

I. Sampling

A. Theory

1. Terms

- a. **Population:** that set of individuals about which we want to gather information
 - (1) Census: a total enumeration of the whole population
- b. **Sample:** the subset of population selected for data gathering and analysis
 - (1) Representative sample: one that accurately mirrors the population on all variables studied
 - (2) Simple random sample: each unit of population has a known and equal chance of selection
 - (3) Probability sample: each unit of the population has a knowable chance of selection
- c. **Confidence level:** the *probability* that the population value we are trying to estimate (e.g., the percentage of the vote that one candidate will win in an election) falls within a specified range of values (called the sampling-error margin) on either side of the observed sample value (e.g., the % of the vote the candidate received among those polled)
 - (1) Most social science polls set that *confidence* level at 95%; i.e., if 100 samples were independently drawn from the population, the *actual population value* will fall within a specific random-sampling error margin on either side of the *observed sample value* 95 times out of 100
 - (2) Or, to say the same thing another way, the *risk* that the observed sample value will fall beyond a specified error margin is less than 5 times out of 100 or 1 out of 20.
- h. **Sampling-error margin:** the specified range that the population value will fall within on either side of the observed sample value (at a specified confidence level)

- 2. Sampling error and sample size in a simple-random sample: Table 2.1, p. 31

Population size is of little consequence above 10,000; i.e., the sample size needed to limit the random-sampling error to $\pm 3\%$ is 1067 respondents regardless of whether the target population is the voters in the city of Wilmington, the state of N.C., or the whole U.S.

$$\sqrt{\text{sample size}} = \frac{(\text{population variability} [\text{probability}]) (\text{confidence level} [z \text{ score}])}{\pm (\text{error margin} [\text{probability}])}$$

$$\sqrt{\text{sample size}} = \frac{(0.5)(1.96)}{\pm (0.03)} = 1067$$

$$\sqrt{\text{sample size}} = \frac{(0.5)(1.96)}{\pm (0.04)} = 600$$

For multi-stage cluster samples, add 1% to each of these error margins

$$\sqrt{\text{sample size}} = \frac{(0.5)(1.96)}{\pm (0.05)} = 384$$

B. Applied sampling and surveys of public opinion (1936 to 1948)

- 1. **Non-representative sampling** (convenience and non-random samples in straw polls)
 - a. *Literary Digest* pre-election polls of 1916, 1920, 1924, 1928, 1932 & 1936
 - (1) Unrepresentativeness of population lists
 - (2) Termination of polling in early September
 - (3) Self-selection bias toward higher classes in mail-in balloting
 - b. Gallup pre-election polls of 1936, 1940, 1944, & 1948
 - (1) Lack of supervision of (middle-class female) interviewers at the last stages of sampling (when they filled their neighborhood and respondent-selection quotas) led to under-representation of poor, working-class, and minority voters
 - (2) Weighting of sample subsets to make up for under-representation
 - (3) Termination of polling two weeks before the election

C. Contemporary (post-1948) **representative sampling** methods

- 1. **Simple-random sampling (SRS):** seldom used – except in exit polls; because there is usually no master list and/or the population is too geographically scattered
- 2. **Multistage-cluster sampling** (face-to-face interviews) slightly higher sampling-error margin
 - a. Division of the population into regions (for nationwide samples, usually 4)
 - b. Random selection within four regions:
 - (1) By 20 Primary Sampling Units (PSUs) within each region 80 PSUs total – SMSAs or counties
 - (2) By 4 or 5 blocks within each PSU 320-400 blocks total
 - (3) By 4 or 5 households within each block 1280-2000 households total
 - (4) One individual within each household 1280-2000 individuals total
 - (a) NES & GSS & most European pollers use a stringent quota method
 - (b) Media polls often use the youngest- or next-birthday selection method
- 3. **Random-digit-dialing or RDD** (telephone interviews)
 - a. Population: non-institutional residential numbers (listed, unlisted, and new-since-last-listing)
 - b. Sampling methods
 - (1) Households
 - (a) Randomly drawn *seed* numbers proportional to area code, exchange, and cluster
 - (b) Plus-one (Sudman) method to randomize the phone numbers used

