Polls
Surveys of the Election Process

> "*How far would Moses have gone, if he had taken a poll in Egypt?*" – Harry S. Truman

Class 2: UCALL Course on Numbers in Everyday Life
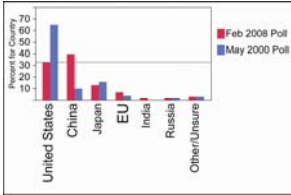**Josef Schmee**

What is a Survey?

- A survey is based on a study of a portion of the population
- Data are collected for a purpose
- Data are typically from individuals
- Data obtained through sampling and with the use of questionnaires

> *Polls are survey of the election process*

Gallup Poll. Feb. 11-14, 2008. N=1,007 adults nationwide. MoE ± 3.

"Which one of the following do you think is the leading economic power in the world today: [see below]**?"** Options rotated

	2/11-14/ 2008	5/18-21/ 2000
China	40%	10%
United States	33%	65%
Japan	13%	16%
European U.	7%	4%
India	2%	–
Russia	2%	2%
Other/Unsure	3%	3%

> *Surveys report opinions of those surveyed, not facts.*

Purpose of a Survey

> *Estimate the characteristics of the universe or population*

- Population Total
  - total number unemployed
- Population mean
  - average household expenditures on food, clothing, etc.
- Population proportion
  - proportion of mortgages under default
  - *proportion in favor of candidate A*
- Population ratio
  - ratio of expenditure on housing to total income

History of Polling

1824	**Harrisburg Pennsylvanian presidential poll**
1848	A. Quetelet: "average man" & Normal dist.
1920s	A. Grossley (radio audience) & G. Gallup (newspaper readers) employ sampling
1936	**Literary Digest versus Sampling**
1941	National Opinion Research Center
1948	Truman defeats Dewey while polls predict otherwise
1948-70s	Research on Reliability (sampling, call-backs, etc.)
1970s –	Research on Validity (non-sampling error, question wording & order, Interview process)

Overview

- Why we need polls
- What is a poll
- How a poll is conducted
- Why polls work
- When to trust a poll
- My Prediction

Why We Need Polls

- What is important to voters versus to politicians, pundits, or special interest groups
  - What voters like and dislike
  - Perceptions of a candidate
  - Who voted for whom and why
- Multi-year comparison of what is on people's mind

Political scientists, historians and participants in politics use polls to explain voting behavior.

Do you feel things in this country are generally going in the right direction or do you feel things have pretty seriously gotten off on the wrong track? (NYT/CBS polls)

Most recent poll:

N = 1,368

Registered Voters = 1,196

Democratic Primary Voters= 510

Republican Primary Voters= 323

Tracking opinion over time

Polling Problems

- Very tight political races
  - Candidates percentages are too close to call
  - Results within the Margin of Sampling Error
- Large proportion of undecided
  - All candidates are equally good or bad
  - Voters may swing either way

Polls will tells us that there are tight races or that there is a large proportion of undecided.

Polls are not good at predicting the outcomes of such political elections.

What Is a Poll

Questionnaire → Select Sample → Complete Questionnaires → Estimate Results → Analyze Results → Communicate Results

- Questions are phrased properly
- Sample is representative
- Answers are unbiased
- Estimation of results reflect voters demographics
- Analysis of results is without bias
- Results are communicated clearly

Questionnaire

Words → Questions → Format → Hypothesis

- Words have multiple meanings or are similar sounding to other words
  - Stock owner: live-stock or financial stock?
  - Medicare versus Medicaid
- Questions can be too broad, ambiguous
  - Question types to avoid (double barreled, loaded)
- Format should avoid confusion
  - Question order, skipping of and branching to questions
- Hypothesis should be specific and focused

How to Ask Questions

- Personal interviews
  - Interviewer and respondent meet
  - Interviewer bias
- **Telephone interviews**
  - **Most polls use landline telephones**
  - **Cell phones ignored**
- Mailed questionnaires
  - Longer questionnaires take longer
  - Non-response
- On-line questionnaires
  - Lower cost alternative, need internet access, often self-selected

How to Write Questions

- Simple words familiar to all respondents
  - Difficult in a diverse population
- Avoid double-barreled questions
  - Two questions in one
  - "Do you plan to sell your house & buy another one?"
  - "Do you support a strong United States and an increase in the defense budget."
- Leading Questions
  - Intentional or unintentional bias
  - See next slide

Question writing is an art requiring experience and hard work

Leading Questions

- Partial mention of alternative
  - "What issue, such as the economy, concerns you most?"
- Emotionally charged words
  - "Do you support a tax on *unfair* oil company profits?"
  - Instead of: "Do you support *President Bush's* decision to send additional troops to Iraq?" ask "Do you favor or oppose sending additional troops to Iraq?"
- Threatening self-esteem of respondent
  - "Do you work?"
- Personalization of questions
  - Instead of "Do you think gambling should be legalized?" ask "Should gambling be legalized?"

