Konfidensintervaller i normalfordelingen (del 1)

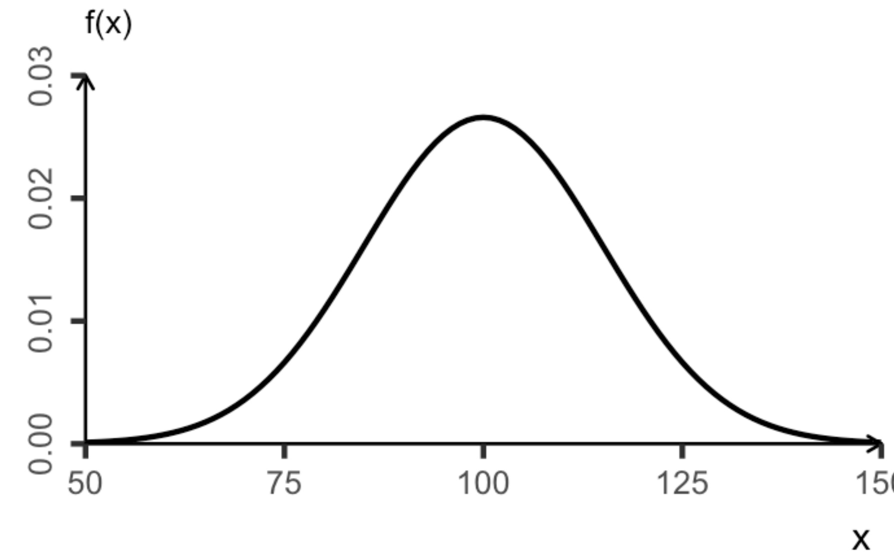
Thea Bjørnland
Institutt for matematiske fag
NTNU

Estimere forventningsverdien μ i normalfordelingen

EKS: IQ

Generelt er IQ-tester konstruert slik at

IQ ~ N(100, 15)   i en bestemt aldersgruppe



Eks: gjennomsnittlig IQ blant alle nordmenn i 20-åra ≈ 100

Forskningsspørsmål:

40 000 studenter

Hva er gjennomsnittlig IQ blant NTNU-studenter?

$$X \sim N(\mu, 15)$$

Antagelse: $\sigma = 15$

IQ til tilfeldig valgt NTNU-student

Ukjent verdi som vi vil estimere

Vi tester IQ til 100 **tilfeldig valgte** NTNU-studenter

Estimere forventningsverdien μ i normalfordelingen

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

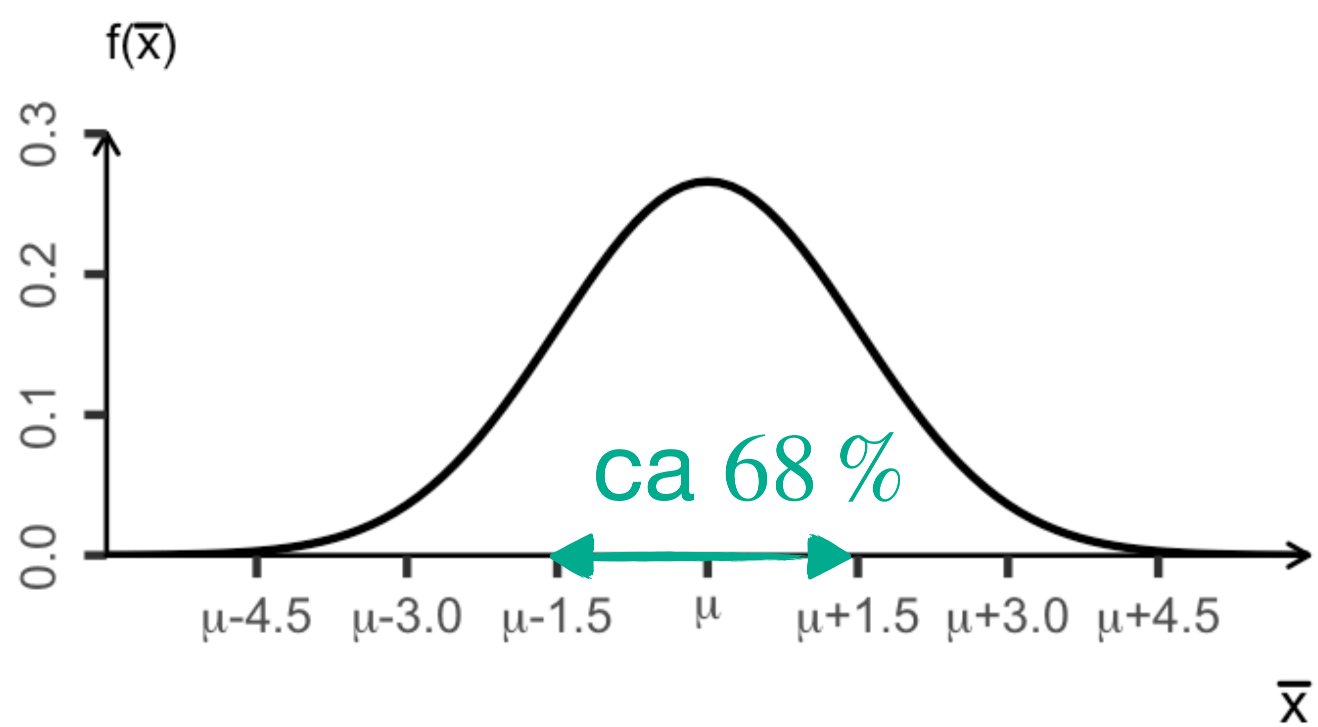
$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

