Thinking Straight Tuesday , May 6

Friday Class will be in Library 2708 that day only

Morning Session

- Review of Exam
- Lecture/discussion on Inductive and Statistical Reasoning

Afternoon Session beginning at 1 pm

Workshop on the ethics of Care/ *Gone Baby Gone*

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 ∴ 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 voters sampled in the poll favored more strict gun control legislation

(likely) About 68 % of the eligible voters in America favor more strict gun control.

Argument with Statistical Premises

(1) About 68 % of the eligible voters in America favor more strict gun control legislation

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

(likely) More strict gun control legislation will ultimately be ratified.

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.

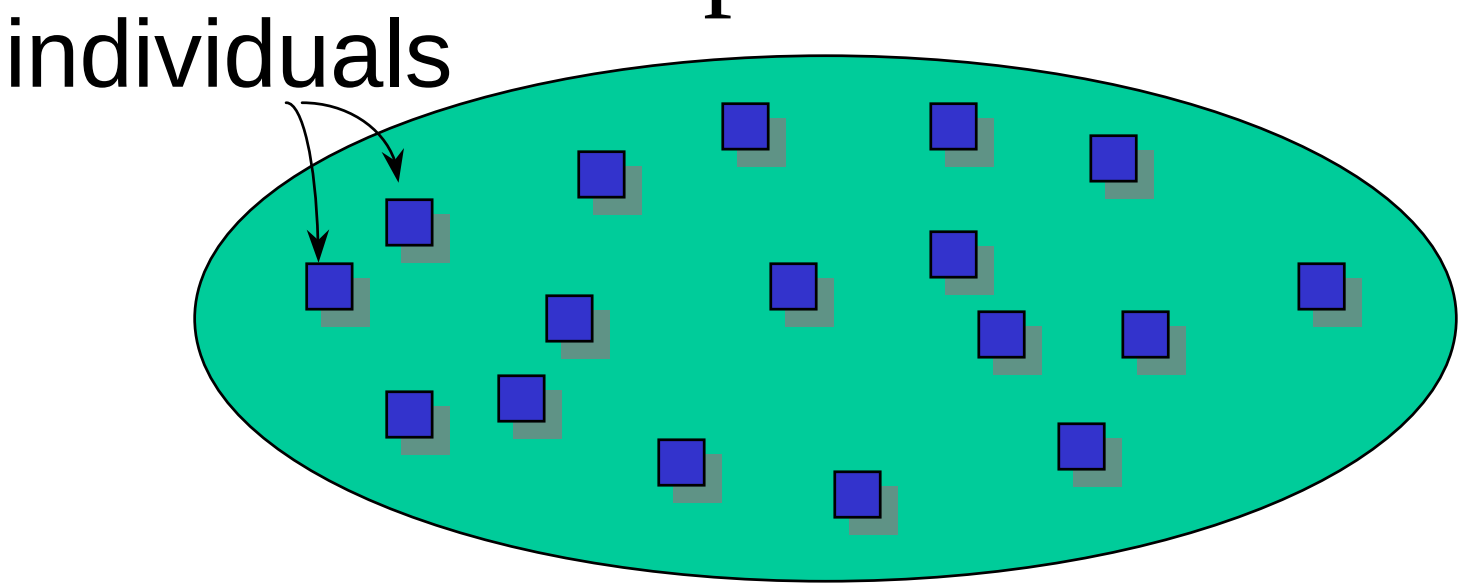
How Data are Obtained

- Observational Study
  - Observes individuals and measures variables of interest but does not attempt to influence the responses
  - Describes some group or situation
- Experiment (next Tuesday)
  - Deliberately imposes some treatment on individuals in order to observe their responses
  - Studies whether the treatment causes change in the response.

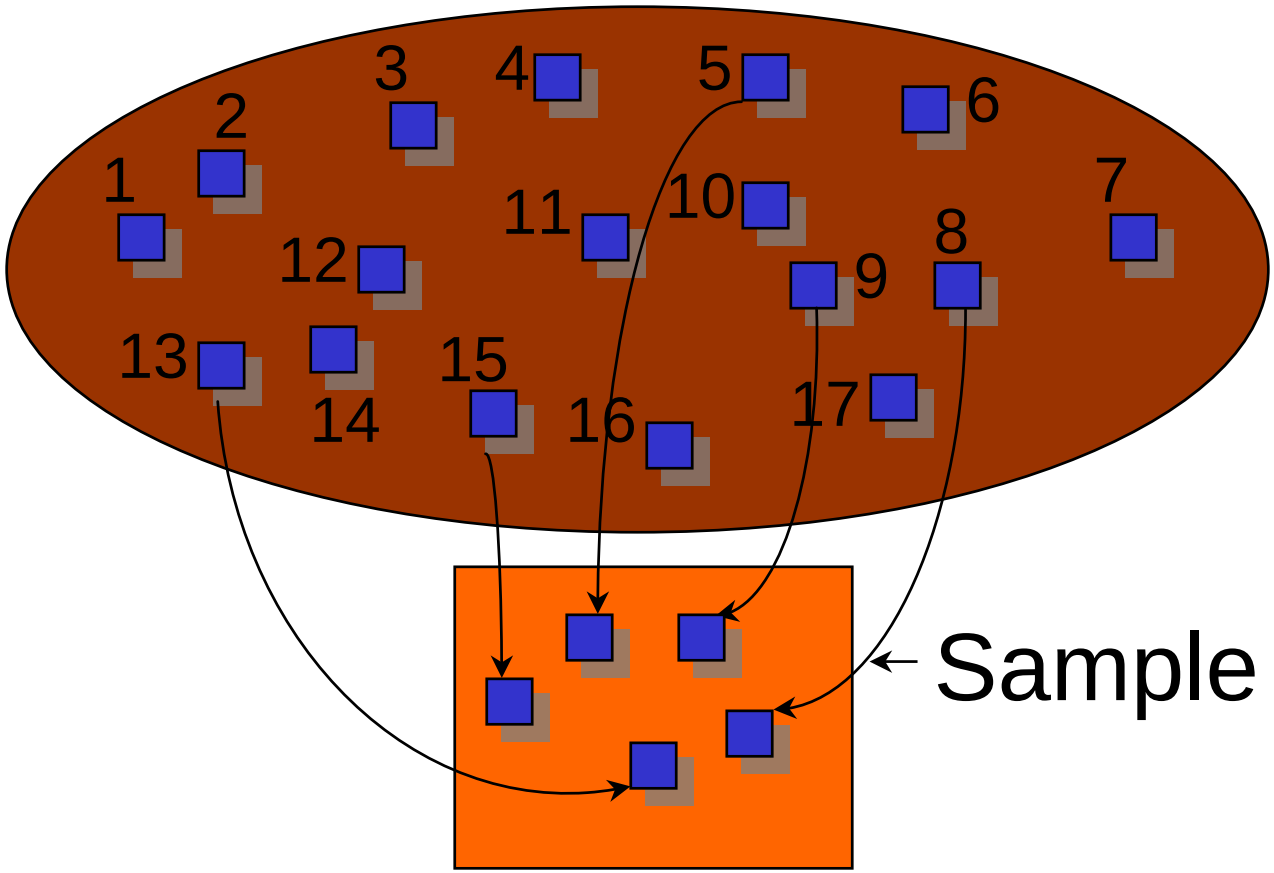
Common Language

- 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



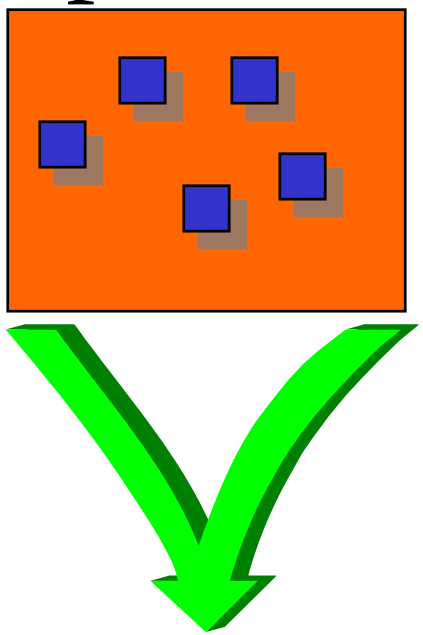
Sampling Frame



List of Individuals

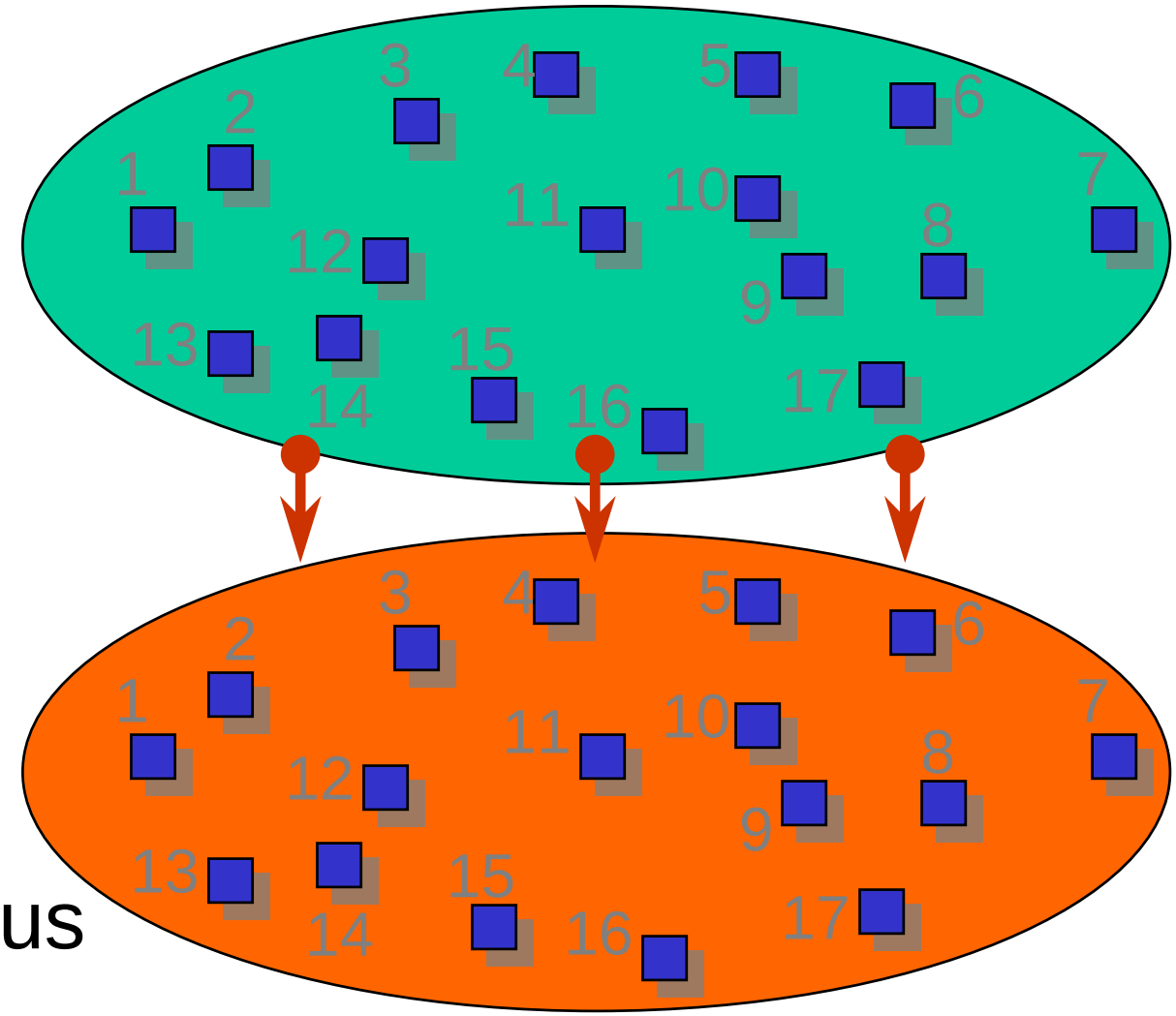
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Sample Survey



measurements
data

Census



Census

List of Individuals

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

Bad Sampling Plans

- Convenience sampling
  - selecting individuals that are easiest to reach
- Voluntary response sampling
  - allowing individuals to choose 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?

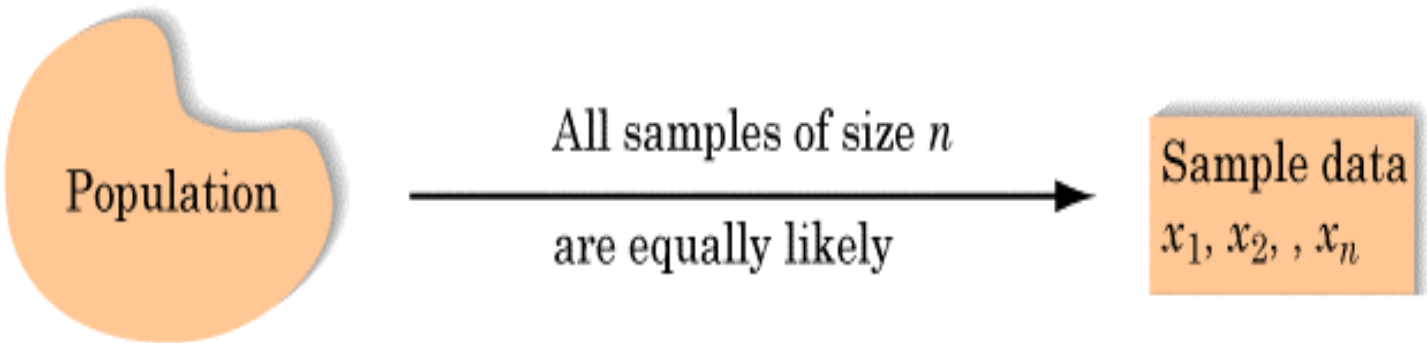
Voluntary Response

- 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

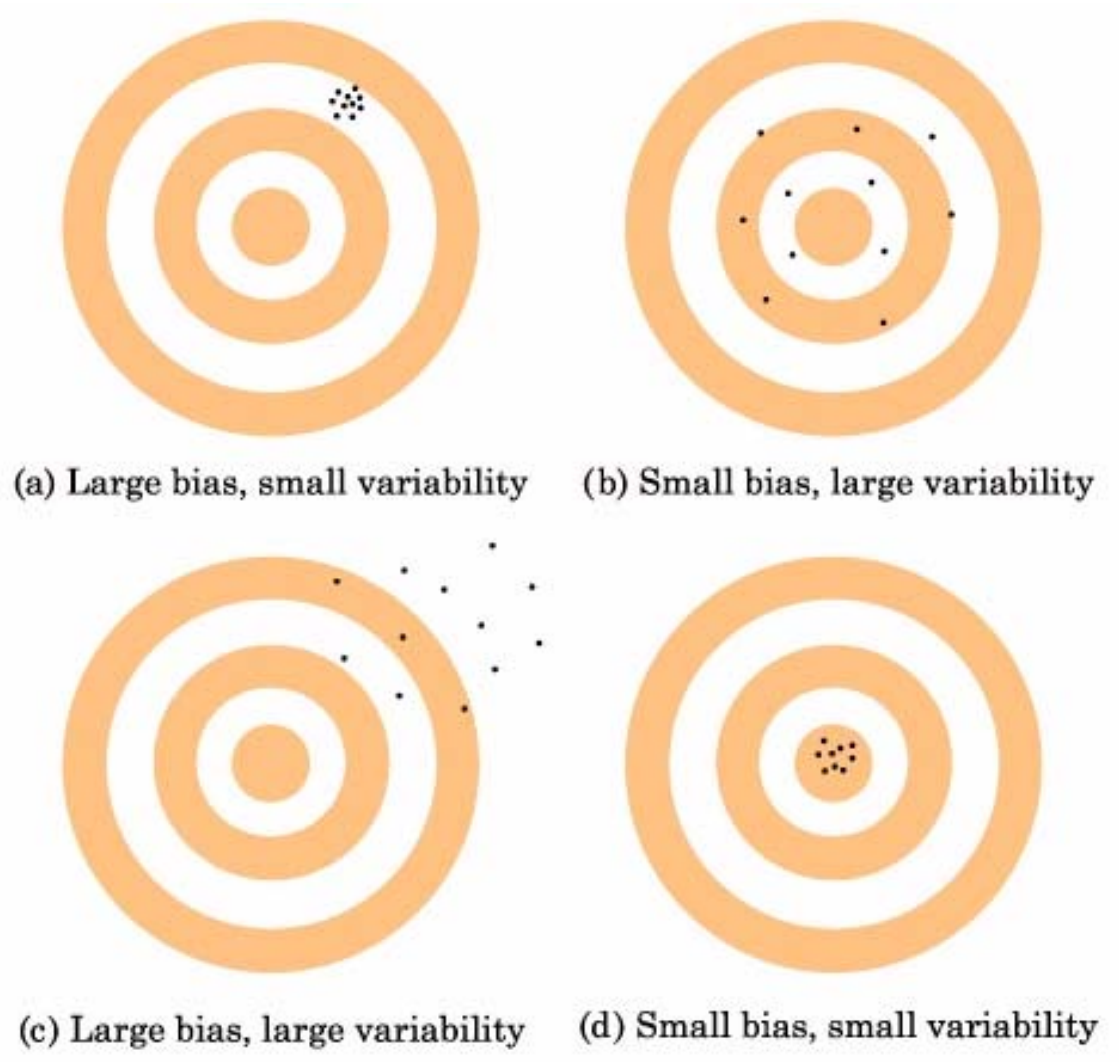
STATISTICS IN SUMMARY
Simple Random Sample



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 random sampling
- To reduce variability, use larger samples
  - 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*")

http://www.usaelectionpolls.com/2008/indiana.html

Barack Obama	48%
Hillary Clinton	40%
Unsure	8%
Other	5%

Zogby Poll
Date: 5/3-4
North Carolina
Added: 5/5/08
Est. MoE = 3.9% [?]

Zogby comments

http://www.usaelectionpolls.com/2008/polls/pdfs/zogby-indiana-north-carolina-may3to4-2008.pdf

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=643

*Assumes a 95% level of confidence

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

For those who have studied some statistics before

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 = $\sqrt{\frac{p*(1-p)}{n}}$

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).

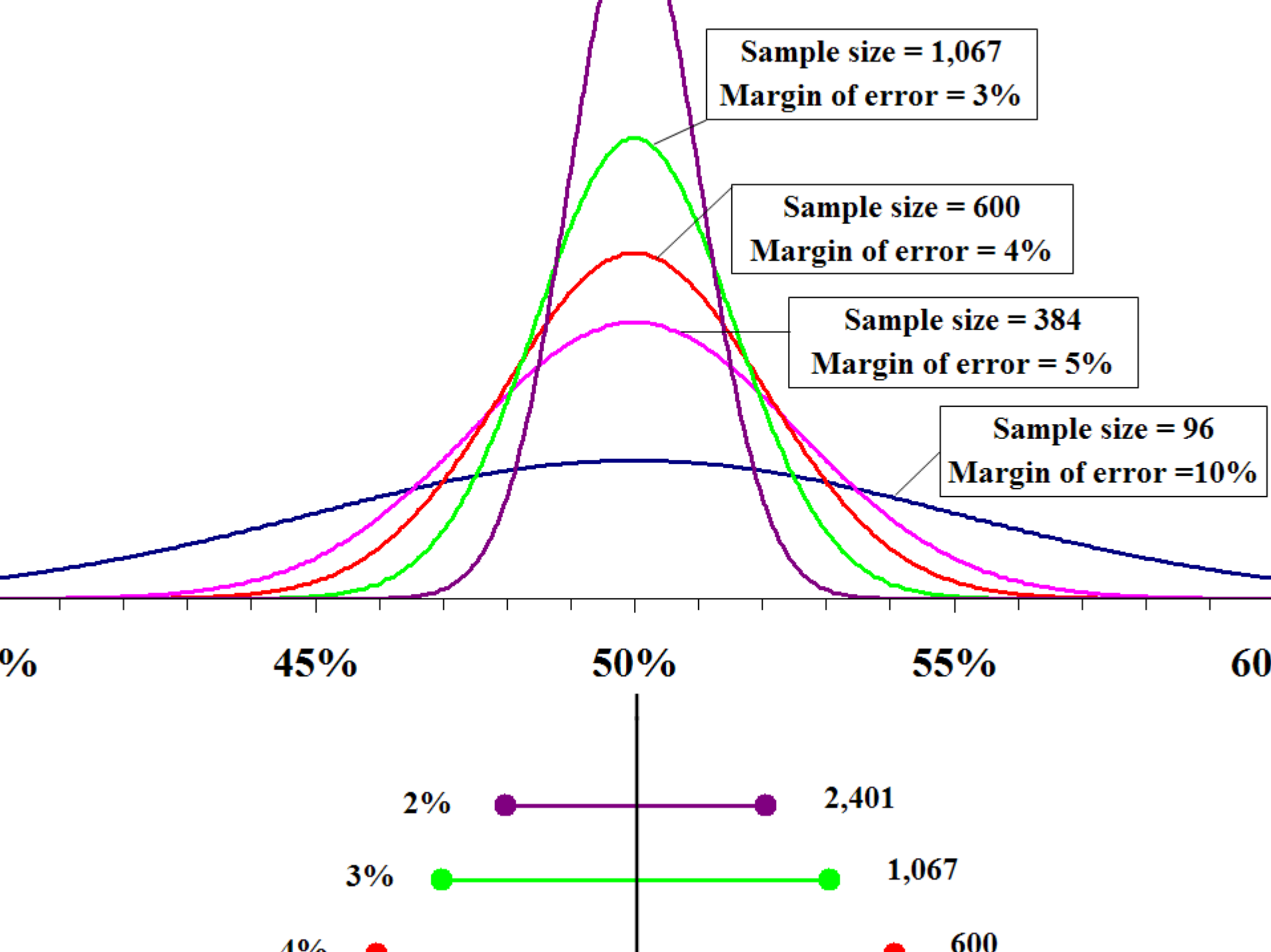
The maximum margin of error occurs when p = 50%.

(Maximum) margin of error (95%) = $1.96*\sqrt{\frac{.5*(1-.5)}{n}} = \frac{.98}{\sqrt{n}} \approx \frac{1}{\sqrt{n}}$

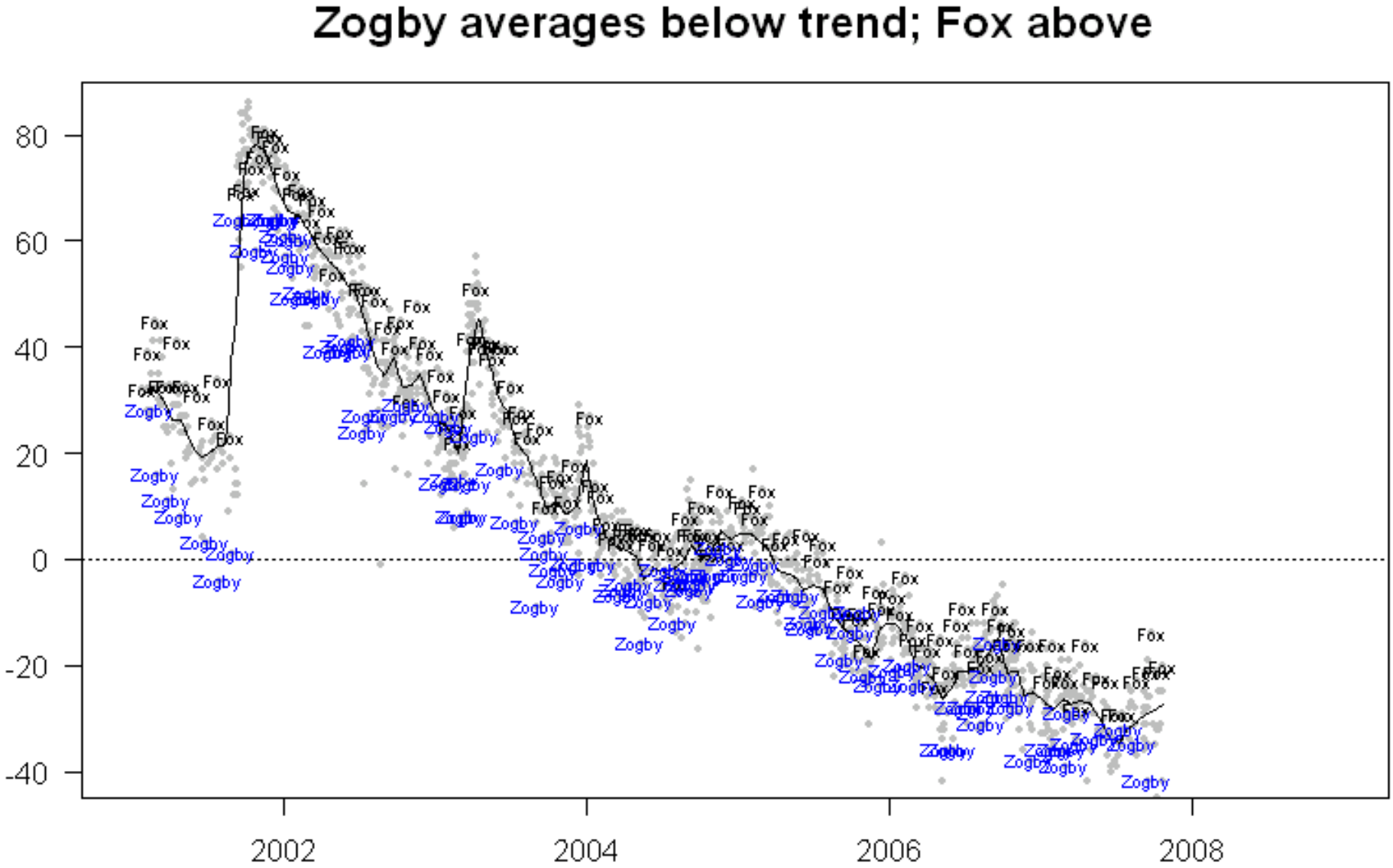
Margin of error at 99% confidence ≈ **1.29/√n**

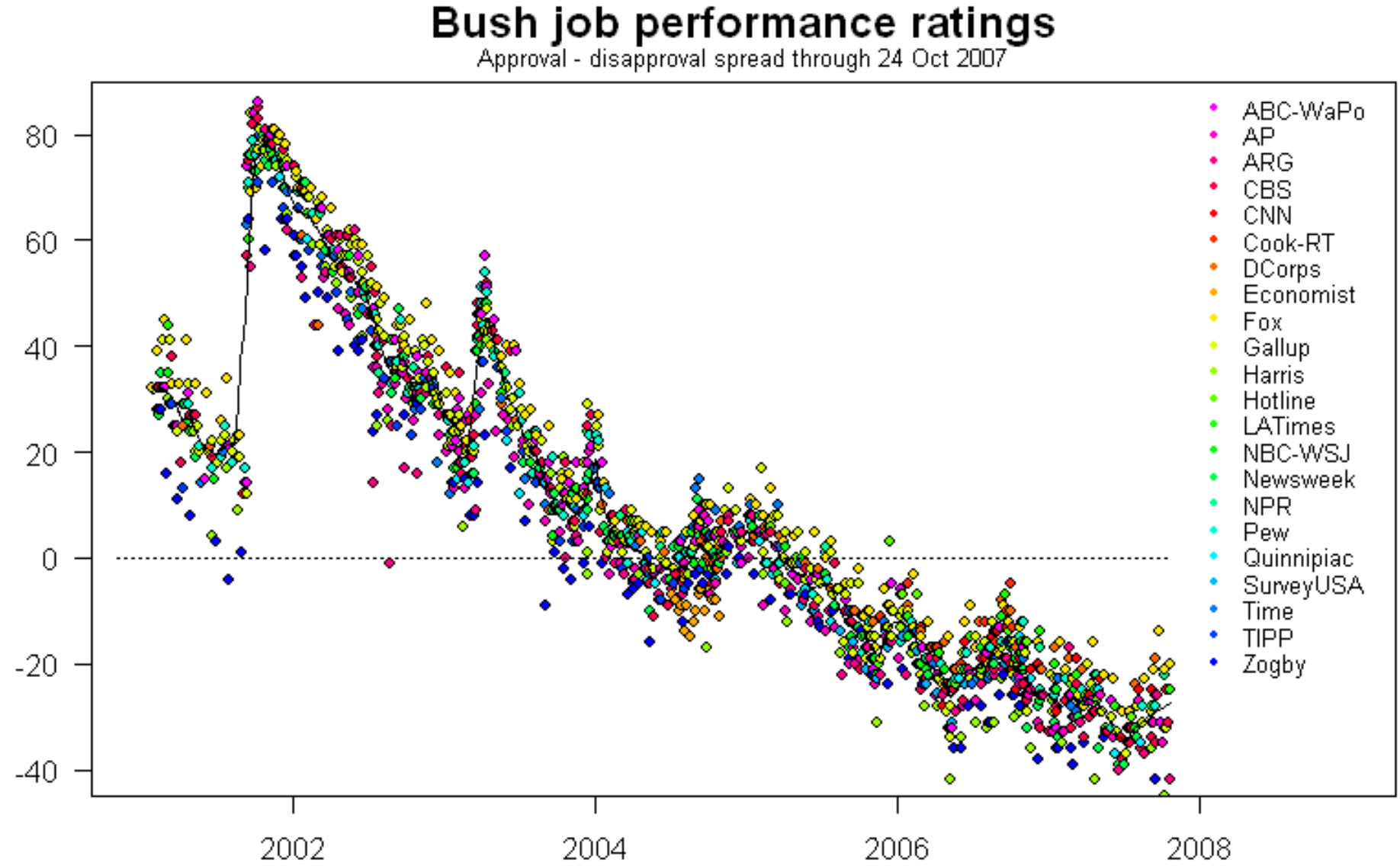
Margin of error at 95% confidence ≈ **.98/√n**

Margin of error at 90% confidence ≈ **.82/√n**


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%
%
45%
50%
55%
60
2%
2,401
3%
1,067
600

Tracking polls on Bush's job approval rating




Bush job performance ratings
Approval - disapproval spread through 24 Oct 2007
80
60
40
20
0
-20
-40
2002
2004
2006
2008
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


That's All Folks