

STANDARD ERROR + CENTRAL LIMIT THEOREM

(... and why we *really* should care...)

Håkon K. Gjessing

Professor/Principal Investigator

Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo

Department of Global Public Health and Primary Care, University of Bergen

Kathmandu

Thursday, 1 October 2026



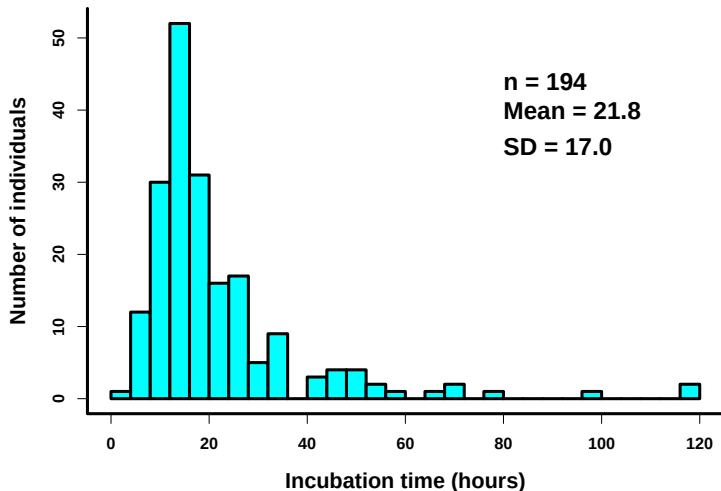
SALMONELLA INFECTIONS

Incubation times for 194 cases of *Salmonella typhimurium* at a medical conference in Wales, 1986

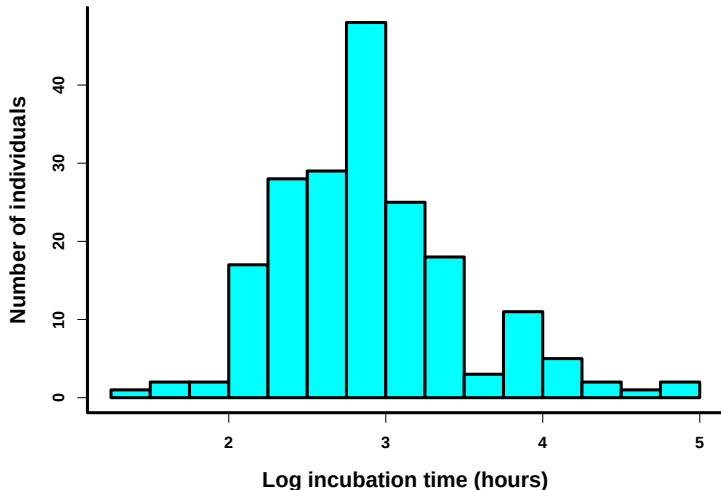
(In hours, sorted by length):

[1]	3.5	4.6	4.7	6.8	7.1	7.4	7.4	7.6	7.7
[10]	7.7	7.7	7.8	7.8	8.1	8.2	8.2	8.4	8.6
[19]	8.6	8.9	9.0	9.2	10.6	10.9	11.1	11.2	11.3
[28]	11.4	11.5	11.5	11.5	11.6	11.6	11.6	11.6	11.6
[37]	11.7	11.8	11.8	11.9	11.9	11.9	11.9	12.0	12.0
[46]	12.1	12.1	12.1	12.1	12.1	12.3	12.3	12.4	12.4
[55]	12.5	12.5	12.5	12.5	12.6	12.9	13.0	13.2	14.7
[64]	14.8	14.9	15.0	15.0	15.1	15.2	15.4	15.4	15.4
[73]	15.4	15.5	15.5	15.6	15.6	15.6	15.6	15.7	15.7
[82]	15.7	15.7	15.7	15.8	15.8	15.8	15.8	15.8	15.8
[91]	15.8	15.9	15.9	15.9	15.9	16.0	16.0	16.0	16.1
[100]	16.1	16.2	16.2	16.3	16.3	16.3	16.3	16.4	16.5
[109]	16.6	16.6	16.6	16.7	16.7	16.7	16.8	17.0	17.3
[118]	18.4	19.4	19.4	19.4	19.5	19.5	19.7	19.9	19.9
[127]	20.0	20.1	20.3	20.3	20.4	20.5	20.7	21.0	23.2
[136]	23.3	23.3	23.5	23.8	23.8	23.9	23.9	24.0	24.2
[145]	24.2	24.2	24.3	24.3	24.4	24.4	24.6	25.3	27.0
[154]	27.1	27.1	27.5	27.5	27.5	27.8	28.1	28.3	30.8
[163]	31.5	31.7	32.0	32.1	32.1	32.1	32.1	32.3	34.9
[172]	35.4	35.7	43.4	43.6	43.8	47.3	47.7	47.7	47.9
[181]	48.0	48.1	48.1	48.4	55.0	55.8	56.1	67.9	68.7
[190]	71.1	79.9	96.2	116.0	116.2				