Estimator → $$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100}$$

Egenskaper:

$$E(\bar{X}) = \mu \qquad Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{15^2}{100} \qquad SD(\bar{X}) = \frac{15}{\sqrt{100}} = 1.5$$

$$\bar{X} \sim N\left(\mu, 1.5\right)$$



Estimere forventningsverdien μ i normalfordelingen

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

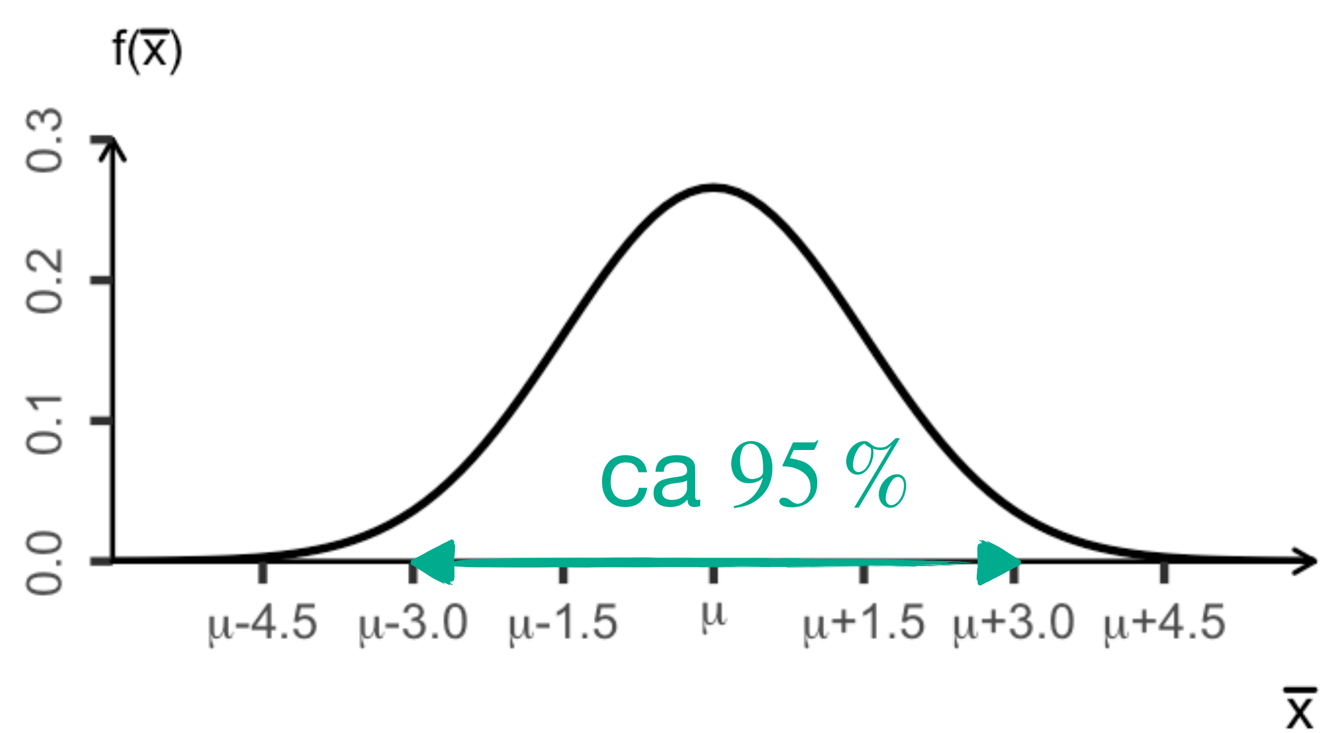
$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

