,835
84	60	4,000	26,895
85	150	7,500	27,045
86	100	5,000	27,145
87	80	4,000	27,225
88	50	3,500	27,275
89	400	31,000	27,675
90	120	6,000	27,795
91	260	13,000	28,055
92	600	24,000	28,655
93	325	11,500	28,980
94	150	7,000	29,130
95	400	22,500	29,530
96	500	26,000	30,030
97	200	8,000	30,230
98	600	23,000	30,830
99	300	12,500	31,130
100	525	20,000	31,655

$K=N/n$
 $K=31,655/10$
 $K=3,165.5$

 $RS=623$
 $2^{nd} = 623+3165.5$
 $=3,788.5$



Obs	Acres	Expenses	Cumulative Acres
1	100	6,000	100
2	50	3,800	150
3	120	6,700	270
4	200	11,000	470
5	80	4,800	550
6	130	7,600	680
7	60	4,000	740
8	220	10,000	960
9	40	2,700	1,000
10	140	8,500	1,140
11	240	12,500	1,380
12	300	12,500	1,680
13	400	17,500	2,080
14	600	22,500	2,680
15	150	8,000	2,830
16	100	5,000	2,930
17	500	18,500	3,430
18	100	6,800	3,530
19	220	12,500	3,750
20	200	20,000	3,950
21	800	56,000	4,750
22	600	48,000	5,350
23	220	22,500	5,570
24	180	19,000	5,750
25	150	17,500	5,900
26	250	21,000	6,150
27	900	65,000	7,050
28	400	39,000	7,450
29	200	23,000	7,650
30	250	28,000	7,900
31	150	7,500	8,050
32	100	45,000	8,150
33	300	12,000	8,450
34	175	18,500	8,625

35	240	26,000	8,865
36	75	4,500	8,940
37	280	15,000	9,220
38	420	22,000	9,640
39	620	31,000	10,260
40	320	17,000	10,580
41	120	5,700	10,700
42	360	19,000	11,060
43	250	22,000	11,310
44	160	17,000	11,470
45	260	21,000	11,730
46	1,200	96,000	12,930
47	750	52,500	13,680
48	80	4,800	13,760
49	50	3,200	13,810
50	200	10,000	14,010
51	175	8,500	14,185
52	400	16,500	14,585
53	100	5,000	14,685
54	150	6,200	14,835
55	50	3,000	14,885
56	200	8,700	15,085
57	80	5,000	15,165
58	400	34,000	15,565
59	90	7,300	15,655
60	500	23,000	16,155
61	140	7,000	16,295
62	75	3,800	16,370
63	300	17,000	16,670
64	160	8,000	16,830
65	460	24,000	17,290
66	300	13,000	17,590
67	200	12,500	17,790
68	125	7,500	17,915

69	500	24,000	18,415
70	1,000	84,000	19,415
71	800	65,000	20,215
72	500	48,000	20,715
73	300	27,500	21,015
74	600	58,000	21,615
75	200	23,000	21,815
76	800	62,500	22,615
77	1,500	120,000	24,115
78	200	19,000	24,315
79	1,200	96,000	25,515
80	400	28,000	25,915
81	600	48,000	26,515
82	200	14,000	26,715
83	120	9,600	26,835
84	60	4,000	26,895
85	150	7,500	27,045
86	100	5,000	27,145
87	80	4,000	27,225
88	50	3,500	27,275
89	400	31,000	27,675
90	120	6,000	27,795
91	260	13,000	28,055
92	600	24,000	28,655
93	325	11,500	28,980
94	150	7,000	29,130
95	400	22,500	29,530
96	500	26,000	30,030
97	200	8,000	30,230
98	600	23,000	30,830
99	300	12,500	31,130
100	525	20,000	31,655

$$K = N/n$$

$$K = 31,655/10$$

$$K = 3,165.5$$

$$RS = 623$$

$$2^{nd} = 623 + 3165.5$$

$$= 3,788.5$$

- 3 - 6,954
- 4 - 10,119.5
- 5 - 13,285
- 6 - 16,450.5
- 7 - 19,616
- 8 - 22,781.5
- 9 - 25,947
- 10 - 29,112.5



