

Sampling People, Records, & Networks

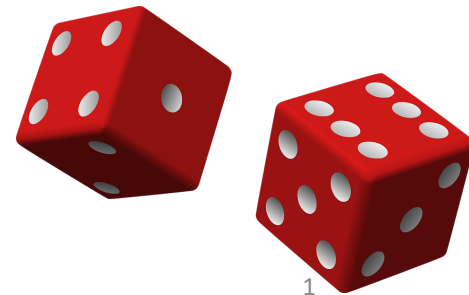
Jim Lepkowski, PhD

Professor & Research Professor *Emeritus*

Institute for Social Research, University of Michigan

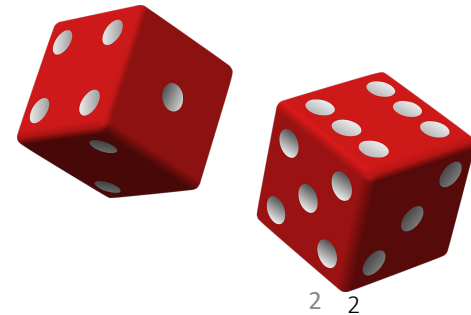
Research Professor,

Joint Program in Survey Methodology, University of Maryland



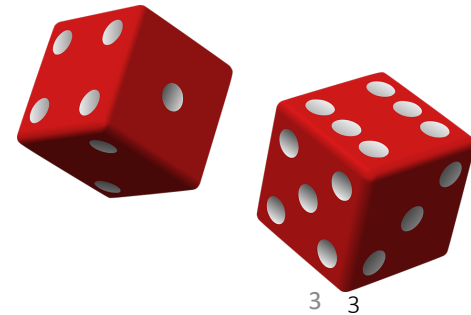
Unit 5

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications

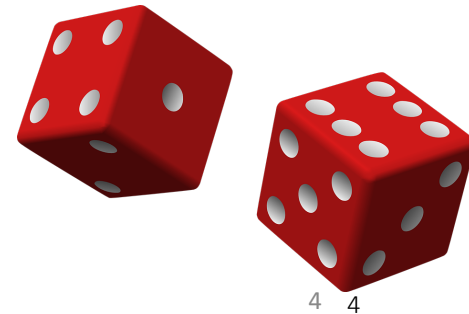


Unit 5

- 1 Systematic selection
 - 2 Intervals with fractions
 - 3 List order
 - 4 Uncertainty estimation
- Unit 1: Sampling as a research tool
 - Unit 2: Mere randomization
 - Unit 3: Saving money
 - Unit 4: Being more efficient
 - **Unit 5: Simplifying sampling**
 - 1 Systematic selection
 - 2 Intervals with fractions
 - 3 List order
 - 4 Uncertainty estimation
 - Unit 6: Some extensions & applications

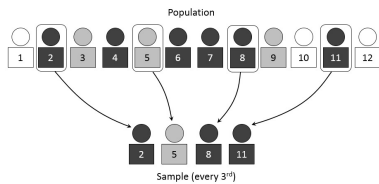


- Interval $k=10$
 - Interval tuned to size
 - Conceptual framework
- Unit 1: Sampling as a research tool
 - Unit 2: Mere randomization
 - Unit 3: Saving money
 - Unit 4: Being more efficient
 - **Unit 5: Simplifying sampling**
 - 1 Systematic selection
 - 2 Intervals with fractions
 - 3 List order
 - 4 Uncertainty estimation
 - Unit 6: Some extensions & applications



- Interval $k=10$
- Interval tuned to size
- Conceptual framework

- A simple method of selecting a sample from a list
- Take every so many elements ...