ABC News/Washington Post Poll. Feb. 28-March 2, 2008. N=629 Democrats and Democratic-leaners nationwide. MoE ± 4. Fieldwork by TNS.

"Who would you like to see win the Democratic nomination for president this year: Hillary Clinton or Barack Obama?" Names rotated. Results include leaners.

	Obama	Clinton	Other	Neither	Unsure
2/28 - 3/2	50%	43%	1%	3%	2%

"As you may know, the Texas and Ohio Democratic primaries will be held next Tuesday. If Clinton wins one of these primaries but loses the other, should she stay in the race, or drop out?" Options rotated

	Stay In	Drop Out	Unsure
2/28 - 3/2	67%	29%	4

"If Clinton loses both the Texas and Ohio primaries, should she stay in the race, or drop out?" Options rotated

	Stay In	Drop Out	Unsure
2/28 - 3/2	45%	51%	4

How a Poll Is Conducted

- Samples are drawn from Sample Frame
- How pollsters select participants
  - Probability based selection is objective
  - Non-probability selection introduces bias
  - Goal is a representative sample
- Simple Random Sample or Panel
- Weighting of subpopulations

Sampling Concepts

- **Elements** are persons which are of interest to the poll results
  - Individual voters
- Target **population** is the collection of all elements of interest in the study
  - All citizens who will vote

≠

- **Sampling units** are a non-overlapping collection of elements.
  - Households
- **Sampling frame** is a list of sampling units.
  - List of registered voters
  - List of telephone exchanges

*A **sample** is a collection of elements put together from sampling units. The sample is drawn from the sampling **frame**, not from the population.*

Problems with Sampling Frames

- Incomplete frames
  - Who is going to vote?
    - Voters not yet registered
    - Registered voters not going to vote
- Multiple incompatible frames
  - Cell phones travel across area codes
- Overlapping frames
  - Land lines tied to location
  - Cell phones tied to person
- Frames do not match time period
  - Outdated lists

1936 Literary Digest Presidential Poll Disaster

Landon, 1,293,669 versus Roosevelt, 972,897

- LD poll correctly predicted election outcomes in 1916, 1920, 1924, 1928, and 1932
- Survey of more 2 million people from subscriber list, car registrations and phone books
- Maine voted in September for Gov and House
  - As Maine goes, so goes the nation!
- Democratic party chairman James Farly in November
  - As Maine goes, so goes Vermont!
- Biased poll of well-to-do group (telephones, car owners)
- Magazine became laughing stock and had to fold
- George Gallup set out to change polling practices

1936 Election Results

Crossley, Gallup, and Elmo Roper predicted correct outcome based on a relatively small sample

Nominee	Franklin D. Roosevelt	Alf Landon
Party	Democratic	Republican
Home State	New York	Kansas
Running mate	John Nance Garner	Frank Knox
Electoral Vote	523	8
States Carried	46	2
Popular Vote	27,752,648	16,681,862
Percentage	60.8%	36.5%

Probability Samples

Each sampling unit has a known probability of being selected into the sample.

- Simple random sample (SRS)
  - Sampling units have equal probability of being selected
  - National SRS of 1050 likely voters
- Stratified Random Sample (Polls by Region)
  - Frame is separated into non-overlapping strata
  - SRS drawn from each stratum
  - Stratum precision is less than overall precision
- Post-stratification
  - Separated into non-overlapping strata after sampling
  - National SRS of 1050 likely voters is stratified into 550 Democrats and 500 Republicans
  - Stratum precision is less than overall precision

A Duel at Gallup

USA Today/Gallup National Poll (Feb 21-24)
- Barack Obama   51%
- Hillary Clinton   39%
- Point spread   12%

Gallup Daily Tracking Poll (Feb 22-24)
- Barack Obama   47%
- Hillary Clinton   45%
- Point spread   2%

USA Today/Gallup National Poll (Feb 21-24)
- John McCain   61%
- M. Huckabee   21%
- Point spread   40%

Gallup Daily Tracking Poll (Feb 22-24)
- John McCain   63%
- M. Huckabee   23%
- Point spread   40%

Despite best efforts, problems occur

Random Digital Dialing

- In USA no list of all adults or voters
- Computer generates phone number from area codes and exchanges
  - Efficient algorithms also get unlisted numbers
  - Sampling unit = phone number
- Land line phone numbers
  - No cell phones
- Telephone number ≠ Voter
  - *Filter to obtain adults, registered voters, likely voters*
  - Asking for specific members of household (birthday)

Non-probability Samples

- Convenience samples
  - TV interview in front of busy store
- Quota samples
  - Fill required number of respondents (quota)
- Internet opt-in surveys
  - Self-selected respondents a la American Idol

Some national polls take random samples from a large pool of volunteers and apply demographic weights to their raw results.

Zogby Interactive Online Polling
(taken from Zogby web-site in March 2008)

How it works:
1. Complete the online registration page and submit it to us.
2. Zogby Interactive will e-mail you and inform you when an online poll is ready for you to take.
3. Complete the survey. (Easy? We thought so.)
4. Zogby International will e-mail you and inform you when results are available to view.
5. See how your opinion stacks up.
6. Be ready to voice your opinion with the leading public opinion pollster in the world.
   If you would like the results of the poll you took (as well as other poll results and Zogby International information), be sure to check the check-box asking if you would like Zogby International press releases and other Zogby information sent to you in the future.
   Thank you for participating in Zogby Interactive's Online Polling. You'll be hearing from us soon!

Weighting of Subpopulations

- ARSE = age, race, sex, education
- Weighting adjusts the raw data of a poll to reflect population characteristics
  - Age groups, ethnic and racial make-up
  - Does not alter answers themselves, but how much answers count
- Adjusting for Selection Probability
- Adjusting for Sample Design
  - Oversample groups to refine estimates on some issue
- Adjusting for Demographics
  - Overrepresented and underrepresented groups

Weighting Example

Subgroup	In Population	In Sample	In Favor	Proportion in Favor
A	0.70	500	250	0.50
B	0.30	500	300	0.60

Unweighted Estimate of % in favor = (250 + 300)/1000 = 55%

Weighted Estimate of % in favor = 0.70•0.5 + 0.30•0.60 = 53%

Weighting by demographic subgroup changes the answers. Pollsters consider weights proprietary information.

Why Polls Work

- Sample Size does not depend on population size
  - At least not for populations larger than 10,000
- Increasing the sample size increases the statistical precision of the poll
- Statistical precision is measured by margin of sampling error
- Margin of sampling error (MoE) is usually expressed in percent %

Polls work because statistical precision increases with sample size and can be measured!

Poll Precision and Accuracy Depend On

- Sample Size
  - Absolute number of people interviewed
- Non-response rate
- How questions were asked
  - Wording of questions
  - Inflection
  - Order of questions
- When the poll was conducted
  - Volatility
  - Differences tend to disappear near election time
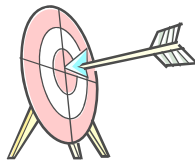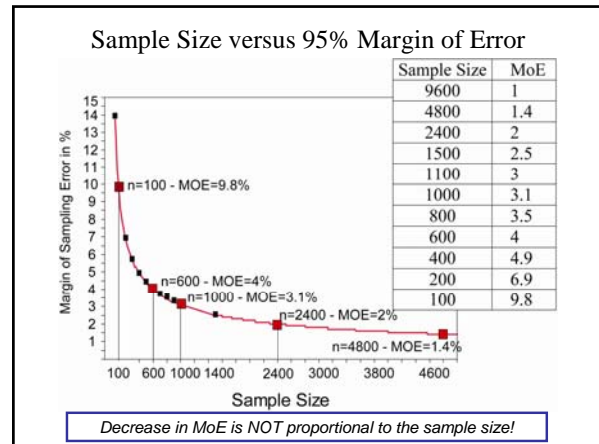
Statistical Margin of Error

- Measures the degree to which the sample results can be expected to differ from the actual population results
- MoE is due to sampling a subset of the population
  - Sampling error
- MoE has a confidence level associated with it
  - Confidence = chance that MoE is correct
  - 95% is gold standard in polling
  - On average, 1 chance in 20 that poll is wrong (statistically)
- MoE not depend on population size
  - Only absolute sample size is important for populations larger than 10,000
- MoE does not include non-sampling errors
  - Sloppiness, wrongly worded questions, wrong answers (lying)

Candidate A: 45% ± 3% ???

- Sample percentage in favor of A is 45%
- Margin of Error (MoE) = ± 3%
- Confidence Interval = estimated % ± MoE
  45% ± 3% for a range 42% to 48%
- Confidence Level in Polls is virtually always 95%
  Rarely mentioned
  Indicates chance that confidence interval is correct
  95% → on average 19 out of 20 polls are correct

With 95% confidence, the approval percentage of candidate A is between 42% and 48%

Sample Size versus 95% Margin of Error

Sample Size	MoE
9600	1
4800	1.4
2400	2
1500	2.5
1100	3
1000	3.1
800	3.5
600	4
400	4.9
200	6.9
100	9.8

Decrease in MoE is NOT proportional to the sample size!

Margin of Error of Subgroups

- MoE depends on sample size in subgroup
- Poll of size 1000 with 10% African American respondents results in subsample of 100
- Margin of Error for this African American subsample is ca. ±10%
  - rather than ±3% as for entire population

MoE for subgroups, e.g., African-Americans, Evangelicals, is smaller than for entire population

---

Problems with Telephone Polls

- Use primarily land lines for Random Digit Dialing
  - Cheap
  - No cell phones
- Multiple users per telephone number
  - Who picks up the phone
- Non-response a hidden problem
  - Different response rates by demographic groups
  - Poor less likely to answer poll questions
  - Call-backs

Zogby Poll for California Primary

- Zogby prediction
  - Obama 49%
  - Clinton 36%
- Actual Result
  - Obama 42.3%
  - Clinton 51.9%
- Zogby Error 23%

- Underestimated turnout among Hispanic voters
  - Screening techniques of "most likely voters" has been confounded
- Weighting of demographic subgroups
  - Considered proprietary
  - Not subject to scrutiny
- Local demographics are different from national

California Primary on February 5, 2008 with 444 Democratic delegates at stake.

New Hampshire Primary

All polls underestimated Clinton's favorable rating!

Poll	Date	Sample	Obama	Clinton	Edwards	Richardson
Final Results	**-**	**-**	**36.4**	**39.0**	**16.9**	**4.6**
RCP Average	**1/6 - 7**	**-**	**38.3**	**30.0**	**18.3**	**5.7**
Suffolk/WHDH	1/6 - 7	500	39	34	15	4
American Res. Group	1/6 - 7	600	40	31	20	4
ReutersC-Span/Zogby	1/5 - 7	862	42	29	17	5
Rasmussen	1/5 - 7	1774	37	30	19	8
CNN/WMUR/	1/5 - 6	599	39	30	16	7
Marist	1/5 - 6	636	36	28	22	7
CBS News	1/5 - 6	323	35	28	19	5

New Hampshire primary on January 8, 2008 with 22 delegates at stake

McCain versus Obama

Poll	Date	Sample	McCain (R)	Obama (D)	Und	Spread
Rasmussen Tracking	04/02 - 05	1700	**48**	**43**	**9**	McCain +5.0
Gallup Tracking	04/02 - 05	4403	**45**	**45**	**4**	Tie
NYTimes/CBS	03/28 – 04/02	1196	**42**	**47**	**7**	Obama +5.0
Hotline/FD	03/28 - 31	799	**46**	**44**	--	McCain +2.0
NBC/WSJ	03/24 - 25	800	**42**	**44**	**7**	Obama +2.0

Beware of the electoral college!

McCain versus Clinton

Poll	Date	Sample	McCain (R)	Clinton (D)	Und	Spread
Rasmussen Tracking	04/02 - 04/05	1700	**48**	**44**	**11**	McCain +4.0
Gallup Tracking	04/02 - 04/05	4403	**47**	**45**	**3**	McCain +2.0
NY Times/CBS	03/28 - 04/02	1196	**43**	**48**	**5**	Clinton +5.0
Hotline/FD	03/28 - 03/31	799	**50**	**41**	--	McCain +9.0
NBC/WSJ	03/24 - 03/25	800	**46**	**44**	**5**	McCain +2.0

Beware of the electoral college!

Exit Polls

- Exit polls are based on actual voters
- Large sample sizes guarantee high precision
- Problems arise (Bush versus Kerry, 2004)
  - Interviewers do not follow careful selection protocol
    - Lengthy questionnaires
    - Legal distance from polling station loses voters
    - Inexperienced interviewers
  - Early voters vote differently from later voters

2004 Exit Polls
National: Kerry 51% to Bush 48%.
Ohio: Kerry lead 51.7 to 48.3 – Bush won 50.8 to 48.7

Decennial Census versus Sampling

- US Census is relied on by Law, but also by businesses and researcher
  - Need for accuracy and current data
- US Census Bureau proposed sampling instead of 100% census
  - Sampling is more accurate with fewer biases than 100% census
    - Problems of undercount in many areas of country
  - Sampling would be cheaper and quicker
    - Census could be performed more frequently
- Some Politicians did not like the idea
- US Supreme Court rejected the proposal

When to Trust a Poll

- Don't, unless ……..
- Sampling is random or at least an imitation of it
- Target population is identified
- Sample size is stated
- Potential sources for bias are accounted for
- Weighting schemes are explained
- Also see slides "Important Questions"

Bias in Polls

- Coverage Bias
  - Non-representative samples
    - Mobile phones
- Response Bias
  - Given answers do not match true beliefs
  - Designed by pollsters to get desired answers
  - Respondents give more extreme positions to influence results, peer-pressure
- Non-Response Bias
  - Selected person refuse to answer questions
  - Culturally dependent

Problems with Polls in Predicting Election Results

- Voters are difficult to identify
  - How to find those that are actually going to vote
  - Response from non-voters is useless
- Non-responses and refusal to participate
  - Demographic factors
- Volatility
  - When voters decide for their favorite candidate only shortly before the election

Important Questions (1)

- Who paid for the poll?
- Why was the poll conducted?
- Who performed the poll?
- How many people participated in poll?
- How were participants selected?
- What was the target population? [Area, groups]
- When were the interviews conducted?
- How were the interviews conducted?

Important Questions (2)

- What questions were asked?
  - Wording was balanced and unbiased?
- What was the ordering of the questions?
  - Earlier questions might influence reponses to later questions
- A reported results based on all participants or on a subset with a different margin of error?
  - How large is he subset?
- Were the data weighted and are the weights known?

New York Times "How the Poll Was Conducted" (1)

The latest NY Times poll is based on telephone interviews conducted **June 10 through June 12** with **1111 adults** throughout the US.

The sample telephone exchanges called was **randomly** selected by a computer from a complete list of more than 42,000 active residential exchanges across the country.

Within each exchange, random digits were added to form a complete telephone number, thus permitting access to listed and unlisted numbers alike. Within each household, one adult was designated by a random procedure to be the respondent for the survey.

New York Times "How the Poll Was Conducted" (2)

The results have been **weighted** to take account of household size and number of telephone lines into the residence and to adjust for variation in the sample relating to geographic region, sex, race, marital status, age and education.

In theory, in **19 cases out of 20**, overall results based on such samples will differ by no more than three percentage points in either direction from what would have been obtained by seeking out all American adults.

New York Times "How the Poll Was Conducted" (3)

For **smaller subgroups** the margin of sampling error is larger.

In addition to sampling error, the practical difficulties of conducting any survey of public opinion may introduce **other sources of error** into the poll. Variation in the wording and order of questions, for example, may lead to somewhat different results.

Complete questions and results are available at nytimes.com/politics

Websites for Polling

- Public Agenda
  - www.publicagenda.org/aboutpa/aboutpa.cfm
- National Council on Public Polls
  - www.ncpp.org/
- The Pew Research Center
  - people-press.org/
- Real Clear Politics
  - www.realclearpolitics.com/
- Gallup Organization
  - www.gallup.com/

My Prediction for 2008 Presidential Election

There is very high probability that the next president
- Is either a Democrat or a Republican
- Is either male or female
- Is either Caucasian or African-American
- Is between 45 and 75 years at the time of inauguration