Congressional Approval Ratings

- Rasmussen Polls: 9% approve of Congress' performance

Congressional Approval Ratings

- Rasmussen Polls: 9% approve of Congress' performance
  - How did they obtain this number?

Congressional Approval Ratings

- Rasmussen Polls: 9% approve of Congress' performance
  - How did they obtain this number?
  - How do they know that its accurate?

Congressional Approval Ratings

- Rasmussen Polls: 9% approve of Congress' performance
  - How did they obtain this number?
  - How do they know that its accurate?
- There is some *true proportion*, p, that approves of Congress' performance

Congressional Approval Ratings

- Rasmussen Polls: 9% approve of Congress' performance
  - How did they obtain this number?
  - How do they know that its accurate?
- There is some *true proportion*, p, that approves of Congress' performance
  - We don't know p (we want to find it)

Congressional Approval Ratings

- Rasmussen Polls: 9% approve of Congress' performance
  - How did they obtain this number?
  - How do they know that its accurate?
- There is some *true proportion*, p, that approves of Congress' performance
  - We don't know p (we want to find it)
- Asking a random person if they approve of Congress is a Bernoulli trial

Congressional Approval Ratings

- Rasmussen Polls: 9% approve of Congress' performance
  - How did they obtain this number?
  - How do they know that its accurate?
- There is some *true proportion*, p, that approves of Congress' performance
  - We don't know p (we want to find it)
- Asking a random person if they approve of Congress is a Bernoulli trial
  - "Do you approve of Congress?" is a yes/no question

Polling

- Ask n randomly selected people "do you approve of Congress"?

Polling

- Ask n randomly selected people "do you approve of Congress"?
  - Say x of them answer "yes"

Polling

- Ask n randomly selected people "do you approve of Congress"?
  - Say x of them answer "yes"
  - Define $\hat{p} = \dfrac{x}{n}$

Polling

- Ask n randomly selected people "do you approve of Congress"?
  - Say x of them answer "yes"
  - Define $\hat{p} = \dfrac{x}{n}$
    - $\hat{p}$ is our estimate for p

Polling

- Ask n randomly selected people "do you approve of Congress"?
  - Say x of them answer "yes"
  - Define $\hat{p} = \dfrac{x}{n}$
    - $\hat{p}$ is our estimate for p
    - Note that we are asking a different question than in previous lectures
      (fraction successes vs. number of successes)

Polling

- Ask n randomly selected people "do you approve of Congress"?
  - Say x of them answer "yes"
  - Define $\hat{p} = \dfrac{x}{n}$
    - $\hat{p}$ is our estimate for p
    - Note that we are asking a different question than in previous lectures
      (fraction successes vs. number of successes)
  - $\hat{p}$ will follow a normal distribution with:

Polling

- Ask n randomly selected people "do you approve of Congress"?
    - Say x of them answer "yes"
    - Define $\hat{p} = \dfrac{x}{n}$
        - $\hat{p}$ is our estimate for p
        - Note that we are asking a different question than in previous lectures
        (fraction successes vs. number of successes)
    - $\hat{p}$ will follow a normal distribution with:
        - $\mu = p$
        - $\sigma = \sqrt{\dfrac{p(1-p)}{n}}$

Polling

- Suppose the true proportion of voters that approve of Congress is 10%

Polling

- Suppose the true proportion of voters that approve of Congress is 10%
- Design a survey that randomly polls 100 people

Polling

- Suppose the true proportion of voters that approve of Congress is 10%
- Design a survey that randomly polls 100 people
  - What is σ?

Polling

- Suppose the true proportion of voters that approve of Congress is 10%
- Design a survey that randomly polls 100 people
    - What is σ?
    - How often will $\hat{p}$ lie between .07 and .13?

Polling Accuracy

- What will happen as n increases?

Polling Accuracy

- What will happen as n increases?
- σ for various n ($p = .1$):

n	100	200	400	800	1600	3200
σ	.0300	.0212	.0150	.0106	.0075	.0053

Polling Accuracy

- What will happen as *n* increases?
- σ for various n ($p = .1$):

n	100	200	400	800	1600	3200
σ	.0300	.0212	.0150	.0106	.0075	.0053

- To cut σ in half, what must you do to *n*?

Polling Accuracy

- What will happen as n increases?
- σ for various n ($p = .1$):

n	100	200	400	800	1600	3200
σ	.0300	.0212	.0150	.0106	.0075	.0053

- To cut σ in half, what must you do to n?
- If you randomly poll 3200 people, how often will will $\hat{p}$ lie between .09 and .11?

