16 – Confidence Intervals

Reference: [ES] 6.1, 6.3

(For pop. mean)

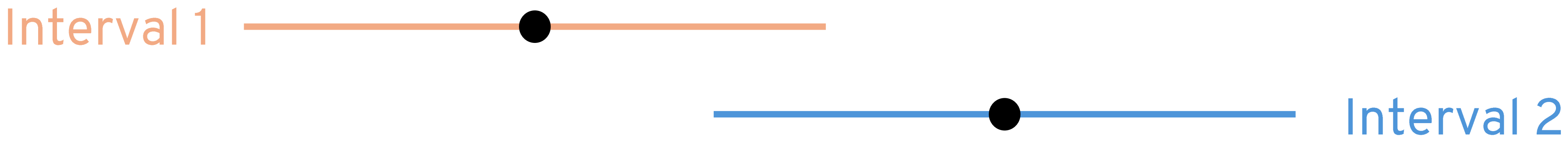
Reference: [ES] 6.1

Fun fact:

Two intervals on the number line have the same length. Then:

1st interval contains the midpoint of 2nd interval

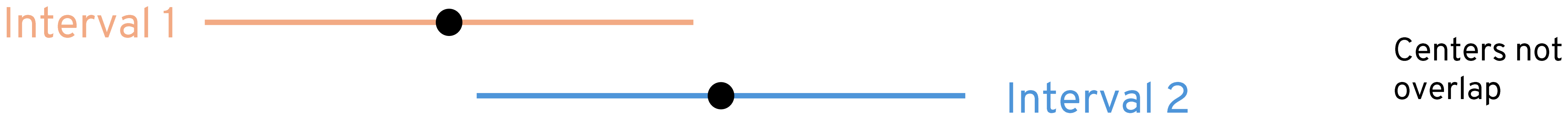
$\leftrightarrow$ 2nd interval contains the midpoint of 1st interval



Centers not overlap

Fun fact:

Two intervals on the number line have the same length. Then:
1st interval contains the midpoint of 2nd interval
$\leftrightarrow$ 2nd interval contains the midpoint of 1st interval

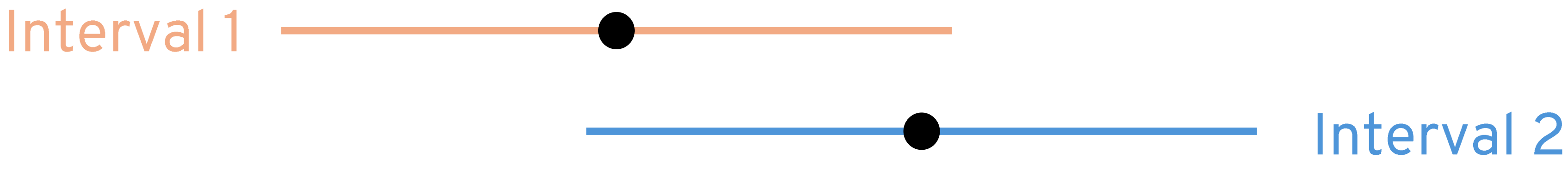


Fun fact:

Two intervals on the number line have the same length. Then:

1st interval contains the midpoint of 2nd interval

↔ 2nd interval contains the midpoint of 1st interval



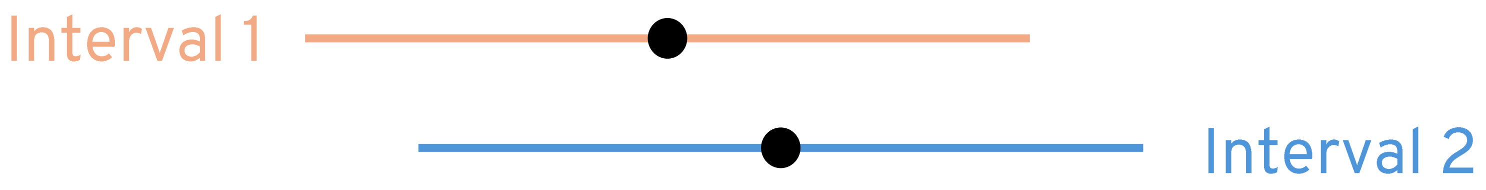
Centers overlap

Fun fact:

Two intervals on the number line have the same length. Then:

1st interval contains the midpoint of 2nd interval

↔ 2nd interval contains the midpoint of 1st interval



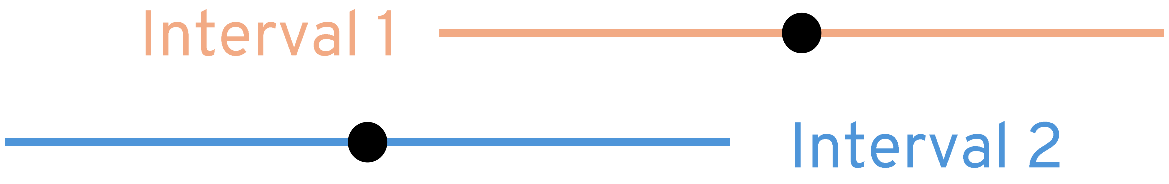
Centers overlap

Fun fact:

Two intervals on the number line have the same length. Then:

1st interval contains the midpoint of 2nd interval

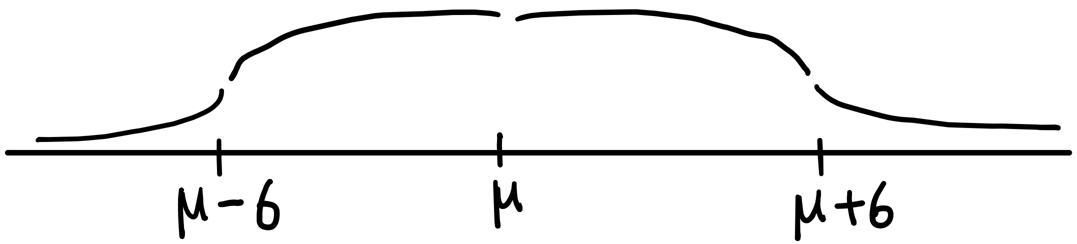
↔ 2nd interval contains the midpoint of 1st interval



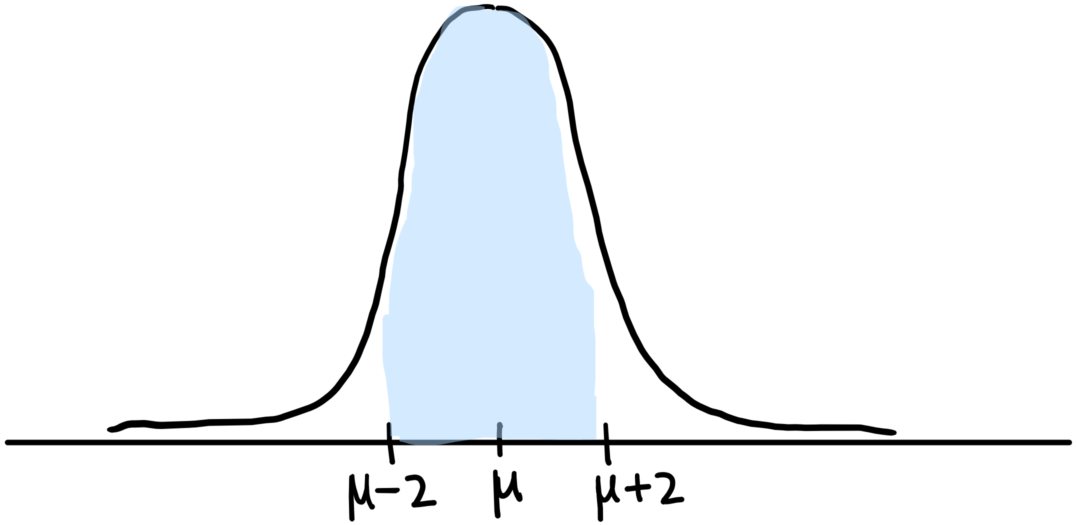
Centers not overlap

Confidence Interval (Intuition)

