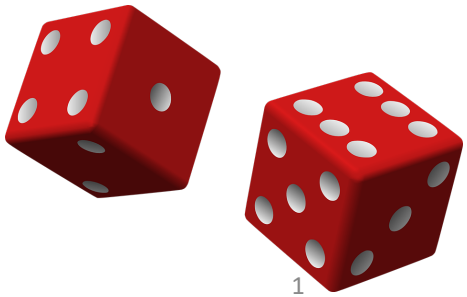
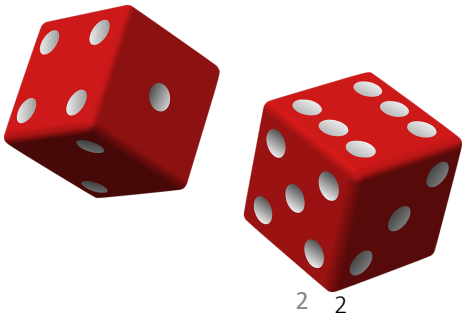
Sampling People, Records, & Networks

Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



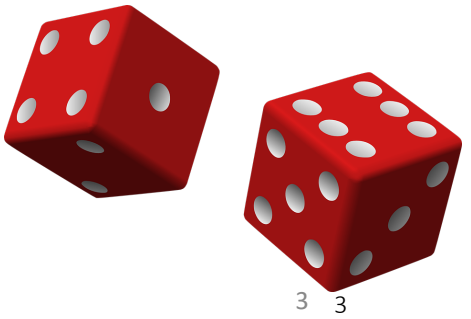
Unit 5

- 1 Systematic selection
- **2 Intervals with fractions**
- 3 List order
- 4 Uncertainty estimation



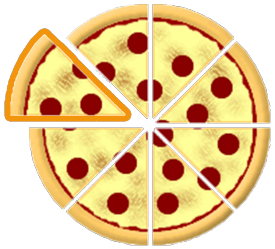
- The problem
- Rounding
- Circular list
- Fractional interval

- 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



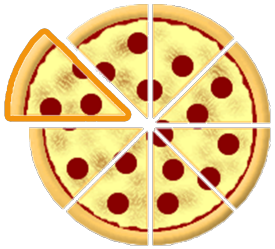
- The problem
- Rounding
- Circular list
- Fractional interval

- To repeat the process, first determine the sampling interval $k = N / n$
- Select a random number (RN) from 1 to k
- Add k repeatedly
- Suppose, for example, there were N = 12,000 dwellings in a city and a sample of n = 500 is required
  - k = 12,000/500 = 24
  - Take a RN from 01 to 24, say 03
  - Take the third dwelling, and every 24th thereafter: 3, 27, 51, *etc.*

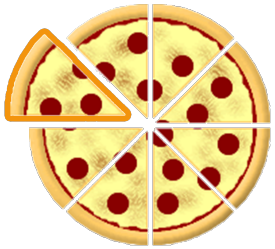


- The problem
- Rounding
- Circular list
- Fractional interval

- But what do we do in the more common situation where *k* is not an integer?
- Examples:
  - *N* = 9, *n* = 2, and *k* = 4.5
  - *N* = 952, *n* = 200, and *k* = 4.76
  - *N* = 170,345, *n* = 1,250, and *k* = 136.272



- But what do we do in the more common situation where *k* is not an integer?
- Examples:
  - *N* = 9, *n* = 2, and *k* = 4.5
  - *N* = 952, *n* = 200, and *k* = 4.76
  - *N* = 170,345, *n* = 1,250, and *k* = 136.272
- Consider three alternatives …



- The problem
- **Rounding**
- Circular list
- Fractional interval

- **First, round the fractional interval**
- **For example, when $N = 9$, $n = 2$, take $k = 4$ or 5**