Estimator → $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_{100}}{100}$

Egenskaper:

$$E(\bar{X}) = \mu \qquad Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{15^2}{100} \qquad SD(\bar{X}) = \frac{15}{\sqrt{100}} = 1.5$$

$$\bar{X} \sim N\left(\mu, 1.5\right)$$



Estimere forventningsverdien μ i normalfordelingen

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

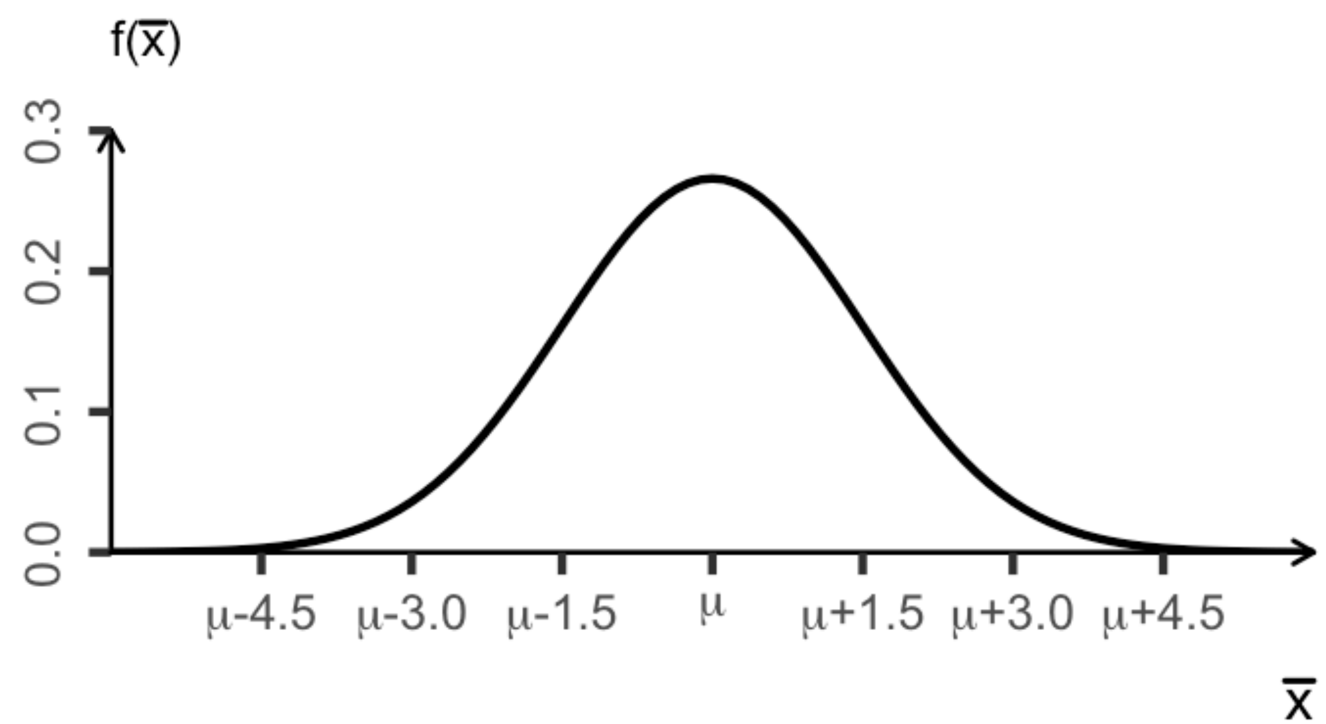
Estimator → $$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100} \qquad \bar{X} \sim N\left(\mu, 1.5\right)$$

2. Vi gjennomfører IQ-tester på n = 100 studenter og regner ut et estimat

$$x_1, x_2, \ldots, x_{100}$$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_{100}}{100}$$

$$\bar{x} = 103.5$$



Hva tror vi egentlig om verdien til parameteren μ?