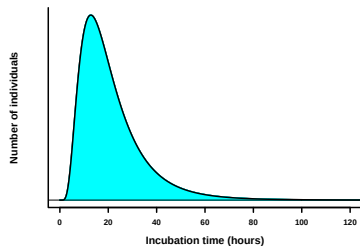
HISTOGRAM OF INCUBATION TIMES FOR 194 INDIVIDUALS XXX



HISTOGRAM OF *log* INCUBATION TIMES FOR 194 INDIVIDUALS



REPEATED SAMPLING FROM A POPULATION

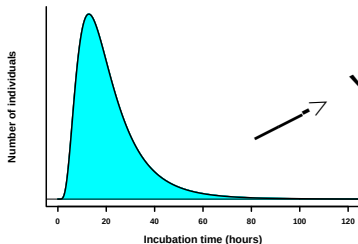


Population (unobserved)

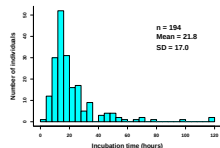
Mean = μ

Std. dev. = σ

REPEATED SAMPLING FROM A POPULATION



Wales, 1986 ($n = 194$)



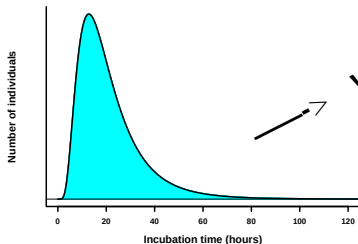
$$\text{Mean} = \bar{X}_1 = 21.8$$

Population (unobserved)

$$\text{Mean} = \mu$$

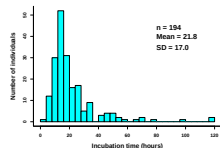
$$\text{Std. dev.} = \sigma$$

REPEATED SAMPLING FROM A POPULATION

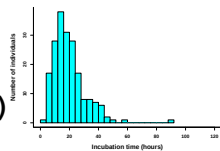


Wales, 1986 ($n = 194$)

($n = 194$)



Mean = $\bar{X}_1 = 21.8$



Mean = $\bar{X}_2 = 19.7$

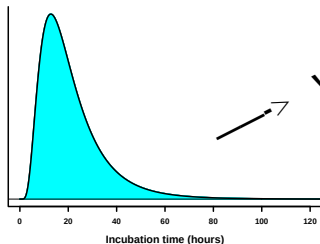
Population (unobserved)

Mean = μ

Std. dev. = σ

REPEATED SAMPLING FROM A POPULATION

Number of individuals



Population (unobserved)

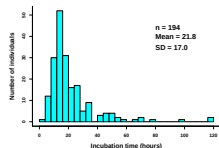
$$\text{Mean} = \mu$$

$$\text{Std. dev.} = \sigma$$

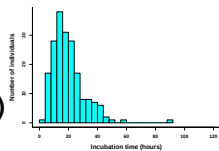
Wales, 1986 ($n = 194$)

($n = 194$)

