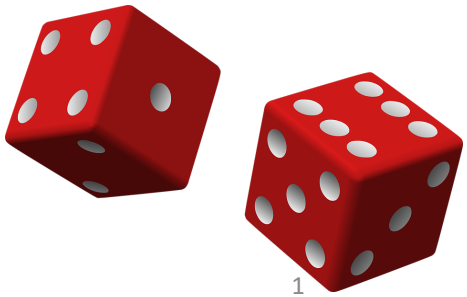
Sampling People, Records, & Networks

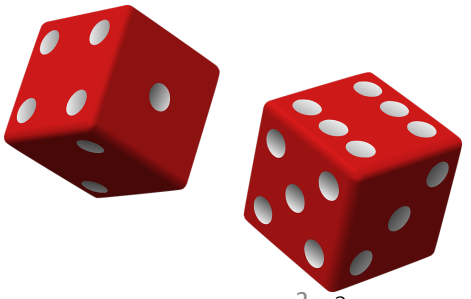
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



Unit 4

- 1 Stratifica-tion
- 2 Sampling variance
- 3 Propor-tionate allocation
- 4 Dispropor-tionate allocations
- 5 Comapring strata
- 6 Number of strata

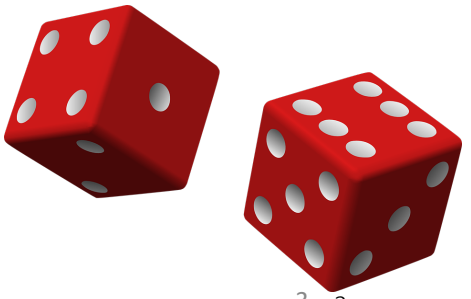
- **Unit 1: Sampling as a research tool**
- **Unit 2: Mere randomization**
- **Unit 3: Saving money**
- **Unit 4: Being more efficient**
- **Unit 5: Simplifying sampling**
- **Unit 6: Some extensions & applications**



Unit 4

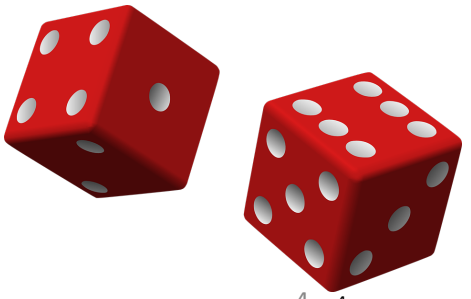
- 1 Forming groups
- 2 Sampling variance
- 3 More on grouping
- 4 Allocate sample
- 5 Other allocations
- 6 Weights

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- **Unit 4: Being more efficient**
  - Forming groups
  - Sampling variance
  - More on grouping
  - Allocate sample
  - Other allocations
  - Weights to combine across strata
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



- The procedure
- Using discrete variables
- Selection
- Combining data across groups

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
  - Forming groups
  - Sampling variance
  - More on grouping
  - Allocate sample
  - Other allocations
  - Weights to combine across strata
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



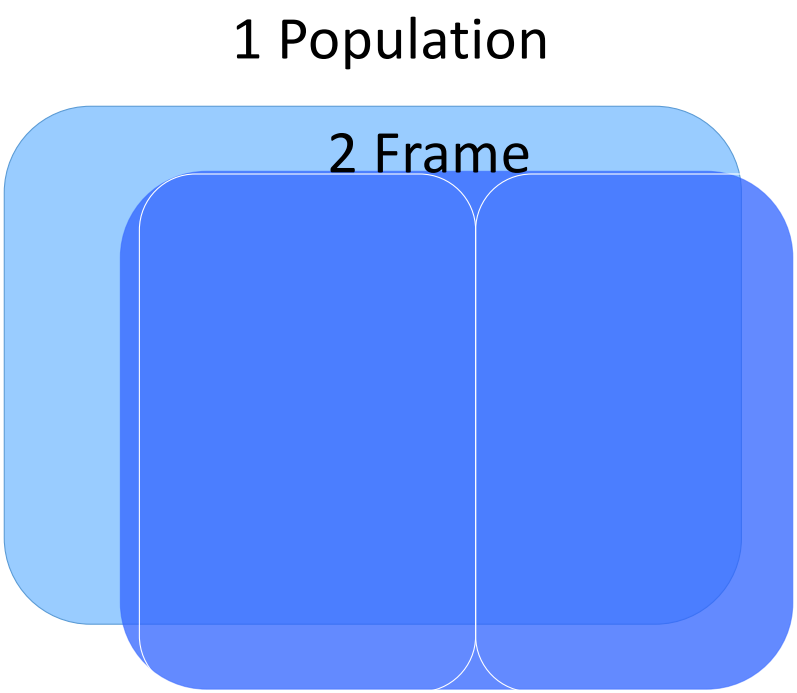
- **The procedure**
- Using discrete variables
- Selection
- Combining data across groups

