

Introduction to Polling Methodology ¹

Loren Collingwood
<mailto:lorenc2@u.washington.edu>
University of Washington

April 24, 2010

¹Department of Political Science

- 1 Define Population and Sampling Frame
- 2 Sample Design
- 3 Questionnaire Design
- 4 Data Collection: Mode Selection
- 5 Post-Survey Adjustments
- 6 Data Analysis

- 1 Define Population and Sampling Frame
- 2 Sample Design
- 3 Questionnaire Design
- 4 Data Collection: Mode Selection
- 5 Post-Survey Adjustments
- 6 Data Analysis

Population and Sampling Frame

- Target Population: Set of units to be studied

Population and Sampling Frame

- Target Population: Set of units to be studied
 - Registered Voters in Washington State

Population and Sampling Frame

- Target Population: Set of units to be studied
 - Registered Voters in Washington State
- Frame Population: Set of target population members that has a chance to be selected into sample.

Population and Sampling Frame

- Target Population: Set of units to be studied
 - Registered Voters in Washington State
- Frame Population: Set of target population members that has a chance to be selected into sample.
 - Most up-to-date list of registered voters with phone matches from Secretary of State's office. Any sample will likely miss people who registered, for instance, yesterday.

Sample Design

- Telephone sampling is currently the standard.

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame
 - Voter File. Greater certainty these are real voters.

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame
 - Voter File. Greater certainty these are real voters.
- Sample Size and Non-Response

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame
 - Voter File. Greater certainty these are real voters.
- Sample Size and Non-Response
 - 30:1 Average; 50:1 Hard to reach population

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame
 - Voter File. Greater certainty these are real voters.
- Sample Size and Non-Response
 - 30:1 Average; 50:1 Hard to reach population
 - Goal: $n = 500$, draw sample of 15,000

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame
 - Voter File. Greater certainty these are real voters.
- Sample Size and Non-Response
 - 30:1 Average; 50:1 Hard to reach population
 - Goal: $n = 500$, draw sample of 15,000
- Statistical Issues and Error in Sample Selection

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame
 - Voter File. Greater certainty these are real voters.
- Sample Size and Non-Response
 - 30:1 Average; 50:1 Hard to reach population
 - Goal: $n = 500$, draw sample of 15,000
- Statistical Issues and Error in Sample Selection
 - Sampling Distributions. Theory

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame
 - Voter File. Greater certainty these are real voters.
- Sample Size and Non-Response
 - 30:1 Average; 50:1 Hard to reach population
 - Goal: $n = 500$, draw sample of 15,000
- Statistical Issues and Error in Sample Selection
 - Sampling Distributions. Theory
 - Simple Random Sampling. Theory

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame
 - Voter File. Greater certainty these are real voters.
- Sample Size and Non-Response
 - 30:1 Average; 50:1 Hard to reach population
 - Goal: $n = 500$, draw sample of 15,000
- Statistical Issues and Error in Sample Selection
 - Sampling Distributions. Theory
 - Simple Random Sampling. Theory
 - Cluster Sampling. Reduced Precision

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame
 - Voter File. Greater certainty these are real voters.
- Sample Size and Non-Response
 - 30:1 Average; 50:1 Hard to reach population
 - Goal: $n = 500$, draw sample of 15,000
- Statistical Issues and Error in Sample Selection
 - Sampling Distributions. Theory
 - Simple Random Sampling. Theory
 - Cluster Sampling. Reduced Precision
 - **Stratified Sampling. Increased Precision and Control.**

Sample Design

- Telephone sampling is currently the standard.
 - Random Digit Dial (RDD). Expanded sampling frame
 - Voter File. Greater certainty these are real voters.
- Sample Size and Non-Response
 - 30:1 Average; 50:1 Hard to reach population
 - Goal: $n = 500$, draw sample of 15,000
- Statistical Issues and Error in Sample Selection
 - Sampling Distributions. Theory
 - Simple Random Sampling. Theory
 - Cluster Sampling. Reduced Precision
 - Stratified Sampling. Increased Precision and Control.
 - Design Effects. $\frac{V_{\gamma}(\bar{y})}{V_{srs}(\bar{y})}$