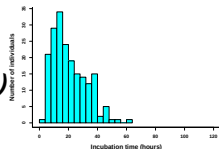
($n = 194$)



$$\text{Mean} = \bar{X}_1 = 21.8$$

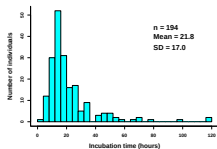


$$\text{Mean} = \bar{X}_2 = 19.7$$

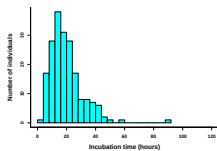


$$\text{Mean} = \bar{X}_3 = 20.5$$

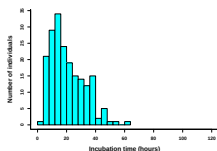
HISTOGRAM OF *averages*



Mean = $\bar{X}_1 = 21.8$

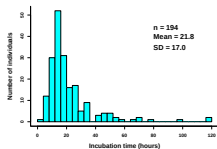


Mean = $\bar{X}_2 = 19.7$

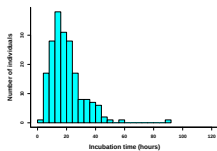


Mean = $\bar{X}_3 = 20.5$

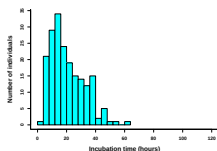
HISTOGRAM OF *averages*



Mean = $\bar{X}_1 = 21.8$

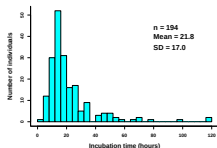


Mean = $\bar{X}_2 = 19.7$

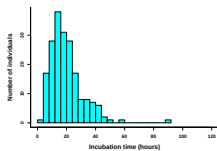


Mean = $\bar{X}_3 = 20.5$

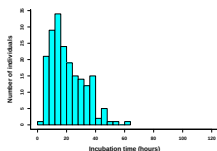
HISTOGRAM OF *averages*



Mean = $\bar{X}_1 = 21.8$

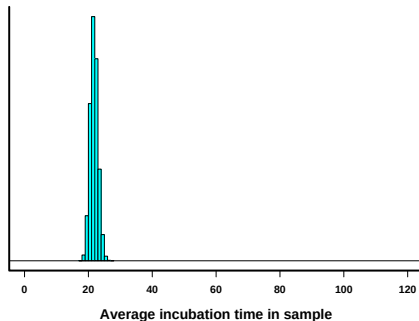


Mean = $\bar{X}_2 = 19.7$



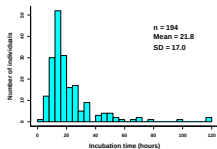
Mean = $\bar{X}_3 = 20.5$

Number of samples (each of size 194)

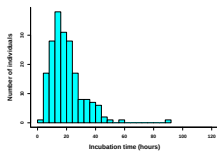


Mean $\approx \mu$

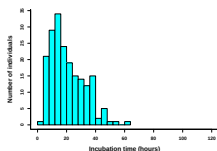
HISTOGRAM OF *averages* (ZOOMED)



Mean = $\bar{X}_1 = 21.8$

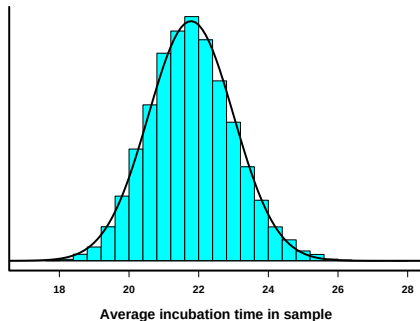


Mean = $\bar{X}_2 = 19.7$



Mean = $\bar{X}_3 = 20.5$

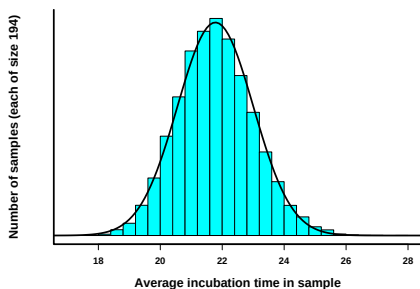
Number of samples (each of size 194)



Mean $\approx \mu$

LAW OF LARGE NUMBERS

Histogram of $\bar{X}$:



$\bar{X}$ (sample) “close to” μ (population)

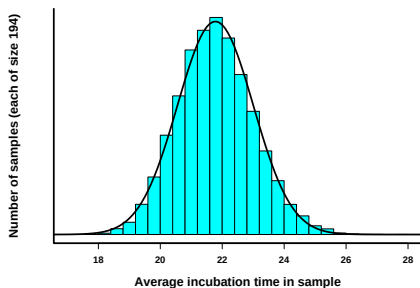
- Standard deviation in **this** distribution:

Standard Error of $\bar{X}$

- Precision = $SE(\bar{X})$

LAW OF LARGE NUMBERS

Histogram of $\bar{X}$:



$\bar{X}$ (sample) “close to” μ (population)

- Standard deviation in **this** distribution:

Standard Error of $\bar{X}$

- Precision = $SE(\bar{X})$

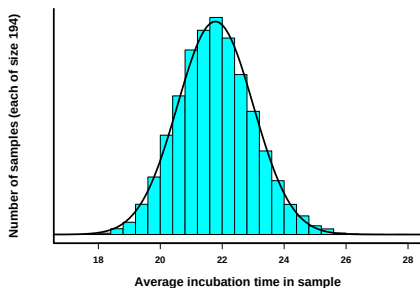
IMPORTANT!

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}} \approx \frac{SD(X)}{\sqrt{n}}$$

$$SE(\bar{X}) \approx \frac{17.0}{\sqrt{194}} = 1.22$$

LAW OF LARGE NUMBERS

Histogram of $\bar{X}$:



$\bar{X}$ (sample) “close to” μ (population)

- Standard deviation in **this** distribution:

Standard Error of $\bar{X}$

- Precision = $SE(\bar{X})$

IMPORTANT!

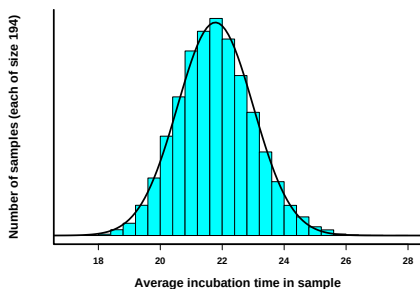
$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}} \approx \frac{SD(X)}{\sqrt{n}}$$

$$SE(\bar{X}) \approx \frac{17.0}{\sqrt{194}} = 1.22$$

- $\sqrt{194} \approx 14$:
 $\bar{X}$ is about 14 times more precise than X

LAW OF LARGE NUMBERS

Histogram of $\bar{X}$:



$\bar{X}$ (sample) “close to” μ (population)

- Standard deviation in **this** distribution:

Standard Error of $\bar{X}$

- Precision = $SE(\bar{X})$

IMPORTANT!

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}} \approx \frac{SD(X)}{\sqrt{n}}$$

$$SE(\bar{X}) \approx \frac{17.0}{\sqrt{194}} = 1.22$$

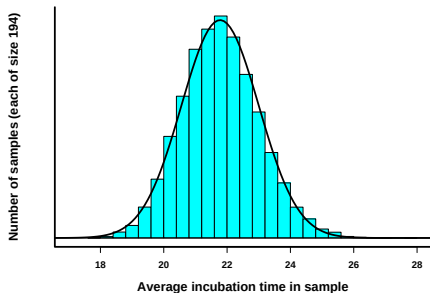
- $\sqrt{194} \approx 14$:
 $\bar{X}$ is about 14 times more precise than X

LAW OF LARGE NUMBERS:

$$\bar{X} \rightarrow \mu \text{ when } n \rightarrow \infty$$

CENTRAL LIMIT THEOREM

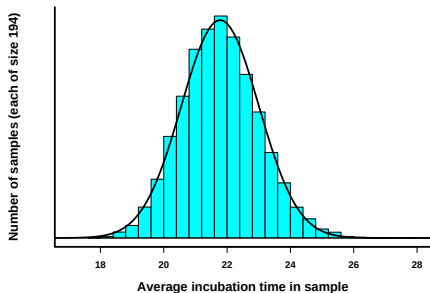
Histogram of $\bar{X}$:



$\bar{X}$ (sample) “close to” μ (population)

