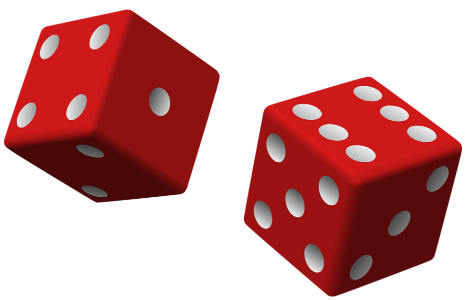
Sampling People, Records, & Networks

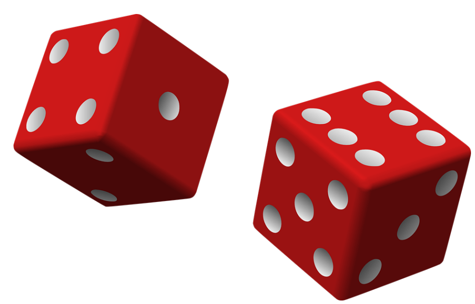
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



Unit 2

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

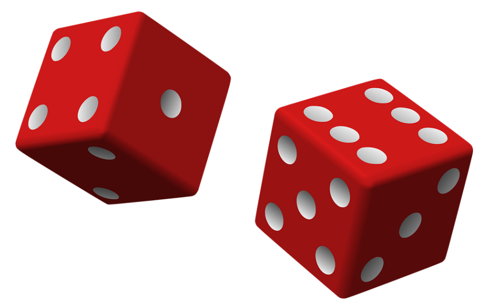
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
  - Lecture 1: Simple Random Sampling (SRS)
  - Lecture 2: A short history
  - Lecture 3: The SRS sampling distribution
  - Lecture 4: Sample size
  - Lecture 5: Margin of error
  - Lecture 6: Sample size & population size
- Unit 3: Saving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



Unit 2

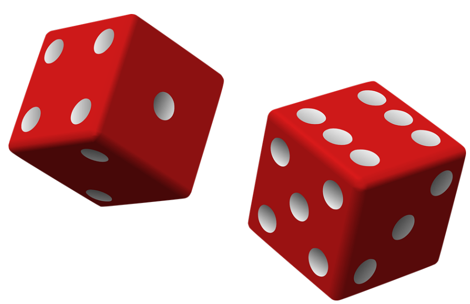
- 1 Simple random sampling
- 2 History
- 3 Sampling distribution
- **4 Sample size**
- 5 Margin of error
- 6 Sample & population size

- 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



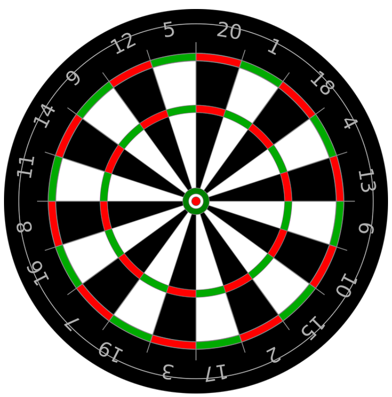
- What we need to know
- Sample size formula
- An example

- 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



- What we need to know
- Sample size formula
- An example

- What sample size do we need to obtain a given standard error of the estimator?
- S^2 population variance known (or guessed)
  - Census
  - Other surveys
  - Administrative records
- Desired standard error
  - Policy requirements in terms of $\sqrt{Var(\overline{y})}$
  - Decision making requirements



- What we need to know
- **Sample size formula**
- An example

- **From previous lecture,** $Var(\bar{y}) = \left(1 - \dfrac{n}{N}\right)\dfrac{S^2}{n}$
- **For an infinitely large population (or for sampling with replacement), this is**

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

- **We can calculate the necessary sample size to achieve** **desired** **variance** $V_d = Var\left(\bar{y}\right)$ **as**

$$n = S^2 / V_d$$

- What we need to know
- **Sample size formula**
- An example

- **In general (that is, not assuming *N* is large), the variance may be expressed as**

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

where

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

- What we need to know
- **Sample size formula**
- An example

- **Let's call n' the necessary sample size –**

$$n' = \frac{S^2}{V_d}$$

- **To calculate the actual *n* needed for a population of a particular size, we adjust --**

$$n = \frac{n'}{1 + \frac{n'}{N}}$$

- What we need to know
- Sample size formula
- An example

- Interested in U.S. population attitudes about how well its current President is doing his or her job:

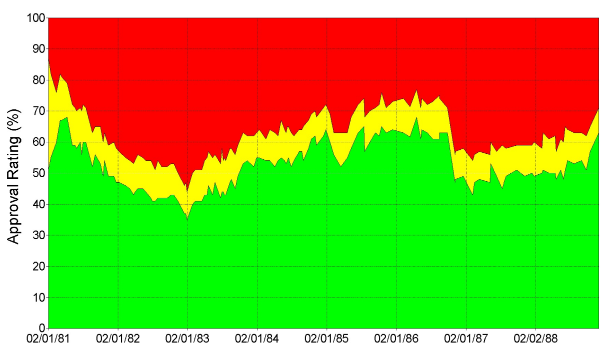
  "Do you approve or disapprove of the job Barrack Obama is doing as President?" (If approve/disapprove, ask:) "Do you approve/disapprove strongly or somewhat?"

- Estimate the proportion *P* approving strongly or somewhat in a new survey

THE PRESIDENT'S SEAL

- What we need to know
- Sample size formula
- An example

- **Suppose *p* = 0.6 in the last survey**
  - **Then** $s^2 = p(1-p) = 0.6(1-0.6) = 0.24$
- **For our new survey about to be conducted, "project" that** $S^2 = 0.24$

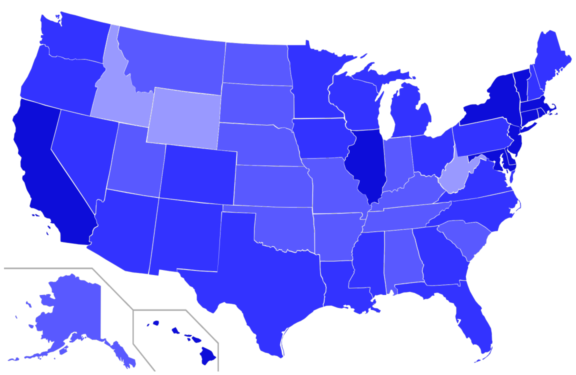


- What we need to know
- Sample size formula
- An example

- **Also need to specify precision of the new survey estimate … in advance … the** $V_d = Var\left(\bar{y}\right)$
- **Let's work backwards …**

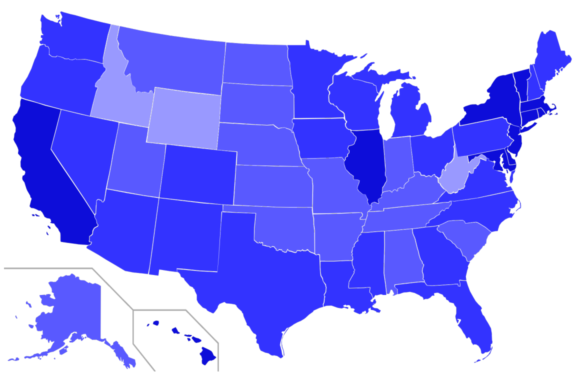
- What we need to know
- Sample size formula
- An example

- **Suppose we would like to end up with an uncertainty statement that says that between 58% and 62% of the U.S. population think President Obama is doing a good job … at a 95% level of confidence.**



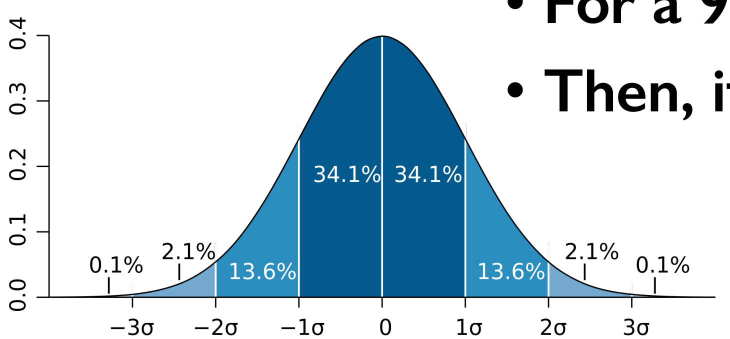
- What we need to know
- Sample size formula
- An example

