How to Read Polls and Surveys

Jorge Luis Romeu, Ph.D.
Emeritus, Mathematics; SUNY Cortland
Research Professor, Syracuse University
Liverpool Masonic Temple
November 20, 2019

Definition of Survey Research

- The collection of data attained by asking individuals questions either in person, on paper, by phone or online …
- It is used to gather the opinions, beliefs and feelings of selected groups of individuals, often chosen for demographic sampling …

Electoral Season brings up *vote intention* polls (or surveys) that candidates, pundits and journalists use, and sometimes misuse, to assess who is ahead in each political contest. We often hear or read that "candidate X is ahead, with 47% of vote intention, to only 45% for his opponent." Such statement is statistically incomplete and incorrect.

For example, many voters use cell phones, often unlisted and difficult to reach by surveys, instead of landlines. Others use answering machines and caller ID devices as screening mechanisms. This limits a survey access to many potential survey subjects, something that can seriously invalidate, or bias, the survey findings.

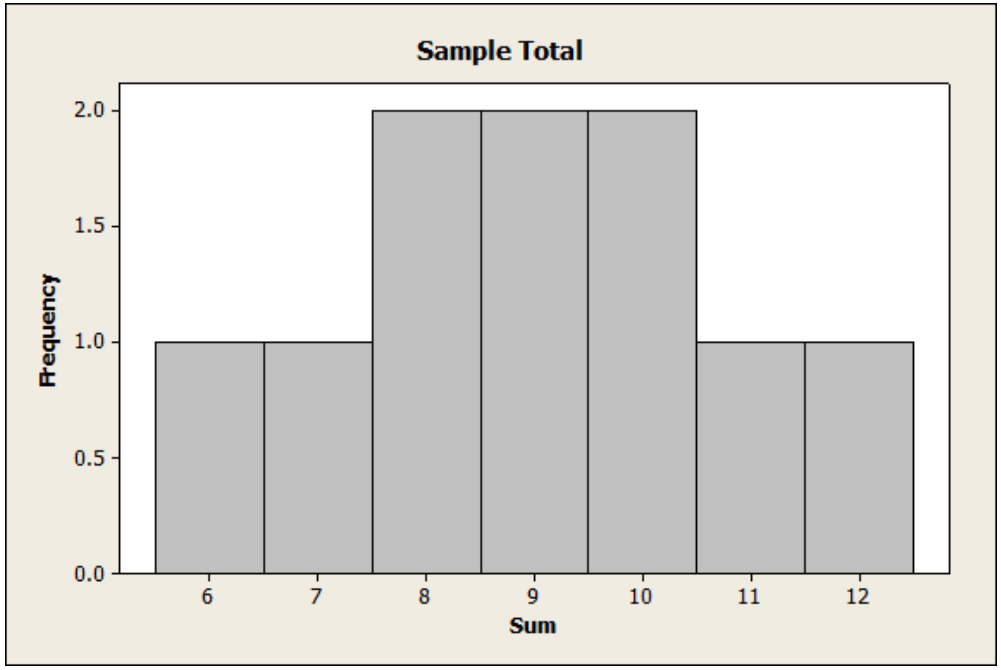
We will use an example of a Poll

- There are TWO containers with beans
  - One is metal (M) and the other one is plastic (P)
  - Different proportions of white/black beans
- Everyone will take a “random” sample of Ten
  - Count the number of BLACK beans in the sample
- Then, you write the results in the sheet
  - And pass the material to the next person

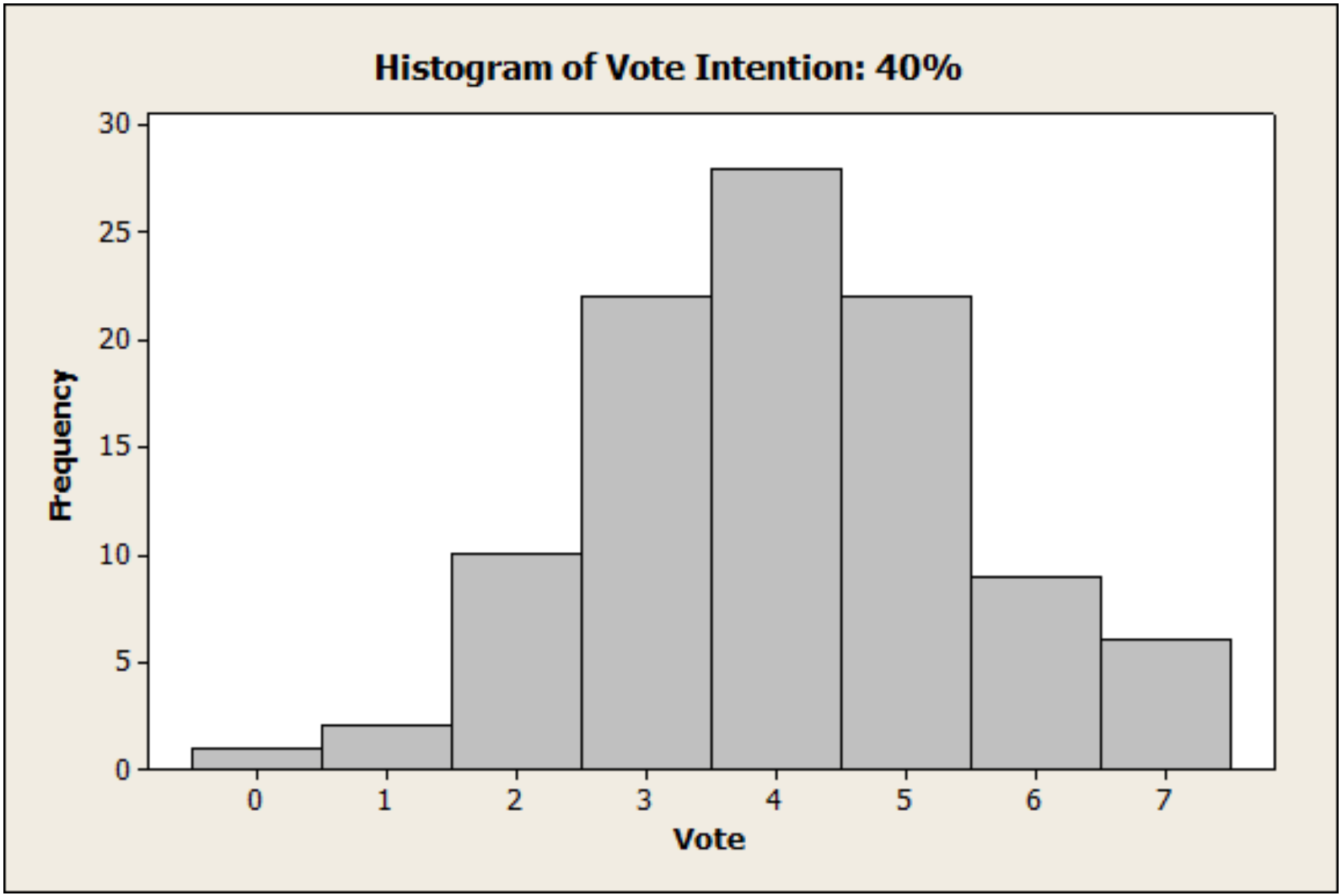
Person	Container	Gender	Black Beans
Number	Plastic /Metal	Male /Female	Out of ten
1	P	M	3
2	M	F	5

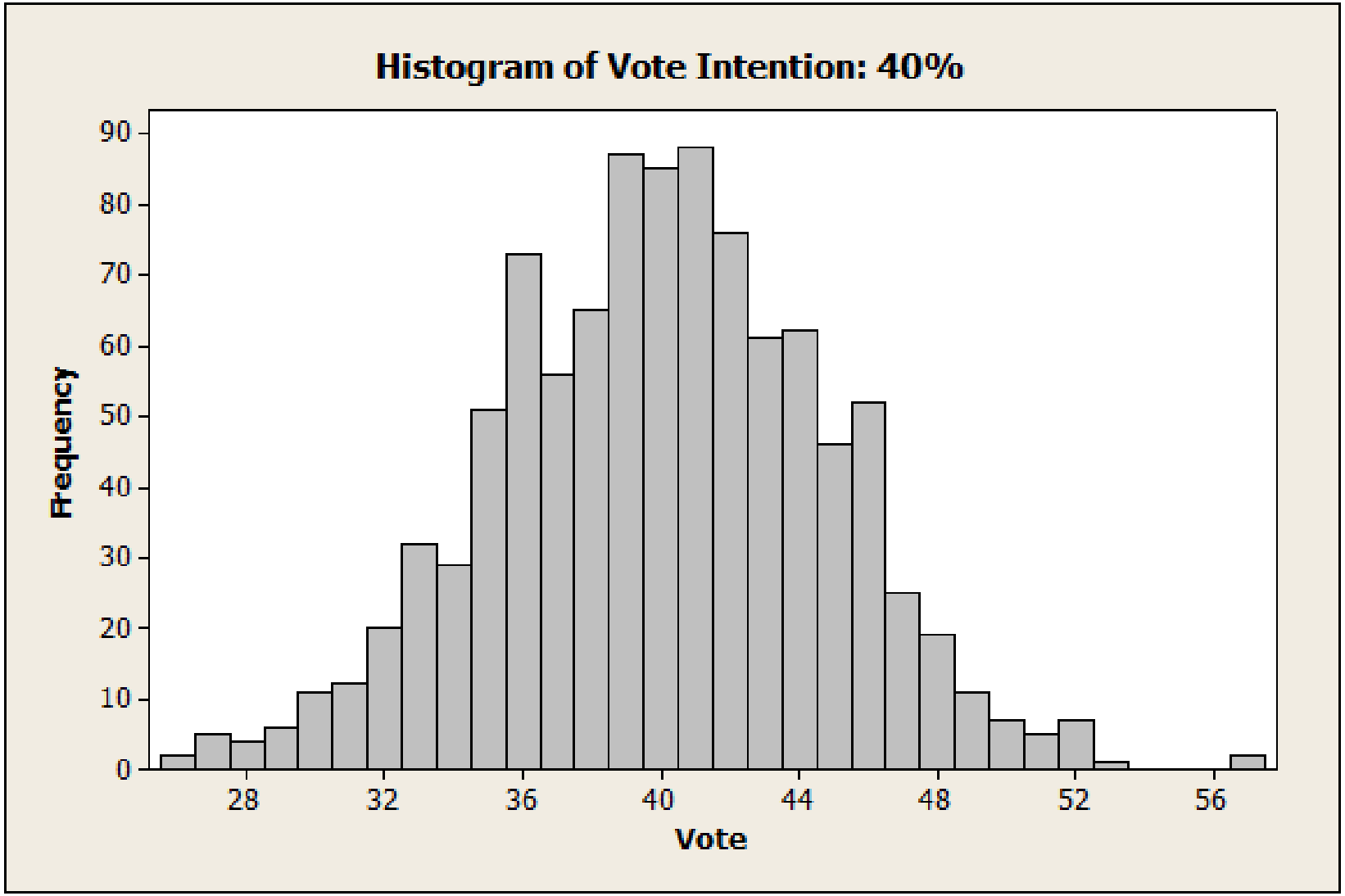
Poll Example: Population 5; Sample 3

First	Second	Third	Sum
1	2	3	6
1	2	4	7
1	2	5	8
1	3	4	8
1	3	5	9
1	4	5	10
2	3	4	9
2	3	5	10
2	4	5	11
3	4	5	12



Above are all possible samples of size 3, from a population of size 5. General Average = 9. If Margin of Error is 2, then 2 out of the 10 possible samples DO NOT cover the True Average of 9.


Histogram of Vote Intention: 40%
Frequency
30
25
20
15
10
5
0
0
1
2
3
4
5
6
7
Vote

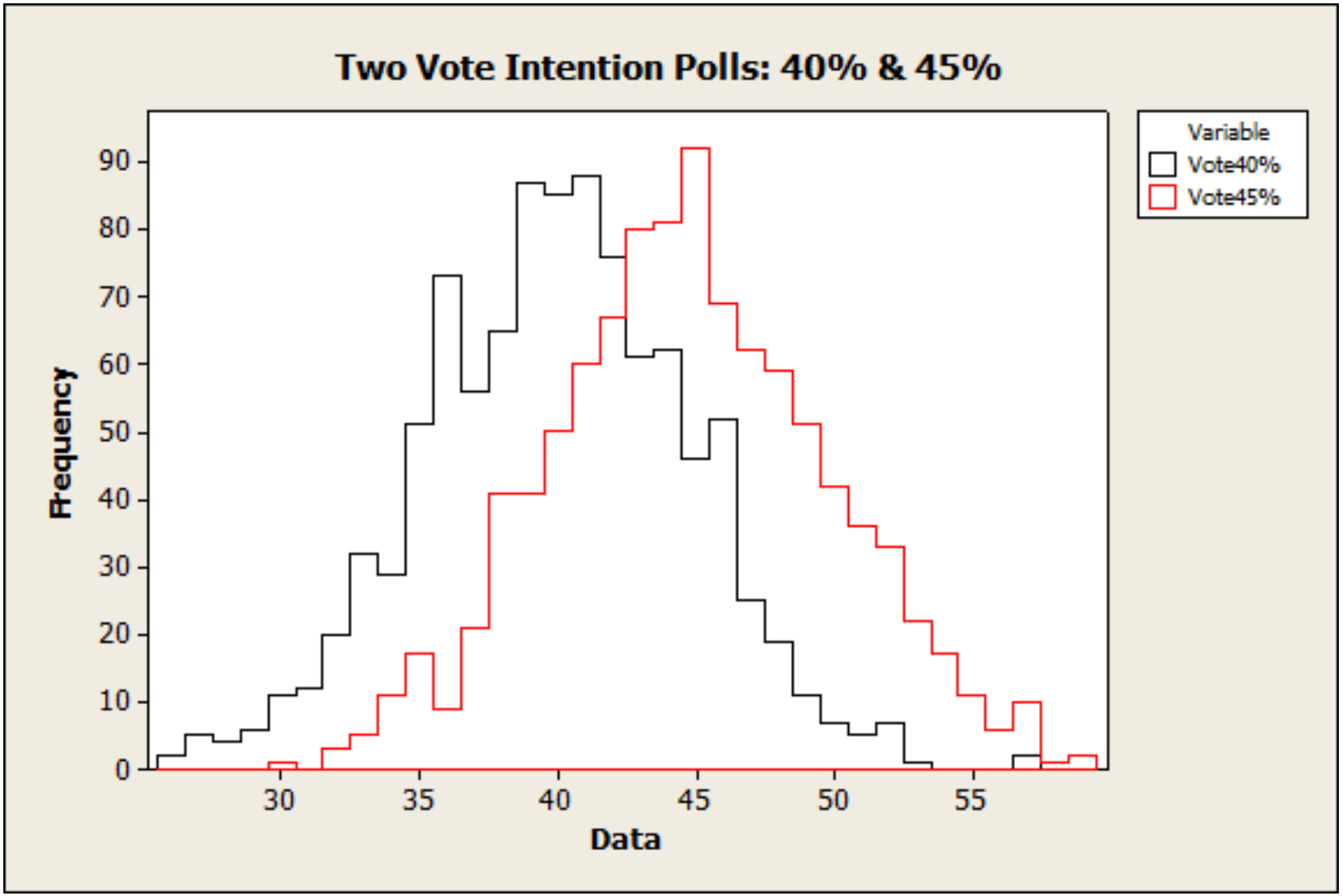

Histogram of Vote Intention: 40%
Frequency
0
10
20
30
40
50
60
70
80
90
28
32
36
40
44
48
52
56
Vote

Technical problems have resulted in some polls grossly overstating or understating the results of an election.

Bad polls include the ones about 2016 U.S. presidential election and the British referendum on Brexit.

Recent polls on the 2020 presidential campaign reached different conclusions, based on the survey sample.

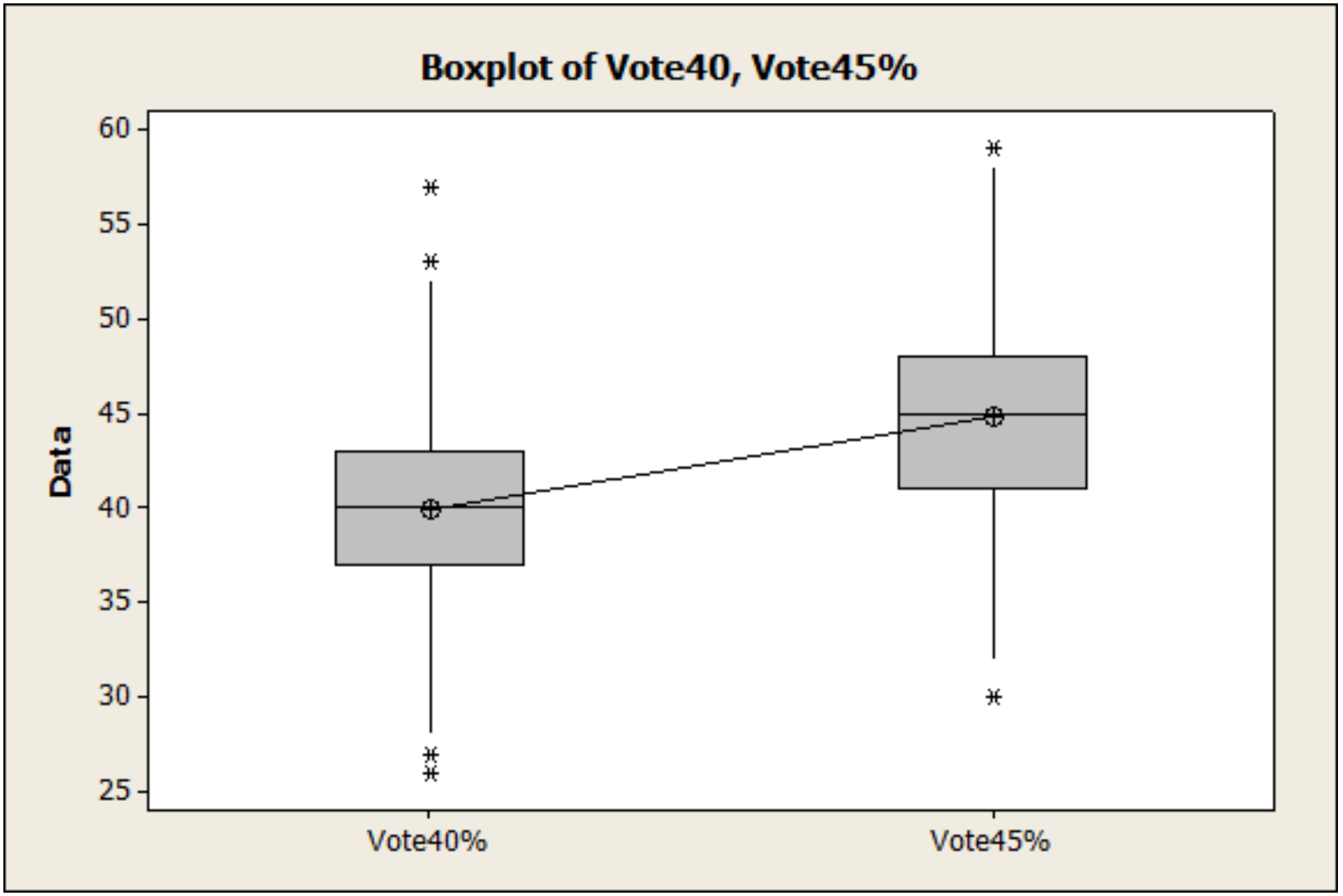
A Washington Post/ABC News national poll found President Donald Trump trailing the five top candidates for the Democratic nomination, while a New York Times/Siena College survey of battleground states released the same week found Trump ahead or tied.


Two Vote Intention Polls: 40% & 45%
Frequency
0
10
20
30
40
50
60
70
80
90
30
35
40
45
50
55
Data
Variable
Vote40%
Vote45%

Confidence: **internal conviction that some situation or statement will hold, is one of them.** For example, we may have 90% confidence that it will rain today, because in 90 out of the last 100 days, having the same humidity conditions as today, it has rained.

Randomness (by chance). In surveys we want to ***estimate the proportion or percent of the final vote***, for a candidate, *based upon a sample* of such voters.

But samples vary. And so do their results. Therefore, instead of giving a single center value for the percentage vote intention, **we derive a *confidence interval***.


Boxplot of Vote40, Vote45%
Data
60
55
50
45
40
35
30
25
Vote40%
Vote45%

Confidence intervals* define *a range of values where the final percent vote intention will be:
given the center value, plus/minus a *margin of error*.

Say a center value of 47%, plus or minus 3%, yielding 47-3=44% to 47+3=50%. The 3% is the *margin of error*, and depends on *sample size* (how many voters have been interviewed), and *coverage probability,* or percentage of times we assume the poll is correct.

We say "candidate X has a vote intention of 47%, with a margin of error of 3%, for a coverage or confidence of 95%." This means that 95% of the times that we take such poll, the final percentage for this candidate will be between 44% and 50%.

Poll results are valid if the sample of voters was taken randomly *from* ***all possible voters***. Otherwise, the poll results are invalid. Such occurred during the 1936 presidential election. A legendary poll result stated that Franklin Delano Roosevelt would be defeated. Pollsters took a telephone-based survey, at a time when only wealthy people could afford to own a phone. And wealthy people hated FDR with a passion. **Poll results inflated** the percent of wealthy voters.

Comparing voter intentions between two candidates: For the previous example, one candidate got 47% voter intention from the poll. And the opposite candidate got 45%. The difference between the two (47-45=2) is **within the *margin of error*** *of* 3%. **Results constitute *a statistical tie***.

Margin of error or *coverage are* ***not always disclosed****.* The Washington Post and do not give *margin of error*. New York Times polls states that Biden leads Trump by 3 percentage points. But, are these within the margin of error?

For two polls of the same sample size, the margin of error increases as we increase coverage probability. For two polls with the same coverage probability, the margin of error decreases as we increase sample size**. Always request margin of error and coverage probability with a poll result**.

Finally, if the election is very close and polls are unable to distinguish which candidate is really ahead, we need to take a larger sample -- or to wait until Election Day to find out!

Questions