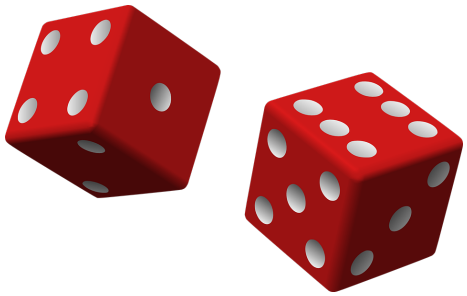
Sampling People, Records, & Networks

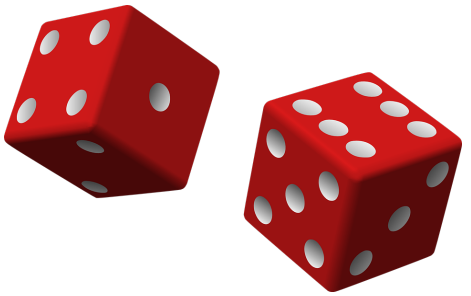
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



Unit 1

- 1 Research designs ...
- 2 Surveys
- 3 Why sample?
- 4 Why randomize?
- 5 Types of sampling
- 6 Evaluating samples
- 7 Units sampled

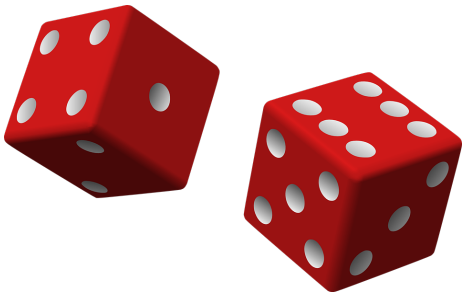
- **Unit 1: Sampling as a research tool**
  - Lecture 1 – Research design & sampling
  - Lecture 2 – Surveys & sampling
  - Lecture 3 -- Why sample at all?
  - Lecture 4 – Why might we randomize, and how do we do it?
  - Lecture 5 – What happens when we randomize?
  - Lecture 6 – How do we evaluate how good the sample is?
  - Lecture 7 – What kinds of things can we sample?
- **Unit 2: Mere randomization**
- **Unit 3: Saving money**
- **Unit 4: Being more efficient**
- **Unit 5: Simplifying sampling**
- **Unit 6: Some extensions & applications**



Unit 1

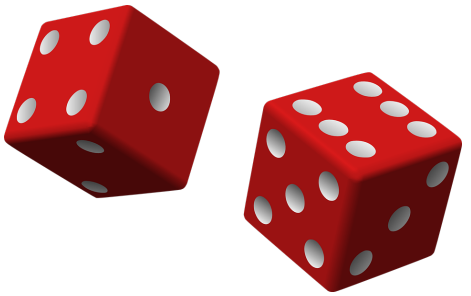
- 1 Research designs ...
- 2 Surveys
- 3 Why sample?
- 4 Why randomize?
- 5 What happens?
- **6 Evaluating samples**
- 7 Units sampled