Seq. No.	ID	Division	Sex	Rank
1	38516070	Eng&Prof	m	3
2	25428686	Medicine	f	3
3	30318994	Medicine	m	3
4	35886147	Medicine	m	1
5	41416693	Eng&Prof	f	2
6	60055684	Lit&SocSci	m	1
7	76731882	Medicine	f	3
8	51765248	Biol&Sci	m	3
9	26471240	Lit&SocSci	f	3
10	25864673	Biol&Sci	m	1
11	23049573	Medicine	m	1
12	12928113	Lit&SocSci	m	1
13	13594590	Lit&SocSci	m	1
14	20820530	Medicine	m	3
15	52026919	Medicine	m	1
16	59283042	Eng&Prof	m	3
17	37941753	Medicine	m	2
18	32498845	Eng&Prof	m	1
19	42120123	Medicine	m	1
20	83562743	Eng&Prof	m	3
21	39834280	Biol&Sci	m	2
22	60683602	Medicine	f	1
23	18186559	Medicine	m	1
24	20110594	Medicine	m	3
25	61862981	Lit&SocSci	m	1

- The procedure
- Using discrete variables
- Selection
- Combining data across groups

- **Stratification procedure**
  - **Population (faculty)**
  - **Frame (faculty list)**
    - **Auxiliary variables: things known about each element in the population before the sample is drawn**
    - **Sequence number, ID, rank, sex, division**

- **The procedure**
- Using discrete variables
- Selection
- Combining data across groups



- The procedure
- Using discrete variables
- Selection
- Combining data across groups

- **Stratification procedure**
  - Population (faculty, step 1)
  - Frame (faculty list, step 2)
    - Auxiliary variables: things known about each element in the population before the sample is drawn
    - Sequence number, ID, rank, sex, division

- **Stratification procedure**
  - Population (faculty, step 1)
  - Frame (faculty list, step 2)
    - Auxiliary variables: things known about each element in the population before the sample is drawn
    - Sequence number, ID, rank, sex, division
  - Divide list into groups based on the auxiliary variables
    - Must be 'discrete' (categorical)
    - Must be known for every element in the list
  - Count up the number of elements in each group, N_h

- The procedure
- **Using discrete variables**
- Selection
- Combining data across groups

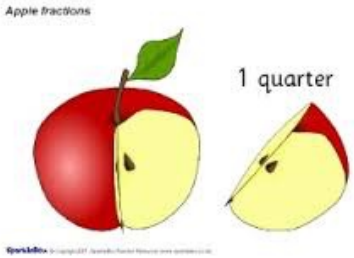
h	Stratum	N_h
1	Assistant	115
2	Associate	75
3	Full	210
Total		400

- **Stratification procedure**
  - **Population (faculty, step 1)**
  - **Frame (faculty list, step 2)**
    - **Auxiliary variables: things known about each element in the population before the sample is drawn**
    - **Sequence number, ID, rank, sex, division**
  - **Divide list into groups based on the auxiliary variables**
    - **Must be 'discrete' (categorical)**
    - **Must be known for every element in the list**
  - **Count up the number of elements in each group N_h**
  - **Compute the fraction of the population in each group W_h**

- The procedure
- **Using discrete variables**
- Selection
- Combining data across groups

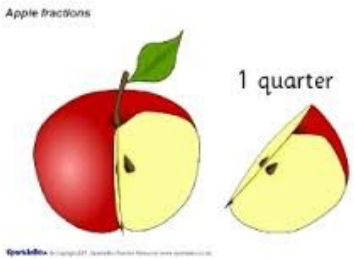
h	Stratum	N_h	W_h
1	Assistant	115	0.2875
2	Associate	75	0.1875
3	Full	210	0.5250
Total		400	1.0000

- **Stratification procedure**
  - Population (faculty, step 1)
  - Frame (faculty list, step 2)
    - Auxiliary variables: things known about each element in the population before the sample is drawn
    - Sequence number, ID, rank, sex, division
  - Divide list into groups based on the auxiliary variables
    - Must be 'discrete' (categorical)
    - Must be known for every element in the list
  - Count up the number of elements in each group N_h
  - Compute the fraction of the population in each group W_h
  - Draw a sample from each group n_h (sample, step 3)
  - Keep track of sampling rates $f_h = n_h / N_h$
  - *sampling fraction*



- The procedure
- Using discrete variables
- **Selection**
- Combining data across groups

h	Stratum	N_h	W_h	n_h	f_h
1	Assistant	115	0.2875	23	0.2
2	Associate	75	0.1875	15	0.2
3	Full	210	0.5250	42	0.2
Total		400	1.0000	80	0.2



- **Stratification procedure - estimation**
  - **Calculate estimate for each group (estimation, step 4a)**
  - **Say means $\bar{y}_1 = \$50, \bar{y}_2 = \$70,\ and\ \bar{y}_3 = \$90$**

$>avg.$

- The procedure
- Using discrete variables
- Selection
- **Combining data across groups**

h	Stratum	N_h	W_h	n_h	f_h	$\bar{y}_h$
1	Assistant	115	0.2875	23	0.2	50
2	Associate	75	0.1875	15	0.2	70
3	Full	210	0.5250	42	0.2	90
Total		400	1.0000	80	0.2	

$>avg.$

- **Stratification procedure - estimation**
  - **Calculate estimate for each group (estimation, step 4a)**
  - **Say means $\bar{y}_1 = \$50, \bar{y}_2 = \$70,\ and\ \bar{y}_3 = \$90$**
  - **Combine estimates across groups (step 4b)**

$$\bar{y} = \sum_{h=1}^{H} W_h \bar{y}_h$$

$>avg.$

- **Stratification procedure - estimation**
  - **Calculate estimate for each group (estimation, step 4a)**
  - **Say means** $\bar{y}_1 = \$50, \bar{y}_2 = \$70, \ and \ \bar{y}_3 = \$90$
  - **Combine estimates across groups (step 4b)**

$$\bar{y} = \sum_{h=1}^{H} W_h \bar{y}_h$$

  - **Or here,** $\bar{y}_w = (0.2875)(\$50) + (0.1875)\$70 + (0.5250)\$90 = \$74.75$

$>avg.$

- The procedure
- Using discrete variables
- Selection
- **Combining data across groups**

h	Stratum	N_h	W_h	n_h	f_h	$\bar{y}_h$
1	Assistant	115	0.2875	23	0.2	50
2	Associate	75	0.1875	15	0.2	70
3	Full	210	0.5250	42	0.2	90
Total		400	1.0000	80	0.2	$74.75

$>avg.$

- **Stratification procedure - estimation**
  - **Calculate estimate for each group (estimation, step 4a)**
  - **Say means** $\bar{y}_1 = \$50, \bar{y}_2 = \$70, \ and \ \bar{y}_3 = \$90$
  - **Combine estimates across groups (step 4b)**

$$\bar{y} = \sum_{h=1}^{H} W_h \bar{y}_h$$

  - **Or here,** $\bar{y} = (0.2875)(\$50) + (0.1875)\$70 + (0.5250)\$90 = \$74.75$
  - **But there are two more steps to go … standard error and confidence interval computation**

>avg.

Unit 4

- 1 Forming groups
- 2 Sampling variance
- 3 More on grouping
- 4 Allocate sample
- 5 Other allocations
- 6 Weights

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
  - Forming groups
  - **Sampling variance**
  - More on grouping
  - Allocate sample
  - Other allocations
  - Weights to combine across strata
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications