PROBABILITY & STATISTICS

Lecture 11 – Confidence Intervals

CONFIDENCE INTERVALS

MOTIVATION

- Point estimation provides a single “best guess”.

MOTIVATION

- Point estimation provides a single “best guess”.
- How confident can we be in our estimation?

MOTIVATION

- Point estimation provides a single “best guess”.
- How confident can we be in our estimation?
- CI proposes a range of plausible values.

DEFINITION

A $1 - \alpha$ confidence interval for a parameter θ is an interval $C_n = (a, b)$ such that $\mathrm{T}_1 = t_1(X_1, \dots, X_n)$, $T_2 = t_2(X_1, \dots, X_n)$ and

$$P(T_1 < \theta < T_2) \geq 1 - \alpha$$

DEFINITION

A $1-\alpha$ confidence interval for a parameter θ is an interval $C_n = (T_1, T_2)$ such that $\mathrm{T}_1 = t_1(X_1, \dots, X_n)$, $T_2 = t_2(X_1, \dots, X_n)$ and

$$P(T_1 < \theta < T_2) \geq 1 - \alpha$$

- **Random** intervals: T_1 and T_2 are functions of random samples.

DEFINITION

A $1-\alpha$ confidence interval for a parameter θ is an interval $C_n = (T_1, T_2)$ such that $\mathrm{T}_1 = t_1(X_1, \dots, X_n)$, $T_2 = t_2(X_1, \dots, X_n)$ and

$$P(T_1 < \theta < T_2) \geq 1-\alpha$$

- **Random** intervals: T_1 and T_2 are functions of random samples.
- θ is unknown, but fixed
  T_1 and T_2 are random

INTERPRETATION

$$C_n = (T_1, T_2):\ \ P(T_1 < \theta < T_2) \geq 1 - \alpha$$

INTERPRETATION

$$C_n = (T_1, T_2):\ \ P(T_1 < \theta < T_2) \geq 1 - \alpha$$

- **Not** a probability statement about θ since it's fixed.

INTERPRETATION

$$C_n = (T_1, T_2): \;\; P(T_1 < \theta < T_2) \geq 1 - \alpha$$

- **Not** a probability statement about θ since it's fixed.
- Common interpretation:

INTERPRETATION

$$C_n = (T_1, T_2): \;\; P(T_1 < \theta < T_2) \geq 1 - \alpha$$

- **Not** a probability statement about θ since it's fixed.

- Common interpretation:

If I repeat the experiment many times, the interval will contain the true value of θ 95% of the time (α=0.05).

CI FOR NORMAL DATA

CI for μ, *known* σ

CI FOR μ, KNOWN σ

$X_1, X_2, \dots, X_n$ − samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ?

CI FOR μ, KNOWN σ

$X_1, X_2, \ldots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ? Consider $\bar{X}$:

CI FOR μ, KNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim$$

CI FOR μ, KNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\qquad\qquad) \quad \Rightarrow$$

CI FOR μ, KNOWN σ

$X_1, X_2, \dots, X_n$ − samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \qquad\qquad) \quad \Rightarrow$$

CI FOR μ, KNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow$$

CI FOR μ, KNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow \qquad\qquad\qquad \sim N(0,1)$$

CI FOR μ, KNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

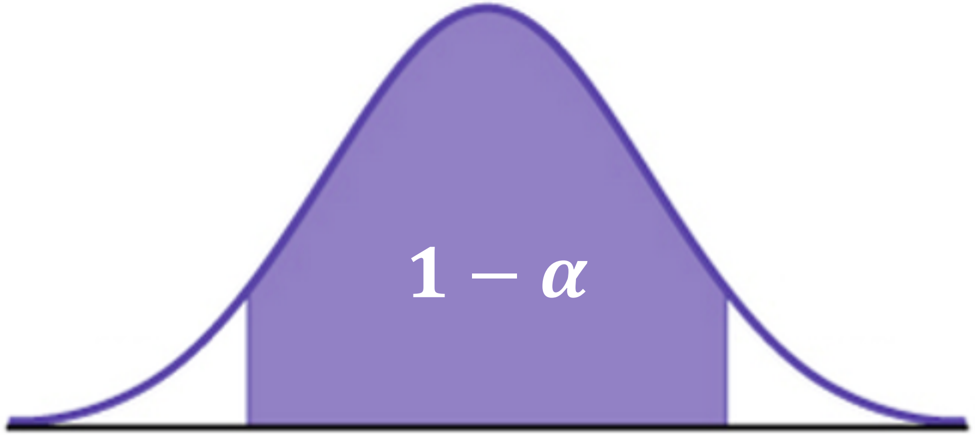
CI FOR μ, KNOWN σ

$X_1, X_2, \ldots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ? Consider $\bar{X}$:

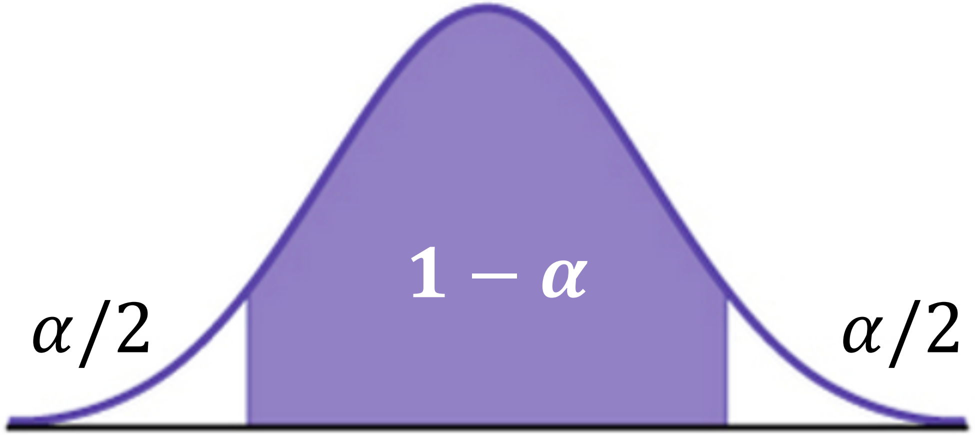
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

$$P\left(\qquad < \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} < \qquad \right) = 1 - \alpha$$