CENTRAL LIMIT THEOREM

Histogram of $\bar{X}$:

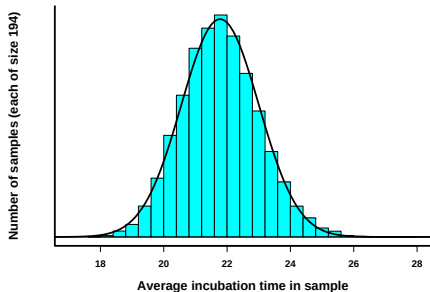


$\bar{X}$ (sample) “close to” μ (population)

- Shape of distribution of $\bar{X}$:
Gaussian (Normal) distribution
(almost, at least)

CENTRAL LIMIT THEOREM

Histogram of $\bar{X}$:

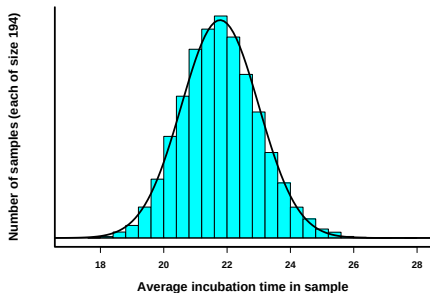


$\bar{X}$ (sample) “close to” μ (population)

- Shape of distribution of $\bar{X}$:
Gaussian (Normal) distribution
(almost, at least)
- Central Limit Theorem:
Regardless of distribution of X ,
distribution of $\bar{X}$ close to Normal

CENTRAL LIMIT THEOREM

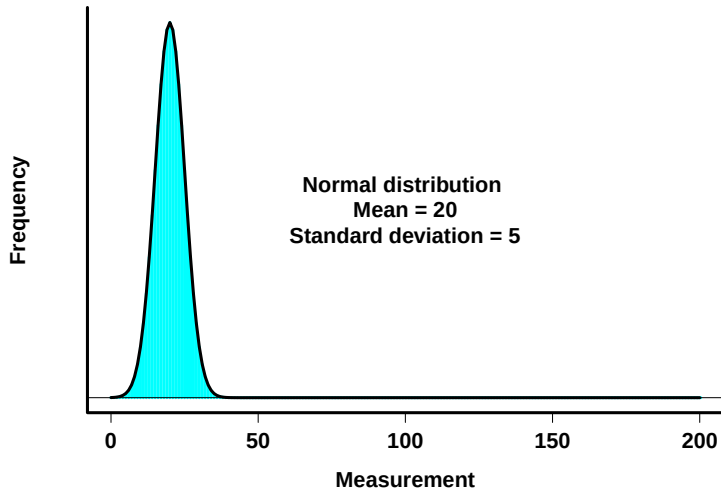
Histogram of $\bar{X}$:



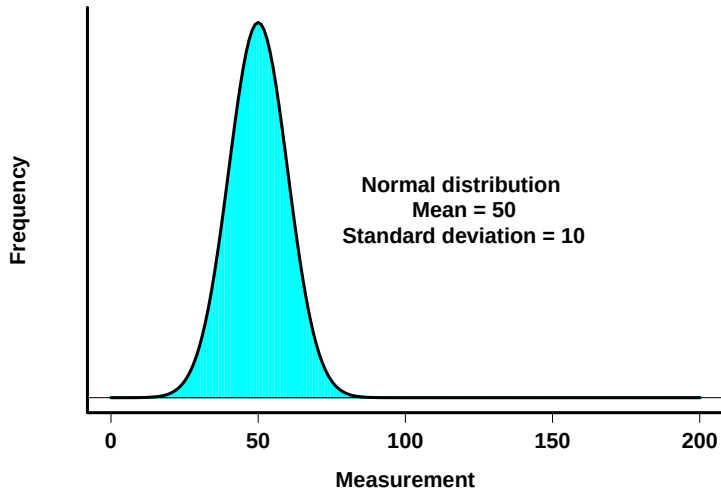
$\bar{X}$ (sample) “close to” μ (population)

- Shape of distribution of $\bar{X}$:
Gaussian (Normal) distribution
(almost, at least)
- Central Limit Theorem:
Regardless of distribution of X ,
distribution of $\bar{X}$ close to Normal
- Larger n (here 194) means
closer to Normal
The more data the better!