- First, round the fractional interval
- For example, when $N = 9$, $n = 2$, take $k = 4$ or 5
  - If $k = 4$ and RN = 1, the sample is the three elements 1, 5, 9.
  - If RN = 2, 3, or 4, the sample has only two elements
  - If $k = 5$ and RN = 1, 2, 3, or 4, the sample has two elements
  - If RN = 5, the sample has only one element

- First, round the fractional interval
- For example, when $N = 9$, $n = 2$, take $k = 4$ or 5
  - If $k = 4$ and RN = 1, the sample is the three elements 1, 5, 9.
  - If RN = 2, 3, or 4, the sample has only two elements
  - If $k = 5$ and RN = 1, 2, 3, or 4, the sample has two elements
  - If RN = 5, the sample has only one element
- What would happen if $N = 952$ and $n = 200$?
  - Rounding k to 5, RN's 1, 2, 3, & 4 select 191, and RN 5 selects 190 – neither sample size is 200!

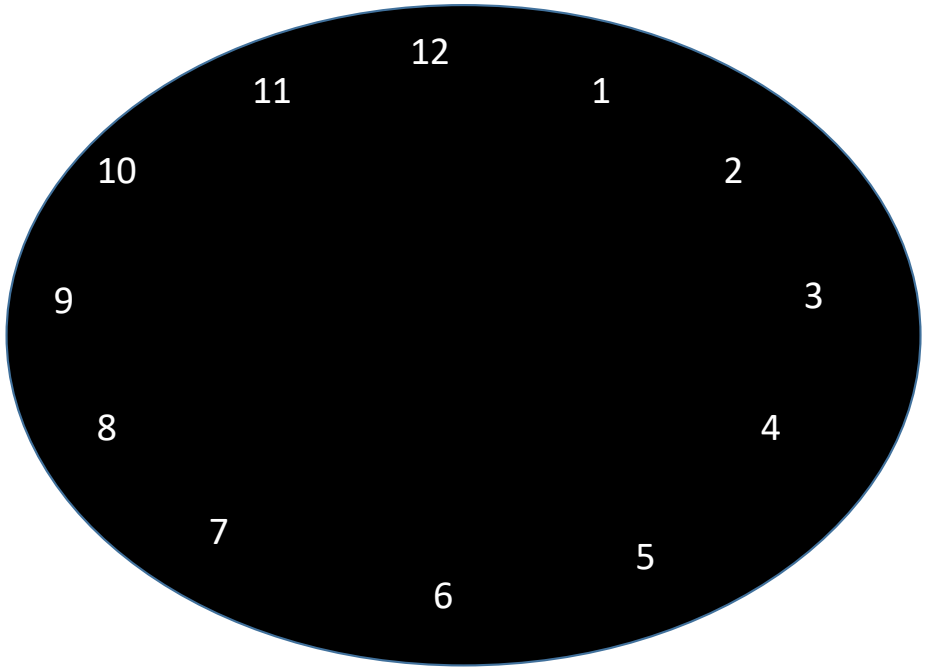
- First, round the fractional interval
- For example, when $N = 9$, $n = 2$, take $k = 4$ or 5
  - If $k = 4$ and RN = 1, the sample is the three elements 1, 5, 9.
  - If RN = 2, 3, or 4, the sample has only two elements
  - If $k = 5$ and RN = 1, 2, 3, or 4, the sample has two elements
  - If RN = 5, the sample has only one element
- What would happen if $N = 952$ and $n = 200$?
  - Rounding k to 5, RN's 1, 2, 3, & 4 select 191, and RN 5 selects 190 – neither sample size is 200!
- What about for $N = 170{,}345$ and $n = 1{,}250$?
  - The sample size can be either 1252 or 1253

- First, round the fractional interval
- For example, when $N = 9$, $n = 2$, take $k = 4$ or 5
  - If $k = 4$ and RN = 1, the sample is the three elements 1, 5, 9.
  - If RN = 2, 3, or 4, the sample has only two elements
  - If $k = 5$ and RN = 1, 2, 3, or 4, the sample has two elements
  - If RN = 5, the sample has only one element
- What would happen if $N = 952$ and $n = 200$?
  - Rounding k to 5, RN's 1, 2, 3, & 4 select 191, and RN 5 selects 190 – neither sample size is 200!
- What about for $N = 170{,}345$ and $n = 1{,}250$?
  - The sample size can be either 1252 or 1253
- Rounding thus has the problem that the sample size is not fixed, and we don't get the target sample size!