Polling Accuracy

- What will happen as n increases?
- σ for various n ($p = .1$):

n	100	200	400	800	1600	3200
σ	.0300	.0212	.0150	.0106	.0075	.0053

- To cut σ in half, what must you do to n?
- If you randomly poll 3200 people, how often will will $\hat{p}$ lie between .09 and .11?
- If the U.S. population suddenly doubled, how accurate would randomly polling 3200 people be?

Polling Accuracy

- What will happen as *n* increases?
- σ for various n ($p = .1$):

n	100	200	400	800	1600	3200
σ	.0300	.0212	.0150	.0106	.0075	.0053

- To cut σ in half, what must you do to *n*?
- If you randomly poll 3200 people, how often will will $\hat{p}$ lie between .09 and .11?
- If the U.S. population suddenly doubled, how accurate would randomly polling 3200 people be?
- In polling, standard deviation is also known as *sampling error*

Polling Accuracy

- If we poll 3200 people, what is the probability that $\hat{p} > .12$?

Polling Accuracy

- If we poll 3200 people, what is the probability that $\hat{p} > .12$?
  - Need to know to know the area to the right of 4 standard deviations

Polling Accuracy

- If we poll 3200 people, what is the probability that $\hat{p} > .12$?
  - Need to know to know the area to the right of 4 standard deviations
    - $\approx .00003$

Polling Accuracy

- If we poll 3200 people, what is the probability that $\hat{p} > .12$?
  - Need to know to know the area to the right of 4 standard deviations
    - $\approx .00003$
  - This could happen due to random circumstances

Polling Accuracy

- If we poll 3200 people, what is the probability that $\hat{p} > .12$?
  - Need to know to know the area to the right of 4 standard deviations
    - $\approx .00003$
  - This could happen due to random circumstances
  - More likely the case that the survey wasn't random

Conducting a Survey

- How should one go about conducting a survey?

Conducting a Survey

- How should one go about conducting a survey?
- A **simple random survey** is:

Conducting a Survey

- How should one go about conducting a survey?
- A **simple random survey** is:
  - unbiased (each individual has the same chance of being polled)

Conducting a Survey

- How should one go about conducting a survey?
- A **simple random survey** is:
  - unbiased (each individual has the same chance of being polled)
  - independent (selection of individual X has no bearing on the selection of individual Y)

Conducting a Survey

- How should one go about conducting a survey?
- A **simple random survey** is:
  - unbiased (each individual has the same chance of being polled)
  - independent (selection of individual X has no bearing on the selection of individual Y)
- We are trying to make sure that asking the question is a Bernoulli trial

Opportunity Bias

- Suppose we ask the first n people who come along.

Opportunity Bias

- Suppose we ask the first n people who come along.
- Problems?

Opportunity Bias

- Suppose we ask the first *n* people who come along.
- Problems?
  - not everyone is equally likely

Opportunity Bias

- Suppose we ask the first n people who come along.
- Problems?
    - not everyone is equally likely
    - not independent: people in groups

Sampling Bias

- Suppose we mail out surveys
  (or randomly phone people)

Sampling Bias

- Suppose we mail out surveys (or randomly phone people)
- Problems?

Sampling Bias

- Suppose we mail out surveys
  (or randomly phone people)
- Problems?
  - some people will choose not to respond

Sampling Bias

- Suppose we mail out surveys
  (or randomly phone people)
- Problems?
  - some people will choose not to respond
  - not everyone is equally likely

Sampling Bias

- Suppose we mail out surveys
  (or randomly phone people)
- Problems?
  - some people will choose not to respond
  - not everyone is equally likely
- **Sampling bias** is a bias where some people are less likely to be polled than others

Roosevelt/Landon Election

- Two polls during the 1936 presidential election:

Roosevelt/Landon Election

- Two polls during the 1936 presidential election:
  - The American Literary Digest collected 2, 000, 000 postal surveys

Roosevelt/Landon Election

- Two polls during the 1936 presidential election:
    - The American Literary Digest collected 2, 000, 000 postal surveys
        - Concluded that Landon would win by a landslide

Roosevelt/Landon Election

- Two polls during the 1936 presidential election:
  - The American Literary Digest collected 2,000,000 postal surveys
    - Concluded that Landon would win by a landslide
    - Collected addresses from readers of the magazine and registered automobile owners

