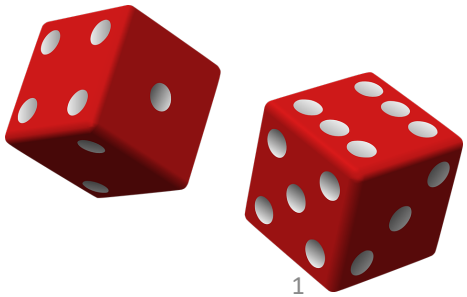
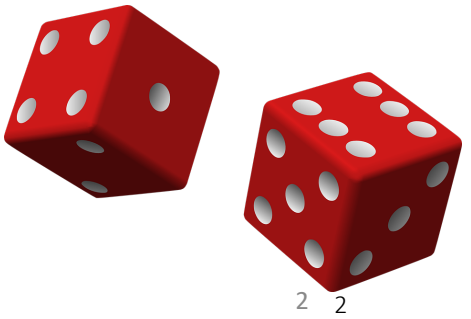
Sampling People, Records, & Networks

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

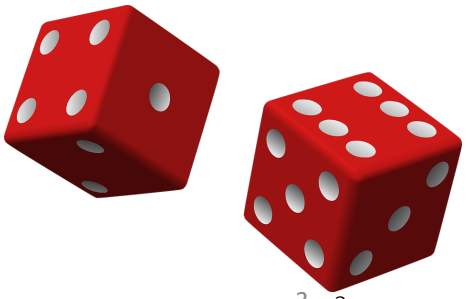


Unit 5

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



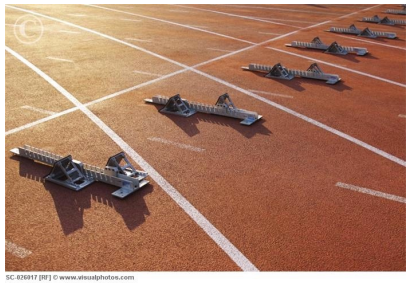
- ...
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- **Unit 5: Simplifying sampling**
  - 1 Systematic selection
  - 2 Intervals with fractions
  - 3 List order
  - **4 Uncertainty estimation**
- Unit 6: Some extensions & applications



- **Estimation of the sample mean:** $\bar{y} = \frac{\sum y_i}{n}$
- **Sampling variance cannot be estimated using only survey data**
  - **Only a single random start used**
  - **Two approaches to dealing with the problem**
    - **Use additional random starts**
    - **Model the variance**
- **Use c random starts and**

$$\bar{y} = \frac{\sum\sum y_{\gamma i}}{cn} = \frac{1}{c}\sum_{\gamma} \bar{y}_{\gamma}$$

$$\text{var}\left(\bar{y}\right) = \frac{1}{c}\sum_{\gamma}\left(\bar{y}_{\gamma} - \bar{y}\right)^2$$



- **Model the population (sample selection process)**
- **SRS model**
  - **Are elements in the list are ordered at random?**
  - **Yes?**

$$\text{var}\left(\bar{y}\right) = \left(1 - f\right)\frac{s^2}{n}$$

- Are elements in the list are ordered at random?
  - Yes?
  - Can we assume homogeneity across 'rows' (zones), in groups of rows?
  - Yes?
  - Assume random ordering within zones
  - Proportionately allocated selection with n_h = 1 selected per zone
- Collapse neighboring zones to create "pseudo strata" that have multiple selections, and using

$$\text{var}\left(\bar{y}\right) = \frac{1-f}{n}\sum W_h s_h^2$$

$$W_h = \frac{n_h}{n}$$

4:3

- **Are elements in the list are ordered at random?**
  - **Yes?**
  - **Can we assume homogeneity across 'rows' (zones), in groups of rows?**
  - **Yes?**
  - **Assume random ordering within zones**
  - **Proportionately allocated selection with n_h = 1 selected per zone**
  - **Is the ordering really almost continuous?**
  - **Yes?**
  - **Stratified random model special case: pair successive rows**

$$\mathrm{var}\left(\bar{y}\right) = \frac{1-f}{n^2}\sum_h \left(y_{h1} - y_{h2}\right)^2$$

Block	# Rental	# HUs	i
240	23	30	1
278	25	33	2
288	42	61	3
377	0	3	4
388	16	27	5
398	37	47	6

- ***Epsem* sample**

$$\bar{y}_{\#rental} = \frac{\sum y_i}{n}$$

$$= (23 + 25 + 42 + 0 + 16 + 37)/6 = 23.83$$

- **Is the list order random?**
- **SRS model**

$$\text{var}(\bar{y}) = (1-f)\frac{s^2}{n} = \left(1-\frac{6}{60}\right)\left(\frac{1}{6}\right)\frac{\left(4543-6*23.83^2\right)}{6-1}$$

$$= (0.90)(0.1667)(226.97) = 34.045$$

- **This list is probably continuously ordered with respect to Y.**
- **Paired selection model, even # elements**

$$\text{var}(\bar{y}) = \frac{(1-f)}{n^2}\sum_{h}^{n/2}(y_{ha} - y_{hb})^2$$

$$= \left(1-\frac{6}{60}\right)\left(\frac{1}{6^2}\right)\left[(23-25)^2 + (42-0)^2 + (16-37)^2\right]$$

$$= (0.9)(0.0278)(4+1764+441) = 55.225$$

Unit 5

- I

- 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**