- **Second solution is one some people prefer**
- **Treat the list as circular**
  - **As beore, calculate the interval *k* = *N* / *n*, and round up or down, say to *k**
  - **Choose a RN anywhere from 1 to *N* at random**
  - **Then start counting every k*th thereafter**
  - **Keep going until exactly *n* elements are selected**
- **But what if you get to the end of the list before you have *n* elements?**

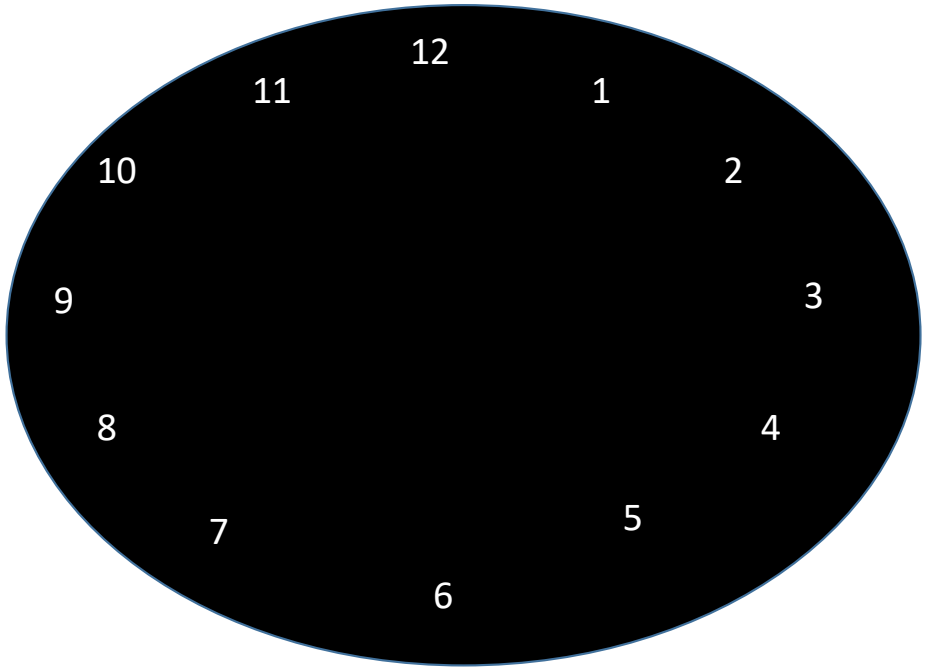
- The problem
- Rounding
- **Circular list**
- Fractional interval

- **Answer: "wrap"**
  - **Think about the list like it is a clock:**



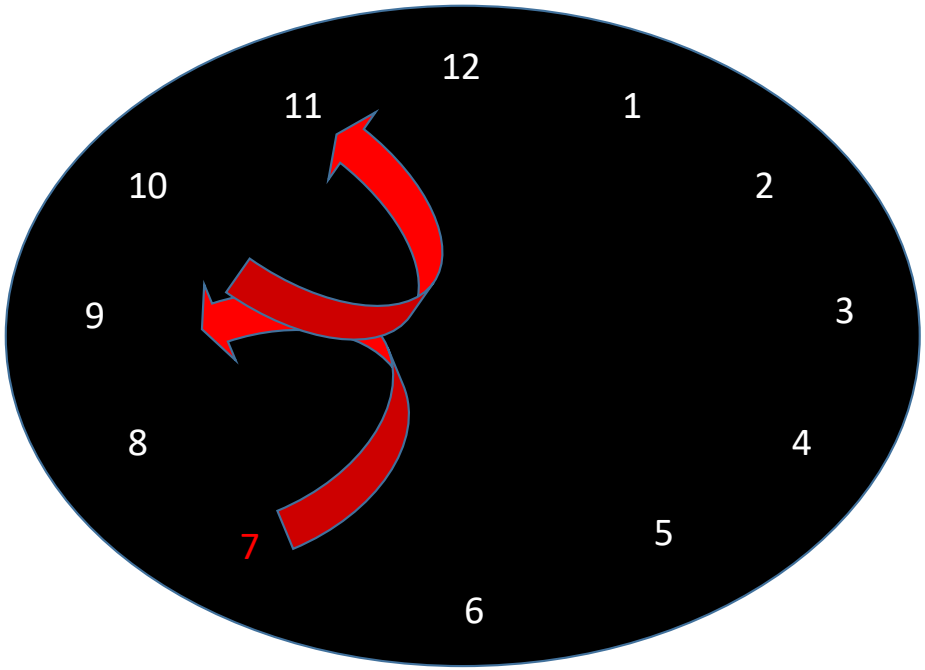
- The problem
- Rounding
- **Circular list**
- Fractional interval

- **Suppose *n* = 5 and *N* = 12, or *k* = 12/5 = 2.4**
  - **Round to [*k*] = 2, and choose random start 7:**



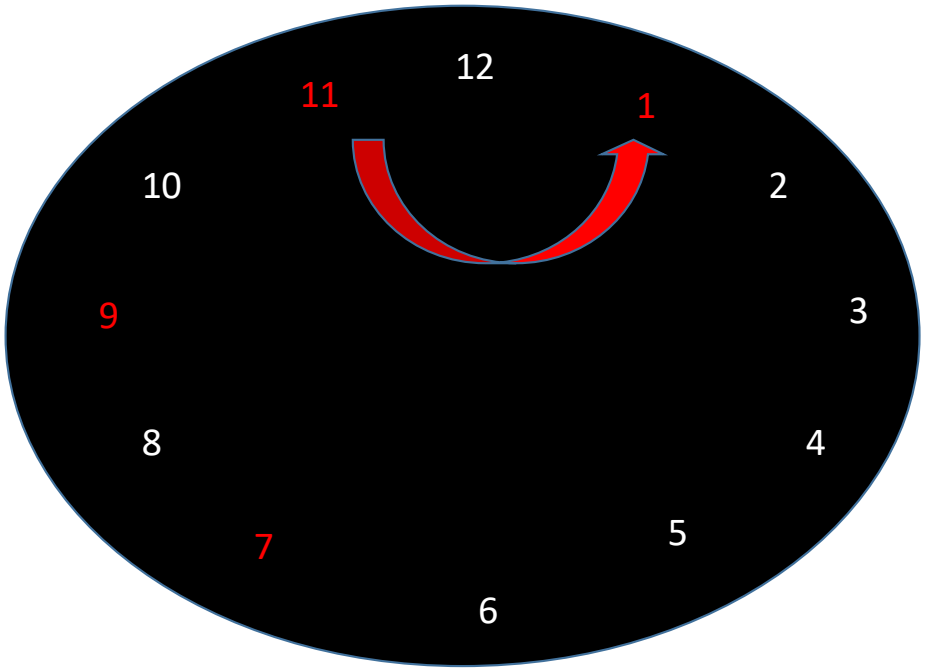
- The problem
- Rounding
- **Circular list**
- Fractional interval

- **Take every 2 after starting at 7 …**
  - **And then 9 … and then 11 … and then …**