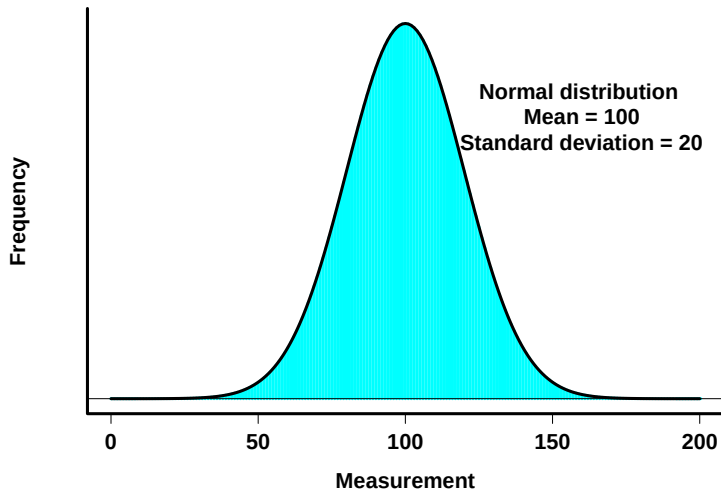
THE NORMAL (GAUSSIAN) DISTRIBUTION



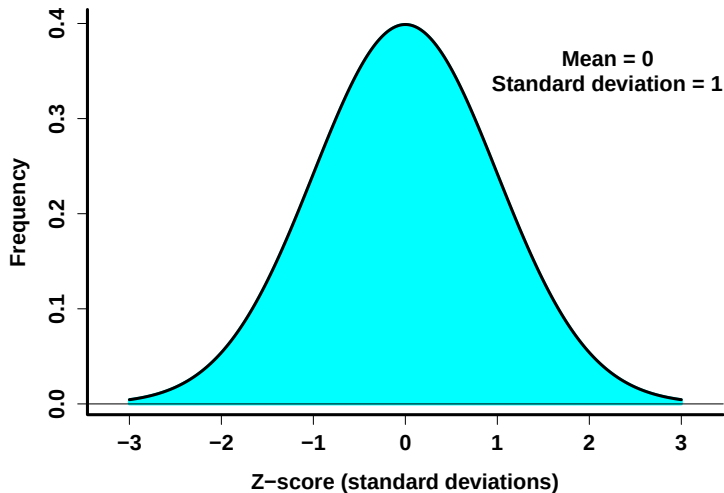
THE NORMAL (GAUSSIAN) DISTRIBUTION



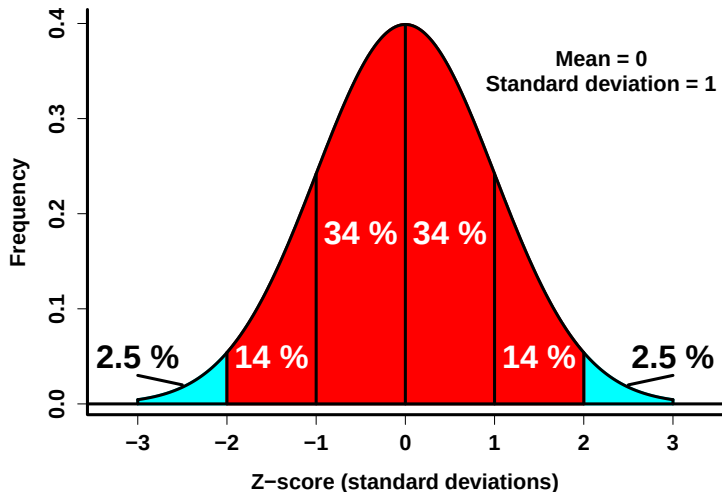
THE NORMAL (GAUSSIAN) DISTRIBUTION



THE *standard* NORMAL DISTRIBUTION



THE *standard* NORMAL DISTRIBUTION



STATISTICAL INFERENCE

Sample		Population
$\bar{X}$	$\rightarrow$	μ
$SD(X)$	$\rightarrow$	σ

- The standard error

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}} \approx \frac{SD(X)}{\sqrt{n}}$$

is the uncertainty (precision) of $\bar{X}$ when estimating μ

How should we use this uncertainty measure??

CONFIDENCE INTERVALS (CI)

Confidence interval based on the Normal distribution:

- $\bar{X}$ is approximately normally distributed ... so the interval

$$(\bar{X} - 1.96 \cdot \text{SE}(\bar{X}), \quad \bar{X} + 1.96 \cdot \text{SE}(\bar{X}))$$

has a 95% chance of covering the “true” μ .

Lower limit:

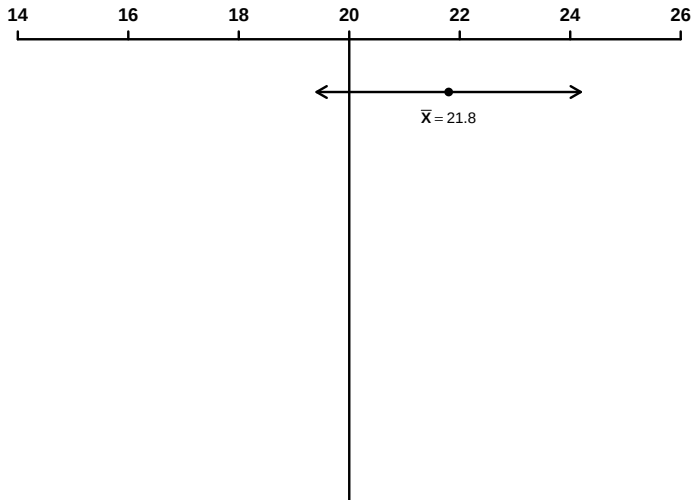
$$\bar{X} - 1.96 \cdot \text{SE}(\bar{X}) = 21.8 - 1.96 \cdot 1.22 = 19.4$$

Upper limit:

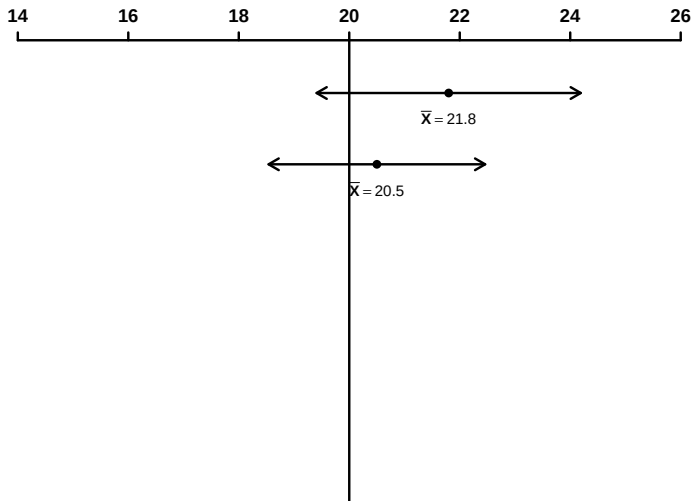
$$\bar{X} + 1.96 \cdot \text{SE}(\bar{X}) = 21.8 + 1.96 \cdot 1.22 = 24.2$$

95% confidence interval: (19.4, 24.2)

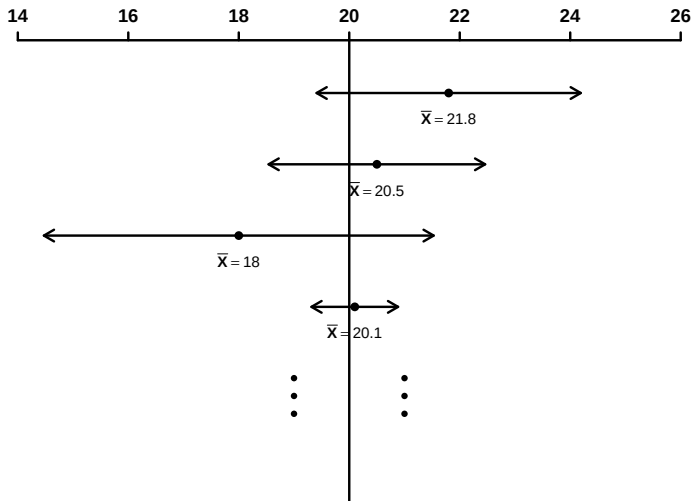
CONFIDENCE INTERVALS IN REPEATED EXPERIMENTS (95% CI)



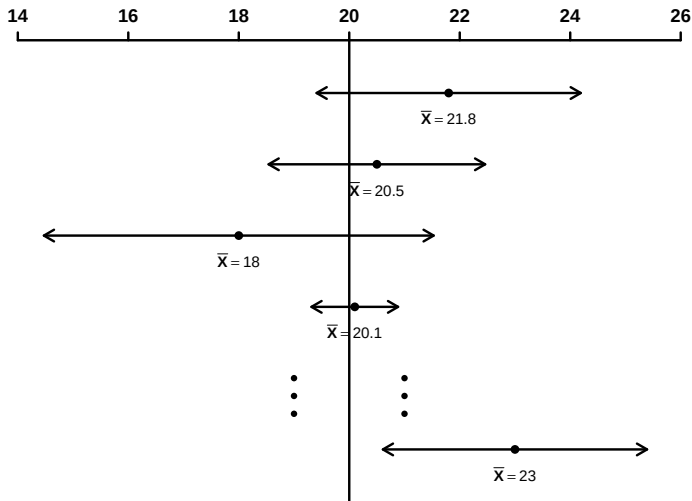
CONFIDENCE INTERVALS IN REPEATED EXPERIMENTS (95% CI)



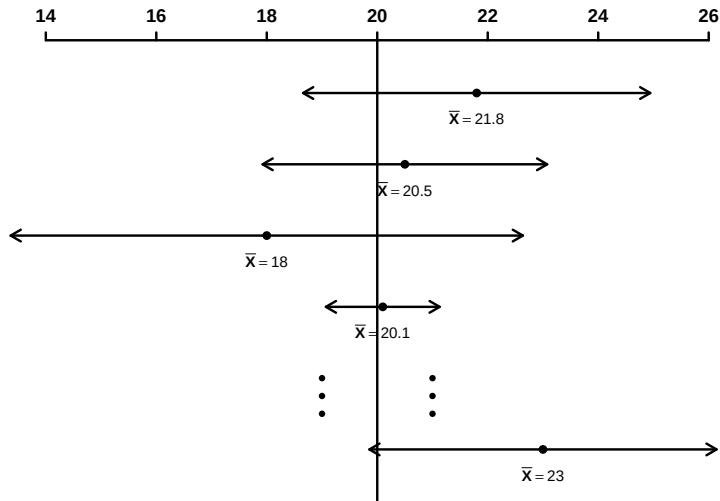
CONFIDENCE INTERVALS IN REPEATED EXPERIMENTS (95% CI)



CONFIDENCE INTERVALS IN REPEATED EXPERIMENTS (95% CI)



CONFIDENCE INTERVALS IN REPEATED EXPERIMENTS (99% CI)



90% CI: 1.64, 95% CI: 1.96, 99% CI: 2.58

SPECIAL NOTE ON THE BINOMIAL DISTRIBUTION

X has a binomial distribution $\text{Bin}(n, p)$

$$X \sim \text{Bin}(n, p)$$

Expected value of X:

$$\mu = E(X) = n \cdot p$$

Standard deviation of X:

$$\sigma = \text{SD}(X) = \sqrt{n \cdot p \cdot (1 - p)}$$

ESTIMATE OF p

$$\hat{p} = \frac{X}{n}$$

STANDARD ERROR OF $\hat{p}$:

$$\text{SE}(\hat{p}) \approx \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

NOTE

A note on notation:

Textbook:

- Population proportion: π
- Sample estimate of proportion: p

Very common in other books (more or less the standard):

- Population proportion: p
- Sample estimate of proportion: $\hat{p}$