Let's Take an Example

- Confidence Interval and Margin of Error, Mean Age WA Poll 2009

$$MOE = \bar{y} + / - 1.96 * \sqrt{v(\bar{y})}, \quad (1)$$

$$MOE = 54.38 + / - 1.96 * \sqrt{\frac{258}{846}} \quad (2)$$

Let's Take an Example

- Confidence Interval and Margin of Error, Mean Age WA Poll 2009

$$MOE = \bar{y} + / - 1.96 * \sqrt{v(\bar{y})}, \quad (1)$$

$$MOE = 54.38 + / - 1.96 * \sqrt{\frac{258}{846}} \quad (2)$$

- 95% Confidence Interval = (53.29, 55.46)

Questionnaire Design

- Question Order, Rotations, Randomizations, and Response Order

Questionnaire Design

- Question Order, Rotations, Randomizations, and Response Order
 - Recency and Primacy Effects

Questionnaire Design

- Question Order, Rotations, Randomizations, and Response Order
 - Recency and Primacy Effects
- Open versus Closed questions

Questionnaire Design

- Question Order, Rotations, Randomizations, and Response Order
 - Recency and Primacy Effects
- Open versus Closed questions
- Balance in Questions. Provide both sides of an argument, for example.

Questionnaire Design

- Question Order, Rotations, Randomizations, and Response Order
 - Recency and Primacy Effects
- Open versus Closed questions
- Balance in Questions. Provide both sides of an argument, for example.
- Middle options and Don't Know/Refused options

Questionnaire Design

- Question Order, Rotations, Randomizations, and Response Order
 - Recency and Primacy Effects
- Open versus Closed questions
- Balance in Questions. Provide both sides of an argument, for example.
- Middle options and Don't Know/Refused options
- **Messaging and Vignettes/Bios.**

Questionnaire Design

- Question Order, Rotations, Randomizations, and Response Order
 - Recency and Primacy Effects
- Open versus Closed questions
- Balance in Questions. Provide both sides of an argument, for example.
- Middle options and Don't Know/Refused options
- Messaging and Vignettes/Bios.
- Cognitive Interviewing, Pre-testing, and Focus Groups.

Data Collection: Depends on Speed, Accuracy, Money

- **Face-to-Face.** American National Election Studies. Higher response rates. Cluster Sampling.

Data Collection: Depends on Speed, Accuracy, Money

- **Face-to-Face.** American National Election Studies. Higher response rates. Cluster Sampling.
- **Telephone Live Interviewer.** Washington Poll.

Data Collection: Depends on Speed, Accuracy, Money

- **Face-to-Face.** American National Election Studies. Higher response rates. Cluster Sampling.
- **Telephone Live Interviewer.** Washington Poll.
- **Telephone Automated Voice Response.** Precision Polling. Rasmussen Poll. Survey USA.

Data Collection: Depends on Speed, Accuracy, Money

- **Face-to-Face.** American National Election Studies. Higher response rates. Cluster Sampling.
- **Telephone Live Interviewer.** Washington Poll.
- **Telephone Automated Voice Response.** Precision Polling. Rasmussen Poll. Survey USA.
- **Internet Panel.** Knowledge Networks. Polimetrix.

Data Collection: Depends on Speed, Accuracy, Money

- **Face-to-Face.** American National Election Studies. Higher response rates. Cluster Sampling.
- **Telephone Live Interviewer.** Washington Poll.
- **Telephone Automated Voice Response.** Precision Polling. Rasmussen Poll. Survey USA.
- **Internet Panel.** Knowledge Networks. Polimetrix.
- **Self Administered Mail.** Census

Post-Survey Adjustments

- Item non-response and imputation. Hot Decking, Multiple Imputation with Chained Equations.