Obs	Acres	Expenses	Cumulative Acres
1	100	6,000	100
2	50	3,800	150
3	120	6,700	270
4	200	11,000	470
5	80	4,800	550
6	130	7,600	680
7	60	4,000	740
8	220	10,000	960
9	40	2,700	1,000
10	140	8,500	1,140
11	240	12,500	1,380
12	300	12,500	1,680
13	400	17,500	2,080
14	600	22,500	2,680
15	150	8,000	2,830
16	100	5,000	2,930
17	500	18,500	3,430
18	100	6,800	3,530
19	220	12,500	3,750
20	200	20,000	3,950
21	800	56,000	4,750
22	600	48,000	5,350
23	220	22,500	5,570
24	180	19,000	5,750
25	150	17,500	5,900
26	250	21,000	6,150
27	900	65,000	7,050
28	400	39,000	7,450
29	200	23,000	7,650
30	250	28,000	7,900
31	150	7,500	8,050
32	100	45,000	8,150
33	300	12,000	8,450
34	175	18,500	8,625

35	240	26,000	8,865
36	75	4,500	8,940
37	280	15,000	9,220
38	420	22,000	9,640
39	620	31,000	10,260
40	320	17,000	10,580
41	120	5,700	10,700
42	360	19,000	11,060
43	250	22,000	11,310
44	160	17,000	11,470
45	260	21,000	11,730
46	1,200	96,000	12,930
47	750	52,500	13,680
48	80	4,800	13,760
49	50	3,200	13,810
50	200	10,000	14,010
51	175	8,500	14,185
52	400	16,500	14,585
53	100	5,000	14,685
54	150	6,200	14,835
55	50	3,000	14,885
56	200	8,700	15,085
57	80	5,000	15,165
58	400	34,000	15,565
59	90	7,300	15,655
60	500	23,000	16,155
61	140	7,000	16,295
62	75	3,800	16,370
63	300	17,000	16,670
64	160	8,000	16,830
65	460	24,000	17,290
66	300	13,000	17,590
67	200	12,500	17,790
68	125	7,500	17,915

69	500	24,000	18,415
70	1,000	84,000	19,415
71	800	65,000	20,215
72	500	48,000	20,715
73	300	27,500	21,015
74	600	58,000	21,615
75	200	23,000	21,815
76	800	62,500	22,615
77	1,500	120,000	24,115
78	200	19,000	24,315
79	1,200	96,000	25,515
80	400	28,000	25,915
81	600	48,000	26,515
82	200	14,000	26,715
83	120	9,600	26,835
84	60	4,000	26,895
85	150	7,500	27,045
86	100	5,000	27,145
87	80	4,000	27,225
88	50	3,500	27,275
89	400	31,000	27,675
90	120	6,000	27,795
91	260	13,000	28,055
92	600	24,000	28,655
93	325	11,500	28,980
94	150	7,000	29,130
95	400	22,500	29,530
96	500	26,000	30,030
97	200	8,000	30,230
98	600	23,000	30,830
99	300	12,500	31,130
100	525	20,000	31,655

$K=N/n$
 $K=31,655/10$
 $K=3,165.5$

 $RS=623$
 $2^{nd} = 623+3165.5$
 $=3,788.5$

3 - 6,954
4 - 10,119.5
5 - 13,285
6 - 16,450.5
7 - 19,616
8 - 22,781.5
9 - 25,947
10 - 29,112.5

Why is PPS Sampling is one of my favs

- Analysts – We want weights as consistent as possible.
- Field Operations – We want this as simple as possible to manage.

Why is PPS Sampling is one of my favs

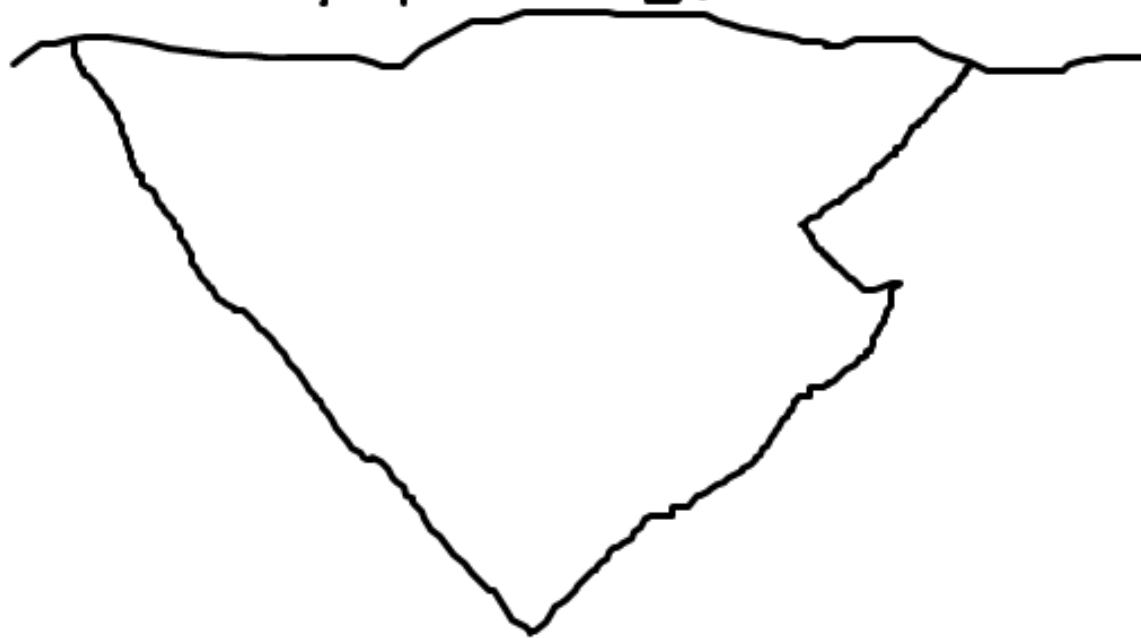
- Analysts – We want weights as consistent as possible.
- Field Operations – We want this as simple as possible to manage.
- Respondents – We are Tax Payers, we want an equal chance to have our voice heard!