- (2) Individuals within households – typically the youngest-member or most-recent-birthday methods
- c. Non-response rates are now approaching 60% of working numbers called
 - (1) Causes – answering-machine screening, refusals, not-at-homes
 - (2) Significance depends on whether those not included have different opinions/behaviors than those interviewed

- II. **Non-sampling error** in scientific polls pp. 40-46
- A. Interviewer effects – gender, age, race, SES, accent
 - B. Response-acquiescence (question-topic) effects
 - 1. Definition: a socially-desirable opinion is given instead of true attitude
 - 2. Examples
 - a. Q topic is too abstract or obscure → a random opinion is given rather than no opinion
 - b. Q topic is too sensitive → a presumed-acceptable opinion is given rather than real attitude
 - C. Question-wording effects
 - 1. Multiple stimuli
 - a. Definition: multidimensional issues lumped into 1 Q produce ambiguity about which dimension is more significant
 - b. Examples
 - (1) Support for free speech for Communists
 - (2) Support for affirmative action and Congress
 - (3) School voucher programs: parental choice versus taxpayer funds
 - (4) Abortion: total ban/medical only/social-economic/morning-after pill/unlimited access
 - 2. Unbalanced choice
 - a. Definition: Q gives only one side or doesn't examine the consequences (e.g. spending & taxes)
 - b. Examples
 - (1) Polar alternatives are better than yes/no to a single choice: "Do you favor or oppose the death penalty?" is better than "Do you favor the death penalty?"
 - (2) Balanced choices are better than Likert scales: "Some people feel that ...; others feel that ..." is better wording than asking how much the respondent agrees or disagrees with one choice
 - 3. No middle position
 - a. Definition: only polar alternatives or balanced choice alternatives given
 - b. Examples
 - (1) No middle position on a scale: e.g., many political ideology scales
 - (2) No "depends" option(s):
 - D. Question-order effects
 - 1. Framing effects
 - a. Definition: previous questions can influence responses to later questions
 - b. Examples
 - (1) US-reporters-in-Russia question before Communist-reporters-in-US question
 - (2) Political-issue questions before presidential popularity questions
 - (3) Vote-choice questions late in polls tend to lower incumbents' vote choice
 - 2. Filter or branching questions
 - a. Definition: preceding Q's used to narrow sample
 - b. Examples
 - (1) Issue-interest questions before issue-preference questions
 - (2) Age, voter-registration, campaign-interest, and past-voting-behavior questions before vote-intention questions
 - E. Non-response effects – to interview or to only selected questions
 - F. Timing-of-data-collection effects

- III. The misuse of surveys
 - A. Modern straw polls – convenience samples over-represent attentive publics
 - 1. Tallies of letters to media or public officials
 - 2. Mass-sample, mail-return surveys
 - 3. 800/900 number "call-in" polls to media or internet sites
 - B. Push polls – loaded questions intended to change rather than to measure opinion

IV. Interpreting scientific surveys

- A. Pre-election surveys
 - 1. Accuracy of pre-election presidential polls (conducted in last week of campaign) has been quite good since 1952 Figure 2-1, p. 50
 - 2. House-effect variations
 - a. Timing of last scheduled poll
 - b. Wording of candidate-choice question(s)
 - c. Allocating undecided respondents
 - d. Weighing non-response rates
 - e. Screening for likely voters – tougher screens tend to give more accurate estimates
- B. Exit polls Voter News Service - (ABC, CBS, NBC, Fox, CNN, AP) has been replaced by the National Election Pool
 - 1. Definition: election-day polls of voters exiting precinct places
 - 2. Problems
 - a. Selecting representative precincts within districts
 - b. Estimating the effects of social desirability
 - c. Weighing refusal rates
 - d. Measuring the effects of releasing exit-poll results on those who have not yet voted