Population is normally distributed with $\mu =?$ and $\sigma = 6$:



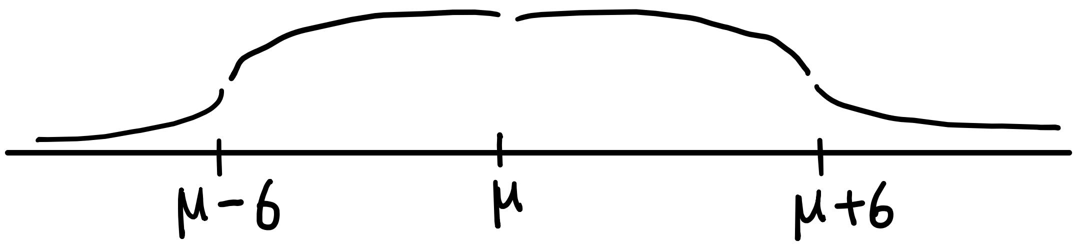
Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:



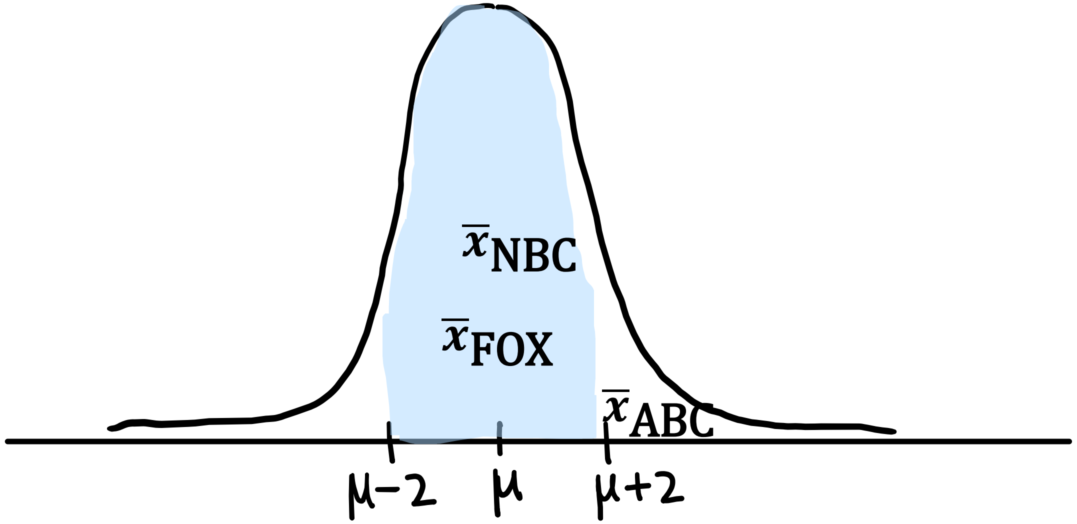
$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.

Confidence Interval (Intuition)

Population is normally distributed with $\mu = ?$ and $\sigma = 6$:



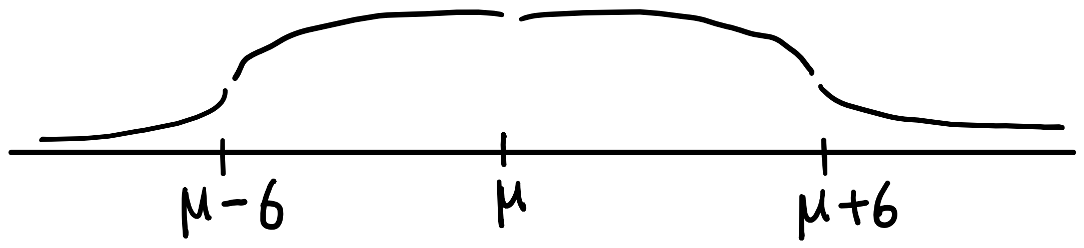
Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:



$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.

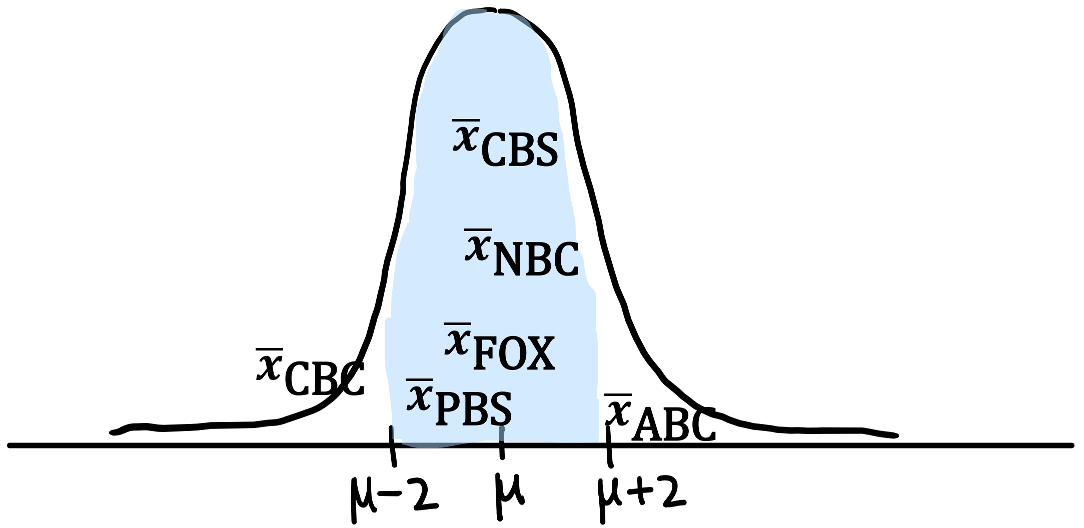
Confidence Interval (Intuition)

Population is normally distributed with $\mu =?$ and $\sigma = 6$:



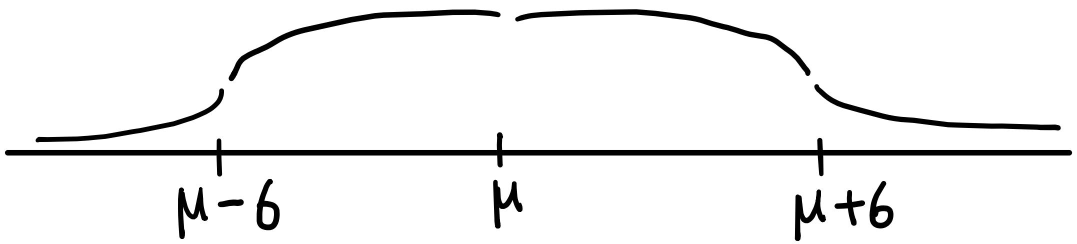
Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:

$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.

