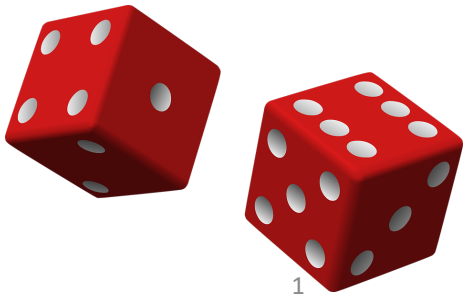
Sampling People, Records, & Networks

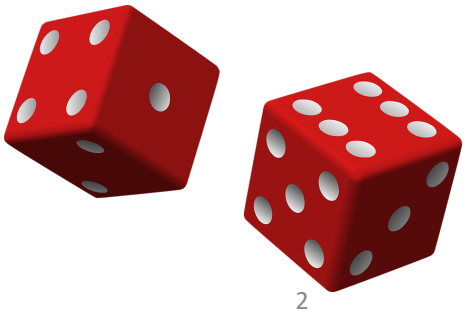
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



Unit 6

1. Stratified multistage sampling
2. Weights for over/under sampling
3. Nonresponse & noncoverage weighting
4. Variance estimation and software
5. Statistical software for sample selection
6. Sampling networks: multiplicity weighting

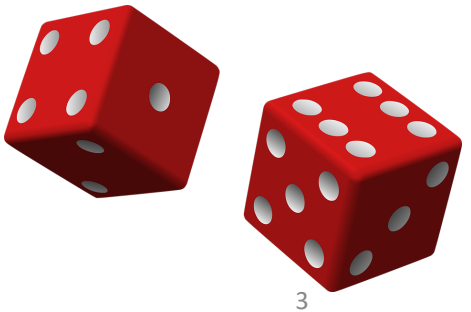
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



Unit 6

1. Stratified multistage sampling
2. Weights for over/under sampling
3. Nonresponse & noncoverage weighting
4. Variance estimation and software
5. Statistical software for sample selection
6. Sampling networks: multiplicity weighting

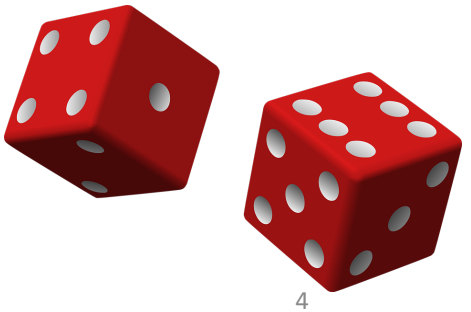
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- **Unit 6: Some extensions & applications**
  1. Statistical software for sample selection
  2. Stratified multistage sampling
  3. Weights for over/under sampling
  4. Nonresponse & noncoverage weighting
  5. Sampling networks: multiplicity weighting
  6. Non-probability sampling



Unit 6

1. Statistical software for sample selection
2. Stratified multistage sampling
3. Weights for over/under sampling
4. Nonresponse & noncoverage weighting
5. Sampling networks: multiplicity weighting
6. **Non-probability sampling**

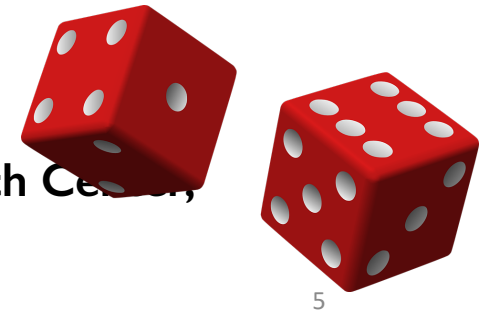
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- **Unit 6: Some extensions & applications**
  1. Statistical software for sample selection
  2. Stratified multistage sampling
  3. Weights for over/under sampling
  4. Nonresponse & noncoverage weighting
  5. Sampling networks: multiplicity weighting
  6. **Non-probability sampling**



- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- Web panels
- Estimation

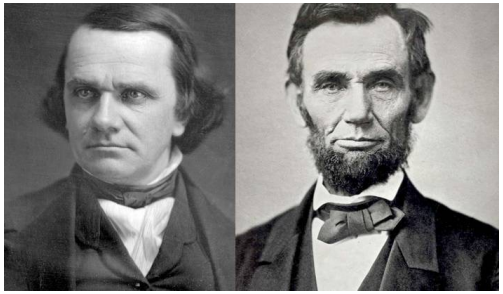
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications
  1. Statistical software for sample selection
  2. Stratified multistage sampling
  3. Weights for over/under sampling
  4. Nonresponse & noncoverage weighting
  5. Sampling networks: multiplicity weighting
  6. **Non-probability sampling**

(Selected material by Sunghee Lee, Survey Research Ce… University of Michigan, used with permission.)



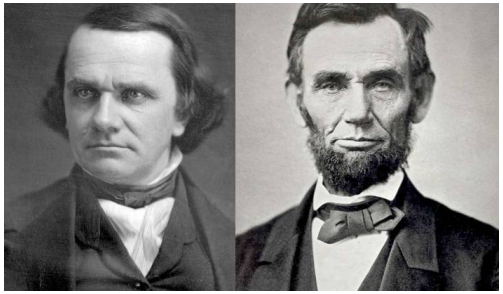
- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- Web panels
- Estimation

- **Leslie Kish (1965)**
  - **"In probability sampling, every element in the population has a known nonzero probability of being selected."**
  - **"Probability sampling requires that the actual selection … be made by a *mechanical procedure* that assigns the desired probabilities. This *randomization* process …"**



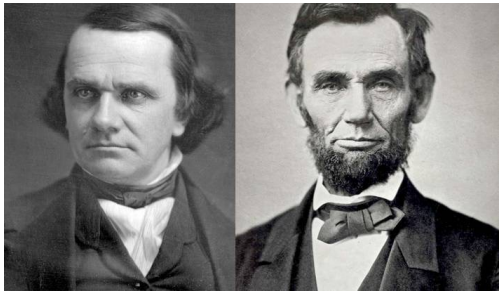
- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- Web panels
- Estimation

- **Leslie Kish (1965)**
  - "In probability sampling, every element in the population has a known nonzero probability of being selected."
  - "Probability sampling requires that the actual selection … be made by a *mechanical procedure* that assigns the desired probabilities. This *randomization* process …"
- **On nonprobability sampling approaches, per Cochran (1977)**
  - "They are not amenable to the development of a sampling theory that is model-free, since no element of random selection is involved. …… Even if a method appears to do well in one, this doe not guarantee that it will do well under different circumstances."



- **Continuing debate**
- Probability "like"
- Snowball & network
- RDS
- Web panels
- Estimation

- **Andrew Gelman (2015)**
  - **"Remember: just about no sample of humans is really a probability sample or even close to a probability sample, and just about no regression model applied to humans is correct or even close to correct."**



- Continuing debate
- **Probability "like"**
- Snowball & network
- RDS
- Web panels
- Estimation

- **Probability sample, nonresponse**
  - **Increased sample size, weights**
  - **Substitution for nonresponse**
    - **Purposive**
    - **Matched**
    - **Random**
    - **Stratified random**

- Continuing debate
- **Probability "like"**
- Snowball & network
- RDS
- Web panels
- Estimation

- **Probability sample, nonresponse**
  - **Increased sample size, weights**
  - **Substitution for nonresponse**
    - **Purposive**
    - **Matched**
    - **Random**
    - **Stratified random**
- **Probability-like sample**
  - **Multistage quota**

- Continuing debate
- **Probability "like"**
- Snowball & network
- RDS
- Web panels
- Estimation

- **Web-based sampling**
  - **Mostly convenient samples**
    - **Email blasts**
    - **Chat rooms**
    - **Instant messengers**
    - **Banner ads**
    - **Social media (e.g., Facebook)**

- Continuing debate
- Probability "like"
- **Snowball & network**
- RDS
- Web panels
- Estimation

