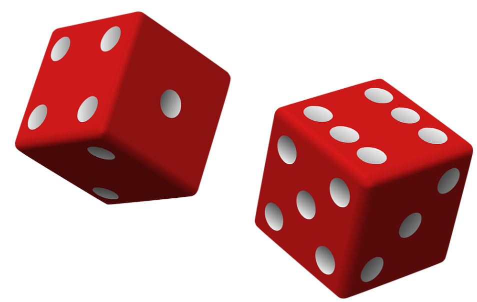
Sampling People, Records, & Networks

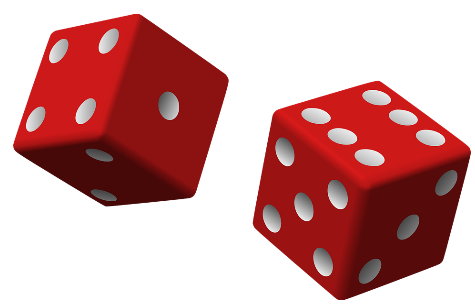
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



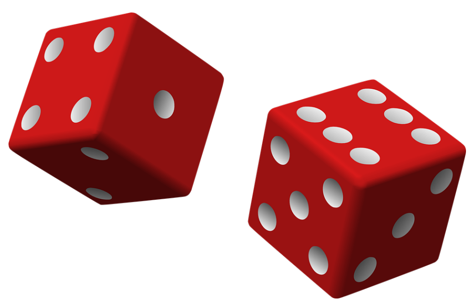
Unit 2

- 1 Simple random sampling
- 2 History
- 3 Sampling distribution
- 4 Sample size
- 5 Margin of error
- 6 Sample & population size

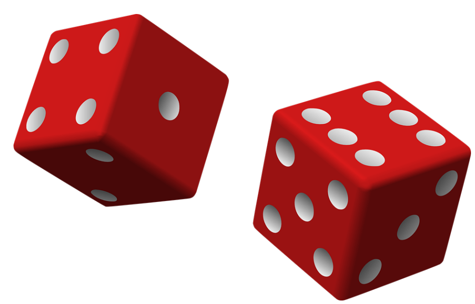
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
  - Lecture 1: Simple Random Sampling (SRS)
  - Lecture 2: A short history
  - Lecture 3: The SRS sampling distribution
  - Lecture 4: Sample size
  - Lecture 5: Margin of error
  - Lecture 6: Sample size & population size
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



Unit 2



- Basic framework
- Properties of the sampling distribution
- Confidence intervals



- **Basic framework**
- Properties of the sampling distribution
- Confidence intervals

- **A sample design for which the unit of selection is the population element**
- **Basic framework: Neyman, 1934**
  - **Must be applicable to all populations**
  - **Must not depend on assumptions about the population structure**
  - **Appropriate for large populations of elements**

- Basic framework
- Properties of the sampling distribution
- Confidence intervals

- **Repeated sampling**
  - **Objective (mechanical) selection of elements**
  - **Consider possible outcomes of the sampling process**
  - **Evaluation of the whole set of possible outcomes**

- Basic framework
- Properties of the sampling distribution
- Confidence intervals

- The set of all possible values of the estimator that can be obtained with a given sample design
  - For a given sample we obtain a particular value, the estimate (such as $\overline{y}$)
- We want to know …
  - … how likely is the estimate to be close to the population value?

- Basic framework
- Properties of the sampling distribution
- Confidence intervals

- In fact, we select just one sample
- The estimate may be correct, or incorrect
- Want to maximize the probability of a satisfactory estimate

- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

- **Unbiasedness**
  - **Expected value (average value):** $E\left(\bar{y}\right)$

- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

- **Unbiasedness**
  - **Expected value (average value):** $E\left(\bar{y}\right)$
  - **Meaning of expected value:**

$$E\left(\bar{y}\right)=\frac{1}{\binom{N}{n}}\sum_{s=1}^{\binom{N}{n}}\bar{y}_s$$

- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

- **Unbiasedness**
  - **Expected value (average value):** $E\left(\bar{y}\right)$
  - **Meaning of expected value:**

$$E\left(\bar{y}\right)=\frac{1}{\binom{N}{n}}\sum_{s=1}^{\binom{N}{n}}\bar{y}_s$$

  - **Meaning of unbiasedness:**

$$E\left(\bar{y}\right)=\frac{1}{\binom{N}{n}}\sum_{s=1}^{\binom{N}{n}}\bar{y}_s=\bar{Y}$$

Standard errors

For our SRS of n = 20,

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

$$=\frac{\left(1-\frac{20}{370}\right)}{20}766.62$$

$$=36.26$$

$$se\left(\bar{y}\right)=\sqrt{\mathrm{var}\left(\bar{y}\right)}=6.02$$

- **Variability from one sample to another**
  - **Variance of the estimator:** $Var(\bar{y})$

- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

- **Variability from one sample to another**
  - **Variance of the estimator:** $Var(\bar{y})$
  - **Meaning of the variance:**

$$Var(\bar{y}) = \frac{1}{\binom{N}{n}} \sum_{s=1}^{\binom{N}{n}} \left(\bar{y}_s - E(\bar{y})\right)^2$$

- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

- **Variability from one sample to another**
  - **Variance of the estimator:** $Var(\bar{y})$
  - **Meaning of the variance:**

$$Var(\bar{y}) = \frac{1}{\binom{N}{n}} \sum_{s=1}^{\binom{N}{n}} \left(\bar{y}_s - E(\bar{y})\right)^2$$