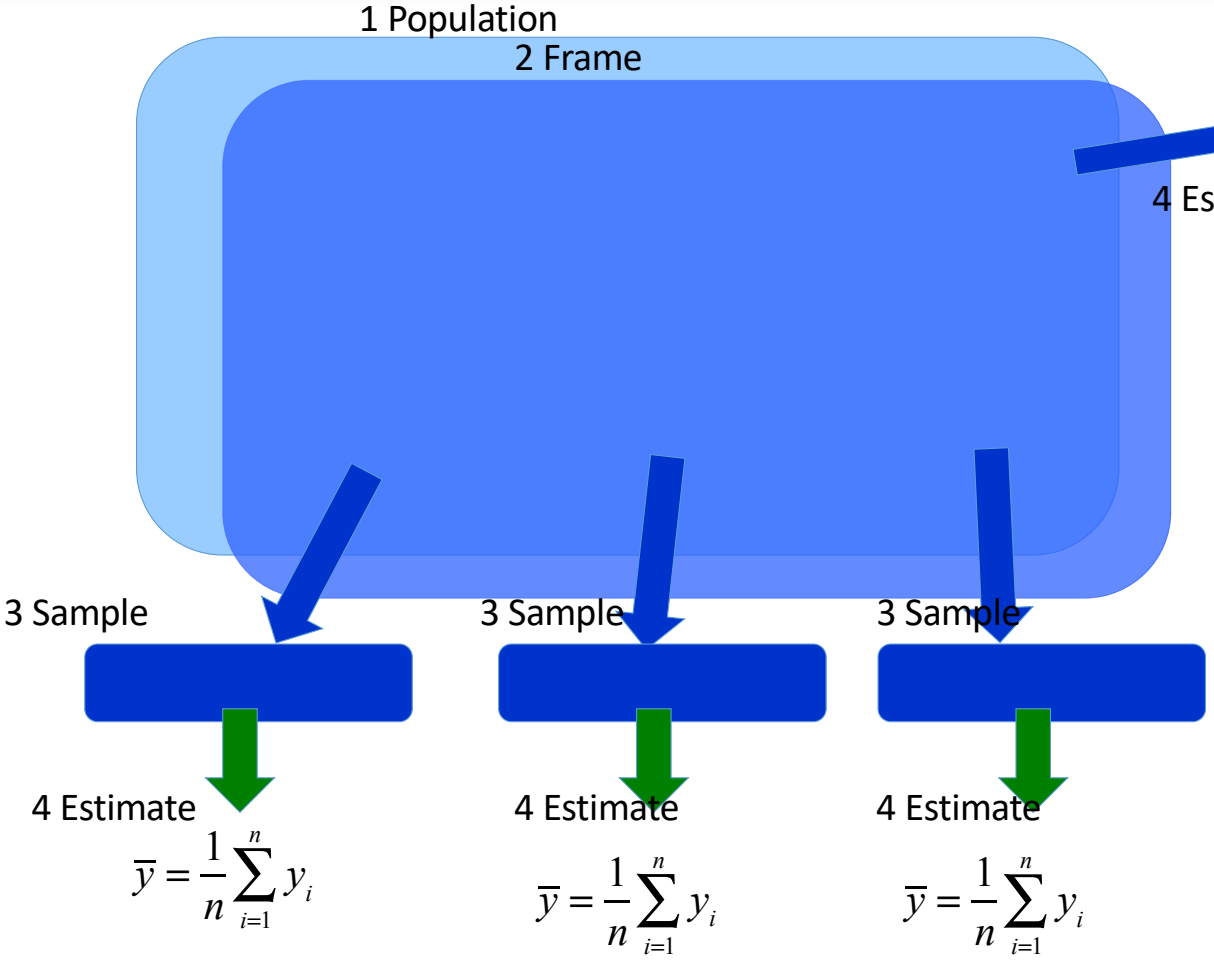
- **Unit 1: Sampling as a research tool**
  - Lecture 1 – Research design & sampling
  - Lecture 2 – Surveys & sampling
  - Lecture 3 -- Why sample at all?
  - Lecture 4 – Why might we randomize, and how do we do it?
  - Lecture 5 – What happens when we randomize?
  - **Lecture 6 – How do we evaluate how good the sample is?**
  - Lecture 7 – What kinds of things can we sample?
- **Unit 2: Mere randomization**
- **Unit 3: Saving money**
- **Unit 4: Being more efficient**
- **Unit 5: Simplifying sampling**
- **Unit 6: Some extensions & applications**



- Sampling distribution
- Standard error
- Confidence interval
- Two measures of data quality

- **Unit 1: Sampling as a research tool**
  - Lecture 1 – Research design & sampling
  - Lecture 2 – Surveys & sampling
  - Lecture 3 -- Why sample at all?
  - Lecture 4 – Why might we randomize, and how do we do it?
  - Lecture 5 – What happens when we randomize?
  - **Lecture 6 – How do we evaluate how good the sample is?**
  - Lecture 7 – What kinds of things can we sample?
- **Unit 2: Mere randomization**
- **Unit 3: Saving money**
- **Unit 4: Being more efficient**
- **Unit 5: Simplifying sampling**
- **Unit 6: Some extensions & applications**