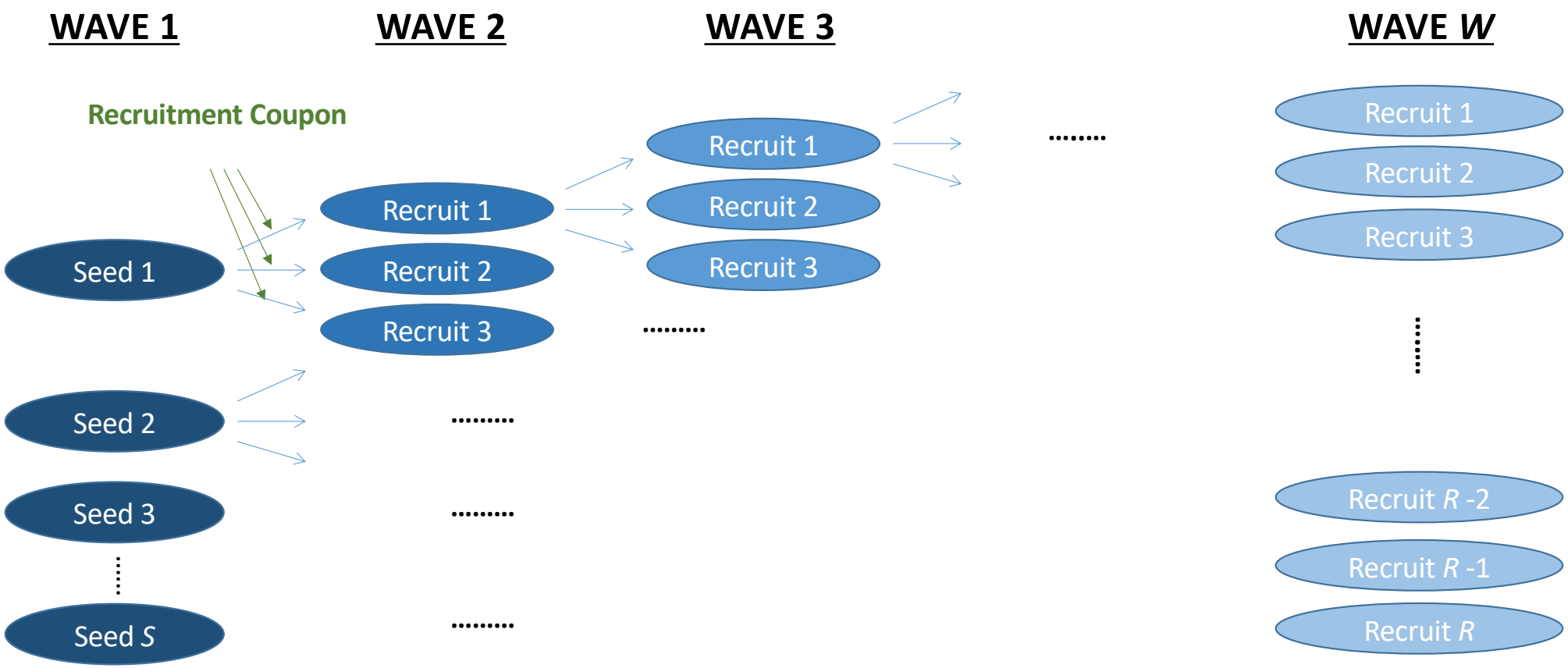
- **Snowball sampling**
  - **Start with a convenient sample of rare group members**
  - **Use "insider knowledge" to locate more members through chain referral**
  - **For qualitative studies**

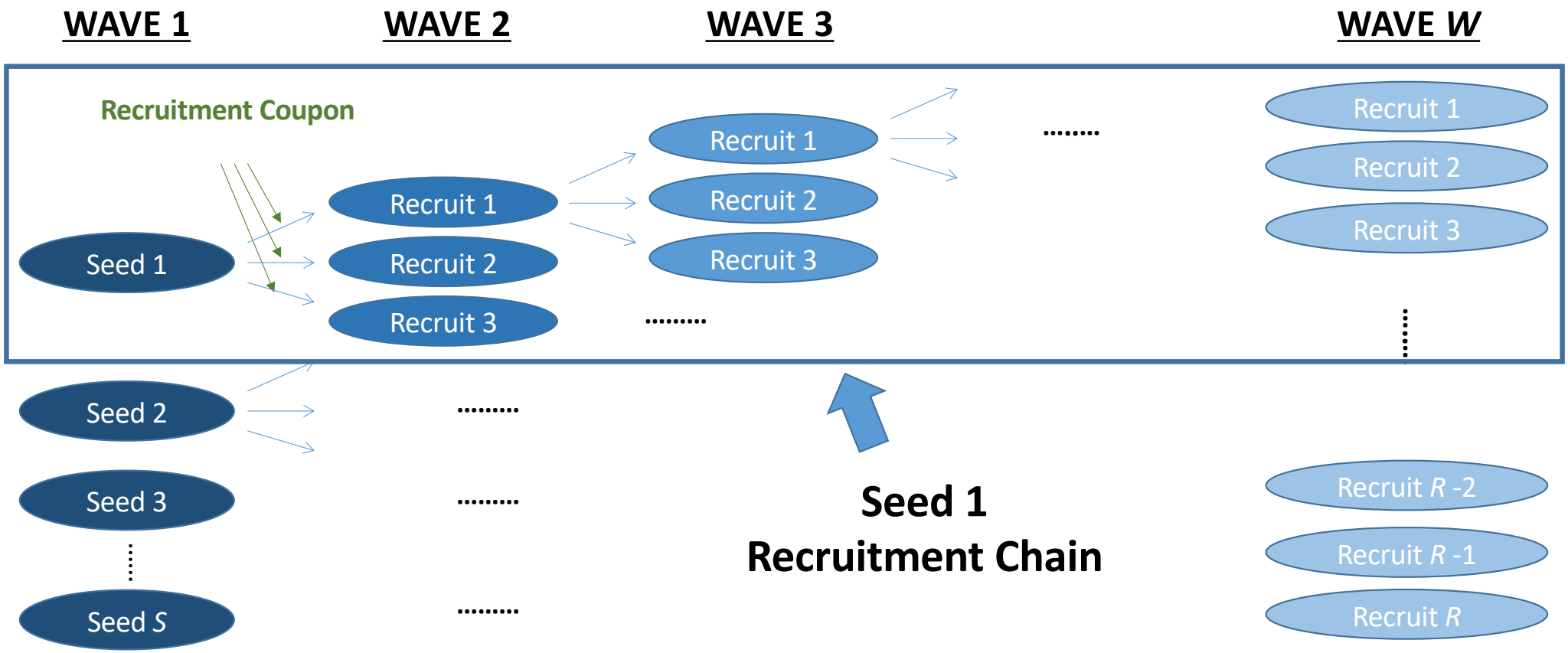
- Continuing debate
- Probability "like"
- **Snowball & network**
- RDS
- Web panels
- Estimation

- **Snowball sampling**
  - **Start with a convenient sample of rare group members**
  - **Use "insider knowledge" to locate more members through chain referral**
  - **For qualitative studies**
- **Network sampling**
  - **Also known as multiplicity sampling (Sirken, 1970)**
  - **Exploits connectedness among people (=networks)**
  - **Counting rule:**
    - **Multiplicity --**
    - **Capitalize on duplicate counting of population elements**
    - **Same person linked to multiple households of their relatives**

- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- Web panels
- Estimation

- **Respondent-driven sampling**
  - **Exploits social networks among rare population members for sampling purposes**
  - **Assuming Markov process in recruitment, claims to produce unbiased estimates (estimate stability=equilibrium points)**
  - **Self-reported network size used as selection probability adjustment weight**


WAVE 1
WAVE 2
WAVE 3
WAVE W
Recruitment Coupon
Seed 1
Seed 2
Seed 3
Seed S
Recruit 1
Recruit 2
Recruit 3
Recruit 1
Recruit 2
Recruit 3
Recruit 1
Recruit 2
Recruit 3
Recruit R -2
Recruit R -1
Recruit R


WAVE 1
WAVE 2
WAVE 3
WAVE W
Recruitment Coupon
Seed 1
Seed 2
Seed 3
Seed S
Recruit 1
Recruit 2
Recruit 3
Recruit 1
Recruit 2
Recruit 3
Recruit 1
Recruit 2
Recruit 3
Recruit R -2
Recruit R -1
Recruit R
Seed 1
Recruitment Chain

- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- **Web panels**
- Estimation

- **Judgment**

- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- **Web panels**
- Estimation

- **Judgment**
- **Probability web panel**
  - **Email roster with reasonable population coverage**
  - **Panel recruited by probability sampling**

- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- **Web panels**
- Estimation

- **Judgment**
- **Probability web panel**
  - **Email roster with reasonable population coverage**
  - **Panel recruited by probability sampling**
- **Nonprobabiliity web panel**
  - **Mass emailing**

- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- **Web panels**
- Estimation

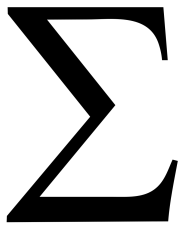
- **Judgment**
- **Probability web panel**
  - **Email roster with reasonable population coverage**
  - **Panel recruited by probability sampling**
- **Nonprobabiliity web panel**
  - **Mass emailing**
  - **River sampling**
    - **Capture visitors of a Web site through banner or pop-up ads to a survey Web site**
    - **Sampling frame: Visitors of other websites**
      - **Who are they? Why do they click on the banners/pop-ups?**
      - **Duplicates?**
      - **Those who do not participate in the survey, no information other than they visited certain Web sites**

- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- Web panels
- Estimation

- **Judgment**
- **Probability web panel**
  - **Email roster with reasonable population coverage**
  - **Panel recruited by probability sampling**
- **Nonprobabiliity web panel**
  - **Mass emailing**
  - **River sampling**
    - **Capture visitors of a Web site through banner or pop-up ads to a survey Web site**
    - **Sampling frame: Visitors of other websites**
      - **Who are they? Why do they click on the banners/pop-ups?**
      - **Duplicates?**
      - **Those who do not participate in the survey, no information other than they visited certain Web sites**
  - **Opt-in/Volunteer panel**

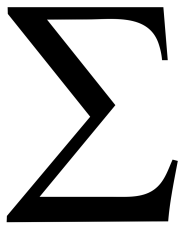
- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- Web panels
- **Estimation**

- **Model based**



- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- Web panels
- **Estimation**

- **Model based**
  - **Sample design model**
    - **SRS**
    - **Cluster sample**
    - **Stratified random**



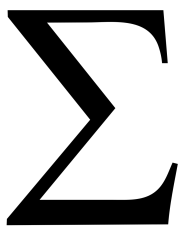
- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- Web panels
- **Estimation**

- **Model based**
  - **Sample design model**
    - **SRS**
    - **Cluster sample**
    - **Stratified random**

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

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

$$\left(\bar{y} - 2 \times se\left(\bar{y}\right), \bar{y} + 2 \times se\left(\bar{y}\right)\right)$$



- Continuing debate
- Probability "like"
- Snowball & network
- RDS
- Web panels
- **Estimation**

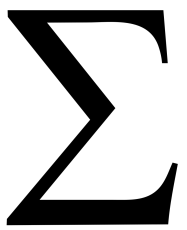
- **Process-based**
  - **Sample design model**
  - **SRS**
  - **Cluster sample**
  - **Stratified random**

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

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

$$\left(\bar{y} - 2 \times se\left(\bar{y}\right), \bar{y} + 2 \times se\left(\bar{y}\right)\right)$$

- **Statistical model-based for population**



Unit 6

1. Stratified multistage sampling
2. Weights for over/under sampling
3. Nonresponse & noncoverage weighting
4. Variance estimation and software
5. Statistical software for sample selection
6. Sampling networks: multiplicity weighting

- 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
  1. Stratified multistage sampling
  2. Weights for over/under sampling
  3. Nonresponse & noncoverage weighting
  4. Variance estimation and software
  5. Statistical software for sample selection
  6. Sampling networks: multiplicity weighting