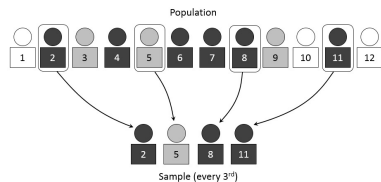


Survey Data Collection & Analytic Specialization

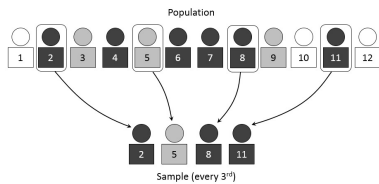
Sampling People, Records, & Networks

- Interval $k=10$
- Interval tuned to size
- Conceptual framework

No.	Date	Time	Reference	Category	Subcategory	Business	Additional information	Amount
1	20160617	13274200	32016172018491	Business services	Advertising services	NA	5556128901	\$833.24
2	20160617	13274203	32016172018492	Merchandise & supplies	Book stores	NA	5552697691	\$26.52
3	20160617	13274210	32016172018493	Other	Utilities	NA	5558518280	\$91.95
4	20160617	13274211	32016172018494	Business services	Mailing & shipping	NA	5554888725	\$17.70
5	20160617	13274211	32016172018495	Business services	Contracting services	NA	5558398606	\$150.66
6	20160617	13274217	32016172018496	Merchandise & supplies	General retail	NA	5552808987	\$80.14
7	20160617	13274218	32016172018497	Merchandise & supplies	Music & video	NA	5558102117	\$59.27
8	20160617	13274220	32016172018498	Merchandise & supplies	Music & video	NA	5557495061	\$85.04
9	20160617	13274230	32016172018499	Business services	Health care services	NA	5559321864	\$624.80
10	20160617	13274236	32016172018500	Restaurant	Restaurant	NA	5551869731	\$56.32
11	20160617	13274239	32016172018501	Merchandise & supplies	Groceries	NA	5554914804	\$82.33
12	20160617	13274240	32016172018502	Business services	Health care services	NA	5552682932	\$842.16
13	20160617	13274241	32016172018503	Business services	Health care services	NA	5551606583	\$971.30
14	20160617	13274242	32016172018504	Fees & adjustments	Fees & adjustments	NA	5558979830	\$73.57
15	20160617	13274249	32016172018505	Business services	Contracting services	NA	5558855254	\$128.74
16	20160617	13274251	32016172018506	Entertainment	Sports events	NA	5557903363	\$122.68
17	20160617	13274258	32016172018507	Travel	Lodging	NA	5554231047	\$241.15
18	20160617	13274266	32016172018508	Other	Miscellaneous	NA	5551339259	\$42.92
19	20160617	13274274	32016172018509	Business services	Health care services	NA	5555502637	\$1,071.25
20	20160617	13274282	32016172018510	Merchandise & supplies	Groceries	NA	5551386371	\$30.51
21	20160617	13274287	32016172018511	Entertainment	Other entertainment	NA	5551508827	\$28.13
22	20160617	13274296	32016172018512	Merchandise & supplies	Department stores	NA	5554661994	\$194.13
23	20160617	13274302	32016172018513	Other	Utilities	NA	5554438826	\$111.73
24	20160617	13274306	32016172018514	Business services	Health care services	NA	5559722356	\$694.48
25	20160617	13274310	32016172018515	Merchandise & supplies	Appliance stores	NA	5553447554	\$294.00



- Interval $k=10$
 - Interval tuned to size
 - Conceptual framework
- Choose the start – first, any, any one up to the interval
 - Say taking every 10th, starting with first ...

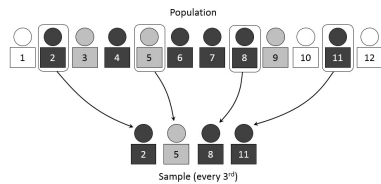


Survey Data Collection & Analytic Specialization

Sampling People, Records, & Networks

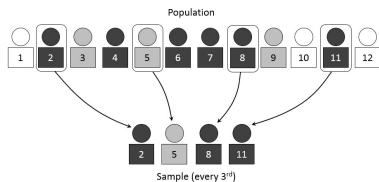
- Interval $k=10$
- Interval tuned to size
- Conceptual framework

No.	Date	Time	Reference	Category	Subcategory	Business	Additional information	Amount
1	20160617	13274200	32016172018491	Business services	Advertising services	NA	5556128901	\$833.24
2	20160617	13274203	32016172018492	Merchandise & supplies	Book stores	NA	5552697691	\$26.52
3	20160617	13274210	32016172018493	Other	Utilities	NA	5558518280	\$91.95
4	20160617	13274211	32016172018494	Business services	Mailing & shipping	NA	5554888725	\$17.70
5	20160617	13274211	32016172018495	Business services	Contracting services	NA	5558398606	\$150.66
6	20160617	13274217	32016172018496	Merchandise & supplies	General retail	NA	5552808987	\$80.14
7	20160617	13274218	32016172018497	Merchandise & supplies	Music & video	NA	5558102117	\$59.27
8	20160617	13274220	32016172018498	Merchandise & supplies	Music & video	NA	5557495061	\$85.04
9	20160617	13274230	32016172018499	Business services	Health care services	NA	5559321864	\$624.80
10	20160617	13274236	32016172018500	Restaurant	Restaurant	NA	5551869731	\$56.32
11	20160617	13274239	32016172018501	Merchandise & supplies	Groceries	NA	5554914804	\$82.33
12	20160617	13274240	32016172018502	Business services	Health care services	NA	5552682932	\$842.16
13	20160617	13274241	32016172018503	Business services	Health care services	NA	5551606583	\$971.30
14	20160617	13274242	32016172018504	Fees & adjustments	Fees & adjustments	NA	5558979830	\$73.57
15	20160617	13274249	32016172018505	Business services	Contracting services	NA	5558855254	\$128.74
16	20160617	13274251	32016172018506	Entertainment	Sports events	NA	5557903363	\$122.68
17	20160617	13274258	32016172018507	Travel	Lodging	NA	5554231047	\$241.15
18	20160617	13274266	32016172018508	Other	Miscellaneous	NA	5551339259	\$42.92
19	20160617	13274274	32016172018509	Business services	Health care services	NA	5555502637	\$1,071.25
20	20160617	13274282	32016172018510	Merchandise & supplies	Groceries	NA	5551386371	\$30.51
21	20160617	13274287	32016172018511	Entertainment	Other entertainment	NA	5551508827	\$28.13
22	20160617	13274296	32016172018512	Merchandise & supplies	Department stores	NA	5554661994	\$194.13
23	20160617	13274302	32016172018513	Other	Utilities	NA	5554438826	\$111.73
24	20160617	13274306	32016172018514	Business services	Health care services	NA	5559722356	\$694.48
25	20160617	13274310	32016172018515	Merchandise & supplies	Appliance stores	NA	5553447554	\$294.00



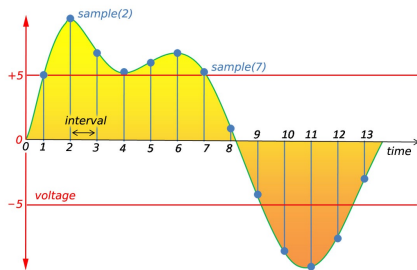
- Interval $k=10$
- Interval tuned to size
- Conceptual framework

- When do we stop?
- When we have n selections?
- This list has N elements
- If $n = 50$, stop with element 50! ... and elements 502 to 1,000 have zero chance of selection
- As do elements 2, 3, 4, ... , 10, and 12, 13, 14, ... , 20, and so on
- Two remedies needed:
 - Spread the selection out over the whole list
 - Vary the selection start



- Interval $k=10$
- Interval tuned to size
- Conceptual framework

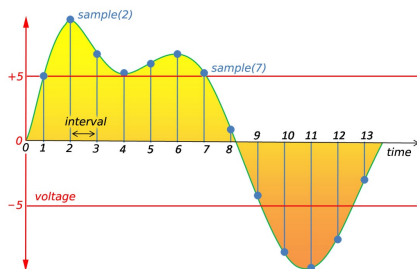
- For example, for $n = 50$, don't take every 10th, but every 20th
- And don't start with the first, but start from any element from 1 to 20 ... at random



Survey Data Collection & Analytic Specialization

Sampling People, Records, & Networks

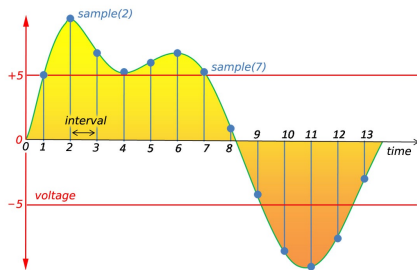
- Interval $k=10$
- Interval tuned to size
- Conceptual framework



No.	Date	Time	Reference	Category	Subcategory	Business	Additional information	Amount
1	20160617	13274200	32016172018491	Business services	Advertising services	NA	5556128901	\$833.24
2	20160617	13274203	32016172018492	Merchandise & supplies	Book stores	NA	5552697691	\$26.52
3	20160617	13274210	32016172018493	Other	Utilities	NA	5558518280	\$91.95
4	20160617	13274211	32016172018494	Business services	Mailing & shipping	NA	5554888725	\$17.70
5	20160617	13274211	32016172018495	Business services	Contracting services	NA	5558398606	\$150.66
6	20160617	13274217	32016172018496	Merchandise & supplies	General retail	NA	5552808987	\$80.14
7	20160617	13274218	32016172018497	Merchandise & supplies	Music & video	NA	5558102117	\$59.27
8	20160617	13274220	32016172018498	Merchandise & supplies	Music & video	NA	5557495061	\$85.04
9	20160617	13274230	32016172018499	Business services	Health care services	NA	5559321864	\$624.80
10	20160617	13274236	32016172018500	Restaurant	Restaurant	NA	5551869731	\$56.32
11	20160617	13274239	32016172018501	Merchandise & supplies	Groceries	NA	5554914804	\$82.33
12	20160617	13274240	32016172018502	Business services	Health care services	NA	5552682932	\$842.16
13	20160617	13274241	32016172018503	Business services	Health care services	NA	5551606583	\$971.30
14	20160617	13274242	32016172018504	Fees & adjustments	Fees & adjustments	NA	5558979830	\$73.57
15	20160617	13274249	32016172018505	Business services	Contracting services	NA	5558855254	\$128.74
16	20160617	13274251	32016172018506	Entertainment	Sports events	NA	5557903363	\$122.68
17	20160617	13274258	32016172018507	Travel	Lodging	NA	5554231047	\$241.15
18	20160617	13274266	32016172018508	Other	Miscellaneous	NA	5551339259	\$42.92
19	20160617	13274274	32016172018509	Business services	Health care services	NA	5555502637	\$1,071.25
20	20160617	13274282	32016172018510	Merchandise & supplies	Groceries	NA	5551386371	\$30.51
21	20160617	13274287	32016172018511	Entertainment	Other entertainment	NA	5551508827	\$28.13
22	20160617	13274296	32016172018512	Merchandise & supplies	Department stores	NA	5554661994	\$194.13
23	20160617	13274302	32016172018513	Other	Utilities	NA	5554438826	\$111.73
24	20160617	13274306	32016172018514	Business services	Health care services	NA	5559722356	\$694.48
25	20160617	13274310	32016172018515	Merchandise & supplies	Appliance stores	NA	5553447554	\$294.00

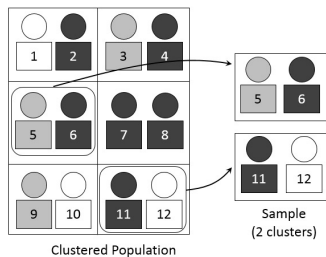
- Interval $k=10$
- Interval tuned to size
- Conceptual framework

- That is, adapt the selection process to the size of sample and size of list
- Calculate an interval of $k = N / n$ ($k = 1,000 / 50 = 20$)
- And choose to start anywhere from 1 to $k = 20 \dots$ at random



- Interval $k=10$
- Interval tuned to size
- Conceptual framework

- Conceptually, this is taking the population, dividing it into k samples, and choosing one of them:



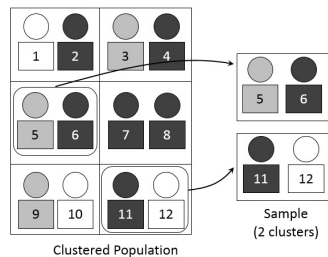
Survey Data Collection & Analytic Specialization

Sampling People, Records, & Networks

- Interval $k=10$
- Interval tuned to size
- Conceptual framework

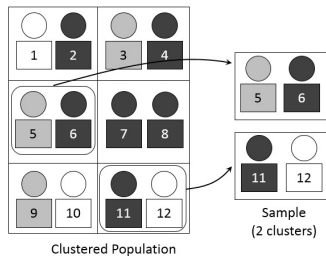
Sample

Selection	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
3	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
4	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
5	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
.
.
.
50	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000



- Interval $k=10$
- Interval tuned to size
- Conceptual framework

- This is equivalent to cluster sampling – each possible systematic sample is a cluster of n elements



Unit 5

- 1 Systematic selection
 - 2 Intervals with fractions
 - 3 List order
 - 4 Uncertainty estimation
- Unit 1: Sampling as a research tool
 - Unit 2: Mere randomization
 - Unit 3: Saving money
 - Unit 4: Being more efficient
 - Unit 5: Simplifying sampling
 - 1 Systematic selection
 - 2 Intervals with fractions
 - 3 List order
 - 4 Uncertainty estimation
 - Unit 6: Some extensions & applications