Post-Survey Adjustments

- Item non-response and imputation. Hot Decking, Multiple Imputation with Chained Equations.
 - **Income. Sensitive questions.**

Post-Survey Adjustments

- Item non-response and imputation. Hot Decking, Multiple Imputation with Chained Equations.
 - Income. Sensitive questions.
- **Weighting.** Calculations are multiplicative generally based on the following elements.

Post-Survey Adjustments

- Item non-response and imputation. Hot Decking, Multiple Imputation with Chained Equations.
 - Income. Sensitive questions.
- Weighting. Calculations are multiplicative generally based on the following elements.
 - **Unit Non-response. Young voters answer phones less often than older women.**

Post-Survey Adjustments

- Item non-response and imputation. Hot Decking, Multiple Imputation with Chained Equations.
 - Income. Sensitive questions.
- Weighting. Calculations are multiplicative generally based on the following elements.
 - Unit Non-response. Young voters answer phones less often than older women.
 - **Differential selection probabilities. May have oversampled a population to increase precision within that group.**

Post-Survey Adjustments

- Item non-response and imputation. Hot Decking, Multiple Imputation with Chained Equations.
 - Income. Sensitive questions.
- Weighting. Calculations are multiplicative generally based on the following elements.
 - Unit Non-response. Young voters answer phones less often than older women.
 - Differential selection probabilities. May have oversampled a population to increase precision within that group.
 - **Post-Stratify to known demographic proportions that are theoretically sensible.**

Frequencies, Crosstabs, Targeting, Modeling, and Report Writing

Response	Frequency
Good thing	71.6
Bad thing	11.9
No difference	13.1
(Don't Know)	3.1
(Refused)	0.2
Total	100.0
N	848

Table: Overall, do you think that statewide initiative elections are a good thing for the state, a bad thing for the state, or they don't make much difference?

Frequencies, Crosstabs, Targeting, Modeling, and Report Writing

- What is the vote share for each candidate by gender, by education, by party?

Frequencies, Crosstabs, Targeting, Modeling, and Report Writing

- What is the vote share for each candidate by gender, by education, by party?
- Target specific demographics for persuasion and mobilization. Do we see movement across votes after messages?

Frequencies, Crosstabs, Targeting, Modeling, and Report Writing

- What is the vote share for each candidate by gender, by education, by party?
- Target specific demographics for persuasion and mobilization. Do we see movement across votes after messages?
- **Statistically model the data.**

Frequencies, Crosstabs, Targeting, Modeling, and Report Writing

- What is the vote share for each candidate by gender, by education, by party?
- Target specific demographics for persuasion and mobilization. Do we see movement across votes after messages?
- Statistically model the data.
 - Model the vote, thermometer ratings. Control for other variables.

Frequencies, Crosstabs, Targeting, Modeling, and Report Writing

- What is the vote share for each candidate by gender, by education, by party?
- Target specific demographics for persuasion and mobilization. Do we see movement across votes after messages?
- Statistically model the data.
 - Model the vote, thermometer ratings. Control for other variables.
 - Likely Voter Modeling.

Frequencies, Crosstabs, Targeting, Modeling, and Report Writing

- What is the vote share for each candidate by gender, by education, by party?
- Target specific demographics for persuasion and mobilization. Do we see movement across votes after messages?
- Statistically model the data.
 - Model the vote, thermometer ratings. Control for other variables.
 - Likely Voter Modeling.
 - $y_i \sim \beta_0 + \beta_1 X_i + \epsilon_i$

Frequencies, Crosstabs, Targeting, Modeling, and Report Writing

- What is the vote share for each candidate by gender, by education, by party?
- Target specific demographics for persuasion and mobilization. Do we see movement across votes after messages?
- Statistically model the data.
 - Model the vote, thermometer ratings. Control for other variables.
 - Likely Voter Modeling.
 - $y_i \sim \beta_0 + \beta_1 X_i + \epsilon_i$
- Put it all together in a narrative report that is easily digestible.