Why is PPS Sampling is one of my favs

- Analysts – We want weights as consistent as possible.
- Field Operations – We want this as simple as possible to manage.
- Respondents – We are Tax Payers, we want an equal chance to have our voice heard!
- Methodology – We just want everyone to be happy.

- Labour Force Survey – 56,000 households across Canada
- Canada split into clusters
- Each cluster has a fixed number of interviews (6 per cluster)
- ...but this causes a problem...

Ottawa



Ottawa









Probability of selection in Cluster A:

$N=500, n=6$

$\text{Prob}=6/500=0.012$

Probability of selection in Cluster B:

$N=1,000, n=6$

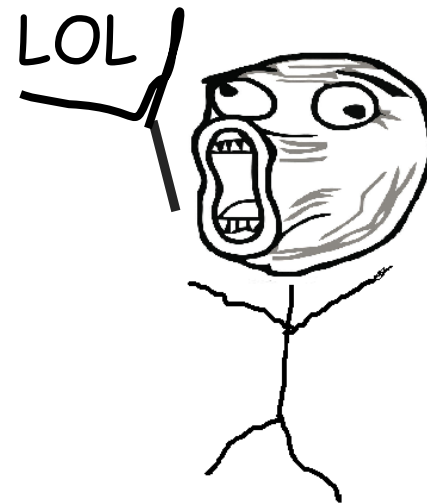
$\text{Prob}=6/1,000=0.006$



I'm a tax payer, why
is it only half as
likely that I get to
share my data?!?!?



Cluster B



Cluster A

Probability of selection in Cluster A:

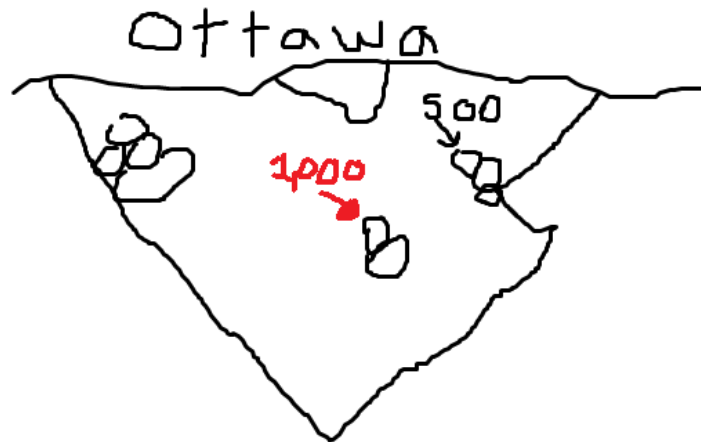
$N=500, n=6$

$\text{Prob}=6/500=0.012$

Probability of selection in Cluster B:

$N=1,000, n=6$

$\text{Prob}=6/1,000=0.006$



Probability of selection in Cluster A:
 $N=500$, $n=6$
 $\text{Prob}=6/500=0.012$

Probability of selection in Cluster B:
 $N=1,000$, $n=6$
 $\text{Prob}=6/1,000=0.006$

Solution:

Make it so the probability of a cluster is based on its size (ahhh!)

If we chose a cluster with:

$$\text{Prob}=X*(\text{clustersize}/1,500)$$

$$\text{ProbA}=X*(500/1,500)=0.3333$$

$$\text{ProbB}=X*(1,000/1,500)=0.66667$$

So, the probability of the 2nd cluster being chosen has doubled, which balances the fact that WITHIN the cluster the households in cluster A have twice the chance of being selected.

Because everyone now has the same chance of being included, everyone has the same design weight (future presentation)



Questions?

You can contact the PRASC team at:
statcan.prasc-prasc.statcan@canada.ca