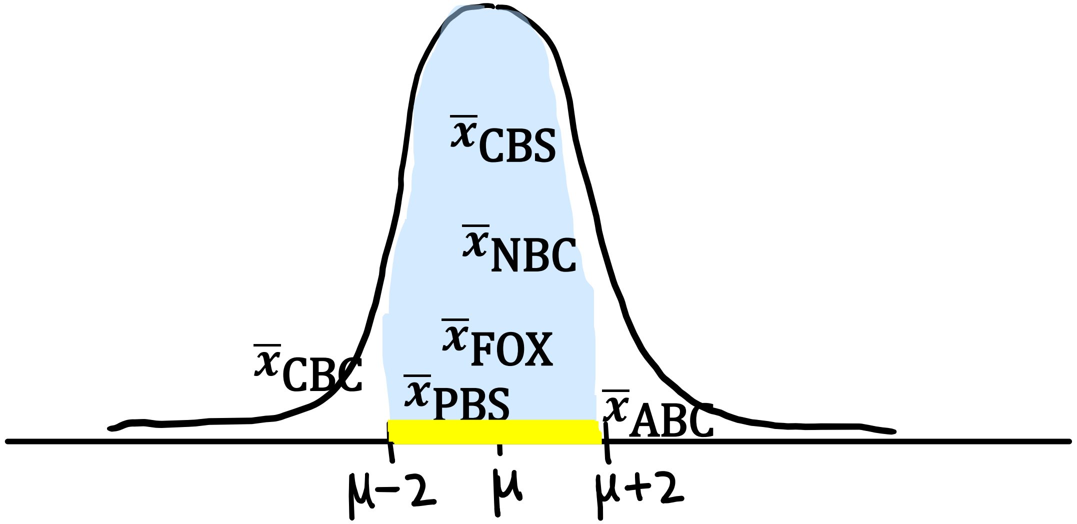


Confidence Interval (Intuition)

Population is normally distributed with $\mu = ?$ and $\sigma = 6$:



Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:

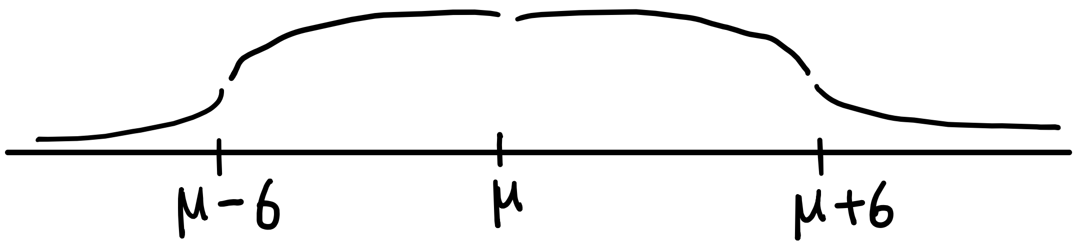


$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.

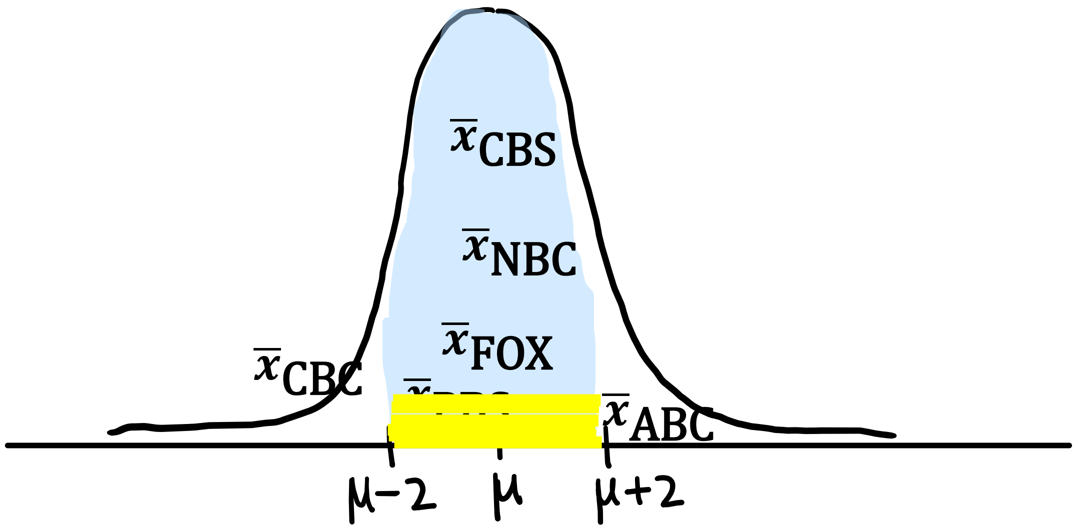
Flipped version: μ lands in interval $\bar{x} \pm 1 \cdot \sigma_{\bar{x}}$ 68% of the time of all resamplings.

Confidence Interval (Intuition)

Population is normally distributed with $\mu = ?$ and $\sigma = 6$:



Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:

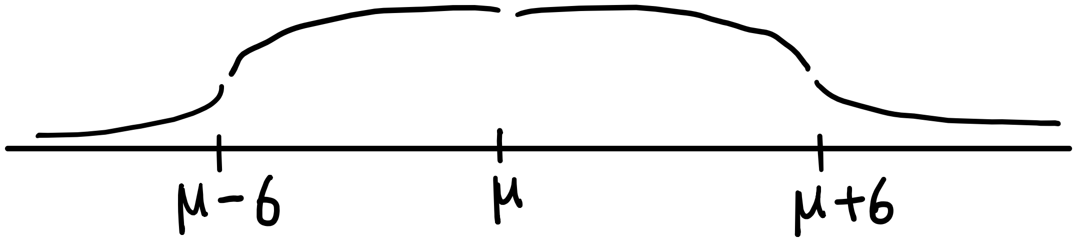


$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.

Flipped version: μ lands in interval $\bar{x} \pm 1 \cdot \sigma_{\bar{x}}$ 68% of the time of all resamplings.

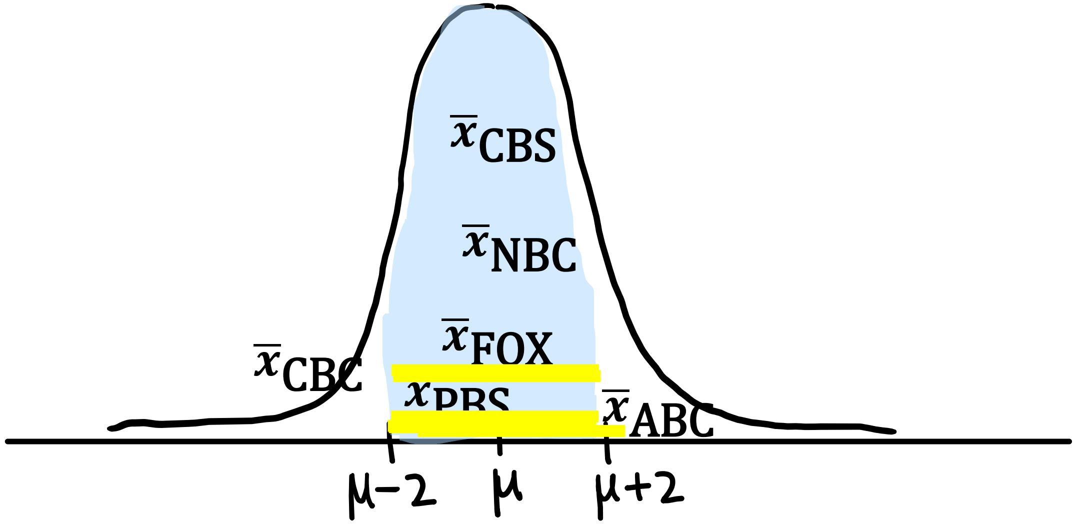
Confidence Interval (Intuition)

Population is normally distributed with $\mu = ?$ and $\sigma = 6$:



Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:

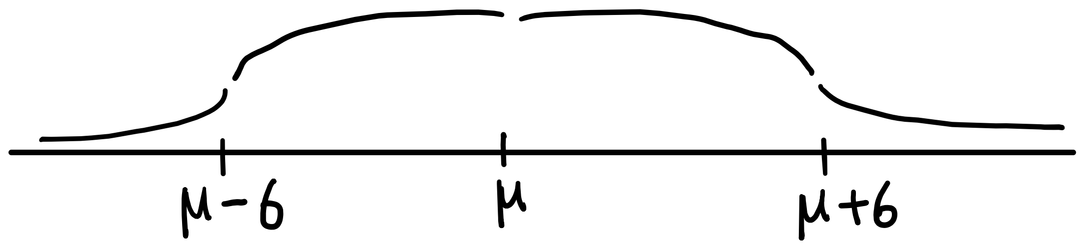
$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.



Flipped version: μ lands in interval $\bar{x} \pm 1 \cdot \sigma_{\bar{x}}$ 68% of the time of all resamplings.

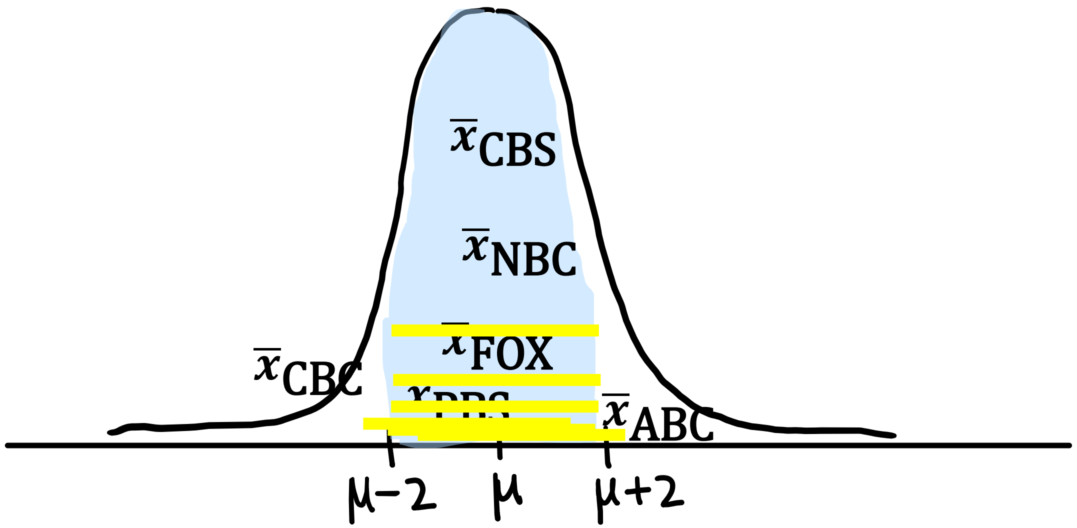
Confidence Interval (Intuition)

Population is normally distributed with
$\mu =?$ and $\sigma = 6$:



Samples of size $n = 9$ have sample mean $\bar{x}$
distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:

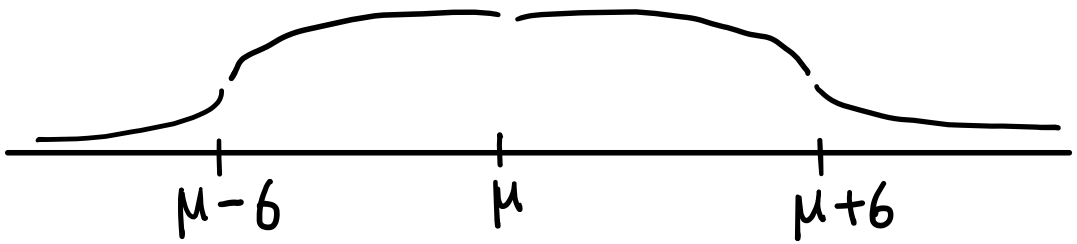
$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of
the time of all resamplings.



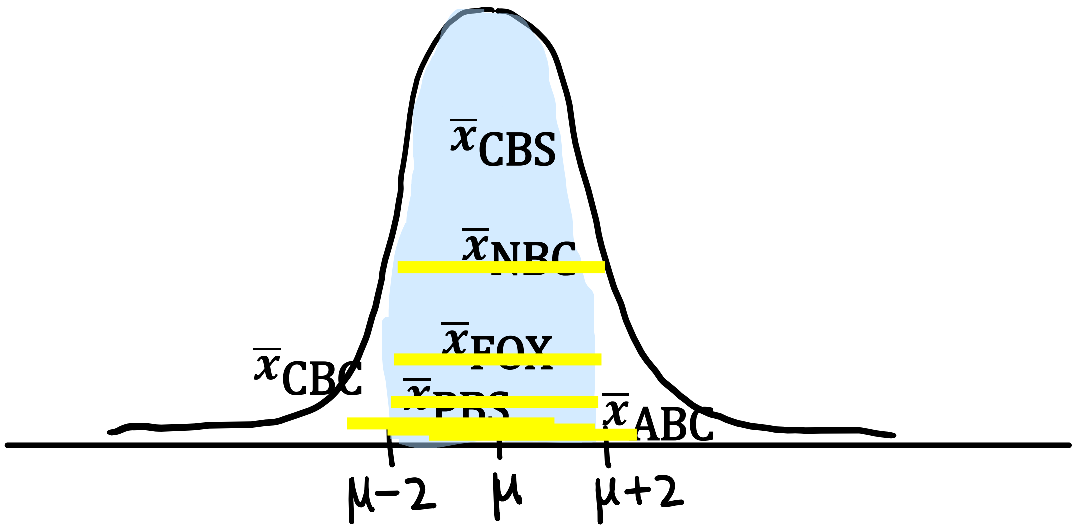
Flipped version: μ lands in interval $\bar{x} \pm 1 \cdot \sigma_{\bar{x}}$
68% of the time of all resamplings.