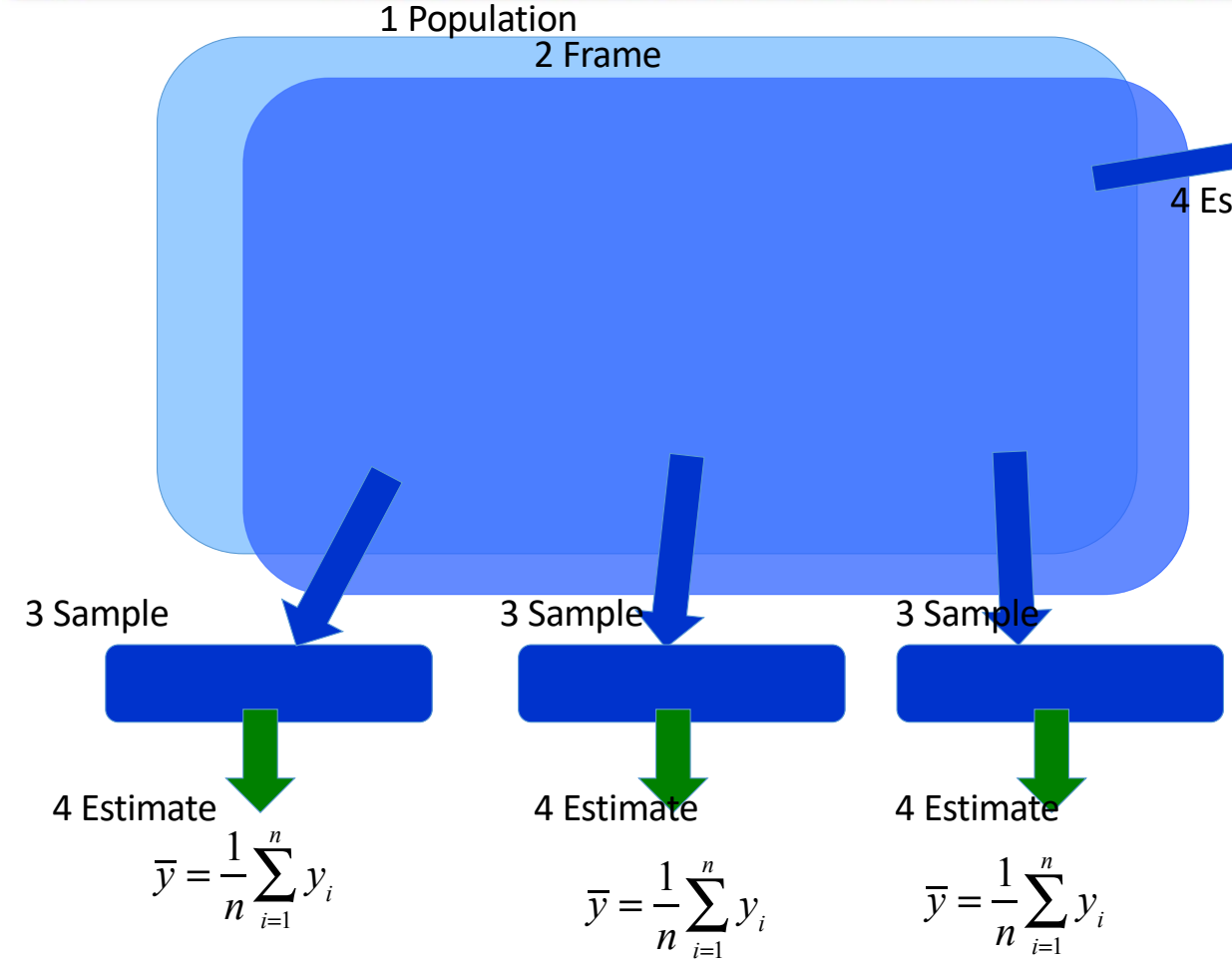
3 Sample

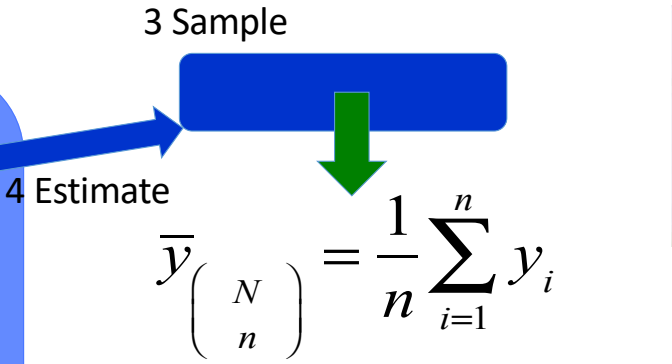
4 Estimate

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

5 Sampling distribution

- **Sampling distribution**
- Standard error
- Confidence interval
- Two measures of data quality


1 Population
2 Frame
3 Sample
3 Sample
3 Sample
4 Estimate
4 Estimate
4 Estimate
$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$


3 Sample
4 Estimate
$\bar{y}_{\binom{N}{n}} = \frac{1}{n}\sum_{i=1}^{n} y_i$
5 Sampling distribution
$Var(\bar{y}) = \frac{1}{\binom{N}{n}} \sum_{s=1}^{\binom{N}{n}} (\bar{y}_s - \bar{\bar{y}})^2$


s

1 Population

2 Frame

3 Sample

4 Estimate

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

3 Sample

4 Estimate

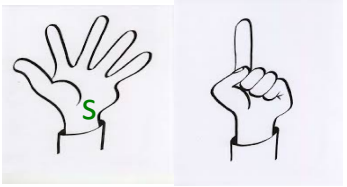
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

3 Sample

4 Estimate

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

3 Sample



4 Estimate

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

5 Sampling distribution

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

6 Standard error

$$se(\bar{y}) = \sqrt{\frac{1-f}{n}s^2}$$

Standard errors

For our SRS of n = 20,

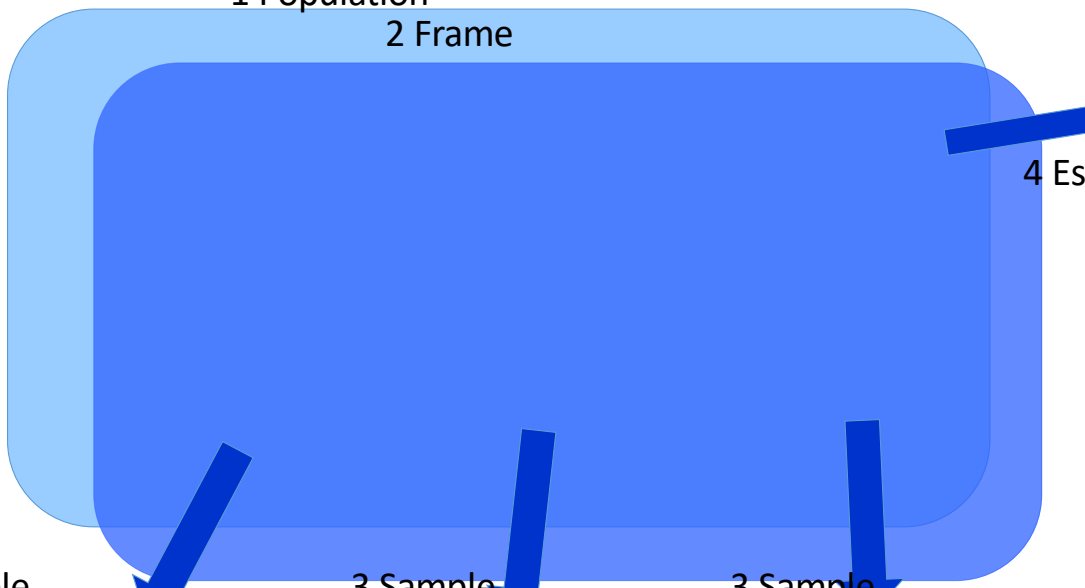
$$Var(\bar{y}) = \frac{(1-f)}{n} S^2$$

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

$$= 36.26$$

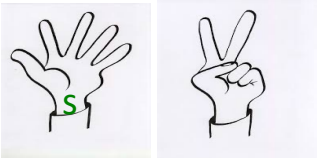
$$SE(\bar{y}) = \sqrt{V(\bar{y})} = 6.02$$

- Sampling distribution
- **Standard error**
- Confidence interval
- Two measures of data quality



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

Probability sampling principles



3 Sample

4 Estimate

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

5 Sampling distribution

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

6 Standard error

$$se(\bar{y}) = \sqrt{\frac{1-f}{n}s^2}$$

7 Confidence interval

$$\bar{y} \pm t_{(0.05,n-1)} \times se(\bar{y})$$

95% Confidence interval

For our SRS of n = 20,

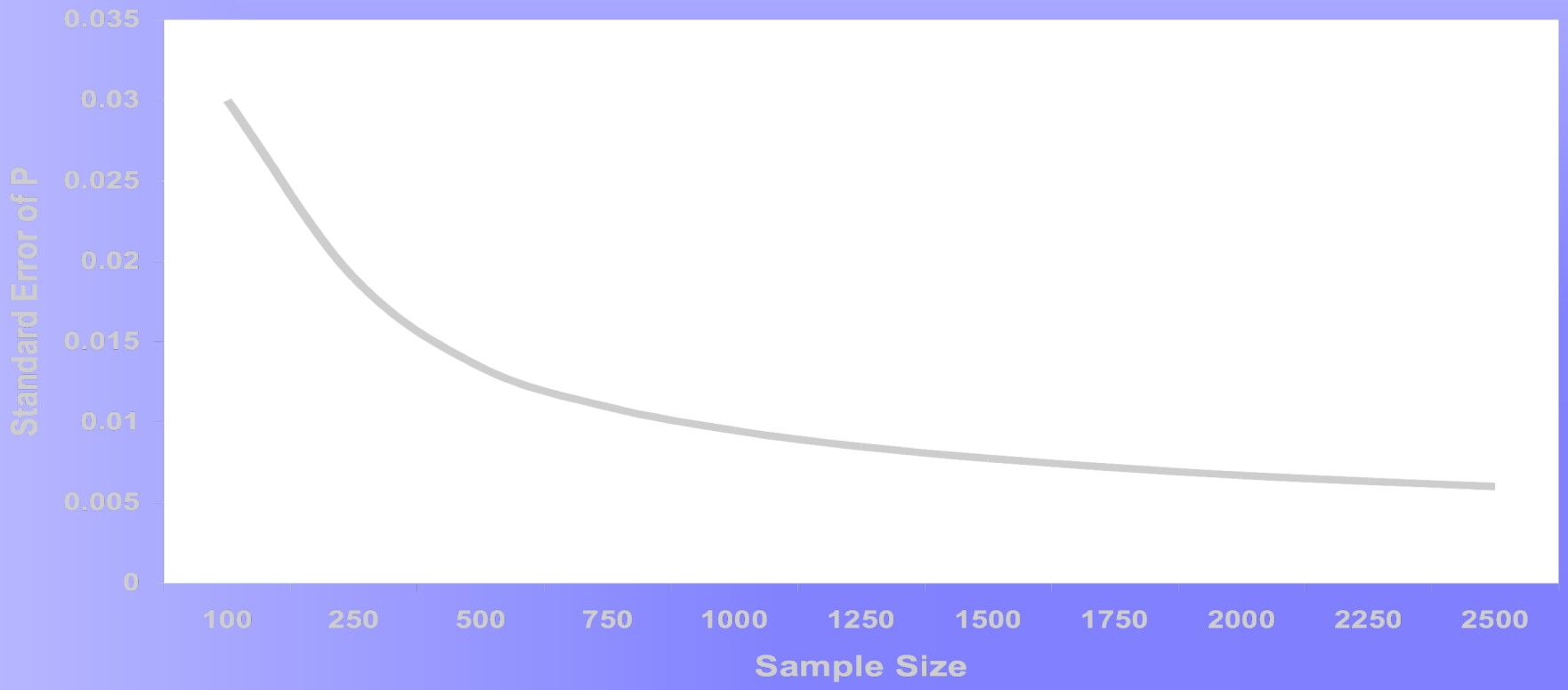
$$\bar{y} \pm t_{(1-\alpha/2,n-1)} SE\left(\bar{y}\right)$$

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

$$78.6 \pm 2.09 \times 6.02$$

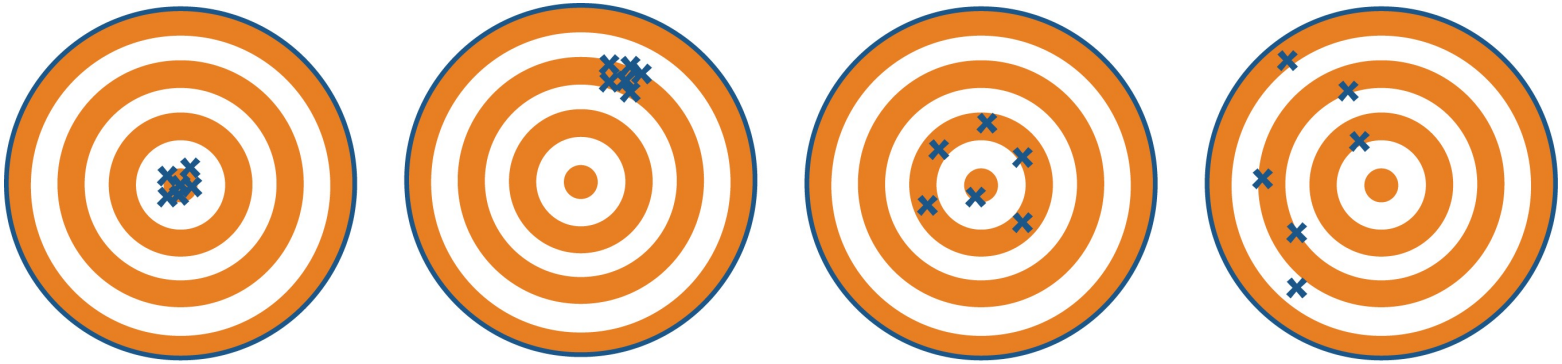
$$\left(66.0, 98.2\right)$$

- Sampling distribution
- Standard error
- **Confidence interval**
- Two measures of data quality


0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Standard Error of P
100
250
500
750
1000
1250
1500
1750
2000
2250
2500
Sample Size

- Sampling distribution
- Standard error
- Confidence interval
- **Two measures of data quality**

- **Two measures of sample quality**



High Accuracy
High Precision

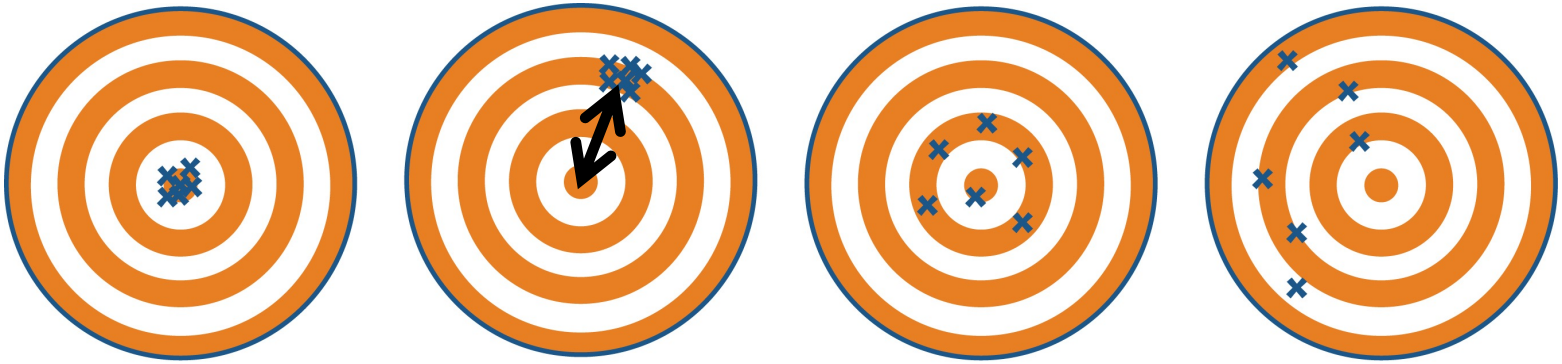
Low Accuracy
High Precision

High Accuracy
Low Precision

Low Accuracy
Low Precision

- Sampling distribution
- Standard error
- Confidence interval
- **Two measures of data quality**

- **Two measures of sample quality**
  - **Bias**



High Accuracy
High Precision

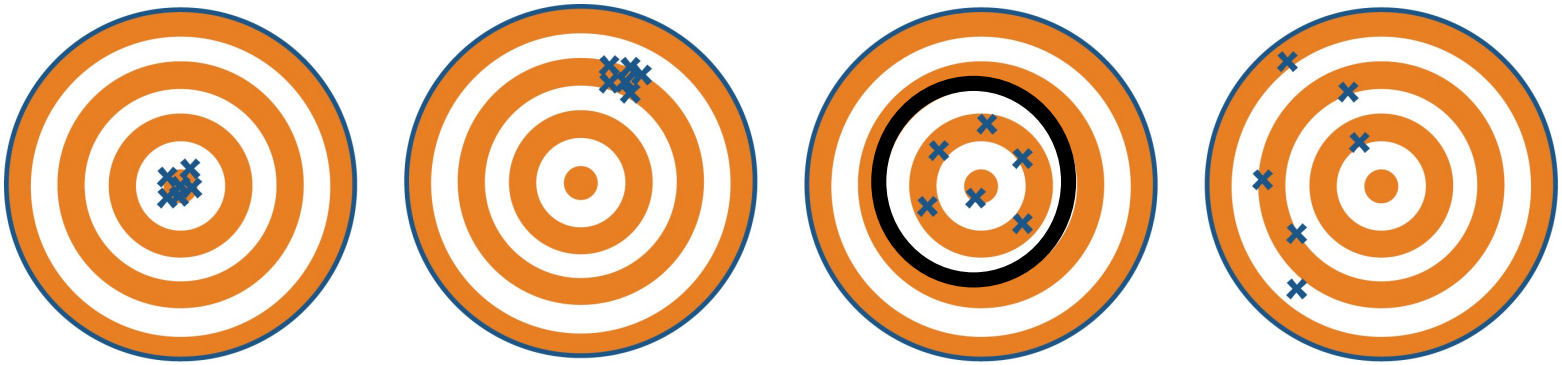
Low Accuracy
High Precision

High Accuracy
Low Precision

Low Accuracy
Low Precision

- Sampling distribution
- Standard error
- Confidence interval
- Two measures of data quality

- **Two measures of sample quality**
  - **Bias**
  - **Variance (standard error)**

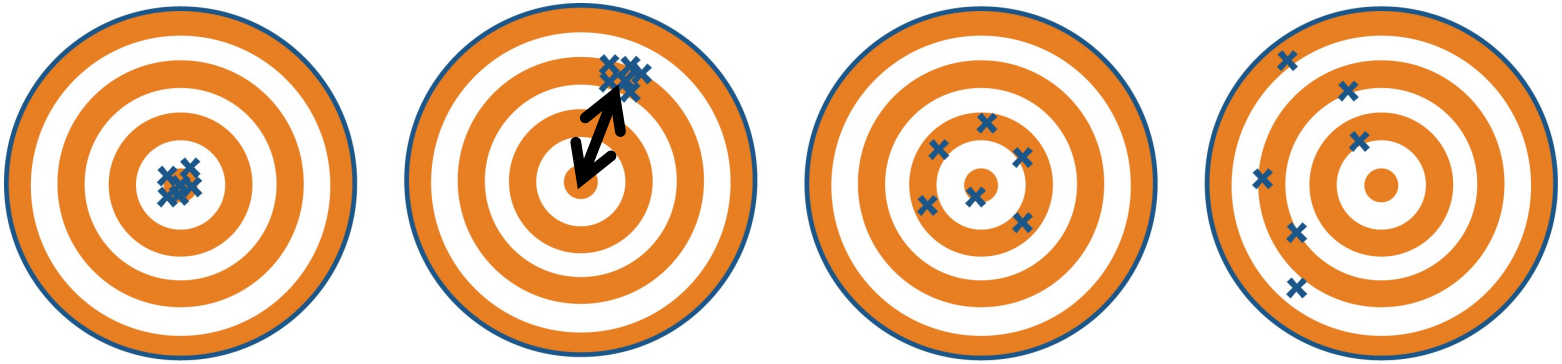


High Accuracy High Precision | **Low Accuracy High Precision** | **High Accuracy Low Precision** | **Low Accuracy Low Precision**

- Sampling distribution
- Standard error
- Confidence interval
- Two measures of data quality

• Two measures of sample quality

- Bias – we can determine theoretically if a sampling technique is unbiased
- Variance (standard error)



High Accuracy
High Precision

Low Accuracy
High Precision

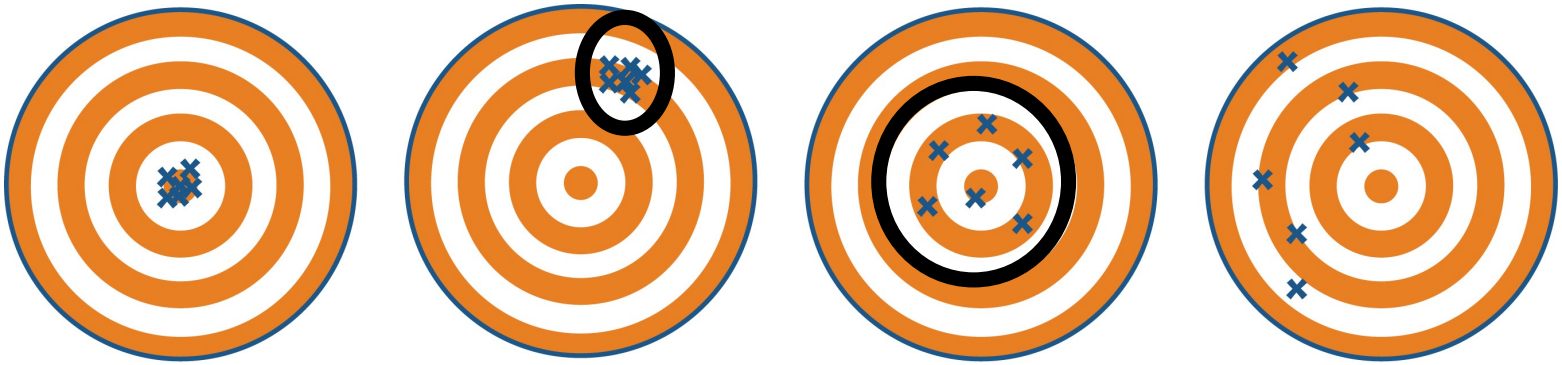
High Accuracy
Low Precision

Low Accuracy
Low Precision

- Sampling distribution
- Standard error
- Confidence interval
- Two measures of data quality

• Two measures of sample quality

- **Bias – we can determine theoretically if a sampling technique is unbiased**
- **Variance (standard error) – we can determine from sample data alone the size of the variance … to compare numerically**

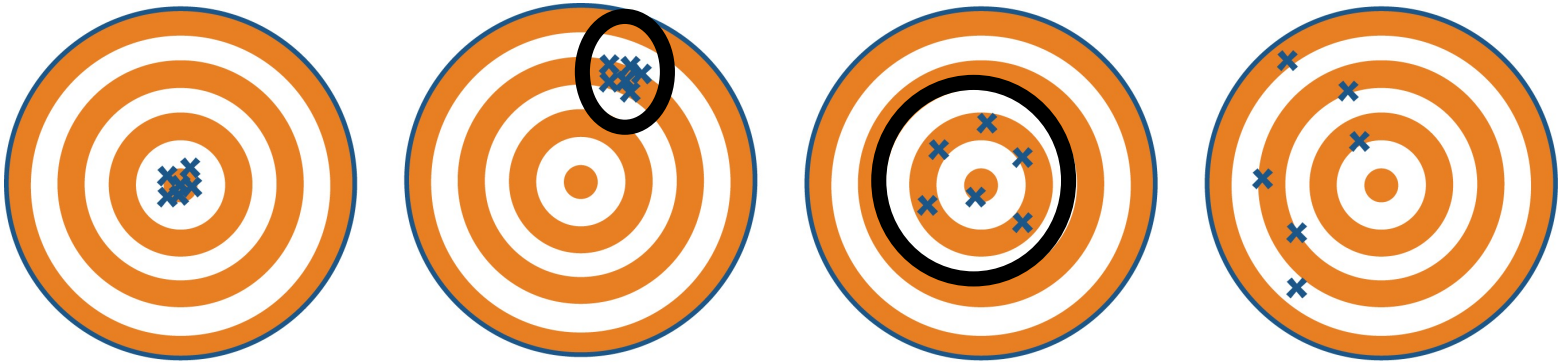


High Accuracy High Precision | **Low Accuracy High Precision** | **High Accuracy Low Precision** | **Low Accuracy Low Precision**

- Sampling distribution
- Standard error
- Confidence interval
- Two measures of data quality

• Two measures of sample quality

- Bias – we can determine theoretically if a sampling technique is unbiased
- Variance (standard error) – we can determine from sample data alone the size of the variance … to compare numerically
- It all depends on the random process to select the sample ...



High Accuracy
High Precision

Low Accuracy
High Precision

High Accuracy
Low Precision

Low Accuracy
Low Precision

- **Thus, two measures …**
  - **Bias**
  - **Variance**
- **And a random process …**
  - **Using random digits applied to a frame to generate, in theory, a large number of possible samples**
  - **And we can measure the variance across all possible samples from just a single randomly drawn sample**
  - **But only random samples allow us to do this without making any assumptions about either ...**
    - **The sampling mechanism**
    - **The population distribution**

Unit 1

- 1 Research designs ...
- 2 Surveys
- 3 Why sample?
- 4 Why randomize?
- 5 Types of sampling
- 6 Evaluating samples
- **7 Units sampled**

- **Unit 1: Sampling as a research tool**
  - Lecture 1 – Research design & sampling
  - Lecture 2 – Surveys & sampling
  - Lecture 3 -- Why sample at all?
  - Lecture 4 – Why might we randomize, and how do we do it?
  - Lecture 5 – What happens when we randomize?
  - Lecture 6 – How do we evaluate how good the sample is?
  - **Lecture 7 – What kinds of things can we sample?**
- **Unit 2: Mere randomization**
- **Unit 3: Saving money**
- **Unit 4: Being more efficient**
- **Unit 5: Simplifying sampling**
- **Unit 6: Some extensions & applications**