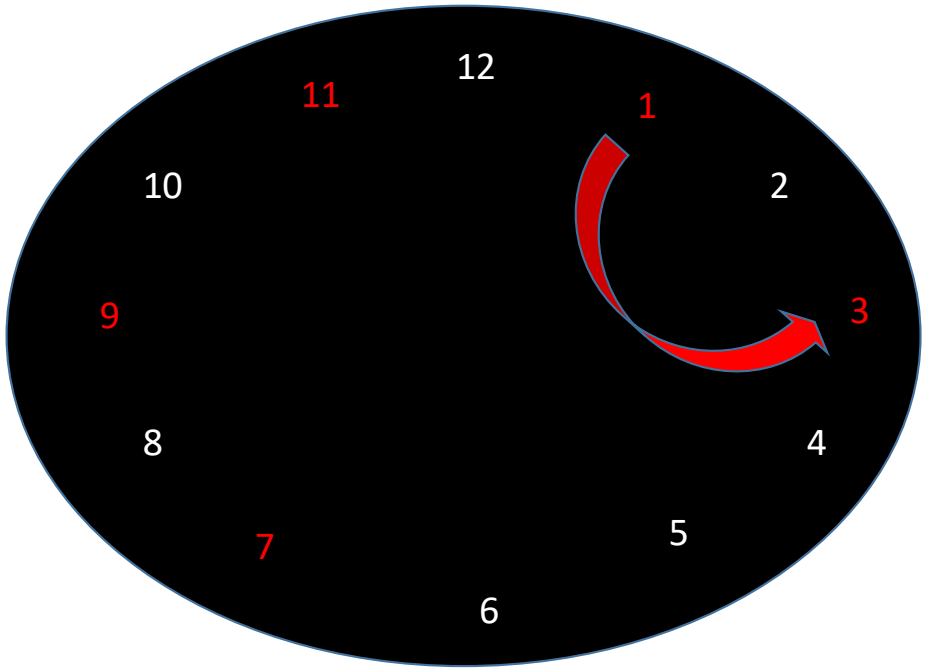
- The problem
- Rounding
- **Circular list**
- Fractional interval

- **… and then … 1**
  - **And then 3 … and then …**



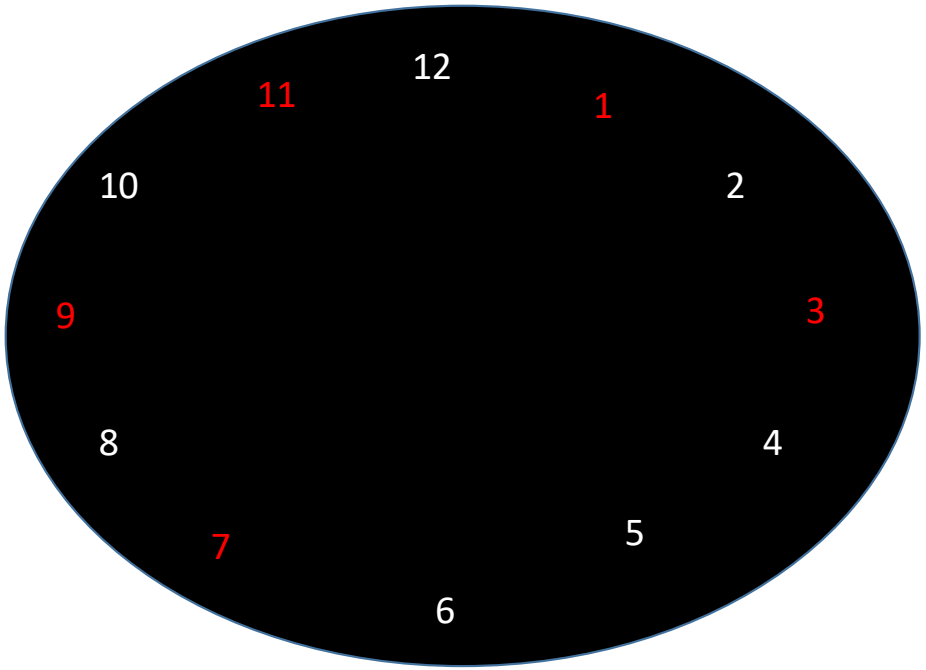
- The problem
- Rounding
- **Circular list**
- Fractional interval