Confidence Interval (Intuition)

Population is normally distributed with $\mu =?$ and $\sigma = 6$:



Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:

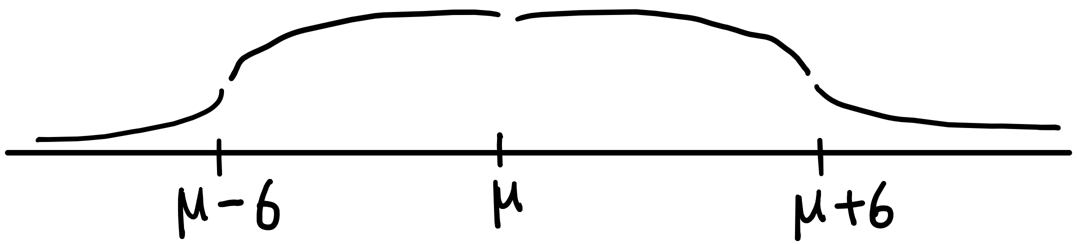


$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.

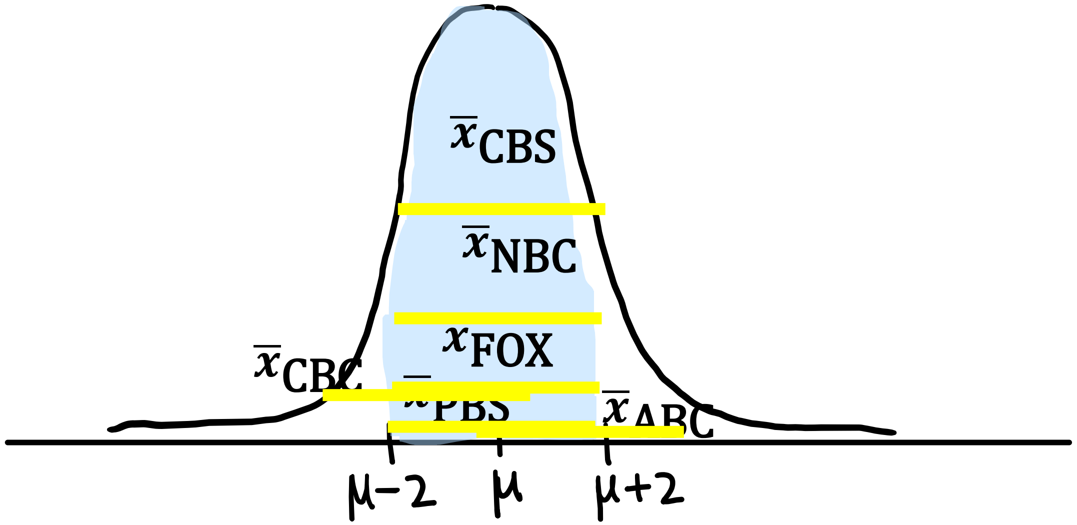
Flipped version: μ lands in interval $\bar{x} \pm 1 \cdot \sigma_{\bar{x}}$ 68% of the time of all resamplings.

Confidence Interval (Intuition)

Population is normally distributed with $\mu =?$ and $\sigma = 6$:



Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:

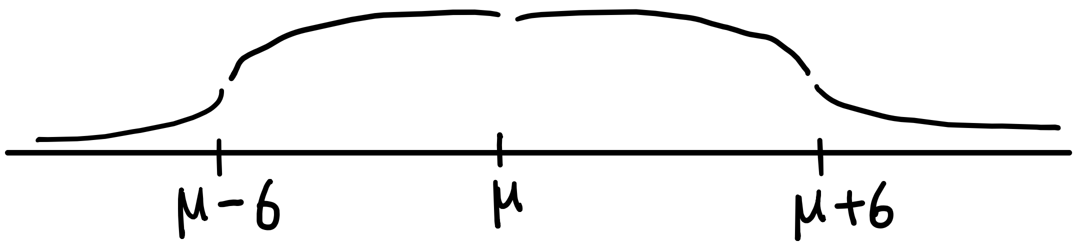


$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.

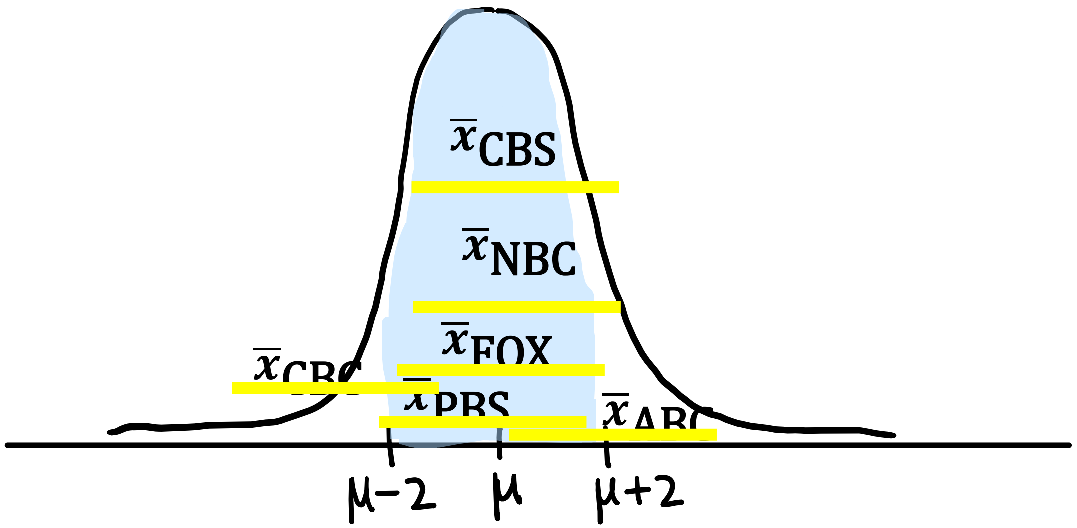
Flipped version: μ lands in interval $\bar{x} \pm 1 \cdot \sigma_{\bar{x}}$ 68% of the time of all resamplings.

Confidence Interval (Intuition)

Population is normally distributed with $\mu =?$ and $\sigma = 6$:



Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:

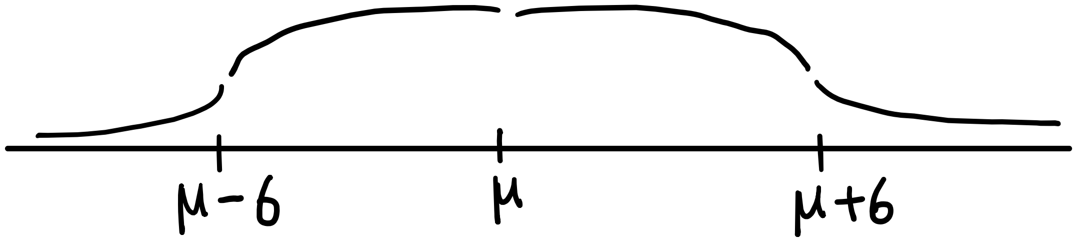


$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.

