WESTMINSTER

International University in Tashkent

Week 6
Point and interval estimation

By

Olmos Isakov

Office hours: Tuesday, 09:00 – 11:00 (ATB 216)

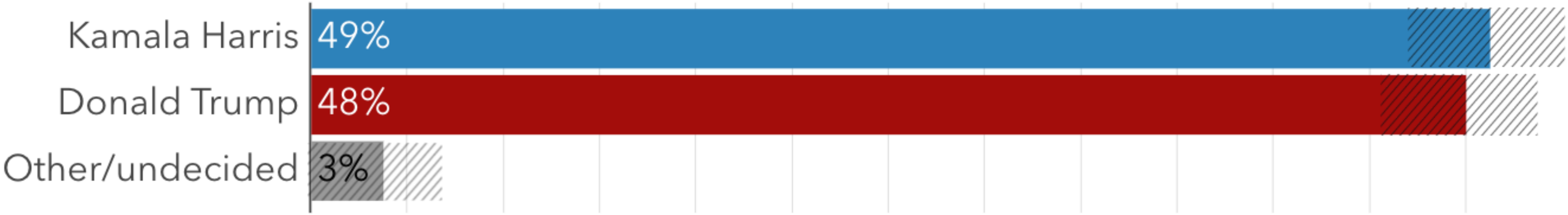
AGENDA

- Point estimation for μ and π;
- Confidence Interval for population mean (μ);
- Confidence Interval for population proportion (π);
- Confidence Interval for population means difference (μ_1-μ_2);
- Confidence Interval for population proportions difference ($\pi_1 - \pi_2$).

2024 Pre-Election poll:

GEORGIA

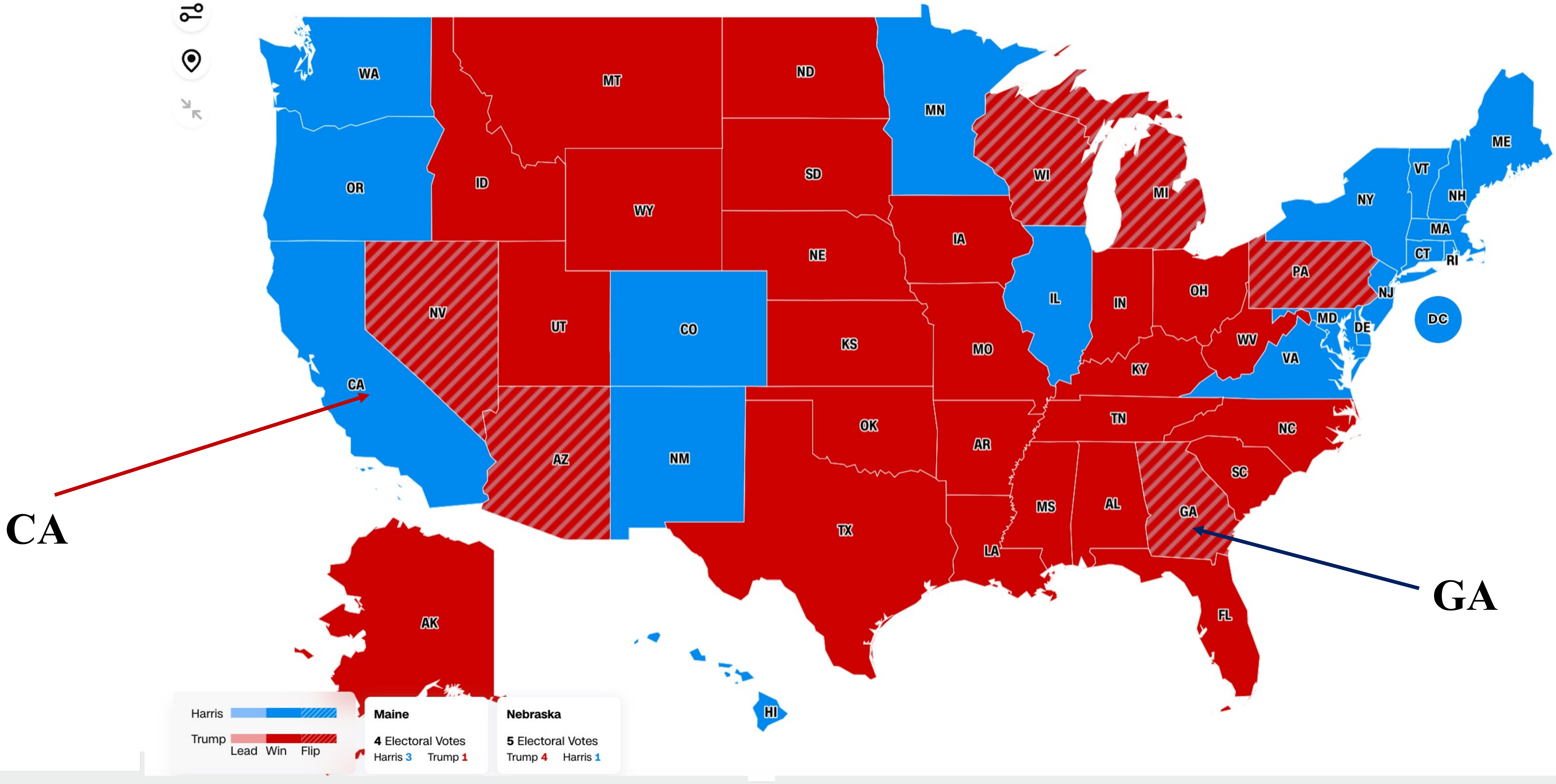
Source: Emerson College Polling

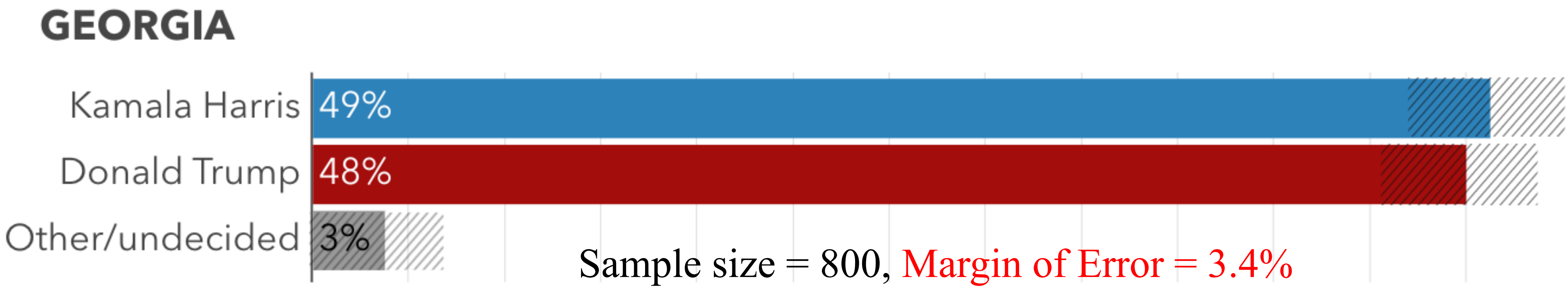


CALIFORNIA

Source	Date	Sample	Harris	Trump	Other
Emerson College	10/16/2024	1,000 LV ±3%	61%	37%	2%

Election results



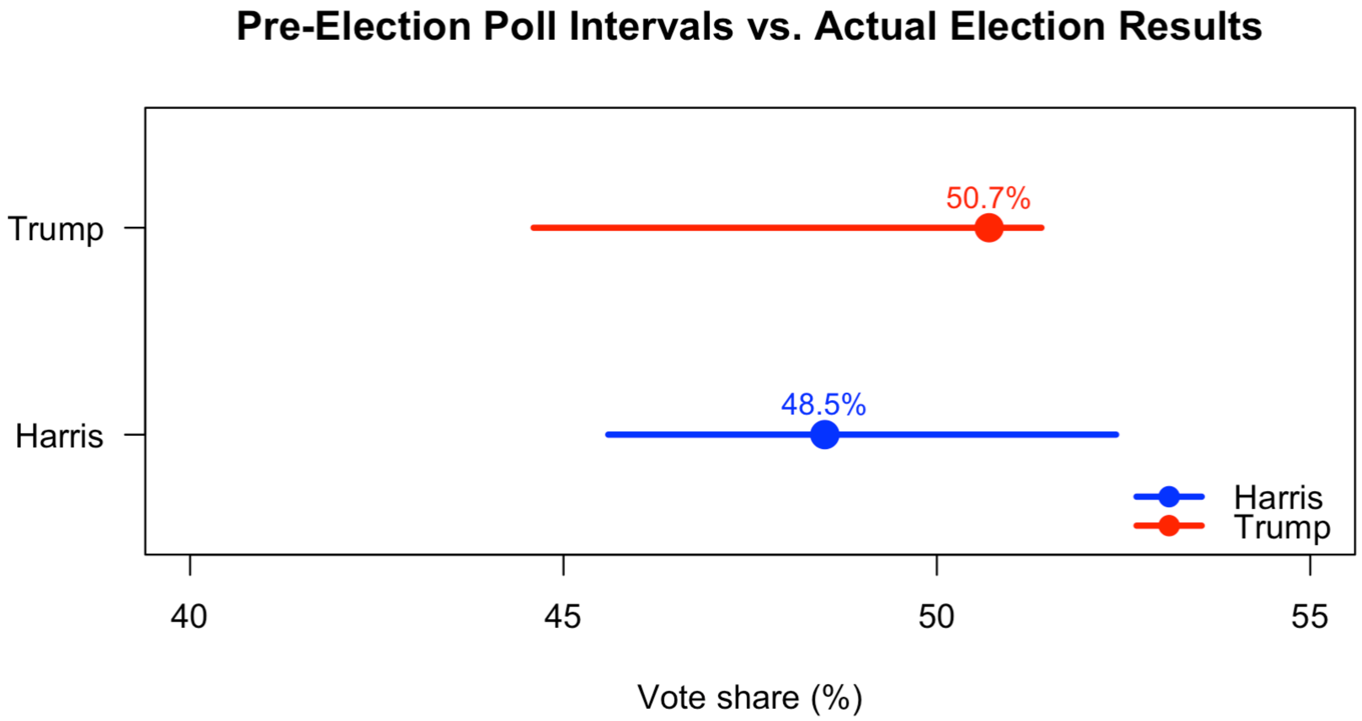


Sample size = 800, Margin of Error = 3.4%
Source: Emerson College Polling

Before election:

% for Harris: 49 ± 3.4 =
(45.6%, 52.4%)

% for Trump: 48 ± 3.4 =
(44.6%, 51.4%)



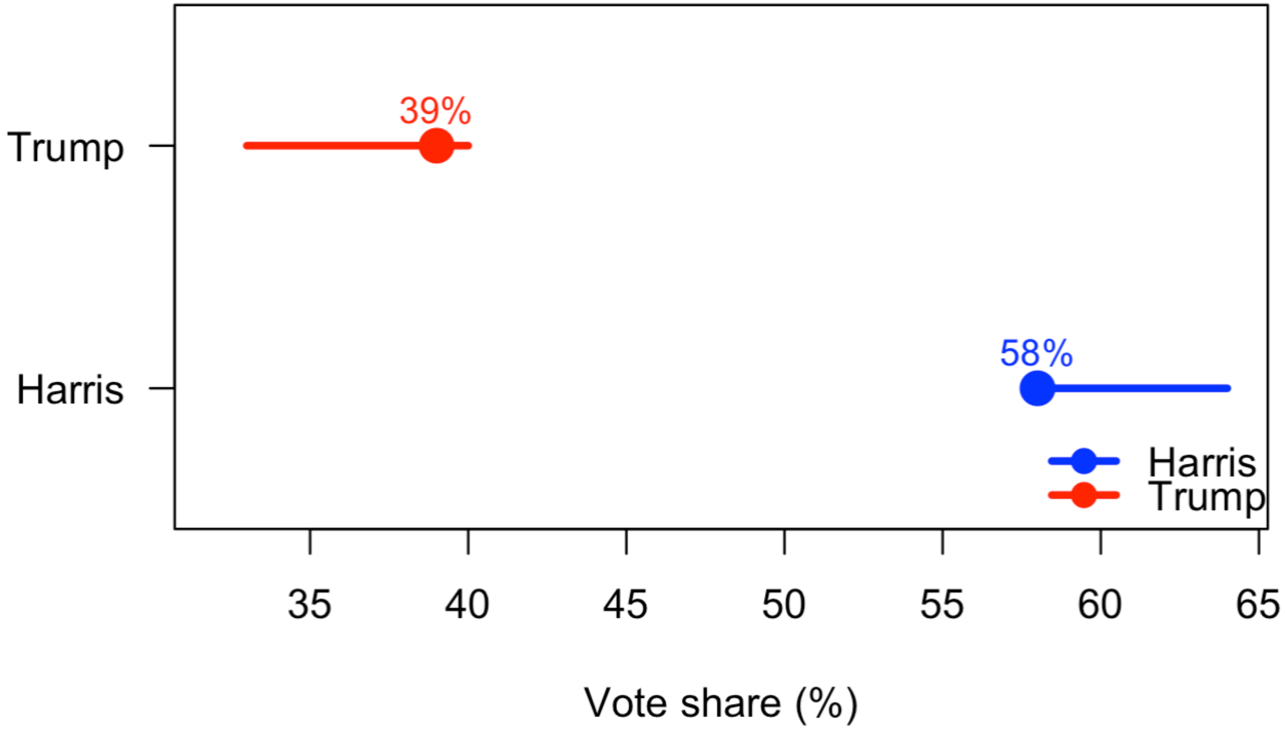
California

Source	Date	Sample	Harris	Trump	Other
Emerson College	10/16/2024	1,000 LV ±3%	61%	37%	2%

Before Election:
Harris: $(61 \pm 3) \Rightarrow (58\%, 64\%)$
Trump: $(37 \pm 3) \Rightarrow (33\%, 40\%)$

After Election:
Harris: 58%
Trump: 39%

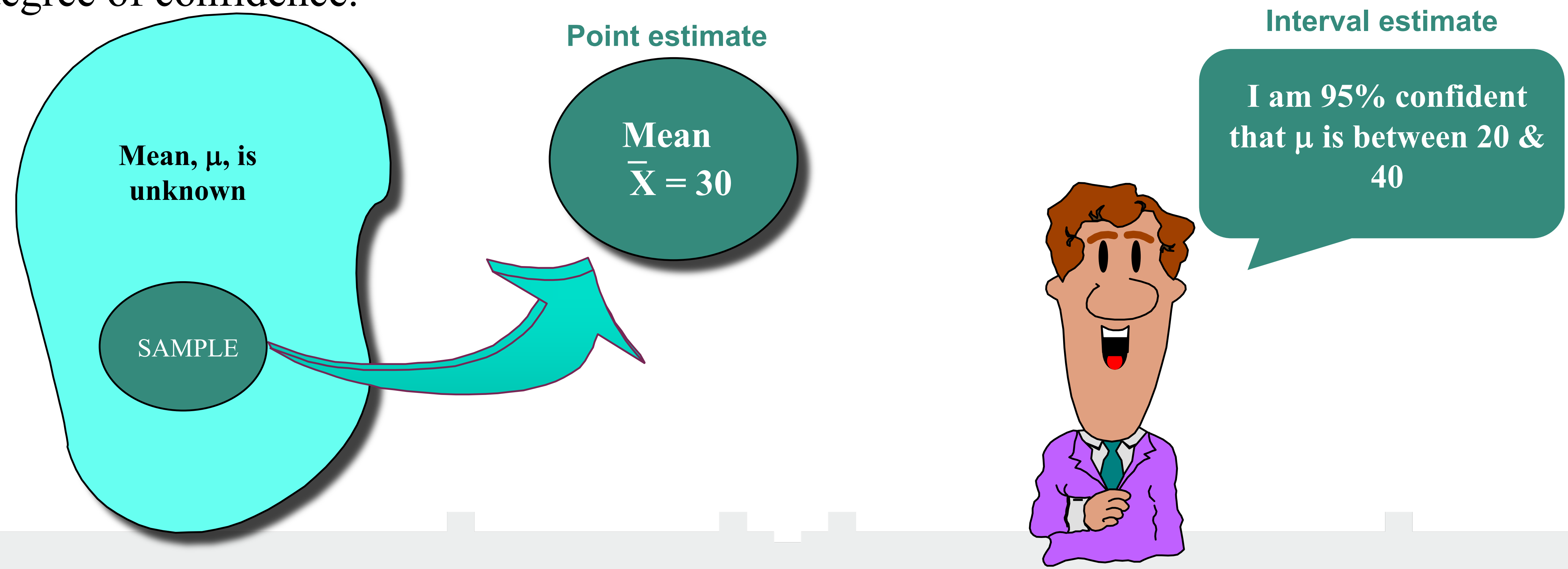
Pre-Election Poll Intervals vs. Actual Election Results



Estimation of the parameter

A single number calculated from the sample that estimates a target population parameter is called a **point estimator**.

Interval estimator - a range of numbers that contain the target parameter with a high degree of confidence.



Confidence interval for population mean (μ)

When σ is known:

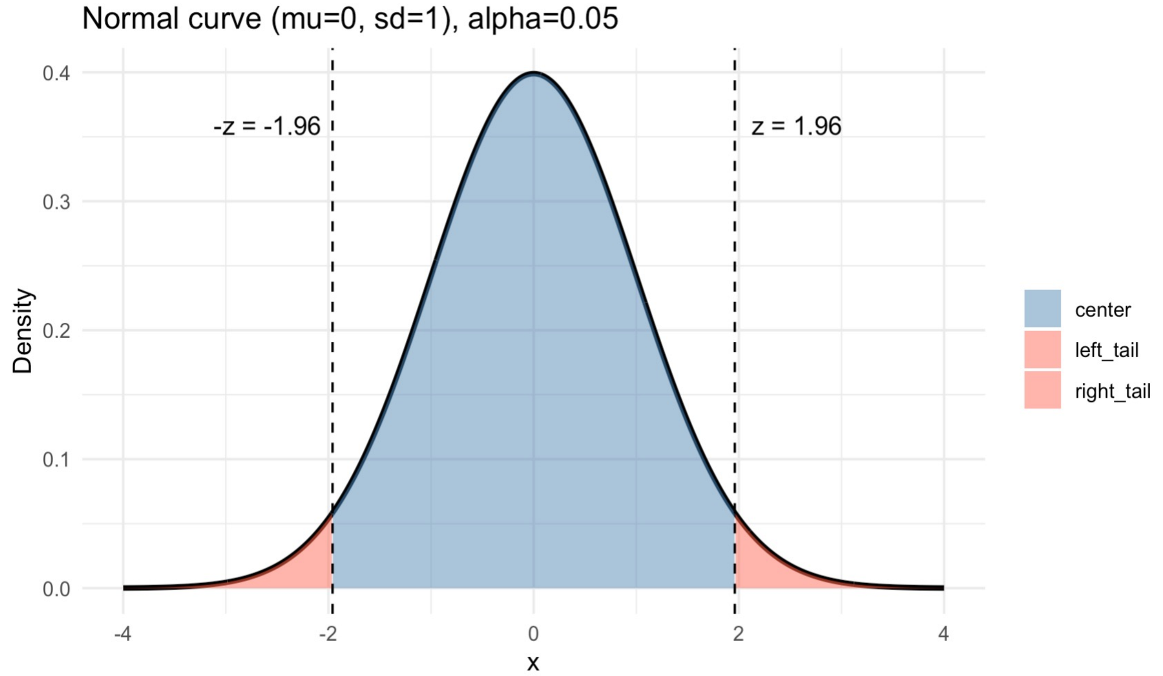
100(1-α)% Confidence Interval for **μ:**

$$\bar{x} \pm \underbrace{z_{\alpha/2}{}^{*} \frac{\sigma}{\sqrt{n}}}_{\text{Margin of Error (E or ME)}}$$

When σ is unknown (n ≥ 30):

100(1-α)% Confidence Interval for **μ:**

$$\bar{x} \pm \underbrace{z_{\alpha/2}{}^{*} \frac{s}{\sqrt{n}}}_{\text{Margin of Error (E or ME)}}$$



Example

You want to study housing costs in Tashkent by sampling recent apartment sales in various (representative) regions of Tashkent city. Your goal is to provide a 95% confidence interval estimate of the housing cost. Previous studies suggest that the population standard deviation is about \$7,200. A random sample of 64 apartments had a mean of \$61,000.

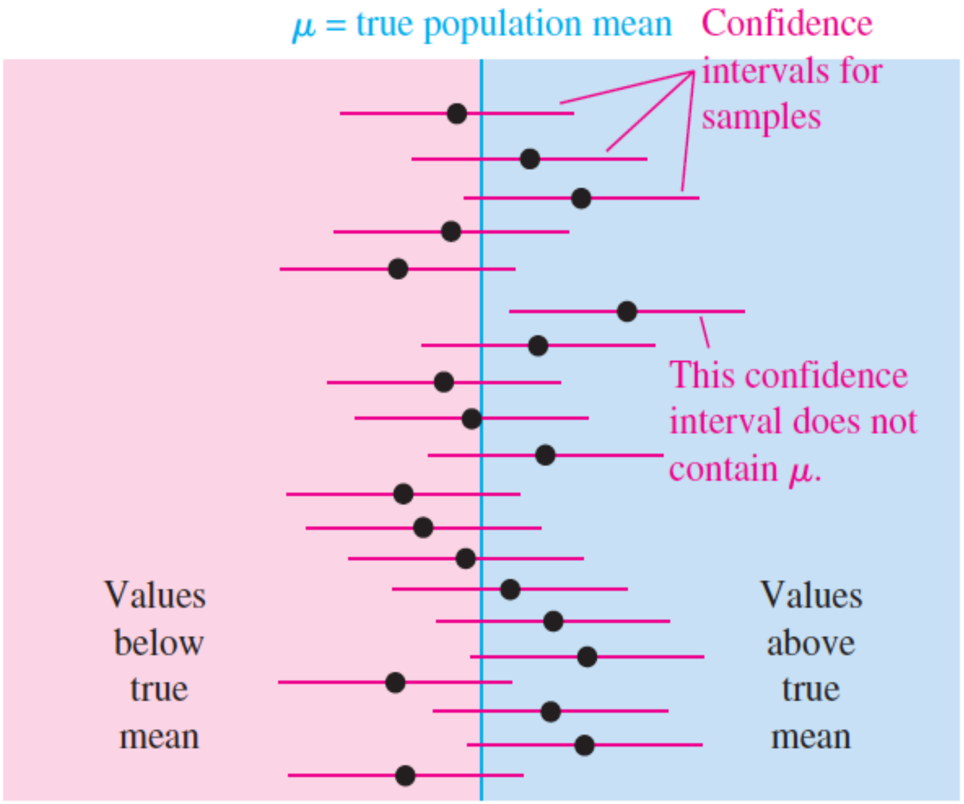
Solution. The ***standard error*** and the ***margin of error*** are as follows:

$$\mathbf{SE} = \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}} = \frac{\mathbf{7200}}{\sqrt{\mathbf{64}}} = \mathbf{900}$$

ME = 1.96*SE = 1.96*900 = 1,764

Lower confidence limit (LCL): 61000 - 1764 = \$59,236
Upper confidence limit (UCL): 61000 + 1764 = \$62,764

Hence, 95% confidence interval for μ: (\$59,236; \$62,764)



Confidence interval for a population proportion (π)

Sampling distribution of the sample proportion estimator, P

The sampling distribution of the sample proportion, P, (which is the estimator of the population proportion, π) is:

$$P \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

approximately, as $n \to \infty$.

Confidence interval endpoints of a single proportion

A $100(1-\alpha)\%$ confidence interval for a single proportion has endpoints:

$$p \pm z_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}} \quad \Rightarrow \quad \left(p - z_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}},\ p + z_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}}\right)$$

where $z_{\alpha/2}$ is the z-value which cuts off $100\alpha/2\%$ probability in the upper tail of the standard normal distribution

Example

A random survey was conducted by Emerson College Polling to find out which candidate was favorite to win the US election in the state of Georgia. Out of 800 surveyed, 392 favored Kamala Harris, 384 favored Donald Trump and the rest indicated the independent candidate. If we denote the population proportion of voters who support Harris by π, and then what is the 95% confidence interval for population proportion?

SOLUTION:

$p = 392/800 = 0.49$ and
$z_{\alpha/2} = 1.96$ for 95% CI.

$$0.49 \pm 1.96 * \sqrt{\frac{0.49 * 0.51}{800}} = 0.49 \pm 0.0346$$

So, 95% CI for π: (0.455, 0.525) or (45.5%, 52.5%)

Choosing minimum sample size

- You want to limit your Margin of Error (ME)
- How many observations (at minimum) do you need for your study?

For mean

$$n = \left(\frac{\sigma * z_{\alpha/2}}{ME}\right)^2$$

For proportion*

$$n = p(1\text{-}p)\left(\frac{z_{\alpha/2}}{ME}\right)^2$$

*if p is unknown, use p = 0.5.

n needs to be rounded up.

Example: sample size

You want to study housing costs in Tashkent by sampling recent house sales in various (representative) districts. Your goal is to provide a 90% confidence interval estimate of the housing cost. Previous studies suggest that the population standard deviation is about $7,200. What sample size (at a minimum) should be used to ensure that the sample mean is within $1000 of the true population mean?

$$n = \left(\frac{7200*1.645}{1000}\right)^2 = 140.28 \qquad n \approx 141\ (\textit{rounded up})$$

Confidence Interval for Two Population mean difference (μ_1 - μ_2)

When σ is known:

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

When σ is unknown,
provided $n_1, n_2 \geq 30$:

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Example: *σ unknown*

Independent random samples of 40 BIS students and 50 FIN students were selected. The average GPA for the random sample of BIS and FIN students were found to be 3.08 and 2.98 respectively. If the sample standard deviations for BIS and FIN students are 0.42 and 0.54 respectively, find a 95% confidence interval for the mean difference (μ_{BIS}-μ_{FIN}).

SOLUTION:

$n_{BIS} = 40 \qquad \bar{x}_{BIS} = 3.08 \qquad \sigma_{BIS} = 0.42$

$n_{FIN} = 50 \qquad \bar{x}_{FIN} = 2.98 \qquad \sigma_{FIN} = 0.54$

$z_{\alpha/2} = z_{0.025} = 1.96$ for 95% CI.

The Confidence Interval for μ_{BIS}-μ_{FIN} :

$$(\overline{x_a} - \overline{x_b}) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}} = (3.08 - 2.98) \pm 1.96\sqrt{\frac{0.42^2}{40} + \frac{0.54^2}{50}} = 0.10 \pm 0.198$$

CI: (-0.098, 0.298)

Confidence Interval for Two Population Proportions difference (π_1-π_2)

Confidence interval endpoints for the difference between two proportions

With point estimates of π_1 and π_2 of $p_1 = r_1/n_1$ and $p_2 = r_2/n_2$, respectively, a $100(1-\alpha)\%$ confidence interval for the difference between two proportions has endpoints:

$$p_1 - p_2 \pm z_{\alpha/2} \times \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

where $z_{\alpha/2}$ cuts off $100\alpha/2\%$ probability in the upper tail of the standard normal distribution

Example: CI for π_1-π_2

A survey is conducted by a bank to estimate the proportion of its customers who would be interested in using a proposed new mobile banking service. In a random sample of 120 bank customers from 18-64 age group, 107 indicated that they were interested. In an independent random sample of 141 customers who were 65 or older, only 73 showed interest in the new service. The respective population proportions are denoted π_A and π_B. Find a 99% confidence interval for the population difference.

SOLUTION:

$$p_A = 107/120 = 0.892 \textit{ and } p_B = 73/141 = 0.518$$

99% confidence interval for $(\pi_A - \pi_B)$:

$$0.892 - 0.518 \pm 2.575 * \sqrt{\frac{0.892(1-0.892)}{120} + \frac{0.518(1-0.518)}{141}}$$

$$\mathbf{0.374 \pm 0.131 \Rightarrow (0.243, 0.505)}$$

REFERENCES

1. Lind et al. (ISBN 978-1-260-18750-2), Chapter 9.
2. McClave & Sincich (ISBN 978-0-321-75593-3), Chapter 7.
3. Ott & Longnecker (ISBN 978-0-495-01758-5), Chapter 6.

Thank You!