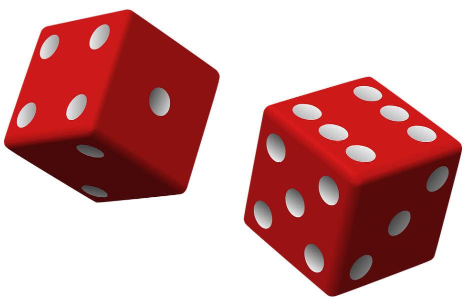
Sampling People, Records, & Networks

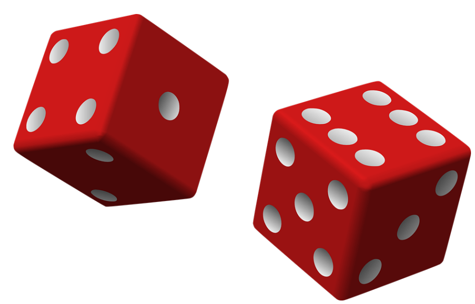
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



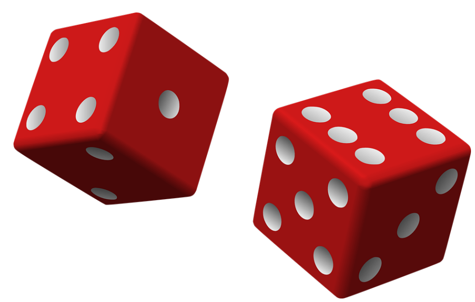
Unit 2

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

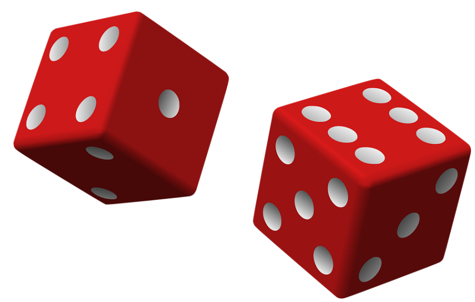
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
  - Lecture 1: Simple Random Sampling (SRS)
  - Lecture 2: A short history
  - Lecture 3: The SRS sampling distribution
  - Lecture 4: Sample size
  - Lecture 5: Margin of error
  - Lecture 6: Sample size & population size
- Unit 3: Saving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



Unit 2



- Sample size revisited
- Leadership approval rating
- China
- U.S.
- Ireland
- Seychelles
- Tuvalu
- The wrong idea



- **Sample size revisited**
- Leadership approval rating
- China
- U.S.
- Ireland
- Seychelles
- Tuvalu
- The wrong idea

- **Recall that sample size for a simple random sample can be calculated, for a margin of error *e*, as**

$$n = \frac{S^2}{\left(\frac{e}{2}\right)^2 + \frac{S^2}{N}}$$

- **For that in the previous example, when *e* = 0.02, for a 95% confidence interval on a proportion *p* = 0.60,**

$$n = \frac{0.24}{\left(\frac{0.02}{2}\right)^2 + \frac{0.24}{250{,}000{,}000}} = 2{,}399.97 = 2{,}400$$

- **Sample size revisited**
- Leadership approval rating
- China
- U.S.
- Ireland
- Seychelles
- Tuvalu
- The wrong idea

- **Recall that sample size for a simple random sample can be calculated, for a margin of error *e*, as**

$$n = \frac{S^2}{\left(\frac{e}{2}\right)^2 + \frac{S^2}{N}}$$

- **The denominator has a term that is the result of the finite population correction**

- **Sample size revisited**
- Leadership approval rating
- China
- U.S.
- Ireland
- Seychelles
- Tuvalu
- The wrong idea

- **Recall that sample size for a simple random sample can be calculated, for a margin of error *e*, as**

$$n = \frac{S^2}{\left(\frac{e}{2}\right)^2 + \frac{S^2}{N}}$$

- **The denominator has a term that is the result of the finite population correction**
- **That term made almost no difference in the example of the voting population in the U.S.**

- **Sample size revisited**
- Leadership approval rating
- China
- U.S.
- Ireland
- Seychelles
- Tuvalu
- The wrong idea

- **Recall that sample size for a simple random sample can be calculated, for a margin of error *e*, as**

$$n = \frac{S^2}{\left(\frac{e}{2}\right)^2 + \frac{S^2}{N}}$$

- **The denominator has a term that is the result of the finite population correction**
- **That term made almost no difference in the example of the voting population in the U.S.**
- **Does it ever make a difference?**

- **Suppose we're evaluating "presidential approval" or "leadership approval" across a number of countries**

- Sample size revisited
- **Leadership approval rating**
- China
- U.S.
- Ireland
- Seychelles
- Tuvalu
- The wrong idea

- **Suppose we're evaluating "presidential approval" across a number of countries and communities**
- **We're not sure what the approval rating will be in each**
  - **Use p = 0.50 or the largest value of $S^2 = p(1-p)$ possible**

- Suppose we're evaluating "presidential approval" across a number of countries and communities
- We're not sure what the approval rating will be in each
  - Use p = 0.50 or the largest value of $S^2 = p(1-p)$ possible
  - This may specify a sample size larger than needed in communities where p is not 0.50
  - In the absence of more precise information about the approval in a community, use the 'conservative' value of p = 0.50 and $S^2 = 0.25$

- Sample size revisited
- Leadership approval rating
- **China**
- U.S.
- Ireland
- Seychelles
- Tuvalu
- The wrong idea

What sample size is needed in China with *N* = 800,000,000 if for a 95% confidence interval *e* = 0.02

$$V_d = \left(\frac{e}{2}\right)^2 = \left(\frac{0.02}{2}\right)^2 = 0.01^2 = 0.0001?$$

- Sample size revisited
- Leadership approval rating
- **China**
- U.S.
- Ireland
- Seychelles
- Tuvalu
- The wrong idea

What sample size is needed in China with *N* = 800,000,000 if for a 95% confidence interval e = 0.02

$$V_d = \left(\frac{e}{2}\right)^2 = \left(\frac{0.02}{2}\right)^2 = 0.01^2 = 0.0001?$$

Calculate

$$n = \frac{S^2}{\left(\frac{e}{2}\right)^2 + \frac{S^2}{N}} = \frac{0.25}{\left(\frac{0.02}{2}\right)^2 + \frac{0.25}{800,000,000}} = 2,499.99 = 2,500$$

What about "approval" in the U.S. with N = 250,000,000 and a 95% confidence interval with e = 0.02?

- Sample size revisited
- Leadership approval rating
- China
- U.S.
- Ireland
- Seychelles
- Tuvalu
- The wrong idea

What about "approval" in the U.S. with *N* = 250,000,000 and a 95% confidence interval with e = 0.02?

$$n = \frac{S^2}{\left(\frac{e}{2}\right)^2 + \frac{S^2}{N}} = \frac{0.25}{\left(\frac{0.02}{2}\right)^2 + \frac{0.25}{250,000,000}} = 2,499.97 = 2,500$$

- Sample size revisited
- Leadership approval rating
- China
- U.S.
- **Ireland**
- Seychelles
- Tuvalu
- The wrong idea

What about "approval" in Ireland with N = 3,200,000?

- Sample size revisited
- Leadership approval rating
- China
- U.S.
- **Ireland**
- Seychelles
- Tuvalu
- The wrong idea

What about "approval" in Ireland with N = 3,200,000?

$$n = \frac{S^2}{\left(\frac{e}{2}\right)^2 + \frac{S^2}{N}} = \frac{0.25}{\left(\frac{0.02}{2}\right)^2 + \frac{0.25}{3{,}200{,}000}} = 2{,}498.04 = 2{,}500$$

- Sample size revisited
- Leadership approval rating
- China
- U.S.
- Ireland
- **Seychelles**
- Tuvalu
- The wrong idea

What about "approval" in the Seychelles with N = 80,000?

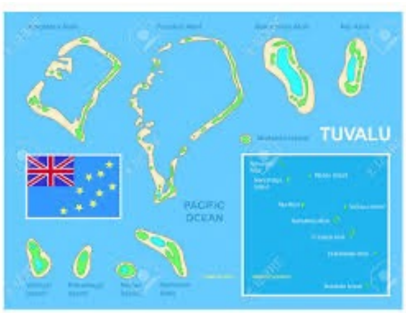
- Sample size revisited
- Leadership approval rating
- China
- U.S.
- Ireland
- **Seychelles**
- Tuvalu
- The wrong idea

What about "approval" in the Seychelles with *N* = 80,000?

$$n = \frac{S^2}{\left(\frac{e}{2}\right)^2 + \frac{S^2}{N}} = \frac{0.25}{\left(\frac{0.02}{2}\right)^2 + \frac{0.25}{80{,}000}} = 2{,}424.24 = 2{,}425$$

- Sample size revisited
- Leadership approval rating
- China
- U.S.
- Ireland
- Seychelles
- **Tuvalu**
- The wrong idea

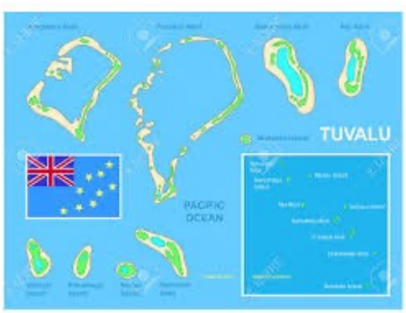
What about "approval" in Tuvalu with N = 8,000?



- Sample size revisited
- Leadership approval rating
- China
- U.S.
- Ireland
- Seychelles
- **Tuvalu**
- The wrong idea

What about "approval" in Tuvalu with *N* = 8,000?

$$n = \frac{S^2}{\left(\frac{e}{2}\right)^2 + \frac{S^2}{N}} = \frac{0.25}{\left(\frac{0.02}{2}\right)^2 + \frac{0.25}{8{,}000}} = 1{,}904.76 = 1{,}905$$



- Sample size revisited
- Leadership approval rating
- China
- U.S.
- Ireland
- Seychelles
- Tuvalu
- **The wrong idea**

Sample size depends on population size, but not in an expected way

It is clearly not proportional:

China	N = 800,000,000	n = 2,500
USA	N = 250,000,000	n = 2,500
Ireland	N = 4,000,000	n = 2,500
Seychelles	N = 80,000	n = 2,424
Tuvalu	N = 8,000	n = 1,904

- Why do we bring this up?
- Because there are textbooks out there that claim sample size should be a fraction of the population, say 10%
  - Thus, the larger the population the larger the sample size
  - Directly proportional

$$f = \frac{n}{N},$$

- Sample size revisited
- Leadership approval rating
- China
- U.S.
- Ireland
- Seychelles
- Tuvalu
- The wrong idea

- Why do we bring this up?
- Because there are textbooks out there that claim sample size should be a fraction of the population, say 10%
  - Thus, the larger the population the larger the sample size
  - Directly proportional
- This is a "common sense" misperception

"How can a sample of only 800 represent the voting public of 250,000,000 in the U.S.?"

- The constant fraction sample size clearly misleads:
  - Studies can't have a relatively small sample size to get any useful results for a large country …

$$f = \frac{n}{N},$$

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