Flipped version: μ lands in interval $\bar{x} \pm 1 \cdot \sigma_{\bar{x}}$ 68% of the time of all resamplings.

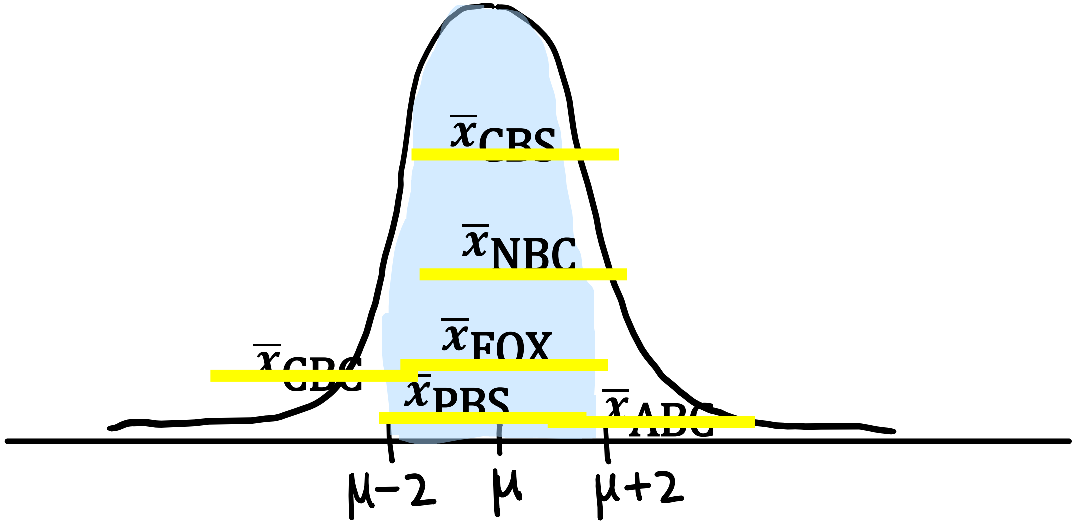
Confidence Interval (Intuition)

Population is normally distributed with $\mu =?$ and $\sigma = 6$:



Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:

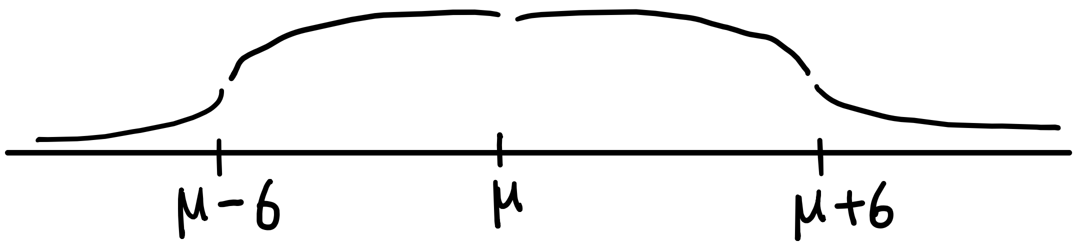
$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.



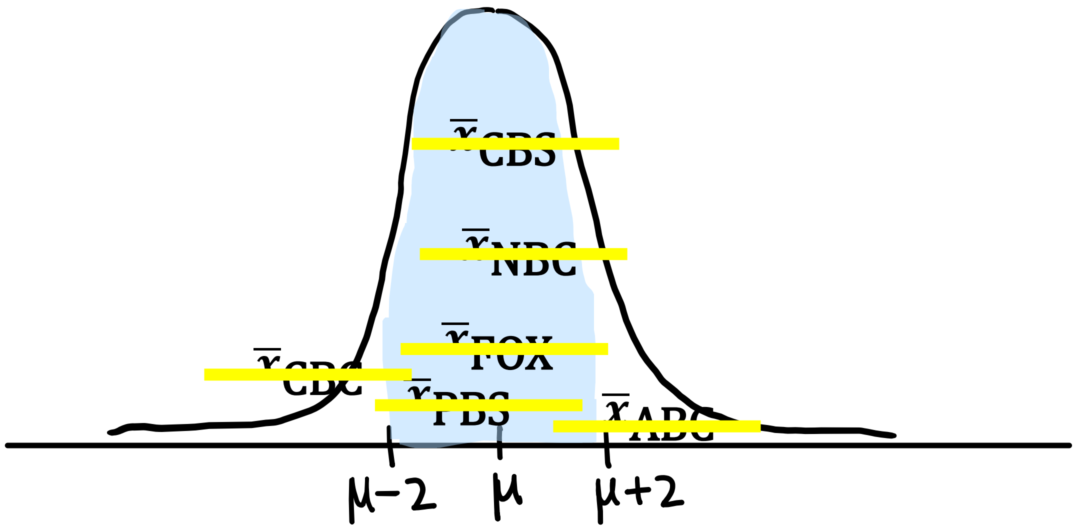
Flipped version: μ lands in interval $\bar{x} \pm 1 \cdot \sigma_{\bar{x}}$ 68% of the time of all resamplings.

Confidence Interval (Intuition)

Population is normally distributed with $\mu =?$ and $\sigma = 6$:



Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:

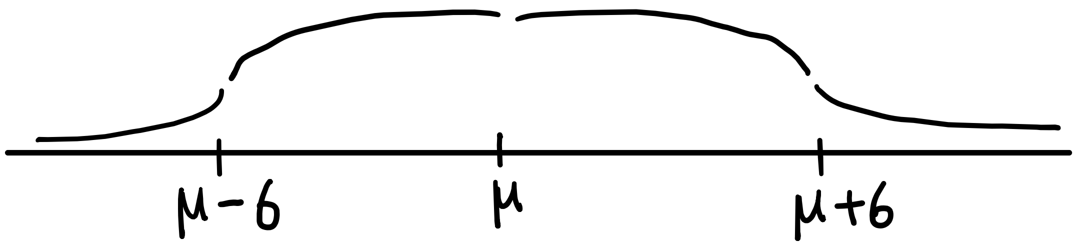


$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.

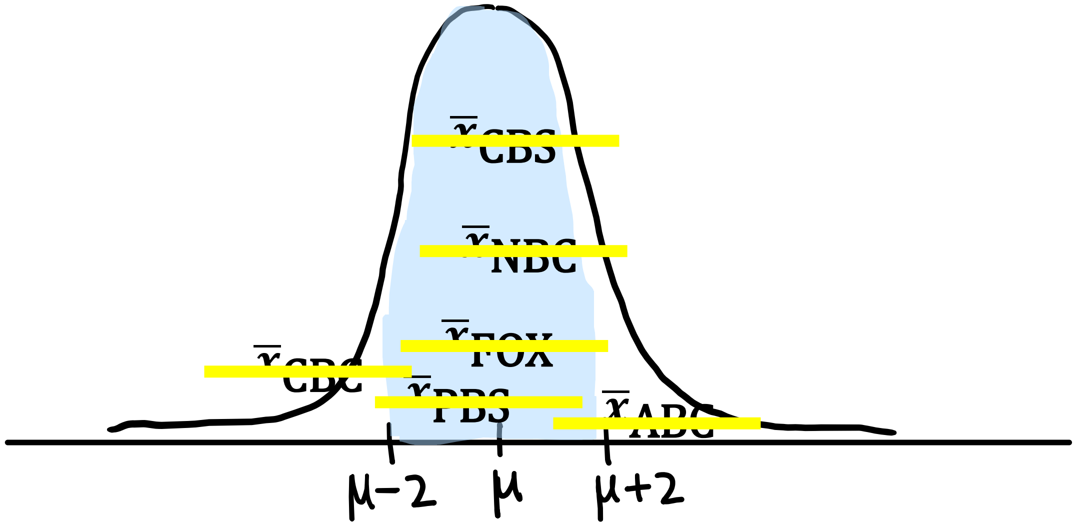
Flipped version: μ lands in interval $\bar{x} \pm 1 \cdot \sigma_{\bar{x}}$ 68% of the time of all resamplings.

Confidence Interval (Intuition)

Population is normally distributed with $\mu =?$ and $\sigma = 6$:



Samples of size $n = 9$ have sample mean $\bar{x}$ distributed like $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{6}{\sqrt{9}} = 2$:



$\bar{x}$ lands in interval $\mu \pm 1 \cdot \sigma_{\bar{x}}$ about 68% of the time of all resamplings.

Flipped version: μ lands in interval $\bar{x} \pm 1 \cdot \sigma_{\bar{x}}$ 68% of the time of all resamplings.

"We are 68% confident that μ is in one of these intervals"

Confidence Interval of Mean with σ known (Definition)

A **point estimate** is a single number estimate of a population parameter:
sample mean $\bar{x}$ estimates population mean μ.

An **interval estimate** is an interval to estimate a population parameter.

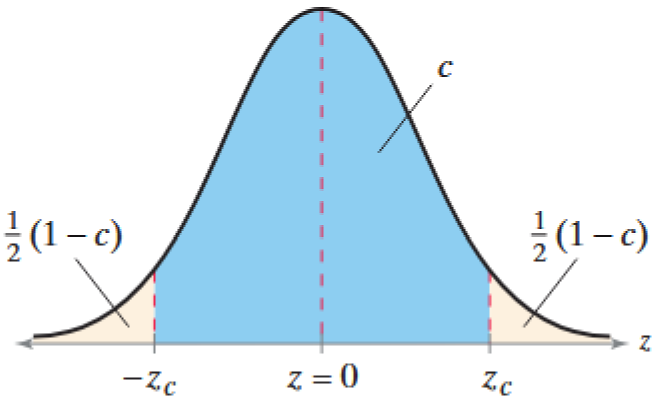
The **level of confidence** c is the probability that the interval estimate contains the population parameter, when we resample many times.

Sampling distribution for $n \geq 30$ is approx. normal so c is the area under the standard normal curve between **critical values** $\pm z_c$:

$z_c = \text{qnorm}\left(\frac{(1-C)}{2}\right)$ ... command in R

Common critical values:

Level of Confidence	z_c
90%	1.645
95%	1.96 ≈ 2
99%	2.575
68%	≈ 1



If $c = 90\%$:	
$c = 0.90$	Area in blue region
$1 - c = 0.10$	Area in yellow regions
$\frac{1}{2}(1 - c) = 0.05$	Area in one tail
$-z_c = -1.645$	Critical value separating left tail
$z_c = 1.645$	Critical value separating right tail

Confidence Interval of Mean with σ known (Facts)

The **c-confidence interval for population mean μ** obtained from an SRS with sample mean $\bar{x}$ is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = z_c\, \sigma_{\bar{x}} = z_c \frac{\sigma}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

"We are *c*% confident that μ is in this interval",

" This is the *c*% confidence interval for μ"

Confidence Interval of Mean with σ known (Example 1)

The **c-confidence interval for population mean μ** obtained from an SRS with sample mean $\bar{x}$ is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = z_c \, \sigma_{\bar{x}} = z_c \frac{\sigma}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

"We are c% confident that μ is in this interval",

" This is the c% confidence interval for μ"

Example 1.
Construct a 95% confidence interval for mean age of all Rowan students given pop. is normal, $\sigma = 2$ years, and an SRS of size $n = 4$ with $\bar{x} = 21$ years.

Answer.
Rowan student pop. is $N \geq 10 \cdot 4$ and pop. distribution is normal so we can proceed.

$$E = 2 \cdot \frac{2}{\sqrt{16}} = 2$$

So the 95% confidence interval is

$$21 - 2 < \mu < 21 + 2$$

Interpretation:
We are 95% confident that the mean age of all Rowan students is between 19 and 23.

Confidence Interval of Mean with σ known (Example 2)

The **c-confidence interval for population mean μ** obtained from an SRS with sample mean $\bar{x}$ is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = z_c\, \sigma_{\bar{x}} = z_c \frac{\sigma}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

"We are c% confident that μ is in this interval",

" This is the c% confidence interval for μ"

Example 2. (Minimal sample size)
Rowan student pop. is normal and $\sigma = 2$ years. Find minimal sample size to be 95% confident the sample mean is within 0.5 year of pop. mean.

Answer.
Using $c = 95\%$ so $z_c = 1.96 \approx 2$, and $E \leq 0.5$

$$0.5 \geq E = 2 \cdot \frac{2}{\sqrt{n}} \Rightarrow \sqrt{n} \geq 8 \Rightarrow n \geq 64$$

Conclusion:
We need 64 students to be in our sample to be 95% confident that the mean age of all Rowan students is within the sample mean.

Confidence Interval (Interpretation)

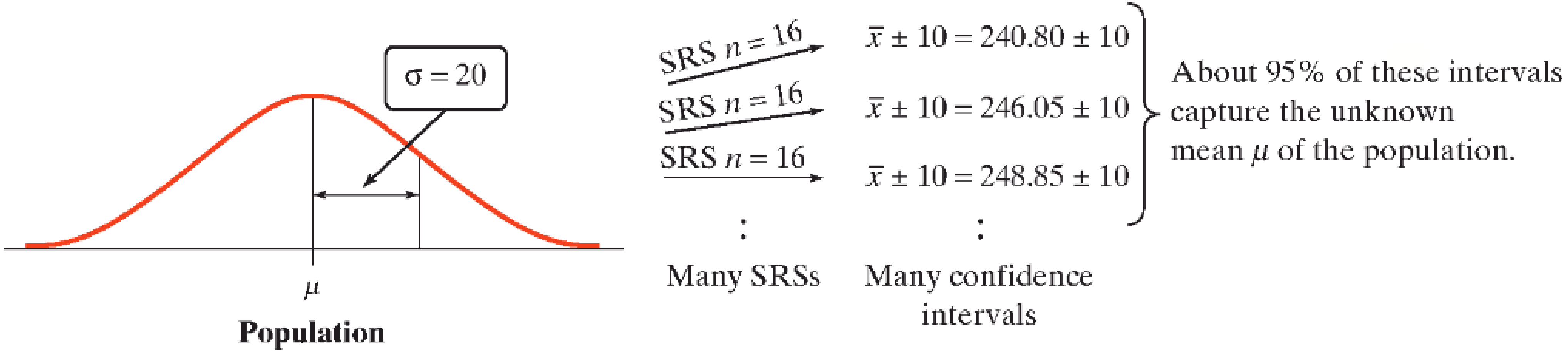


FIGURE 8.4 To say that $\bar{x} \pm 10$ is a 95% confidence interval for the population mean μ is to say that, in repeated samples, about 95% of these intervals capture μ.

Confidence Interval (Interpretation)

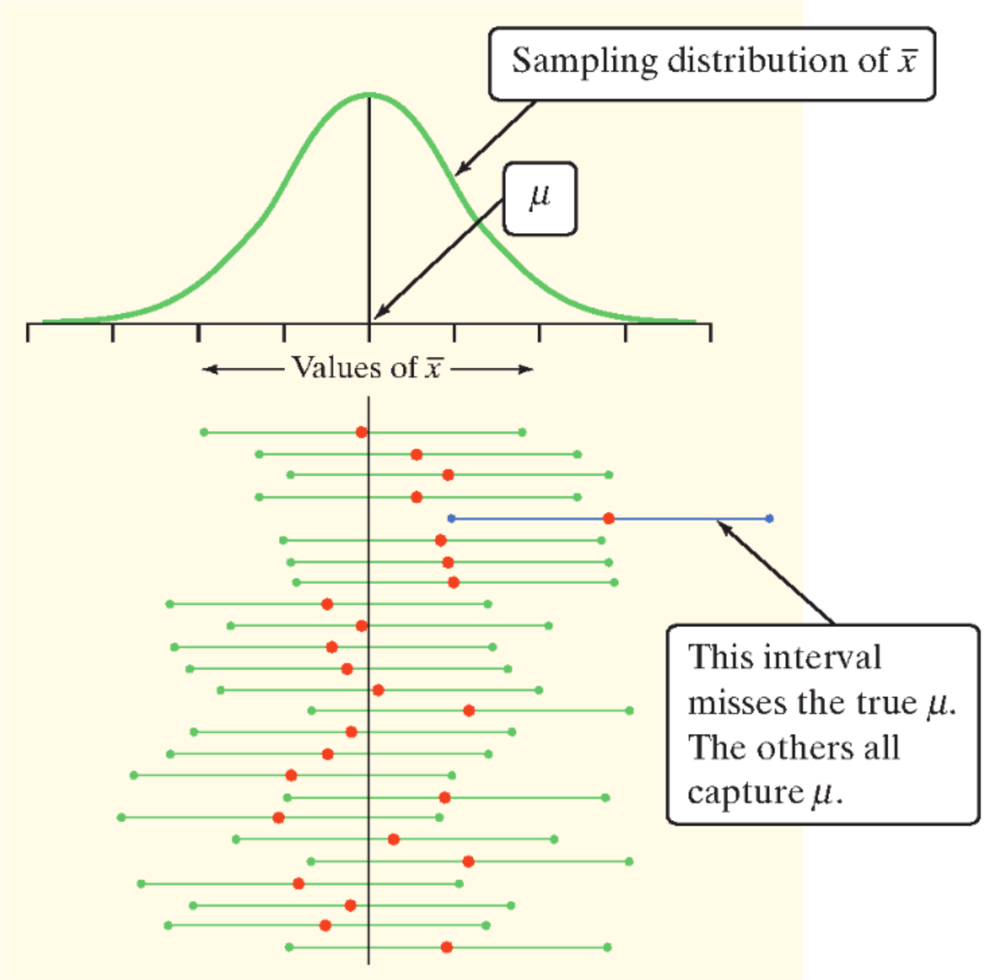


FIGURE 8.5 Twenty-five samples of the same size from the same population gave these 95% confidence intervals. In the long run, about 95% of samples give an interval that captures the population mean μ.

Being *95% confident* is shorthand for:

"If we take many samples of the same size from this population, about 95% of them will result in an interval that captures the true parameter value μ."

(For pop. proportion)

Reference: [ES] 6.3

Confidence Interval of Proportion (Facts)

The $\boldsymbol{c}$**-confidence interval for pop. proportion** $\boldsymbol{p}$ obtained from an SRS with sample proportion $\hat{p}$ is

$$\hat{p} - E < p < \hat{p} + E$$

with **margin of error**

$$E = z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Large counts:** $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$

Confidence Interval of Proportion (Facts)

The **c-confidence interval for pop. proportion p** obtained from an SRS with sample proportion $\hat{p}$ is

$$\hat{p} - E < p < \hat{p} + E$$

with **margin of error**

$$E = z_c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Large counts:** $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$

Example 1.
A survey of 1550 U.S. adults finds 1054 use FB. Find a 95% confidence interval for the population proportion.

Answer.

- 10% Condition: Population is $\geq 10 \cdot 1550$.
- Large counts: $1054 \geq 10$ and $1550 - 1054 \geq 10$

So we can proceed.

$$E = 1.96\sqrt{\frac{0.68(0.32)}{1550}} \approx 0.023$$

So

$$0.68 - 0.023 < p < 0.68 + 0.023$$

We are 95% confident the population proportion of U.S. adults who use FB is between 65.7% and 70.3%.

Confidence Interval of Proportion (Facts)

The **c-confidence interval for pop. proportion p** obtained from an SRS with sample proportion $\hat{p}$ is

$$\hat{p} - E < p < \hat{p} + E$$

with **margin of error**

$$E = z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Large counts:** $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$

To find **minimal sample size** n to ensure $E <$ some number, solve for n. If $\hat{p}$ is unknown, use $\hat{p} = 0.5$.

Confidence Interval of Proportion (Facts)

The **c-confidence interval for pop. proportion p** obtained from an SRS with sample proportion $\hat{p}$ is

$$\hat{p} - E < p < \hat{p} + E$$

with **margin of error**

$$E = z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Large counts:** $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$

To find **minimal sample size** n to ensure $E <$ some number, solve for n. If $\hat{p}$ is unknown, use $\hat{p} = 0.5$.

Example 2. (Minimum sample size)
Find min. sample size to be 95% confident the sample proportion is within 2% of true value.

Answer.

Confidence Interval of Proportion (Facts)

The **c-confidence interval for pop. proportion p** obtained from an SRS with sample proportion $\hat{p}$ is

$$\hat{p} - E < p < \hat{p} + E$$

with **margin of error**

$$E = z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

when the following conditions are met:
- **10% Condition:** $N \geq 10n$.
- **Large counts:** $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$

To find **minimal sample size** n to ensure $E <$ some number, solve for n. If $\hat{p}$ is unknown, use $\hat{p} = 0.5$.

Example 2. (Minimum sample size)
Find min. sample size to be 95% confident the sample proportion is within 2% of true value.

Answer.

$$0.02 > E = 1.96 \sqrt{\frac{0.5(0.5)}{n}}$$

$$\Rightarrow \sqrt{n} \geq \frac{1.96 \cdot 0.5}{0.02} = 49$$

So

$$n \geq 2401$$