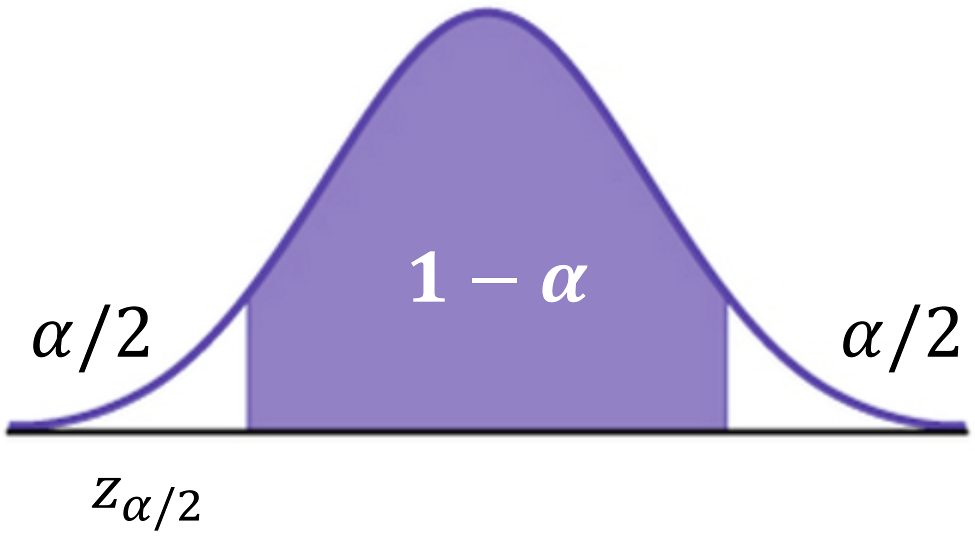
CI FOR μ, KNOWN σ



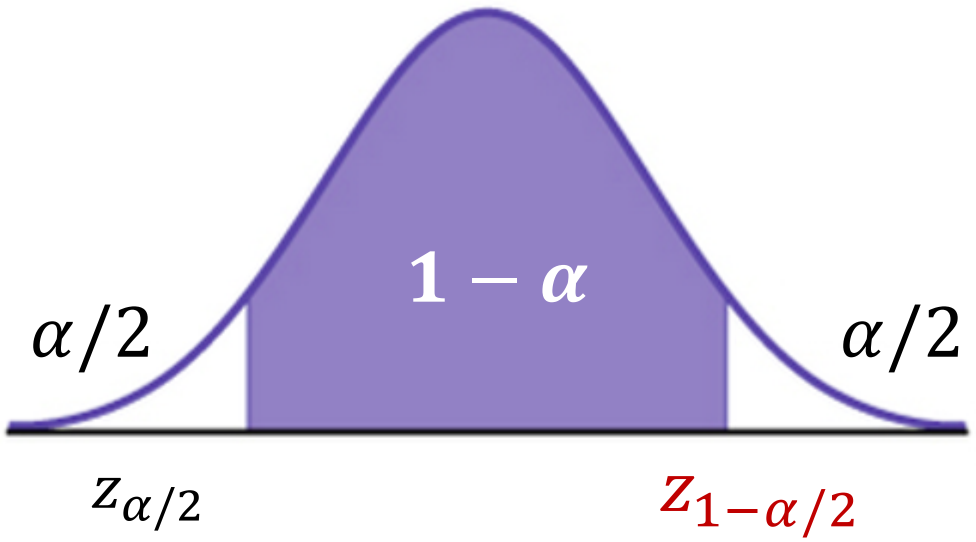
CI FOR μ, KNOWN σ



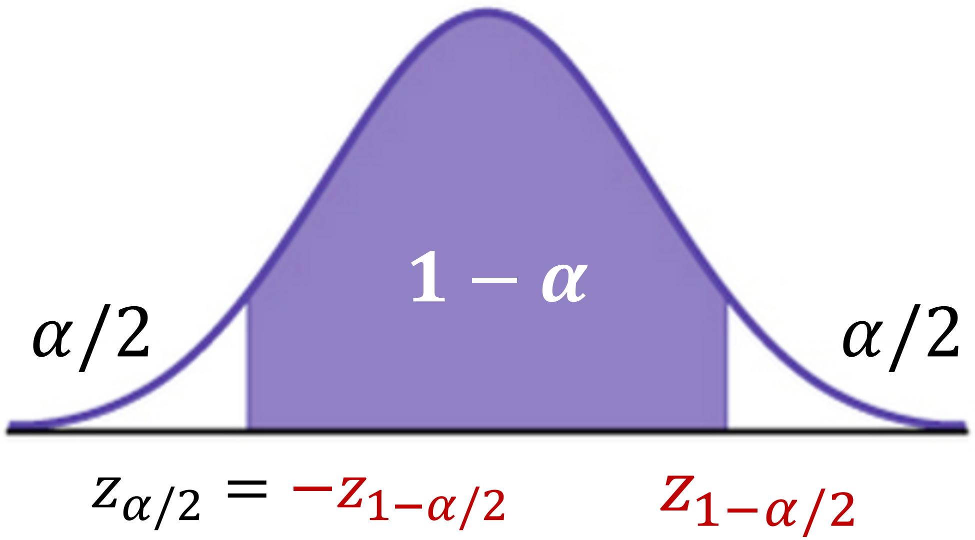
CI FOR μ, KNOWN σ



CI FOR μ, KNOWN σ



CI FOR μ, KNOWN σ



CI FOR μ, KNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} < z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

CI FOR μ, KNOWN σ

$X_1, X_2, \ldots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is known.
How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} < z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\boldsymbol{\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}}\right) = 1-\alpha$$

CI FOR μ, KNOWN σ: EXAMPLE 1

- $X_1, X_2, \ldots, X_{100}$ – samples from $N(\mu, 3^2)$, $\bar{X} = 12.$
- Give the 95%-CI for μ.

CI FOR μ, KNOWN σ: EXAMPLE 1

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, 3^2)$, $\bar{X} = 12.$
- Give the 95%-CI for μ.

- Interval:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

CI FOR μ, KNOWN σ: EXAMPLE 1

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, 3^2)$, $\bar{X} = 12.$
- Give the 95%-CI for μ.

- Interval:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05,$$

CI FOR μ, KNOWN σ: EXAMPLE 1

- $X_1, X_2, \ldots, X_{100}$ – samples from $N(\mu, 3^2)$,
- Give the 95%-CI for μ.

- Interval:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

Quantile (p)	$\Phi^{-1}(p, 0, 1)$
0.995	2.58
0.99	2.33
0.975	1.96
0.95	1.64
0.9	1.28

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05, \qquad z_{1-0.025} =$$

CI FOR μ, KNOWN σ: EXAMPLE 1

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, 3^2)$,
- Give the 95%-CI for μ.

Quantile (p)	$\Phi^{-1}(p, 0, 1)$
0.995	2.58
0.99	2.33
0.975	1.96
0.95	1.64
0.9	1.28

- Interval:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05, \qquad z_{1-0.025} =$$

CI FOR μ, KNOWN σ: EXAMPLE 1

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, 3^2)$,
- Give the 95%-CI for μ.

Quantile (p)	$\Phi^{-1}(p, 0, 1)$
0.995	2.58
0.99	2.33
0.975	1.96
0.95	1.64
0.9	1.28

- Interval:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05, \qquad z_{1-0.025} = 1.96$$

CI FOR μ, KNOWN σ: EXAMPLE 1

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, 3^2)$, $\bar{X} = 12.$
- Give the 95%-CI for μ.

- Interval:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05, \qquad z_{1-0.025} = 1.96$$

$$12 \pm \frac{3}{\sqrt{100}} \cdot 1.96$$

CI FOR μ, KNOWN σ: EXAMPLE 2

- $X_1, X_2, \ldots, X_{100}$ – samples from $N(\mu, 1^2)$, $\quad \bar{X} = 5.$
- Give the 90%-CI for μ.

CI FOR μ, KNOWN σ: EXAMPLE 2

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, 1^2)$, $\bar{X} = 5$.
- Give the 90%-CI for μ.

- Interval:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

Quantile (p)	$\Phi^{-1}(p, 0, 1)$
0.995	2.58
0.99	2.33
0.975	1.96
0.95	1.64
0.9	1.28

CI FOR μ, KNOWN σ: EXAMPLE 2

- $X_1, X_2, \ldots, X_{100}$ – samples from $N(\mu, 1^2)$, $\bar{X} = 5.$
- Give the 90%-CI for μ.

- Interval:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

Quantile (p)	$\Phi^{-1}(p, 0, 1)$
0.995	2.58
0.99	2.33
0.975	1.96
0.95	1.64
0.9	1.28

CI FOR μ, KNOWN σ: EXAMPLE 2

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, 1^2)$, $\quad \bar{X} = 5.$
- Give the 90%-CI for μ.

- Interval:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

$$1 - \alpha = 0.90 \Rightarrow \alpha = 0.1, \qquad z_{1-0.05} = 1.64$$

CI FOR μ, KNOWN σ: EXAMPLE 2

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, 1^2)$, $\bar{X} = 5$.
- Give the 90%-CI for μ.

- Interval:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

$$1 - \alpha = 0.90 \Rightarrow \alpha = 0.1, \qquad z_{1-0.05} = 1.64$$

$$5 \pm \frac{1}{\sqrt{100}} \cdot 1.64$$

PRACTICE!

Google Classroom -> z-intervals

CONFIDENCE INTERVALS: RECAP

- NORMAL DISTRIBUTION:

$$X_1, X_2, \dots, X_n - \text{i.i.d. samples}$$

  - CI for μ, σ is known: **z-interval**

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}, \quad z_{1-\alpha/2} - \text{quantile from } N(0,1)$$

CI FOR NORMAL DATA

CI for μ, *unknown* σ

CI FOR μ, UNKNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ?

CI FOR μ, UNKNOWN σ

$X_1, X_2, \ldots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

CI FOR μ, UNKNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$$

CI FOR μ, UNKNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

CI FOR μ, UNKNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

CI FOR μ, UNKNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

Approximate with sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$:

CI FOR μ, UNKNOWN σ

$X_1, X_2, \ldots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

Approximate with sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$:

$$\frac{(\bar{X}-\mu)\sqrt{n}}{s}$$

CI FOR μ, UNKNOWN σ

$X_1, X_2, \dots, X_n$ − samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

Approximate with sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$:

$$\frac{(\bar{X}-\mu)\sqrt{n}}{s} \sim t(n-1)$$ − Student distribution

STUDENT DISTRIBUTION

William Sealy Gosset

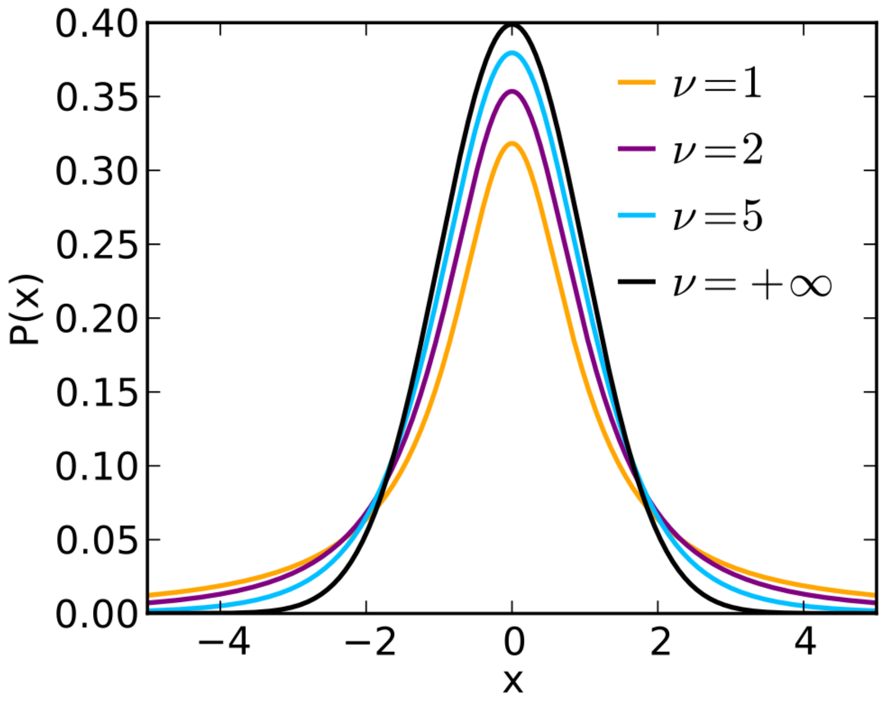
STUDENT DISTRIBUTION

$$\frac{(\bar{X}_n - \mu)\sqrt{n}}{s} \sim t(n-1)$$

William Sealy Gosset

STUDENT DISTRIBUTION

$$\frac{(\bar{X}_n - \mu)\sqrt{n}}{s} \sim t(n-1)$$



William Sealy Gosset

CI FOR μ, UNKNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\frac{(\bar{X}-\mu)\sqrt{n}}{s} \sim t(n-1) \text{ – Student distribution}$$

CI FOR μ, UNKNOWN σ

$X_1, X_2, \ldots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\frac{(\bar{X}-\mu)\sqrt{n}}{s} \sim t(n-1) \text{ – Student distribution}$$

Before: $\mathrm{P}\left(-z_{1-\alpha/2} < \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} < z_{1-\alpha/2}\right) = 1-\alpha$

CI FOR μ, UNKNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\frac{(\bar{X}-\mu)\sqrt{n}}{s} \sim t(n-1) \text{ – Student distribution}$$

Before: $$\mathrm{P}\left(-z_{1-\alpha/2} < \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} < z_{1-\alpha/2}\right) = 1-\alpha$$

Now:

$$\mathrm{P}\left(\qquad < \frac{(\bar{X}-\mu)\sqrt{n}}{s} < \qquad \right) = 1-\alpha$$

CI FOR μ, UNKNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\frac{(\bar{X}-\mu)\sqrt{n}}{s} \sim t(n-1) \text{ – Student distribution}$$

Before: $$\mathrm{P}\left(-z_{1-\alpha/2} < \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} < z_{1-\alpha/2}\right) = 1-\alpha$$

Now: same logic, replace $z_{1-a/2}$ with $t_{1-a/2}$:

$$\mathrm{P}\left(\qquad < \frac{(\bar{X}-\mu)\sqrt{n}}{s} < \qquad \right) = 1-\alpha$$

CI FOR μ, UNKNOWN σ

$X_1, X_2, \dots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\frac{(\bar{X}-\mu)\sqrt{n}}{s} \sim t(n-1) \text{ – Student distribution}$$

Before: $$\mathrm{P}\left(-z_{1-\alpha/2} < \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} < z_{1-\alpha/2}\right) = 1-\alpha$$

Now: same logic, replace $z_{1-\alpha/2}$ with $t_{1-\alpha/2}$:

$$\mathrm{P}\left(-t_{1-\alpha/2} < \frac{(\bar{X}-\mu)\sqrt{n}}{s} < t_{1-\alpha/2}\right) = 1-\alpha$$

CI FOR μ, UNKNOWN σ

$X_1, X_2, \ldots, X_n$ – samples from $N(\mu, \sigma^2)$, σ is unknown.

How to construct a CI for μ? Consider $\bar{X}$:

$$\frac{(\bar{X}-\mu)\sqrt{n}}{s} \sim t(n-1) \text{ – Student distribution}$$

Before: $$\mathrm{P}\left(-z_{1-\alpha/2} < \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} < z_{1-\alpha/2}\right) = 1-\alpha$$

Now: same logic, replace $z_{1-\alpha/2}$ with $t_{1-\alpha/2}$:

$$P\left(\boldsymbol{\bar{X}} - \frac{\boldsymbol{s}}{\sqrt{\boldsymbol{n}}}\boldsymbol{t_{1-\alpha/2}} < \boldsymbol{\mu} < \boldsymbol{\bar{X}} + \frac{\boldsymbol{s}}{\sqrt{\boldsymbol{n}}}\boldsymbol{t_{1-\alpha/2}}\right) = 1-\alpha$$

CI FOR μ, UNKNOWN σ: EXAMPLE

- $X_1, X_2, \dots, X_{20}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- $\bar{X} = 42, \quad s^2 = 36$

- Give the 95%-CI for μ.

CI FOR μ, UNKNOWN σ: EXAMPLE

- $X_1, X_2, \dots, X_{20}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- $\bar{X} = 42, \quad s^2 = 36$

- Give the 95%-CI for μ.

$$\mu = \bar{X} \pm \frac{s}{\sqrt{n}} t_{1-\alpha/2}$$

CI FOR μ, UNKNOWN σ: EXAMPLE

- $X_1, X_2, \dots, X_{20}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- $\bar{X} = 42, \quad s^2 = 36$

- Give the 95%-CI for μ.

$$\mu = \bar{X} \pm \frac{s}{\sqrt{n}} t_{1-\alpha/2}$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05,$$

CI FOR μ, UNKNOWN σ: EXAMPLE

- $X_1, X_2, \dots, X_{20}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- $\bar{X} = 42, \quad s^2 = 36$

- Give the 95%-CI for μ.

$$\mu = \bar{X} \pm \frac{s}{\sqrt{n}} t_{1-\alpha/2}$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05, \qquad t_{1-0.025} =$$

CI FOR μ, UNKNOWN σ: EXAMPLE

- $X_1, X_2, \dots, X_{20}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- $\bar{X} = 42, \quad s^2 = 36$

- Give the 95%-CI for μ.

$$\mu = \bar{X} \pm \frac{s}{\sqrt{n}} t_{1-\alpha/2}$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05, \qquad t_{1-0.025} = 2.093$$

CI FOR μ, UNKNOWN σ: EXAMPLE

- $X_1, X_2, \dots, X_{20}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- $\bar{X} = 42, \quad s^2 = 36$

- Give the 95%-CI for μ.

$$\mu = \bar{X} \pm \frac{s}{\sqrt{n}} t_{1-\alpha/2}$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05, \qquad t_{1-0.025} = 2.093$$

$$\mu = 42 \pm \frac{6}{\sqrt{20}} \cdot 2.093$$

CI FOR NORMAL DATA

CI for σ, unknown μ, σ

CI FOR σ, UNKNOWN μ, σ

- $X_1, X_2, \ldots, X_{100}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- How to construct a CI for σ?

CI FOR σ, UNKNOWN μ, σ

- $X_1, X_2, \ldots, X_{100}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- How to construct a CI for σ? Consider S^2:

CI FOR σ, UNKNOWN μ, σ

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- How to construct a CI for σ? Consider S^2:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

CI FOR σ, UNKNOWN μ, σ

- $X_1, X_2, \ldots, X_{100}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- How to construct a CI for μ? Consider S^2:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\frac{(n-1)s^2}{\sigma^2}$$

CI FOR σ, UNKNOWN μ, σ

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- How to construct a CI for μ? Consider S^2:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) -$$

Chi-squared distribution

CHI-SQUARED DISTRIBUTION

$$Z_1, Z_2, \dots, Z_n \ - i.i.d.\ N(0,1)$$

CHI-SQUARED DISTRIBUTION

$$Z_1, Z_2, \dots, Z_n \ - i.i.d.\ N(0,1)$$

$$\Rightarrow$$

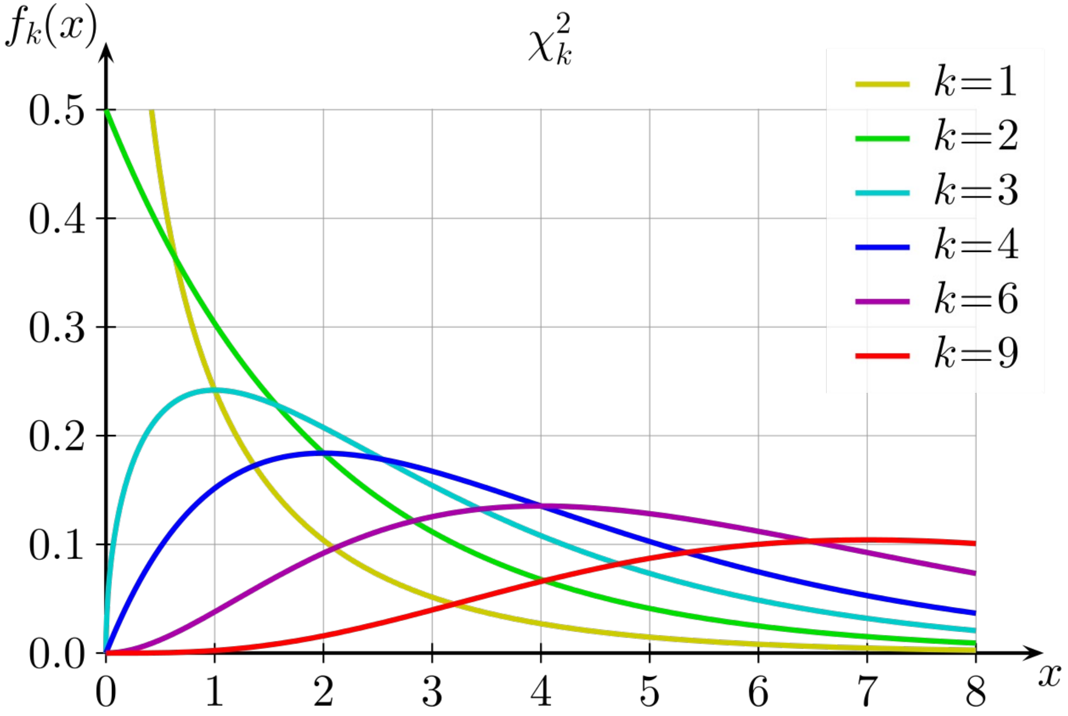
$$Q = Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n-1)$$

CHI-SQUARED DISTRIBUTION

$$Z_1, Z_2, \dots, Z_n \; - i.i.d. \; N(0,1)$$

$$\Rightarrow$$

$$Q = Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n-1)$$



CI FOR σ, UNKNOWN μ, σ

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- How to construct a CI for σ? Consider S^2:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

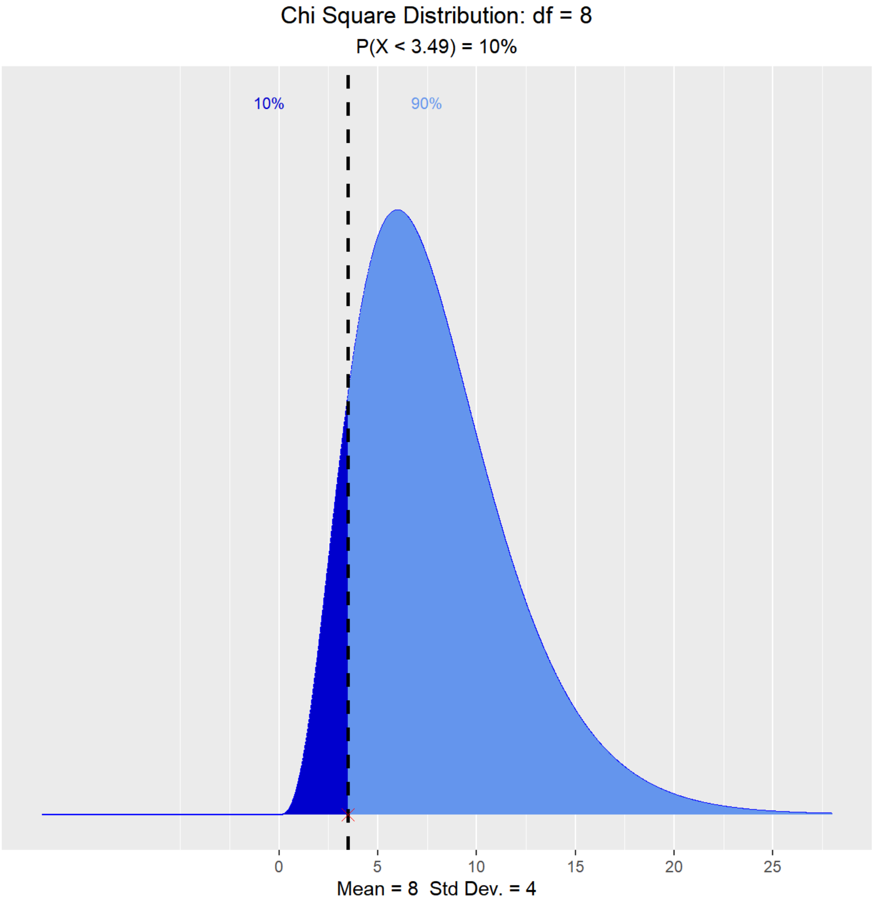
CI FOR σ, UNKNOWN μ, σ

- $X_1, X_2, \ldots, X_{100}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- How to construct a CI for σ? Consider S^2:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

$$P\left(\qquad < \frac{(n-1)s^2}{\sigma^2} < \qquad \right) = 1 - \alpha$$

CI FOR σ, UNKNOWN μ, σ



CI FOR σ, UNKNOWN μ, σ

- $X_1, X_2, \ldots, X_{100}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- How to construct a CI for σ? Consider S^2:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

$$P\left(c_{\alpha/2} < \frac{(n-1)s^2}{\sigma^2} < c_{1-\alpha/2} \right) = 1 - \alpha$$

CI FOR σ, UNKNOWN μ, σ

- $X_1, X_2, \dots, X_{100}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- How to construct a CI for σ? Consider S^2:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

$$P\left(c_{\alpha/2} < \frac{(n-1)s^2}{\sigma^2} < c_{1-\alpha/2}\right) = 1-\alpha$$

$$P\left(\boldsymbol{\frac{(n-1)s^2}{c_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{c_{\alpha/2}}}\right) = 1-\alpha$$

CI FOR σ, UNKNOWN μ, σ: EXAMPLE

- $X_1, X_2, \dots, X_{20}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- $\bar{X} = 42, \quad s^2 = 36$

CI FOR σ, UNKNOWN μ, σ: EXAMPLE

- $X_1, X_2, \ldots, X_{20}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- $\bar{X} = 42, \quad s^2 = 36$

- Give the 95%-CI for σ.

CI FOR σ, UNKNOWN μ, σ: EXAMPLE

- $X_1, X_2, \dots, X_{20}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- $\bar{X} = 42, \quad s^2 = 36$

- Give the 95%-CI for σ.

$$\left[\frac{(n-1)s^2}{c_{1-\alpha/2}}; \frac{(n-1)s^2}{c_{\alpha/2}}\right]$$

CI FOR σ, UNKNOWN μ, σ: EXAMPLE

- $X_1, X_2, \dots, X_{20}$ – samples from $N(\mu, \sigma^2)$, σ is unknown.
- $\bar{X} = 42, \quad s^2 = 36$

- Give the 95%-CI for σ.

$$\left[\frac{(n-1)s^2}{c_{1-\alpha/2}}; \frac{(n-1)s^2}{c_{\alpha/2}}\right]$$

$$\left[\frac{19 \cdot 36}{32.85}; \frac{19 \cdot 36}{8.91}\right]$$

LARGE SAMPLES

Using Central Limit Theorem

CI FOR LARGE SAMPLES

- Typical task: estimating the mean of a distribution.
- Suppose $X_1, \ldots X_n$ is drawn from an unknown distribution.
- How to construct a CI?

- CLT:

If $\mu, \sigma^2 < \infty$ and if n is sufficiently large, then:

$$\frac{(\bar{X} - \mu)\sqrt{n}}{s} \approx N(0,1)$$

CI FOR LARGE SAMPLES

- Typical task: estimating the mean of a distribution.
- Suppose $X_1, \dots X_n$ is drawn from an unknown distribution.
- How to construct a CI?

- CLT:

If $\mu, \sigma^2 < \infty$ and if n is sufficiently large, then:

$$\frac{(\bar{X} - \mu)\sqrt{n}}{s} \approx N(0,1) \quad \Rightarrow \quad \mu \approx$$

CI FOR LARGE SAMPLES

- Typical task: estimating the mean of a distribution.
- Suppose $X_1, \dots X_n$ is drawn from an unknown distribution.
- How to construct a CI?

- CLT:

If $\mu, \sigma^2 < \infty$ and if n is sufficiently large, then:

$$\frac{(\bar{X} - \mu)\sqrt{n}}{s} \approx N(0,1) \quad \Rightarrow \quad \mu \approx \bar{X} \pm \frac{s}{\sqrt{n}} z_{1-\alpha/2}$$

CI FOR
BERNOULLI DISTRIBUTION

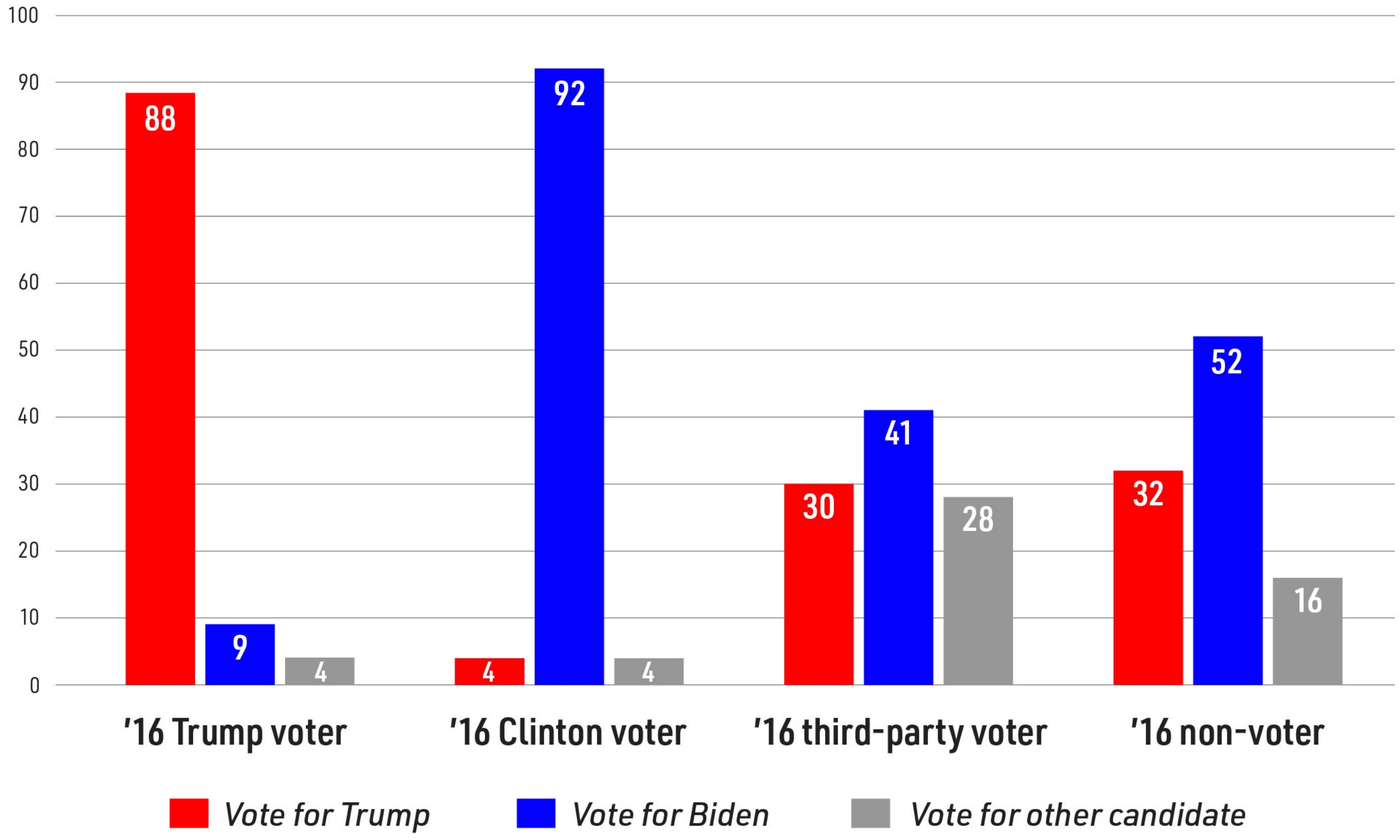
POLITICAL POLLS

- Political polls are often reported as a value with a margin-of-error.

52% favor candidate A with a margin-of-error of ±5%.

Presidential Election Preview:

How Clinton, Trump Voters from 2016 Plan to Vote in 2020



A total of 1,510 eligible voters, who are adult members of the USC Dornsife Center for Economic and Social Research's Understanding America Study internet panel, participated from August 11 – 16, 2020. Margin of sampling error for this preliminary sample is +/-3 percentage points. Tracking graphs will be updated every day. For full question text, methodology, and other information, visit https://uasdata.usc.edu/index.php

POLITICAL POLLS

- Political polls are often reported as a value with a margin-of-error.

52% favor candidate A with a margin-of-error of ±5%.

- The actual precise meaning:

POLITICAL POLLS

- Political polls are often reported as a value with a margin-of-error.

52% favor candidate A with a margin-of-error of ±5%.

- The actual precise meaning:

If p is the proportion of the population that supports A, then

POLITICAL POLLS

- Political polls are often reported as a value with a margin-of-error.

52% favor candidate A with a margin-of-error of ±5%.

- The actual precise meaning:

If p is the proportion of the population that supports A, then the point estimate for p is $\hat{p} = 0.52$

POLITICAL POLLS

- Political polls are often reported as a value with a margin-of-error.

52% favor candidate A with a margin-of-error of ±5%.

- The actual precise meaning:

If p is the proportion of the population that supports A, then
the point estimate for p is $\hat{p} = 0.52$
and the 95% - CI is $52\% \pm 5\%$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \ldots, X_n$.

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

$$\hat{p} = \bar{X}$$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \ldots, X_n$.

$$\text{CLT:} \quad \hat{p} = \bar{X} \approx N\left(\quad , \quad \quad \right)$$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

CLT: $\hat{p} = \bar{X} \approx N\left(p, \qquad \right)$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

$$\text{CLT:} \quad \hat{p} = \bar{X} \approx N\left(p, \frac{p(1-p)}{n}\right)$$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

CLT: $$\hat{p} = \bar{X} \approx N\left(p, \frac{p(1-p)}{n}\right)$$

Approximation: $\frac{p(1-p)}{n} \approx \frac{\hat{p}(1-\hat{p})}{n}$ (not valid for small samples)

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

CLT: $\hat{p} = \bar{X} \approx N\left(p, \frac{p(1-p)}{n}\right)$

Approximation: $\frac{p(1-p)}{n} \approx \frac{\hat{p}(1-\hat{p})}{n}$ (not valid for small samples)

$\approx N(0,1)$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

$$\text{CLT:} \quad \hat{p} = \bar{X} \approx N\left(p, \frac{p(1-p)}{n}\right)$$

$$\text{Approximation:} \quad \frac{p(1-p)}{n} \approx \frac{\hat{p}(1-\hat{p})}{n} \text{ (not valid for small samples)}$$

$$\frac{(\hat{p} - p)\sqrt{n}}{\sqrt{\hat{p}(1 - \hat{p})}} \approx N(0,1)$$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

$$\frac{(\hat{p}-p)\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}} \approx N(0,1)$$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

$$\frac{(\hat{p}-p)\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}} \approx N(0,1)$$

$$P\left(-z_{1-\alpha/2} < \frac{(\hat{p}-p)\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}} < z_{1-\alpha/2}\right) = 1-\alpha$$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

$$\frac{(\hat{p}-p)\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}} \approx N(0,1)$$

$$P\left(-z_{1-\alpha/2} < \frac{(\hat{p}-p)\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}} < z_{1-\alpha/2}\right) = 1-\alpha$$

$$P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1-\alpha$$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \ldots, X_n$.

$$P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

$$P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$$

- Example: toss a coin 100 times, 60 heads. CI for p ($\alpha = 0.05$):

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \dots, X_n$.

$$P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$$

- Example: toss a coin 100 times, 60 heads. CI for p ($\alpha = 0.05$):

$$p \in (0.6 - 1.96 \cdot 0.049\,; 0.6 + 1.96 \cdot 0.049)$$

CI FOR P (BERNOULLI)

- Construct a confidence interval for the parameter p of the Bernoulli distribution based on samples $X_1, \ldots, X_n$.

$$P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$$

- Example: toss a coin 100 times, 60 heads. CI for p ($\alpha = 0.05$):

$$p \in (0.6 - 1.96 \cdot 0.049\,; 0.6 + 1.96 \cdot 0.049)$$

$$p \in (0.504; 0.696)$$

THE 95%-CI: RULE-OF-THUMB

$$p \approx \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

THE 95%-CI: RULE-OF-THUMB

$$p \approx \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

What is the maximum possible value of $\hat{p}(1-\hat{p})$?

THE 95%-CI: RULE-OF-THUMB

$$p \approx \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

What is the maximum possible value of $\hat{p}(1-\hat{p})$? 1/4!

THE 95%-CI: RULE-OF-THUMB

$$p \approx \hat{p} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

What is the maximum possible value of $\hat{p}(1-\hat{p})$? 1/4!

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq$$

THE 95%-CI: RULE-OF-THUMB

$$p \approx \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

What is the maximum possible value of $\hat{p}(1-\hat{p})$? 1/4!

$$z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq z_{1-\frac{\alpha}{2}} \cdot \frac{1}{2\sqrt{n}} =$$

THE 95%-CI: RULE-OF-THUMB

$$p \approx \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

What is the maximum possible value of $\hat{p}(1-\hat{p})$? 1/4!

$$z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq z_{1-\frac{\alpha}{2}} \cdot \frac{1}{2\sqrt{n}} = 1.96 \cdot \frac{1}{2\sqrt{n}} \approx$$

THE 95%-CI: RULE-OF-THUMB

$$p \approx \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

What is the maximum possible value of $\hat{p}(1-\hat{p})$? 1/4!

$$z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq z_{1-\frac{\alpha}{2}} \cdot \frac{1}{2\sqrt{n}} = 1.96 \cdot \frac{1}{2\sqrt{n}} \approx \frac{1}{\sqrt{n}}$$

THE 95%-CI: RULE-OF-THUMB

$$p \approx \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

What is the maximum possible value of $\hat{p}(1-\hat{p})$? 1/4!

$$z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq z_{1-\frac{\alpha}{2}} \cdot \frac{1}{2\sqrt{n}} = 1.96 \cdot \frac{1}{2\sqrt{n}} \approx \frac{1}{\sqrt{n}}$$

$$p \approx$$

THE 95%-CI: RULE-OF-THUMB

$$p \approx \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

What is the maximum possible value of $\hat{p}(1-\hat{p})$? 1/4!

$$z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq z_{1-\frac{\alpha}{2}} \cdot \frac{1}{2\sqrt{n}} = 1.96 \cdot \frac{1}{2\sqrt{n}} \approx \frac{1}{\sqrt{n}}$$

$$p \approx \hat{p} \pm \frac{1}{\sqrt{n}}$$

EXAMPLE: POLITICAL POLLS

- Two polls:
  - Fast and First:
    polls 40 random voters and finds 22 support A.
  - Quick but Cautious:
    polls 400 random voters and finds 190 support A.

EXAMPLE: POLITICAL POLLS

- Two polls:
  - Fast and First:

    polls 40 random voters and finds 22 support A.
  - Quick but Cautious:

    polls 400 random voters and finds 190 support A.

- Let's find the point estimates and 95% rule-of-thumb confidence intervals for each poll:

EXAMPLE: POLITICAL POLLS

- Two polls:
  - Fast and First:

    polls 40 random voters and finds 22 support A.
  - Quick but Cautious:

    polls 400 random voters and finds 190 support A.

- Let's find the point estimates and 95% rule-of-thumb confidence intervals for each poll:

$$p = \hat{p} \pm \frac{1}{\sqrt{n}}$$

EXAMPLE: POLITICAL POLLS

- Two polls:
  - Fast and First:

    polls 40 random voters and finds 22 support A.
  - Quick but Cautious:

    polls 400 random voters and finds 190 support A.

- Poll 1:

$$\hat{p} =$$

- Poll 2:

$$\hat{p} =$$

EXAMPLE: POLITICAL POLLS

- Two polls:
  - Fast and First:
    polls 40 random voters and finds 22 support A.
  - Quick but Cautious:
    polls 400 random voters and finds 190 support A.

- Poll 1:

$$\hat{p} = 22/40$$

- Poll 2:

$$\hat{p} =$$

EXAMPLE: POLITICAL POLLS

- Two polls:
  - Fast and First:

    polls 40 random voters and finds 22 support A.
  - Quick but Cautious:

    polls 400 random voters and finds 190 support A.

- Poll 1:

$$\hat{p} = 22/40$$

- Poll 2:

$$\hat{p} = 190/400$$

EXAMPLE: POLITICAL POLLS

- Two polls:
  - Fast and First:

    polls 40 random voters and finds 22 support A.
  - Quick but Cautious:

    polls 400 random voters and finds 190 support A.

- Poll 1:

$$\hat{p} = 22/40 \qquad p = \hat{p} \pm \frac{1}{\sqrt{n}} =$$

- Poll 2:

$$\hat{p} = 190/400 \qquad p = \hat{p} \pm \frac{1}{\sqrt{n}} =$$

EXAMPLE: POLITICAL POLLS

- Two polls:
  - Fast and First:
    polls 40 random voters and finds 22 support A.
  - Quick but Cautious:
    polls 400 random voters and finds 190 support A.

- Poll 1:

$$\hat{p} = 22/40 \qquad p = \hat{p} \pm \frac{1}{\sqrt{n}} = 0.55 \pm 0.16$$

- Poll 2:

$$\hat{p} = 190/400 \qquad p = \hat{p} \pm \frac{1}{\sqrt{n}} =$$

EXAMPLE: POLITICAL POLLS

- Two polls:
  - Fast and First:
    polls 40 random voters and finds 22 support A.
  - Quick but Cautious:
    polls 400 random voters and finds 190 support A.

- Poll 1:

$$\hat{p} = 22/40 \qquad p = \hat{p} \pm \frac{1}{\sqrt{n}} = 0.55 \pm 0.16$$

- Poll 2:

$$\hat{p} = 190/400 \qquad p = \hat{p} \pm \frac{1}{\sqrt{n}} = 0.475 \pm 0.05$$

EXAMPLE: POLITICAL POLLS

- Two polls:
  - Fast and First:
    polls 40 random voters and finds 22 support A.
  - Quick but Cautious:
    polls 400 random voters and finds 190 support A.

- Poll 1:

$$\hat{p} = 22/40 \qquad p = \hat{p} \pm \frac{1}{\sqrt{n}} = 0.55 \pm 0.16$$

- Poll 2:

$$\hat{p} = 190/400 \qquad p = \hat{p} \pm \frac{1}{\sqrt{n}} = 0.475 \pm 0.05$$

SAMPLE SIZE

How large should my sample be?

SAMPLE SIZE DETERMINATION

- Data collection is difficult.
- How much is 'just enough'?

- Example: estimating CI for the mean, σ is know.

SAMPLE SIZE DETERMINATION

- Data collection is difficult.
- How much is 'just enough'?

- Example: estimating CI for the mean, σ is know.

$$\mu \in \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

SAMPLE SIZE DETERMINATION

- Data collection is difficult.
- How much is 'just enough'?

- Example: estimating CI for the mean, σ is know.

$$\mu \in \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

SAMPLE SIZE DETERMINATION

- Data collection is difficult.
- How much is 'just enough'?

- Example: estimating CI for the mean, σ is know.

$$\mu \in \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

Limit the width of the interval: $\frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \epsilon$

SAMPLE SIZE DETERMINATION

- Data collection is difficult.
- How much is ‘just enough’?

- Example: estimating CI for the mean, σ is know.

$$\mu \in \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

Limit the width of the interval: $\frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \epsilon$

$$n \geq$$

SAMPLE SIZE DETERMINATION

- Data collection is difficult.
- How much is ‘just enough’?

- Example: estimating CI for the mean, σ is know.

$$\mu \in \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

Limit the width of the interval: $\frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \epsilon$

$$n \geq \frac{\sigma^2 z_{1-\alpha/2}^2}{\epsilon^2}$$

SAMPLE SIZE DETERMINATION

- At a juice factory machine is set up as follows:
  - the average content of juice per bottle equals μ;
  - σ = 5cl.

SAMPLE SIZE DETERMINATION

- At a juice factory machine is set up as follows:
  - the average content of juice per bottle equals μ;
  - σ = 5cl.

- What sample size is required to estimate the average contents to within 0.5cl at the 95% confidence level?

SAMPLE SIZE DETERMINATION

- At a juice factory machine is set up as follows:
  - the average content of juice per bottle equals μ;
  - σ = 5cl.

- What sample size is required to estimate the average contents to within 0.5cl at the 95% confidence level?

$$n \geq \frac{\sigma^2 z^2_{1-\frac{\alpha}{2}}}{\epsilon^2} =$$

SAMPLE SIZE DETERMINATION

- At a juice factory machine is set up as follows:
  - the average content of juice per bottle equals μ;
  - σ = 5cl.

- What sample size is required to estimate the average contents to within 0.5cl at the 95% confidence level?

$$n \geq \frac{\sigma^2 z^2_{1-\frac{\alpha}{2}}}{\epsilon^2} = \frac{5^2 \cdot 1.96^2}{0.5^2} \approx$$

SAMPLE SIZE DETERMINATION

- At a juice factory machine is set up as follows:
  - the average content of juice per bottle equals μ;
  - σ = 5cl.

- What sample size is required to estimate the average contents to within 0.5cl at the 95% confidence level?

$$n \geq \frac{\sigma^2 z^2_{1-\frac{\alpha}{2}}}{\epsilon^2} = \frac{5^2 \cdot 1.96^2}{0.5^2} \approx 384.16$$

SAMPLE SIZE DETERMINATION

- At a juice factory machine is set up as follows:
  - the average content of juice per bottle equals μ;
  - σ = 5cl.

- What sample size is required to estimate the average contents to within 0.5cl at the 95% confidence level?

$$n \geq \frac{\sigma^2 z^2_{1-\frac{\alpha}{2}}}{\epsilon^2} = \frac{5^2 \cdot 1.96^2}{0.5^2} \approx 384.16$$

$$n^* = 385$$

SAMPLE SIZE DETERMINATION

- Similar reasoning applies to the CI for Bernoulli.

- *How many customers should be surveyed in order to estimate the share of satisfied customers to within 3% at the 95% confidence level?*

SAMPLE SIZE DETERMINATION

- Similar reasoning applies to the CI for Bernoulli.

- *How many customers should be surveyed in order to estimate the share of satisfied customers to within 3% at the 95% confidence level?*

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

SAMPLE SIZE DETERMINATION

- Similar reasoning applies to the CI for Bernoulli.

- *How many customers should be surveyed in order to estimate the share of satisfied customers to within 3% at the 95% confidence level?*

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$$

SAMPLE SIZE DETERMINATION

- Similar reasoning applies to the CI for Bernoulli.

- *How many customers should be surveyed in order to estimate the share of satisfied customers to within 3% at the 95% confidence level?*

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$$

Let $p_0 = 0.5$ – the worst possible variance

SAMPLE SIZE DETERMINATION

- Similar reasoning applies to the CI for Bernoulli.

- *How many customers should be surveyed in order to estimate the share of satisfied customers to within 3% at the 95% confidence level?*

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$$

Let $p_0 = 0.5$ – the worst possible variance

$$1.96\sqrt{\frac{0.25}{n}} \leq 0.03 \iff$$

SAMPLE SIZE DETERMINATION

- Similar reasoning applies to the CI for Bernoulli.

- *How many customers should be surveyed in order to estimate the share of satisfied customers to within 3% at the 95% confidence level?*

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$$

Let $p_0 = 0.5$ – the worst possible variance

$$1.96\sqrt{\frac{0.25}{n}} \leq 0.03 \iff n \geq 1067.1$$