  - **Algebraically equivalent formula:**

$$Var(\bar{y}) = \frac{1}{\binom{N}{n}} \sum_{s=1}^{\binom{N}{n}} \left(\bar{y}_s - E(\bar{y})\right)^2 = \left(1 - \frac{n}{N}\right)\frac{S^2}{n}$$

- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

- **Variability from one sample to another**
  - **Components:**

$$S^2$$

$$\left(1-\frac{n}{N}\right)=\left(1-f\right)$$

$$\frac{1}{n}$$

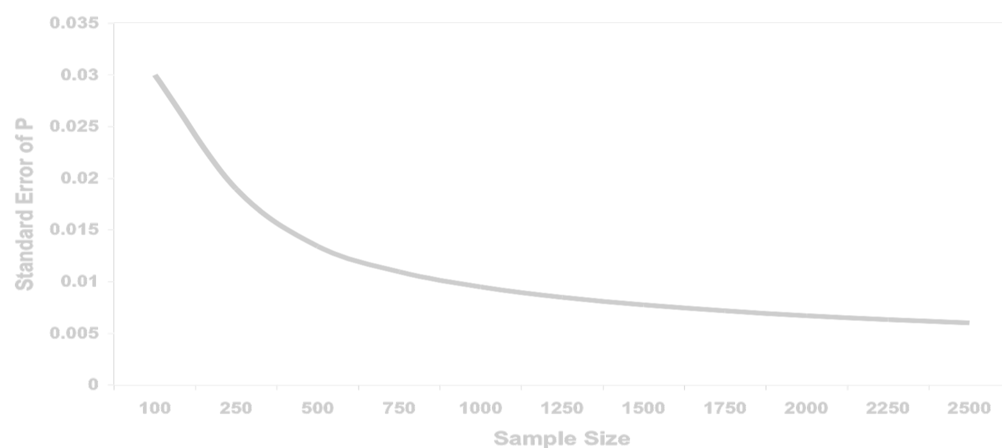
- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

- **Variability from one sample to another**
  - **Components:**

$$S^2$$

$$\left(1-\frac{n}{N}\right)=\left(1-f\right)$$

$$\frac{1}{n}$$



- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

- **Variability from one sample to another**
  - **Scale conversion:**

$$SE(\bar{y}) = \sqrt{Var(\bar{y})} = \sqrt{\left(1-\frac{n}{N}\right)\frac{S^2}{n}} = \frac{S}{\sqrt{n}}\sqrt{\left(1-\frac{n}{N}\right)}$$

- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

- **Estimating variability from one sample to another**
  - **Element variance:** S^2
  - **Estimated element variance:**

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \qquad \left(= \frac{n}{n-1}p\left(1-p\right)\right)$$

- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

- **Estimating variability from one sample to another**
  - **Element variance:** S^2
  - **Estimated element variance:**

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \qquad \left(= \frac{n}{n-1}p\left(1-p\right)\right)$$

  - **Estimated sampling variance & standard error**

$$var(\bar{y}) = \left(1-\frac{n}{N}\right)\frac{s^2}{n} \qquad \left(= \left(1-\frac{n}{N}\right)\frac{p\left(1-p\right)}{n-1}\right)$$

- Basic framework
- **Properties of the sampling distribution**
- Confidence intervals

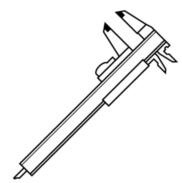
- **Estimating variability from one sample to another**
  - **Element variance:** S^2
  - **Estimated element variance:**

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \qquad \left(= \frac{n}{n-1} p\left(1-p\right) \right)$$

  - **Estimated sampling variance & standard error:**

$$var(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n} \qquad \left(= \left(1 - \frac{n}{N}\right)\frac{p\left(1-p\right)}{n-1} \right)$$

$$se\left(\bar{y}\right) = \frac{s}{\sqrt{n}}\sqrt{\left(1 - \frac{n}{N}\right)}$$



PRECISION

- Basic framework
- Properties of the sampling distribution
- Confidence intervals

- **For large samples, the sampling distribution of $\bar{y}$ is Normal**
  - **Law of large numbers or Central limit theorem**

- Basic framework
- Properties of the sampling distribution
- **Confidence intervals**

- **For large samples, the sampling distribution of $\bar{y}$ is Normal**
  - **Law of large numbers or Central limit theorem**
- **Form an interval around $\bar{y}$:**

$$\bar{y} \pm z_{(1-\alpha/2)} \times se\left(\bar{y}\right)$$

- Basic framework
- Properties of the sampling distribution
- Confidence intervals

- **For large samples, the sampling distribution of $\bar{y}$ is Normal**
  - **Law of large numbers or Central limit theorem**
- **Form an interval around $\bar{y}$:**

$$\bar{y} \pm 1.96 \times se\left(\bar{y}\right)$$

- $\left(1-\alpha\right)\% \; or \; 95\%$ **confidence interval**

- **A statement of uncertainty about our estimated mean**

95% Confidence interval

For our SRS of n = 20,

$$\bar{y} \pm t_{(1-\alpha/2, n-1)} se(\bar{y})$$

$$78.6 \pm t_{(0.975, 19)} \times 6.02$$

$$78.6 \pm 2.09 \times 6.02$$

$$(66.0, 98.2)$$

Unit 2

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
  - Lecture 1: Simple Random Sampling (SRS)
  - Lecture 2: A short history
  - Lecture 3: The SRS sampling distribution
  - **Lecture 4: Sample size**
  - Lecture 5: Margin of error
  - Lecture 6: Sample size & population size
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications