Friday May 6, 2011 Schedule for the Day

Am: Critical Reasoning
     --Discussion of Today's Assignment from Exercise 8.1
     --Discussion and Workshop on new material: Sampling

 Pm: Ethical Reasoning
     --Virtue Ethics

Comments on Critical Reasoning Assignment for Friday, April 1

Determine which of the following statements are generalizations and which are descriptions of particular states of affairs. Indicate which of the generalizations are *universal* (apply to all, every, no or none) and which are *statistical* (apply to some—that is, not all or none). Remember that statistical generalizations can include terms such as *most, many, few,* and *some percentage.* It might be debated whether some statements are general or particular.

A2. Most people don't trust government.   **Generalization, Statistical**

A4. Dale always parties on Friday nights.   **Generalization, Universal**

A6. People don't get everything they like.   **Generalization, Universal**

A 8. Few people enjoy having their gallbladder removed.   **Generalization, Statistical**

A10. Every animal with a heart is an animal with kidneys.   **Generalization, Universal**

A 12 Brenda voted Republican and Mike voted Democratic.   **Particular**

Discussion of Chapter 2 Exercise 8.1 B2, B4, B6, B8, B10

Label the arguments in the following passages as deductive or inductive. (One passage contains both kinds of argument.) Among the inductive, note which are sampling arguments (particular-to-general arguments), which are arguments with statistical premises (general-to-particular arguments), and which include both.

B2. The outlook for education in America is bleak. Educational disaster will be avoided only if people give up their desire for more tax cuts. But Americans are not willing to do that

Deductive

B4. Alvin should pay at least half his income in taxes because everybody who has more than a $1 million income, whether from wages or some other source, should pay at least 50 percent in taxes no matter what his or her deductions.

Deductive

B6. Any time population increases in a state, the housing demand increases as well. Population has been increasing in Arizona, Washington, and Nevada. So we can expect the demand for housing to increase in those states.

Deductive

B8 Willie was late on Monday and late on Tuesday. We shouldn't expect him to be on time today.

Inductive, Statistical

B10. "The Survey of Consumer Finances shows that the top 1 percent of households controls roughly 42 percent of the country's nonhome wealth in 2004. If so, a targeted marketing campaign designed to induce these individuals to buy Widgets should improve the bottom line for Widget, Inc."

Federal Reserve 2004 Survey of Consumer Finances

Deductive

A Sample

It is decision time at Widget, Inc. The company president says: "Our market research department has just completed a test of the new and improved Widget in three test market areas: Dallas, Detroit, and Denver. In all three cities the consumers preferred the new Widget over the old two to one. I think we should go for it."

Deductive main argument [If most consumers generally prefer the new Widget to the old one, we should start marketing the new Widget. Most customers generally prefer the new Widget.   Therefore, we should start marketing the new Widget. Where the second premise is supported by the sampling argument:
  In the sample most customers generally preferred the new Widget.
 (likely) Most customers generally prefer the new Widget.]

Perhaps there is an inductive, argument with a statistical premise at the end of the passage*: (1) In cities generally, consumers will prefer the new Widget.*
 (2) The areas we target for marketing the new Widget are cities.
 (likely) In areas we target for marketing, consumers will prefer the new Widget.
Where the first premise is supported by the sampling argument:
 In the test market samples in 3 cities, consumers preferred the new Widget.
 (likely) In cities generally, consumers will prefer the new Widget.
Note that the two possible sampling arguments have different units of analysis: customers in one, cities in the other.

Example 8.1   *Inductive Argument (Particular-to-General)*

Premise   (1) In studies of 5,000 people, those who had more exposure to environmental smoke had a higher frequency of lung cancer.

Conclusion   (likely)People who have more exposure to environmental smoke generally have a higher frequency of lung cancer.

Example 8.2   *Deductive Argument (Modus Ponens)*

Premise   (1)   People who have more exposure to environmental smoke generally have a higher frequency of lung cancer.

Premise   (2) If (1), then we should restrict smoking in public places.

Conclusion $\therefore$ We should restrict smoking in public places.

Inductive Argument (Particular-to-General=Sampling Argument)

(1) The first two layers of strawberries contain many ripe ones.
___

(likely) All layers of strawberries contain many ripe ones.

Inductive Argument ("Classic "Inductive Argument, Past to Future)
Variation of particular-to-general argument

(1) In the 1960s measures to combat inflation led to increased unemployment.

(2) In the 1980s measures to combat inflation led to increased unemployment.

(3) In the 1990s measures to combat inflation led to increased unemployment.
___

(likely)Measures to combat inflation will always lead to increased unemployment.

Inductive Argument (General-to-Particular)

    (1) Most 103-year-old persons who have major surgery suffer serious complications.

    (2) Didi is a 103-year-old person who has had major surgery.

_____

(likely)   Didi will suffer serious complications.

On our view some inductive arguments go from the general to the particular

Deductive

(1) All God's creatures need potassium in their diets.

(2) Alvin is one of God's creatures.

∴ Alvin needs potassium in his diet

Inductive with Statistical Premise

(1) Most adults can tolerate moderate amounts of sugar in their diets.

(2) Alvin is an adult.

(likely) Alvin can tolerate moderate amounts of sugar in his diet.

Jerry must be pretty well off. Lexus owners have higher- than-average incomes and Jerry owns a Lexus.

Deductive Version

(1) All Lexus owners have higher-average incomes.
(2) Jerry owns a Lexus.

∴ Jerry has a higher-than-average income.

Inductive Version

(1) Most Lexus owners have higher-average incomes.
(2) Jerry owns a Lexus.

(likely) Jerry has a higher-than-average income.

A recent poll of a random sample of Americans of voting age indicated that 68 percent favored a constitutional amendment aimed at assuring a balanced budget. With such a large approval rating, it is only a matter of time before a balanced budget amendment is ultimately passed into law. This is because most proposed additions to the Constitution that have substantial public support ultimately gain ratification.

Reconstruction

Sampling Argument

68% of the eligible (likely) voters sampled in the poll favored a balanced
budget constitutional amendment

(likely) About 68 % of the eligible (likely) voters in America favor a
balanced budget constitutional amendment.

Argument with Statistical Premises

(1) About 68 % of the eligible (likely) voters in America favor a balanced budget
constitutional amendment.

 (2) Most measures supported by a large portion of the American Public become law.

(likely) A balanced budget constitutional amendment.

1. Attacking the evidence. Is the evidence cited in the premise true or can the data be disputed

2. Questioning the representativeness of the sample.
       (a) Size of Sample
       (b) Sample Selection

3. Pointing to a shift in the unit of analysis

4. Challenging the truth of the conclusion.

How Data are Obtained

- ## Observational Study
  - Observes individuals and measures variables of interest but does not attempt to influence the responses
  - Describes some group or situation

- ## Experiment
  - Deliberately imposes some treatment on individuals in order to observe their responses
  - Studies whether the treatment causes change in the response.

Common Language

- Population
- Individual (case)
- Sampling Frame
  - individuals that could *possibly* be selected for the sample (not necessarily the same as the population)
- Sample
- Sample Survey
  - type of observational study; data collected on a sample
- Census
- Variable
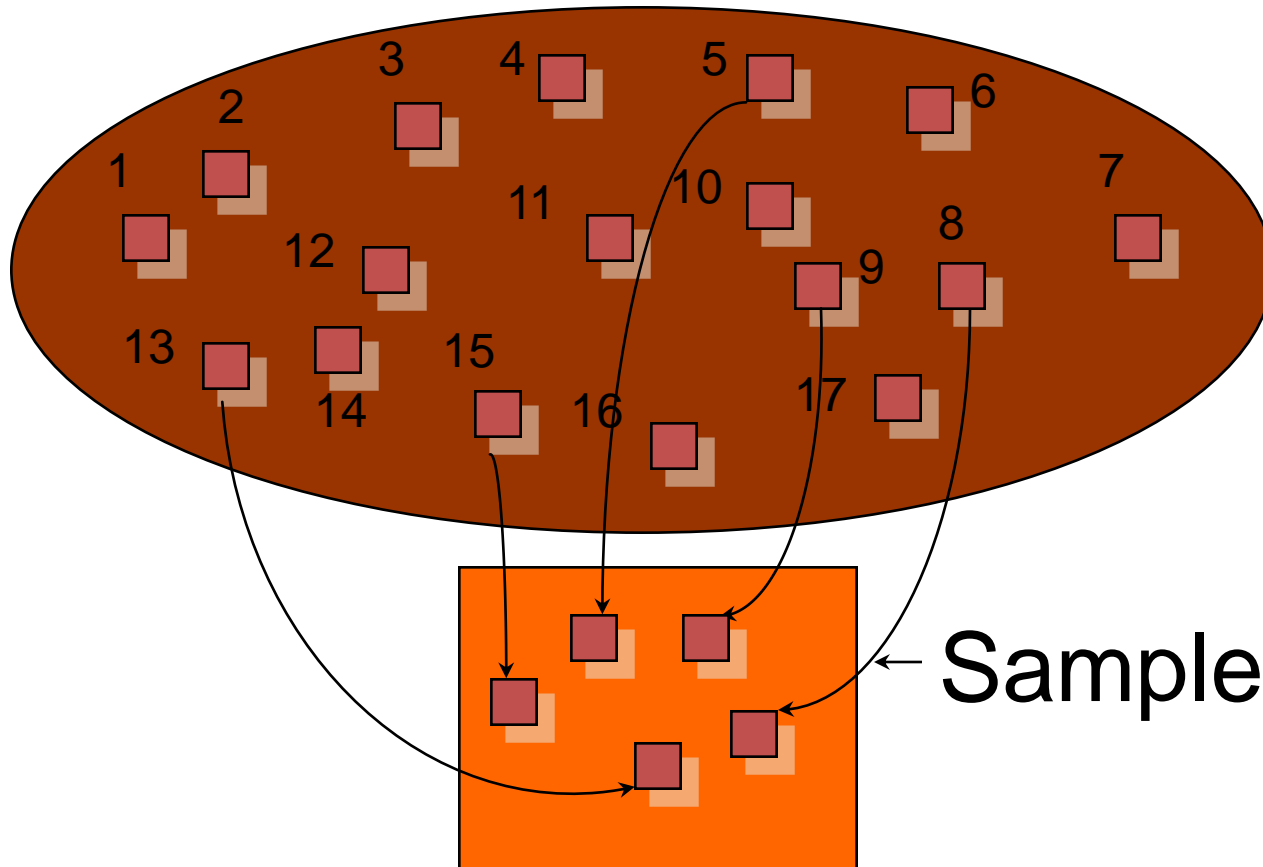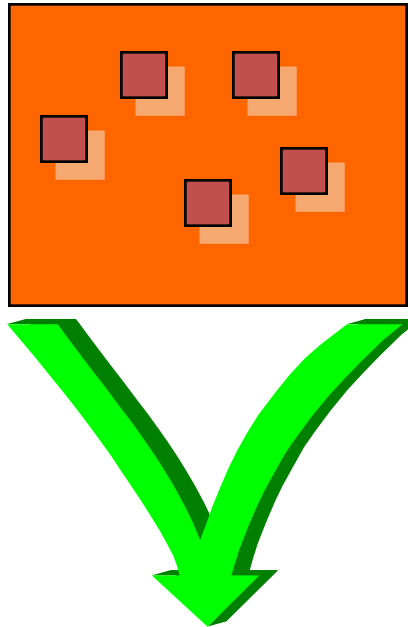 - characteristic of an individual

Population

individuals
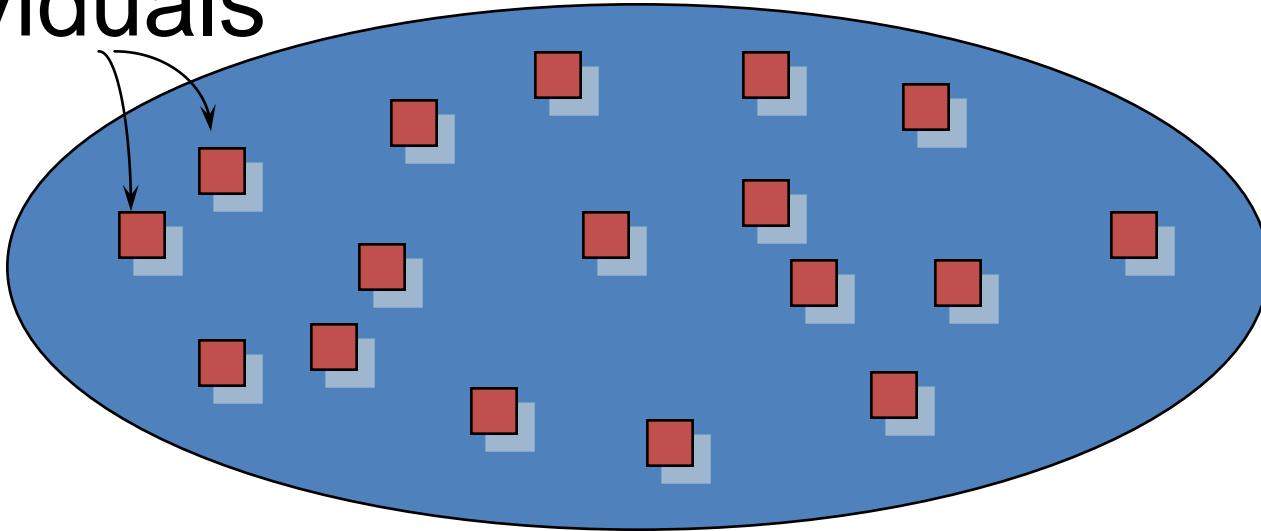
Sampling Frame

List of Individuals

1
2
3
4
✓ 5
6
7
✓ 8
✓ 9
10
11
12
✓ 13
14
✓ 15
16
17

Sample

Sample Survey

measurements
data

Census

Census

List of Individuals

Bad Sampling Plans

- Convenience sampling
  - selecting individuals that are easiest to reach

- Voluntary response sampling
  - allowing individuals to <u>choose</u> to be in the sample

- Both of these techniques are **biased**
  - systematically favor certain outcomes

Convenience Sampling

- Sampling mice from a large cage to study how a drug affects physical activity
  - lab assistant reaches into the cage to select the mice one at a time until 10 are chosen

- Which mice will likely be chosen?
  - could this sample yield biased results?

Voluntary Response

- To prepare for her book *Women and Love*, Shere Hite sent questionnaires to 100,000 women asking about love, sex, and relationships.
  - 4.5% responded
  - Hite used those responses to write her book

- Moore (*Statistics: Concepts and Controversies*, 1997) noted:
  - respondents "were fed up with men and eager to fight them…"
  - "the anger became the theme of the book…"
  - "but angry women are more likely" to respond

Simple Random Sampling

- Each individual in the population has the same chance of being chosen for the sample
- Each group of individuals (in the population) of the required size (n) has the same chance of being the sample actually selected
- Random selection:
  - "drawing names out of a hat"
  - random number table or computer software

STATISTICS IN SUMMARY
Simple Random Sample

Population

All samples of size n

are equally likely

Sample data
$x_1, x_2, , x_n$
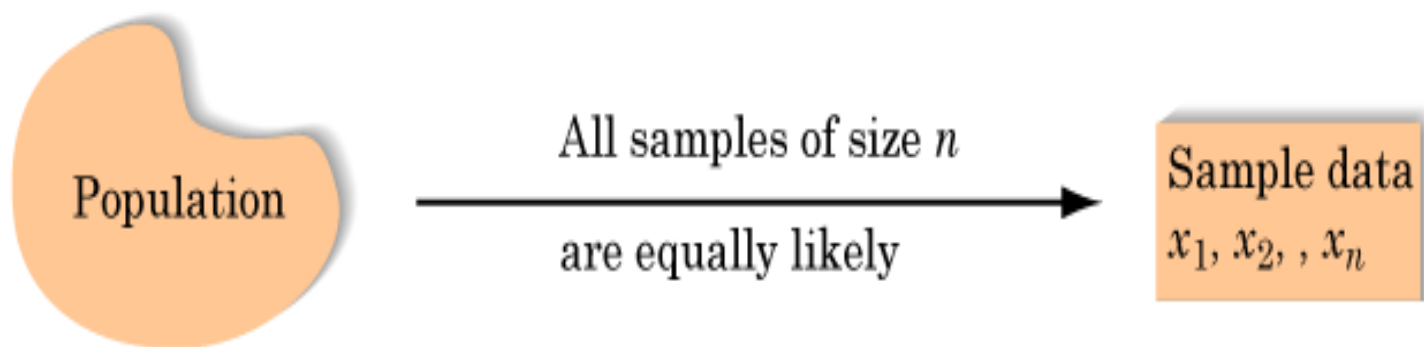
Stratified Random Sample

- first divide the population into groups of similar individuals, called **strata**

- second, choose a separate simple random sample in <u>each</u> stratum

- third, combine these simple random samples to form the full sample

  - if only certain strata are (randomly) chosen to be used, and <u>all</u> subjects in these strata make up the sample, then we have a *cluster sample*

Sampling Terminology

- ## Parameter
  - fixed, unknown number that describes the population

- ## Statistic
  - known value calculated from a sample
  - a statistic is used to estimate a parameter

- ## Bias
  - in repeated samples, the sample statistic consistently misses the population parameter in the same direction

- ## Variability
  - different samples from the same population may yield different values of the sample statistic

Bias and Variability

(a) Large bias, small variability

(b) Small bias, large variability

(c) Large bias, large variability

(d) Small bias, small variability

Sampling Strategy

- To reduce bias, use <u>random sampling</u>

- To reduce variability, use <u>larger samples</u>
  - estimates from random samples will be closer to the true values in the population if the samples are larger
  - how close will they be?
    - margin of error

Margin of Error

- The amount by which the proportion obtained from the sample ($\hat{p}$) will differ from the true population proportion (p) rarely exceeds the *margin of error*.

- Typical margin of error: 1/sqrt(n)
  - In 95% of surveys, the sample proportion will not differ from the population proportion by any more than the margin of error. ("*95% confidence*")

Field Date Oct 18-28, 2010

Q.11 Thinking ahead to the November election for U.S. Senate, are you planning to vote for, (ROTATE ORDER NAMES READ) Patty Murray, who prefers the Democratic Party, or Dino Rossi, who prefers the Republican party?

46% Patty Murray -- certain
2%   Patty Murray -- could change
1%   Undecided -- lean Murray
42% Rossi -- certain
2%   Rossi -- could change
1%   Undecided -- lean Rossi
6%   Undecided/Don't know/Can't Remember/Didn't Vote

COMBINED SENATE VOTE
49% Murray
45% Rossi
6% Undecided

LIKELY VOTERS (across waves 1 & 2)
51% Murray
45% Rossi
4% Undecided

N=695, ME= $\pm$ 3.7% ,= 51 $\pm$ 3.7   95% confident that true proportion is between 46.7% and 54.7%

The poll was fielded by telephone, using live callers, and called registered voters statewide at random. Phone numbers were matched to the voter record, and a combination of landlines and cell phones were contacted. If a respondent was not home, or busy, up to 5 call-backs were scheduled with each number. Overall 500 surveys were completed in wave 1, and an additional cross-section of 500 surveys were completed in wave 2. Each wave has a margin of error of +/- 4.3%.

Rough ME $\approx$ $1/\sqrt{n}$ **1/ square root of (695) = 1/26.363 = .0379**

Adobe Acrobat Document

About the Poll

The KCTS9/KPLU/Washington Poll is a non-partisan, academic survey research project sponsored by the Washington Institute for the Study of Ethnicity & Race (WISER), a research center at the University of Washington in the School of Social Sciences. Dr. Matt Barreto, an Associate Professor of Political Science is the Director of WISER and principal investigator on this survey.

The survey was administered by telephone, by the Center for Survey Research at the UW, based on a randomly selected list of phone numbers using a list of registered voters. The survey was in the field from October 18 – 28, 2010. A total of 500 registered voters throughout the state of Washington were interviewed, yielding in a 4.3% margin of error. Likely voters are drawn from both waves 1 and 2, and a total of 695 likely voters were polled, for a margin of error of +/- 3.7%. Frequency percentages may not add up to 100% due to rounding.

The current poll was a collaborative effort between KCTS-9, KPLU and the Washington Poll.

Questions/Comments: mbarreto@uw.edu, lorenc2@uw.edu

United States, U.S. Senator (Partisan office, 6-year term)Statewide

Candidate	Vote	Vote %
Patty Murray (Prefers Democratic Party)	1,314,930	52.36 %
Dino Rossi (Prefers Republican Party)	1,196,164	47.64 %
Total Votes	2,511,094	100.00%

LIKELY VOTERS (across waves 1 & 2)

51% Murray   51% + 2 %   = 53%

45% Rossi   45 % + 2 %   = 47%

4% Undecided   **Allocated Equally**

Actual Percentage

52.36%

47.64%

Very Close

Other polls

1* If the 2010 Election for United States senate were held today would you vote for Republican Dino Rossi or Democrat Patty Murray?

	9/25/10	10/9/10	10/30/10	
Dino Rossi (R)	47%	47%	47%	+2= 49
Patty Murray (D)	48%	46%	49%	+2= 51
Some other candidate	2%	7%	4%	
Not sure	3%	0%	0%	

NOTE: Margin of Sampling Error, +/- 3 percentage points with a 95% level of confidence

Actual Percentage

52.36% Murray

47.64% Rossi

Public Policy Polling (PPP)   a union oriented polling organization   results

Raleigh, N.C. – Every time PPP has polled the Washington Senate race this year it's found the race to be within 2 or 3 points and our final poll there is no exception. But there is one twist- for the first time we find Dino Rossi leading Patty Murray, by a 50-48 margin.

PPP surveyed 2,055 likely Washington voters from October 29th to 31st. The margin of error for the survey is +/- 2.2%. Other factors, such as refusal to be interviewed and weighting, may introduce additional error that is more difficult to quantify.

Q1   The candidates for US Senate are Democrat Patty Murray and Republican Dino Rossi. If the election was today, who would you vote for?

Patty Murray .... 49 =1+ 48%

Dino Rossi .... 51 =1+ 50%

Undecided... 2%

More about Sample Size

Survey Sample Size	Margin of Error Percent* % ±	Margin of Error Proportion* prop ±
100000	0.3	0.003
20000	1	0.007
10000	1	0.010
2000	2	0.022
1500	3	0.026
1000	3	0.032
900	3	0.033
800	3	0.035
700	4	0.038
600	4	0.041
500	4	0.045
400	5	0.050
300	6	0.058
200	7	0.071
100	10	0.100
50	14	0.141

⬅ n=695

*Assumes a 95% level of confidence

1/ square root of (695) = 1/26.363 = .0379

Sample Size	of Error Percent* % ±	of Error Percent** % ±	of Error Percent*** % ±
100000	0.3	0.4	0.3
20000	1	1	1
10000	1	1	1
2000	2	3	2
1500	3	3	2
1000	3	4	3
900	3	4	3
800	3	5	3
700	4	5	3
600	4	5	3
500	4	6	4
400	5	6	4
300	6	7	5
200	7	9	6
100	10	13	8
50	14	18	12
	*Assumes a 95% level of confidence	**Assumes a 99% level of confidence	***Assumes a 90% level of confidence

The margin of error reported with poll results is what is considered the 95% confidence level range. Meaning a pollster has a 95% confidence that the true measurement lies within the margin of error.

The standard error equation is shown below.

Standard error $=$, $\sqrt{\dfrac{p*(1-p)}{n}}$

where p represent the support level of the poll and n is the number of voters polled.

And the 95% confidence interval is 1.96 * (standard error).

The maximum margin of error occurs when p = 50%.

(**Maximum) margin of error (95%) =** $\quad 1.96 * \sqrt{\dfrac{.5*(1-.5)}{n}} = \dfrac{.98}{\sqrt{n}} \approx \dfrac{1}{\sqrt{n}}$
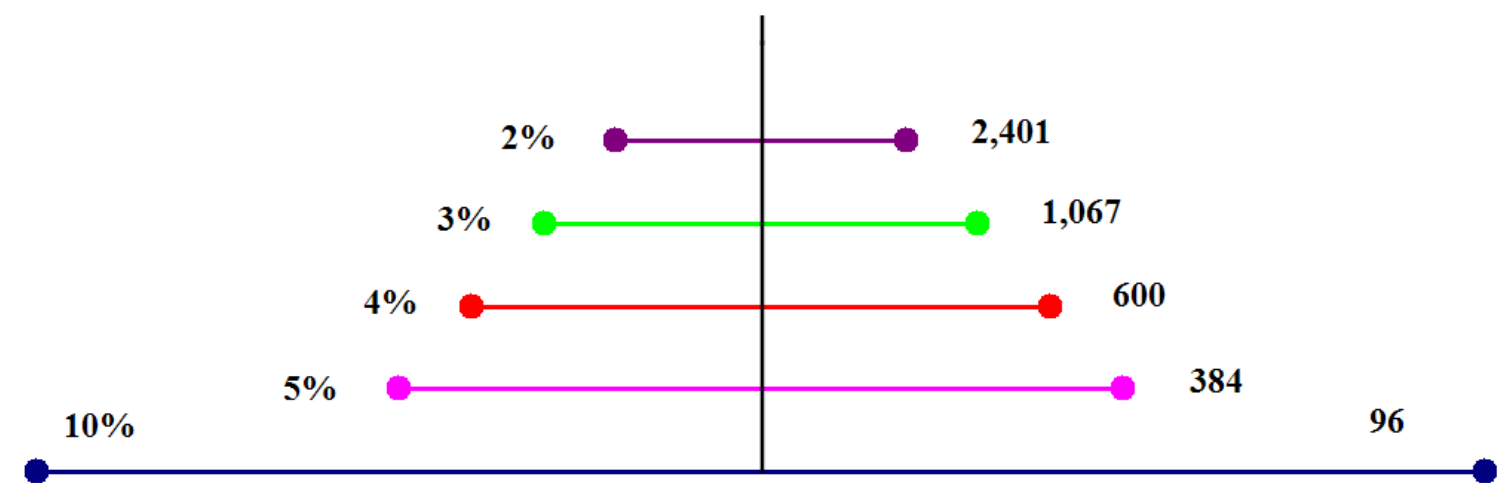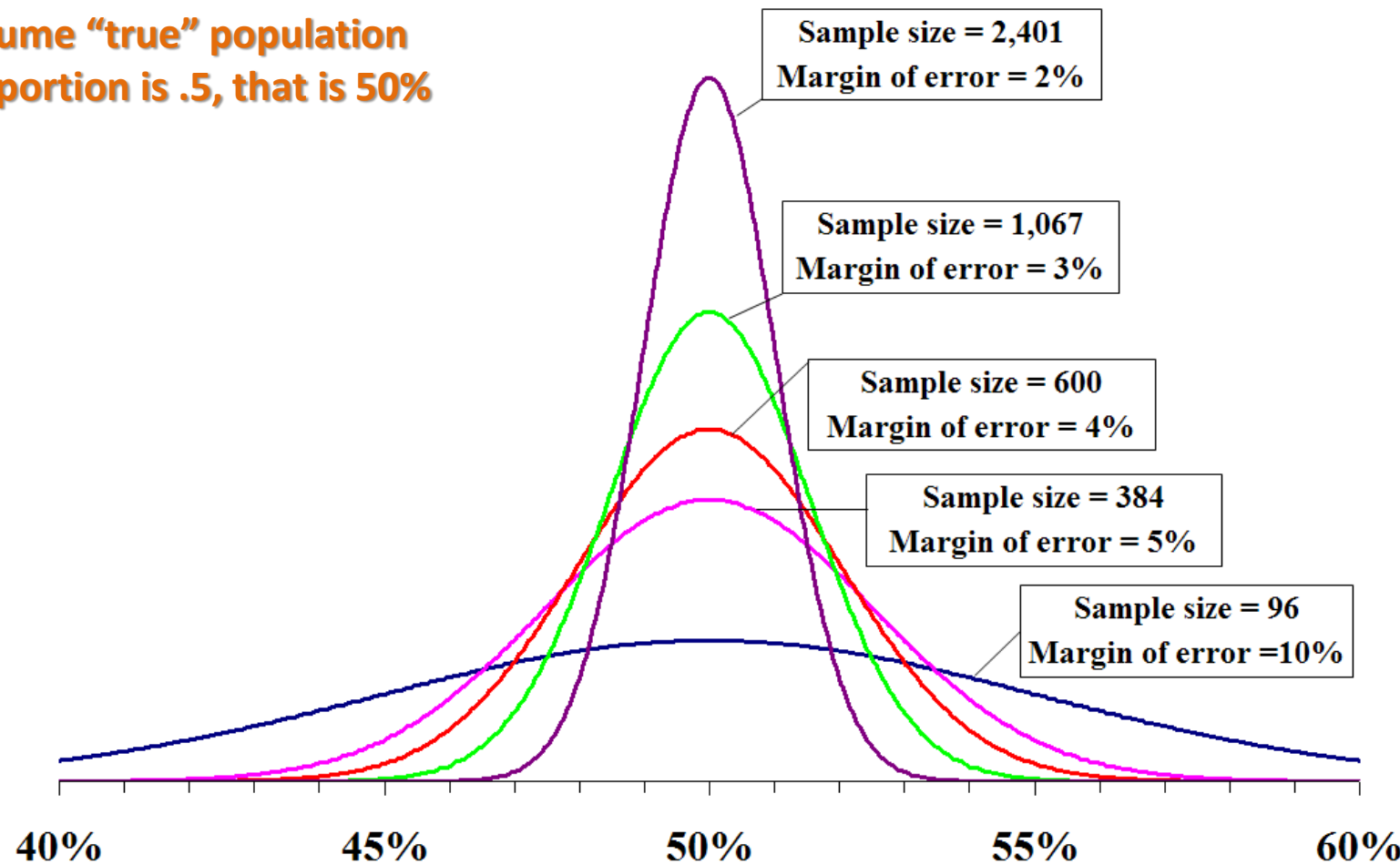
Margin of error at 99% confidence $\approx$ **1.29/√n**

Margin of error at 95% confidence $\approx$ **.98/√n** ⟸ **Here is were the rough and ready formula comes in 1/√n**

Margin of error at 90% confidence $\approx$ **.82/√n**

Assume "true" population proportion is .5, that is 50%

Sample size = 2,401
Margin of error = 2%

Sample size = 1,067
Margin of error = 3%

Sample size = 600
Margin of error = 4%

Sample size = 384
Margin of error = 5%

Sample size = 96
Margin of error =10%

40%   45%   50%   55%   60%

2%   2,401
3%   1,067
4%   600
5%   384
10%   96

Errors in Sample Surveys

- Random sampling errors
  - measured by *margin of error*
    This can be estimated and taken into account

- Sampling errors

- Nonsampling errors

Likely to undermine the inference from sample to population
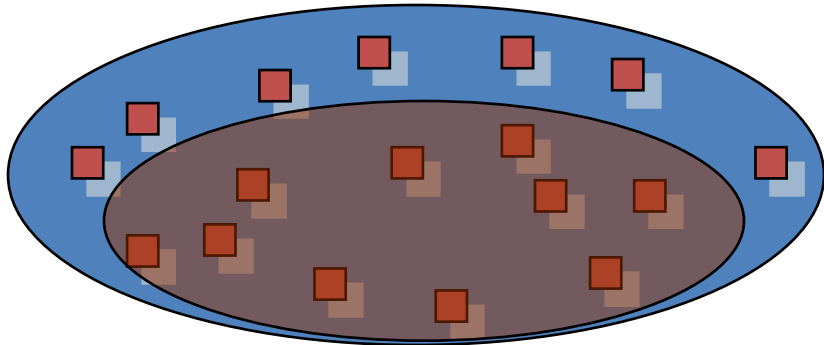
Sampling Errors

- Difficulties
  - Using the wrong sampling frame (next slide)

- Disasters
  - Using voluntary response (volunteer sample)
  - Using a convenience or haphazard sample
  - ❖ cannot extend results to the population of interest (need a broad cross-section of the population) Lack **"external validity"**

Using the Wrong Sampling Frame
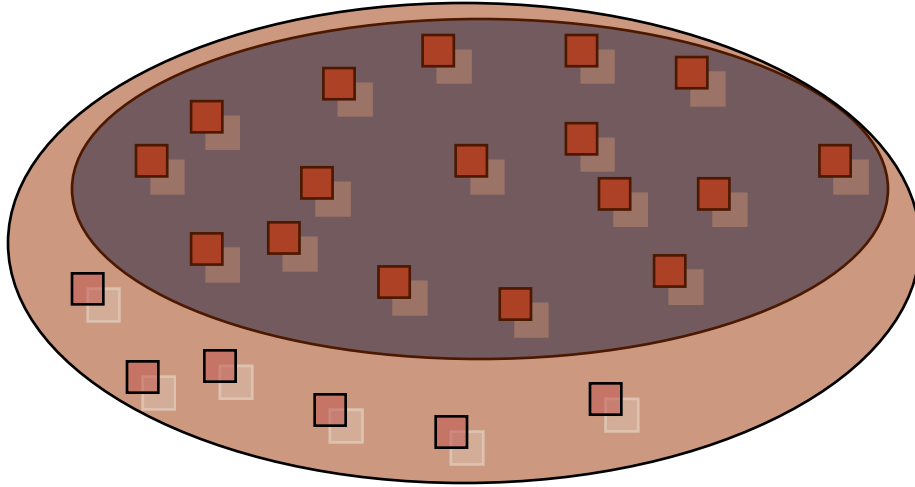
Including some units not in the population.

Undercoverage: Excluding some units in the population.

Nonsampling Errors

- Difficulties
  - Processing errors (data entry, calculations)
  - Wording of questions / Response error

- Disasters
  - Nonresponse (cannot contact subjects or they do not respond)

A Common Source of Nonsampling error: The Pitfalls of Asking Survey Questions

- Deliberate bias

- Unintentional bias

- Desire to please

- Asking the uninformed

- Unnecessary complexity

- Ordering of questions

- Confidentiality and anonymity

Deliberate Bias

- "If you found a wallet with $20 in it, would you return the money?"

- "If you found a wallet with $20 in it, would you do the right thing and return the money?"

Unintentional Bias

- "I have taught several students over the past few years."
  - How many students do you think I have taught?
  - How many years am I referring to?

- "Over the past few days, how many servings of fruit have you eaten?"
  - How many days are you considering?
  - What constitutes a serving?

Desire to Please

- "Is your instructor doing a good job presenting the course material in a clear and interesting way?"
    - Yes
    - No

Asking the Uninformed
A Case Study
Washington Post National Weekly Edition (April 10-16, 1995, p. 36)

- A 1978 poll done in Cincinnati asked people whether they "favored or opposed repealing the 1975 Public Affairs Act."

  – There was no such act!

  – About one third of those asked expressed an opinion about it.

Unnecessary Complexity

- "Do you sometimes find that you have arguments with your family members and co-workers?"
  - Arguments with family members
  - Arguments with co-workers

Ordering of Questions

- "How often do you normally go out on a date? about ___ times a month."
- "How happy are you with life in general."
  - Strong association between these questions.
  - If the ordering is reversed, then there would be no strong association between these questions

Confidentiality and Anonymity

- Confidential answer
  - respondent is known, but the information is a secret

- Anonymous answer
  - the respondent is not known, or cannot be linked to his/her response

Criticism of Sampling Arguments

√ 1. Attacking the evidence. Is the evidence cited in the premise true or can the data be disputed

√ 2. Questioning the representativeness of the sample.
   (a) Size of Sample
   (b) Sample Selection

3. Pointing to a shift in the unit of analysis

4. Challenging the truth of the conclusion.

Shifting the Unit of Analysis

Suppose a student newspaper carried out a survey about whether courses at the university included an exam. Most of the courses surveyed did. The newspaper concluded that most teachers give exams. We can reconstruct the argument in this way:

Example 8.8:

(1) Most **courses** sampled at the university give exams

(likely) Most **teachers** in the university give exams

It could be that most courses included exams, but that these courses were taught by a relatively small number of the junior faculty, whereas most faculty taught more advanced courses and offered research opportunities that used papers rather than exams.

Is there a problem here?

Example 8.9

(1) 20 percent of **schools** sampled across the United States fail to meet the Average Yearly requirement of the No Child Left Behind Act.
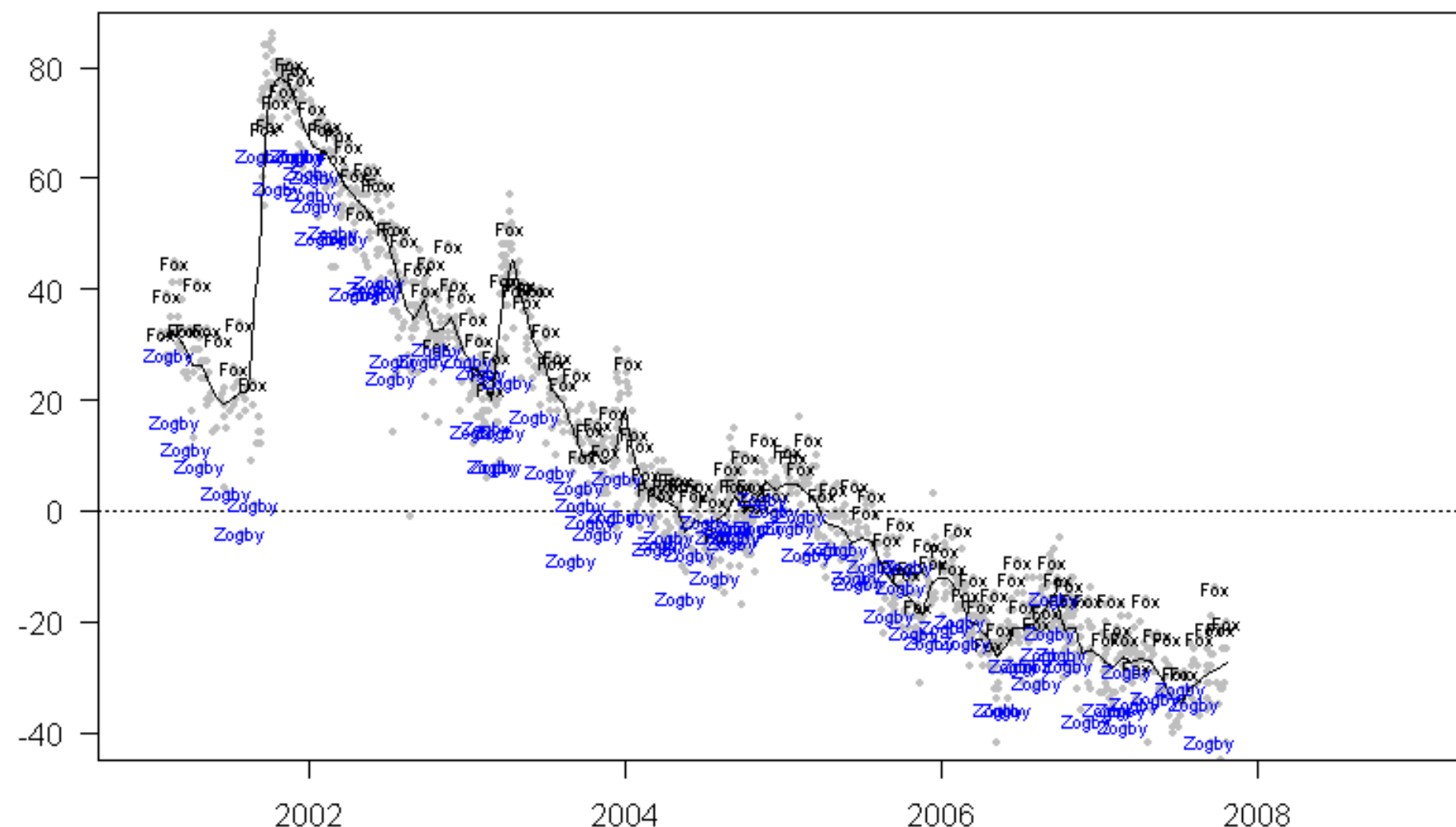
(likely) 20 percent of **schools districts** across the United States fail to meet the Average Yearly requirement of the No Child Left Behind Act.

Directly Challenging the Truth of Conclusion

This is inappropriate for deductive arguments   It might include counter evidence from other samples or other studies
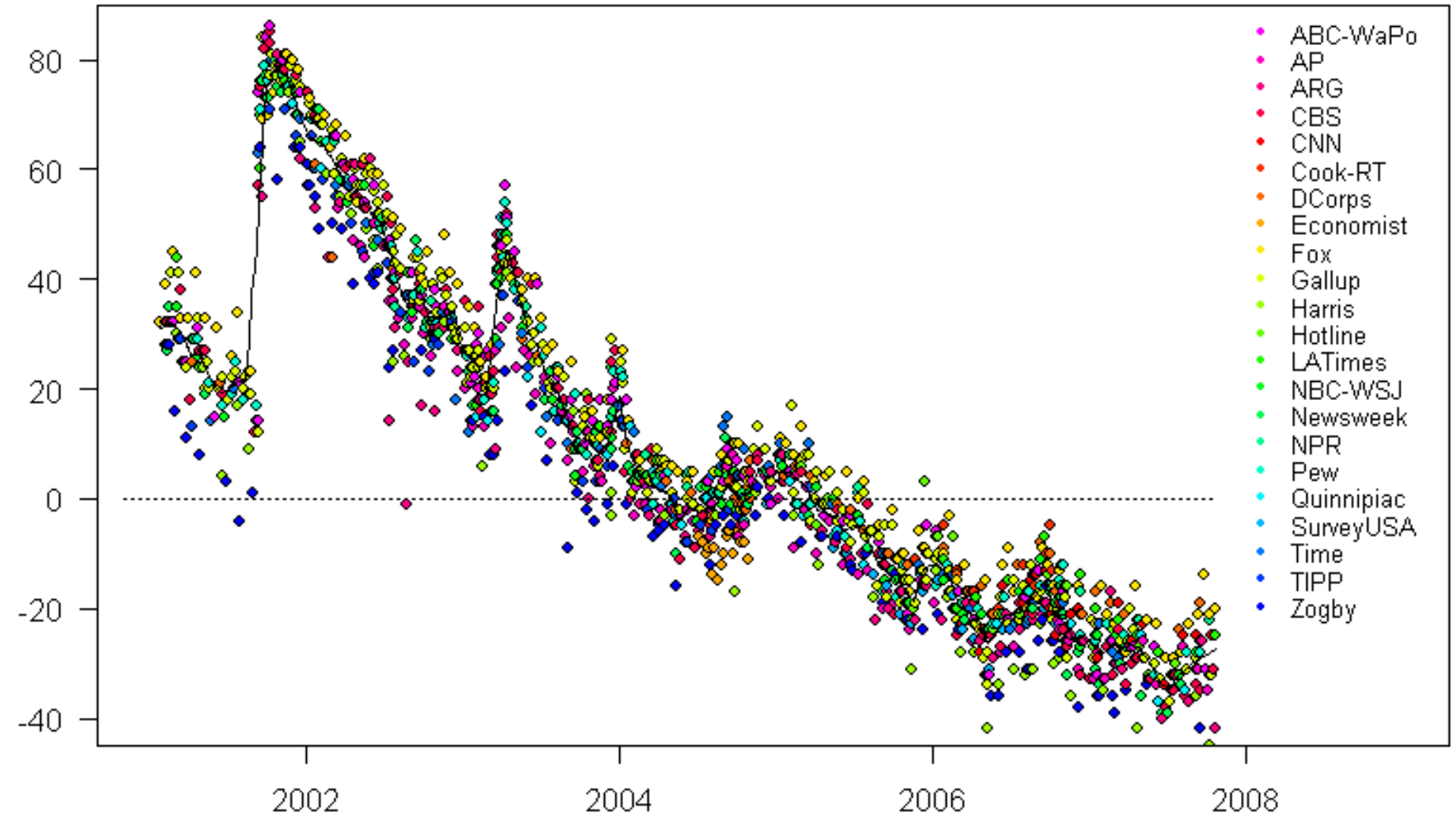
Tracking polls on Bush's job approval rating

Zogby averages below trend; Fox above

Bush job performance ratings
Approval - disapproval spread through 24 Oct 2007

Legend:
- ABC-WaPo
- AP
- ARG
- CBS
- CNN
- Cook-RT
- DCorps
- Economist
- Fox
- Gallup
- Harris
- Hotline
- LATimes
- NBC-WSJ
- Newsweek
- NPR
- Pew
- Quinnipiac
- SurveyUSA
- Time
- TIPP
- Zogby

II 1

(1) An insufficient number of those sampled said that they would vote for Roosevelt.
(likely)An insufficient number of voters would vote for Roosevelt to elect him.

The sample is not representative; the less well-off would be less likely to read *Literary Digest* or to have a telephone or an automobile, especially during the Depression. A random sample of addresses for registered voters (those likely to vote) would have produced better results, but the data would have been difficult to obtain for the country as a whole. This is a classic example of a very large sample that fails to support an inductive inference because it is not representative

II 2

(1) Being female was not associated with greater life expectancy
*   than being male*.[in the sample of three countries: Uganda, Rwanda and Nigeria]
(likely) Being female is not associated with greater life expectancy than being male (in general).

The sample might not be representative (and is small). These African countries have had high rate of HIV/AIDS and periods of warfare, which could have affected females as much or more than men or made them more susceptible to death in childbirth.

II 3

(1) (In the sample) The number of bank robberies increased
*   (likely) The number of bank robbers increased.*

Shift in the unit of analysis. The premise concerns the number of robberies; the conclusion changes to the number of robbers. But the number of robbers could remain the same or even go down if they were committing more robberies

II 4

(1) A very small percentage of the unvaccinated Cook County children cared for
*   by Homefirst in Chicago and unvaccinated Amish children in rural Lancaster*
*   Pennsylvania become autistic.*
(likely) A very small percentage of (all) unvaccinated children become autistic.

Although the sample is large, two communities might not be representative of the whole population to which the passage generalizes. Those children born while under the care of a health organization promoting home birth and those born into the Amish lifestyle, could well have mothers who provided a prenatal or early childhood environment (besides resisting vaccination) that warded off autism.