- **… and then … 1**
  - **And then 3 … and then … STOP because *n* = 5**



- The problem
- Rounding
- **Circular list**
- Fractional interval

- **Remember, start anywhere on the list …**
  - **… and wrap**



- Third, there's another more widely used technique that uses the fraction
- Suppose that $N = 23$ and $n = 5$, so that $k = 23/5 = 4.6$
- As we've seen, if $[k] = 5$ is applied, $n = 4$ or 5

	1	2	3	4	5
1	1	2	3	4	5
2	6	7	8	9	10
3	11	12	13	14	15
4	16	17	18	19	20
5	21	22	23		

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- **Fractional interval**

- **Here, use the fractional interval …**
  - **Choose a random start from 0.1 to 4.6**
  - **But how do you do that?**

$$\frac{3}{8}+\frac{6}{4}$$

- Here, use the fractional interval …
  - Choose a random start from 0.1 to 4.6
  - But how do you do that?

- One way is with a table of random numbers
- Since we need a number from 0.1 to 4.6, why not choose a random number from 01 to 46
- Suppose the number is 35
- "Insert" a decimal to make it fractional: 3.5

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- Fractional interval

- Here, use the fractional interval …
  - Choose a random start from 0.1 to 4.6
  - But how do you do that?

- One way is with a table of random numbers
- Since we need a number from 0.1 to 4.6, why not choose a random number from 01 to 46
- Suppose the number is 35
- "Insert" a decimal to make it fractional: 3.5

$$\frac{3}{8}+\frac{6}{4}$$

- Alternatively, generate a UNIFORM random number from zero to 1 in statistical software, say 0.76087
- Multiply by 4.6, and get 3.5

- The problem
- Rounding
- Circular list
- **Fractional interval**

- **But then what?**
  - **Do systematic counting …**
  - **But "count" every 4.6 …**

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- **Fractional interval**

- **But then what?**
  - **Do systematic counting …**
  - **But "count" every 4.6 …**
  - **Starting with (1) 3.5, we 'count' to (2) 3.5 + 4.6 = 8.1**

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- **Fractional interval**

- **But then what?**
  - **Do systematic counting …**
  - **But "count" every 4.6 …**
  - **Starting with (1) 3.5, we 'count' to (2) 3.5 + 4.6 = 8.1 …**
  - **And again, (3) 8.1 + 4.6 = 12.7 …**

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- **Fractional interval**

- **But then what?**
  - **Do systematic counting …**
  - **But "count" every 4.6 …**
  - **Starting with (1) 3.5, we 'count' to (2) 3.5 + 4.6 = 8.1 …**
  - **And again, (3) 8.1 + 4.6 = 12.7 …**
  - **And again, (4) 12.7 + 4.6 = 17.3 …**

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- **Fractional interval**

- **But then what?**
  - **Do systematic counting …**
  - **But "count" every 4.6 …**
  - **Starting with (1) 3.5, we 'count' to (2) 3.5 + 4.6 = 8.1 …**
  - **And again, (3) 8.1 + 4.6 = 12.7 …**
  - **And again, (4) 12.7 + 4.6 = 17.3 …**
  - **And again, (5) 17.3 + 4.6 = 21.9 …**

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- Fractional interval

- But then what?
  - Do systematic counting …
  - But "count" every 4.6 …
  - Starting with (1) 3.5, we 'count' to (2) 3.5 + 4.6 = 8.1 …
  - And again, (3) 8.1 + 4.6 = 12.7 …
  - And again, (4) 12.7 + 4.6 = 17.3 …
  - And again, (5) 17.3 + 4.6 = 21.9 …
  - And just to be sure, one more time gives us, (5) 21.9 + 4.6 = 26.5!