Roosevelt/Landon Election

- Two polls during the 1936 presidential election:
  - The American Literary Digest collected 2, 000, 000 postal surveys
    - Concluded that Landon would win by a landslide
    - Collected addresses from readers of the magazine and registered automobile owners
    - Bias towards car owners (rich)

Roosevelt/Landon Election

- Two polls during the 1936 presidential election:
    - The American Literary Digest collected $2,000,000$ postal surveys
        - Concluded that Landon would win by a landslide
        - Collected addresses from readers of the magazine and registered automobile owners
        - Bias towards car owners (rich)
    - Gallup Poll selected $50,000$ people using U.S. Census data, and physically tracked them down

Roosevelt/Landon Election

- Two polls during the 1936 presidential election:
  - The American Literary Digest collected $2,000,000$ postal surveys
    - Concluded that Landon would win by a landslide
    - Collected addresses from readers of the magazine and registered automobile owners
    - Bias towards car owners (rich)
  - Gallup Poll selected $50,000$ people using U.S. Census data, and physically tracked them down
    - Concluded that Roosevelt would win by a landslide

Dewey/Truman Election

- For the 1948 election, the Chicago Tribune wanted to go to press early

Dewey/Truman Election

- For the 1948 election, the Chicago Tribune wanted to go to press early
- Trusted a phone survey claiming that Dewey would win by a landslide

Dewey/Truman Election

- For the 1948 election, the Chicago Tribune wanted to go to press early
- Trusted a phone survey claiming that Dewey would win by a landslide
- Problems?

Dewey/Truman Election

- For the 1948 election, the Chicago Tribune wanted to go to press early
- Trusted a phone survey claiming that Dewey would win by a landslide
- Problems?
    - Telephones were not widespread

Dewey/Truman Election

- For the 1948 election, the Chicago Tribune wanted to go to press early
- Trusted a phone survey claiming that Dewey would win by a landslide
- Problems?
  - Telephones were not widespread
  - Bias towards the rich and urban-dwellers

Rasmussen Polls

- During 2012 election, Gallup and Rasmussen ran polls

Rasmussen Polls

- During 2012 election, Gallup and Rasmussen ran polls
  - Only called landlines

Rasmussen Polls

- During 2012 election, Gallup and Rasmussen ran polls
  - Only called landlines
  - Rasmussen did not reattempt a missed call

Rasmussen Polls

- During 2012 election, Gallup and Rasmussen ran polls
  - Only called landlines
  - Rasmussen did not reattempt a missed call
  - Gallup reattempted missed calls

Rasmussen Polls

- During 2012 election, Gallup and Rasmussen ran polls
  - Only called landlines
  - Rasmussen did not reattempt a missed call
  - Gallup reattempted missed calls
- Problems?

Rasmussen Polls

- During 2012 election, Gallup and Rasmussen ran polls
  - Only called landlines
  - Rasmussen did not reattempt a missed call
  - Gallup reattempted missed calls
- Problems?
- Both companies had skewed results

Dealing with Bias

- After an election, we know the bias for every poll

Dealing with Bias

- After an election, we know the bias for every poll
- FiveThirtyEight's model:

Dealing with Bias

- After an election, we know the bias for every poll
- FiveThirtyEight's model:
    - Use previous elections to determine a "bias factor"

Dealing with Bias

- After an election, we know the bias for every poll
- FiveThirtyEight's model:
  - Use previous elections to determine a "bias factor"
  - Take weighted average of all polls
  - (count more recent polls more strongly)

Dealing with Bias

- After an election, we know the bias for every poll
- FiveThirtyEight's model:
  - Use previous elections to determine a "bias factor"
  - Take weighted average of all polls
  - (count more recent polls more strongly)
- Can use extra data to weight answers:

Dealing with Bias

- After an election, we know the bias for every poll
- FiveThirtyEight's model:
  - Use previous elections to determine a "bias factor"
  - Take weighted average of all polls
  - (count more recent polls more strongly)
- Can use extra data to weight answers:
  - Suppose one polls 80% men and 20% women

Dealing with Bias

- After an election, we know the bias for every poll
- FiveThirtyEight's model:
  - Use previous elections to determine a "bias factor"
  - Take weighted average of all polls
  - (count more recent polls more strongly)
- Can use extra data to weight answers:
  - Suppose one polls 80% men and 20% women
  - Can weight men's answers by .25 to get a better representation