- **Suppose we would like to end up with an uncertainty statement that says that between 58% and 62% of the U.S. population think President Obama is doing a good job ... at a 95% level of confidence.**
- **Recall that the upper confidence limit, the 62% value, is the proportion, 60% in this case, plus a multiplier times the standard error.**



- What we need to know
- Sample size formula
- An example

- **Suppose we would like to end up with an uncertainty statement that says that between 58% and 62% of the U.S. population think President Obama is doing a good job ... at a 95% level of confidence.**
- **Recall that the upper confidence limit, the 62% value, is the proportion, 60% in this case, plus a multiplier times the standard error.**
- **That is, 62% = 60% + z x se(60%)**
- **For a 95% confidence interval, z = 1.96, say, z = 2**
- **Then, if 62% = 60% + 2 x se(60%), se(60%) = 1%**



- **If that's the kind of confidence interval we want, then we want a standard error of 1%**
- **Of course, se(p) = 0.01 is another way to say this, in terms of what we want to have happen**

- If that's the kind of confidence interval we want, then we want a standard error of 1%
- Of course, se(p) = 0.01 is another way to say this, in terms of what we want to have happen
- Proportions are better to work with than percentages
- We need the square of the standard error, or the variance:

$$V_d = Var(p) = (SE(p))^2$$

$$\frac{1}{4} = 0.25 = 25\%$$

- What we need to know
- Sample size formula
- An example

- If that's the kind of confidence interval we want, then we want a standard error of 1%
- Of course, se(p) = 0.01 is another way to say this, in terms of what we want to have happen
- Proportions are better to work with than percentages
- We need the square of the standard error, or the variance:

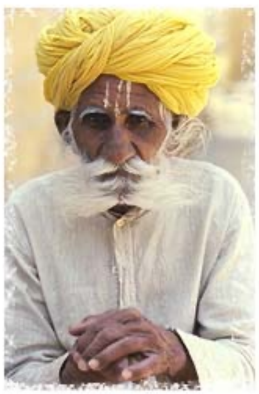
$$V_d = Var(p) = \left(SE(p)\right)^2$$

- That is, $V_d = (0.01)^2 = 0.0001$

- What we need to know
- Sample size formula
- **An example**

- **Hence, we have $S^2 = 0.24$ and $V_d = 0.0001$**

- What we need to know
- Sample size formula
- **An example**

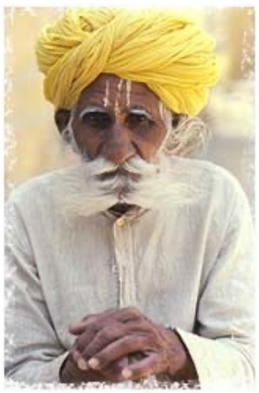


A wise man once said: "Never begin data collection without calculating the necessary sample size!" (See story left.)

- **Hence, we have $S^2 = 0.24$ and $V_d = 0.0001$**
- **This yields a necessary sample size of**

$$n' = \frac{S^2}{V_d} = \frac{0.24}{0.0001} = 2{,}400$$

- What we need to know
- Sample size formula
- **An example**



A wise man once said: "Never begin data collection without calculating the necessary sample size!" (See story left.)

- **Hence, we have** $S^2 = 0.24$ **and** $V_d = 0.0001$
- **This yields a necessary sample size of**

$$n' = \frac{S^2}{V_d} = \frac{0.24}{0.0001} = 2{,}400$$

- **Adjustment for the finite population:**

$$n = \frac{n'}{1 + \frac{n'}{N}} = \frac{2{,}400}{1 + \frac{2{,}400}{250{,}000{,}000}} = 2{,}399.97 = 2{,}400$$

- What we need to know
- Sample size formula
- An example

- **There are two questions we'll address in the next lectures related to sample size ...**

Is there a more direct way to figure this out from a projected confidence interval?

Why doesn't the population size have a big effect on the sample size?

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