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- **Fractional interval**

- **But then what?**
  - **Do systematic counting …**
  - **But "count" every 4.6 …**
  - **Starting with (1) 3.5, we 'count' to (2) 3.5 + 4.6 = 8.1 …**
  - **And again, (3) 8.1 + 4.6 = 12.7 …**
  - **And again, (4) 12.7 + 4.6 = 17.3 …**
  - **And again, (5) 17.3 + 4.6 = 21.9 …**
  - **And just to be sure, one more time gives us, (5) 21.9 + 4.6 = 26.5!**
  - **Oops! We're off the list!**

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- **Fractional interval**

- **But before we got off the list, we had *n*= 5 "selections"**
  - **3.5, 8.1, 12.7. 17.3, and 21.9**
- **What do we do with the decimals, though?**

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- **Fractional interval**

- **But before we got off the list, we had *n*= 5 "selections"**
  - **3.5, 8.1, 12.7. 17.3, and 21.9**
- **What do we do with the decimals, though?**
  - **Truncate to the whole number**
  - **That is, our selections are**
  - **3~~.5~~, 8~~.1~~, 12~~.7~~. 17~~.3~~, and 21~~.9~~ -**

$$\frac{3}{8}+\frac{6}{4}$$

- But before we got off the list, we had *n*= 5 "selections"
  - 3.5, 8.1, 12.7. 17.3, and 21.9
- What do we do with the decimals, though?
  - Truncate to the whole number
  - That is, our selections are
  - 3~~.5~~, 8~~.1~~, 12~~.7~~. 17~~.3~~, and 21~~.9~~ -
- We have chosen elements 3, 8, 12, 17, 21!

$$\frac{3}{8}+\frac{6}{4}$$

RS	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107
	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153
	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199
	231	232	233	234	235	236	237	238	239						

. . . . . .

RS	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92
	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138
	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184
	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230

RS	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107
	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153
	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199
	231	232	233	234	235	236	237	238	239						

RS	32	33	34	**35**	36	37	38	39	40	41	42	43	44	45	46
	78	79	80	**81**	82	83	84	85	86	87	88	89	90	91	92
	124	125	126	**127**	128	129	130	131	132	133	134	135	136	137	138
	170	171	172	**173**	174	175	176	177	178	179	180	181	182	183	184
	216	217	218	**219**	220	221	222	223	224	225	226	227	228	229	230

RS	01	02	03	04	**05**	06	07	08	09	10	11	12	13	14	15
	47	48	49	50	**51**	52	53	54	55	56	57	58	59	60	61
	93	94	95	96	**97**	98	99	100	101	102	103	104	105	106	107
	139	140	141	142	**143**	144	145	146	147	148	149	150	151	152	153
	185	186	187	188	**189**	190	191	192	193	194	195	196	197	198	199
	231	232	233	234	**235**	236	237	238	239						

. . . . . .

RS	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92
	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138
	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184
	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230

- The problem
- Rounding
- Circular list
- **Fractional interval**

What does this all mean?

Element	Random start	No. of RS's	*f*
1	10-19	10	1/4.6
2	20-29	10	1/4.6
3	30-39	10	1/4.6
4	40-46,01-03	10	1/4.6
5	04-13	10	1/4.6
…	…	…	…
23	01-09,46	10	1/4.6

$$\frac{3}{8}+\frac{6}{4}$$

- The problem
- Rounding
- Circular list
- **Fractional interval**

What does this all mean?
Simple method that is *epsem*

Element	Random start	No. of RS's	*f*
1	10-19	10	1/4.6
2	**20-29**	**10**	**1/4.6**
3	30-39	10	1/4.6
4	40-46,01-03	10	1/4.6
5	04-13	10	1/4.6
…	…	…	…
23	01-09,46	10	1/4.6

$$\frac{3}{8}+\frac{6}{4}$$

Unit 5

- 1 Systematic selection
- 2 Intervals with fractions
- **3 List order**
- 4 Uncertainty estimation