Estimere forventningsverdien μ i normalfordelingen

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

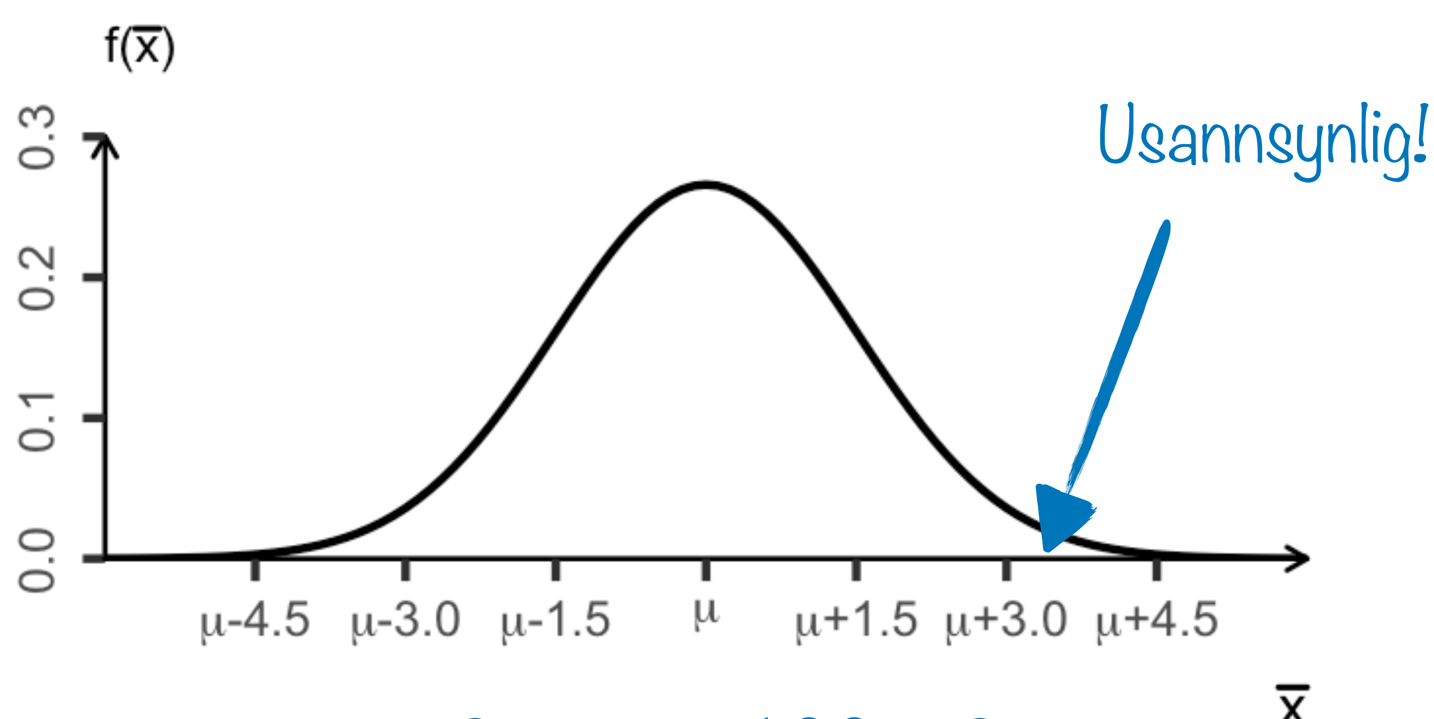
Estimator → $$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100} \qquad \bar{X} \sim N\left(\mu, 1.5\right)$$

2. Vi gjennomfører IQ-tester på n = 100 studenter og regner ut et estimat

$$x_1, x_2, \ldots, x_{100}$$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_{100}}{100}$$

$$\bar{x} = 103.5$$



Hva tror vi egentlig om verdien til parameteren μ?

? $\mu = 100$?

Estimere forventningsverdien μ i normalfordelingen

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

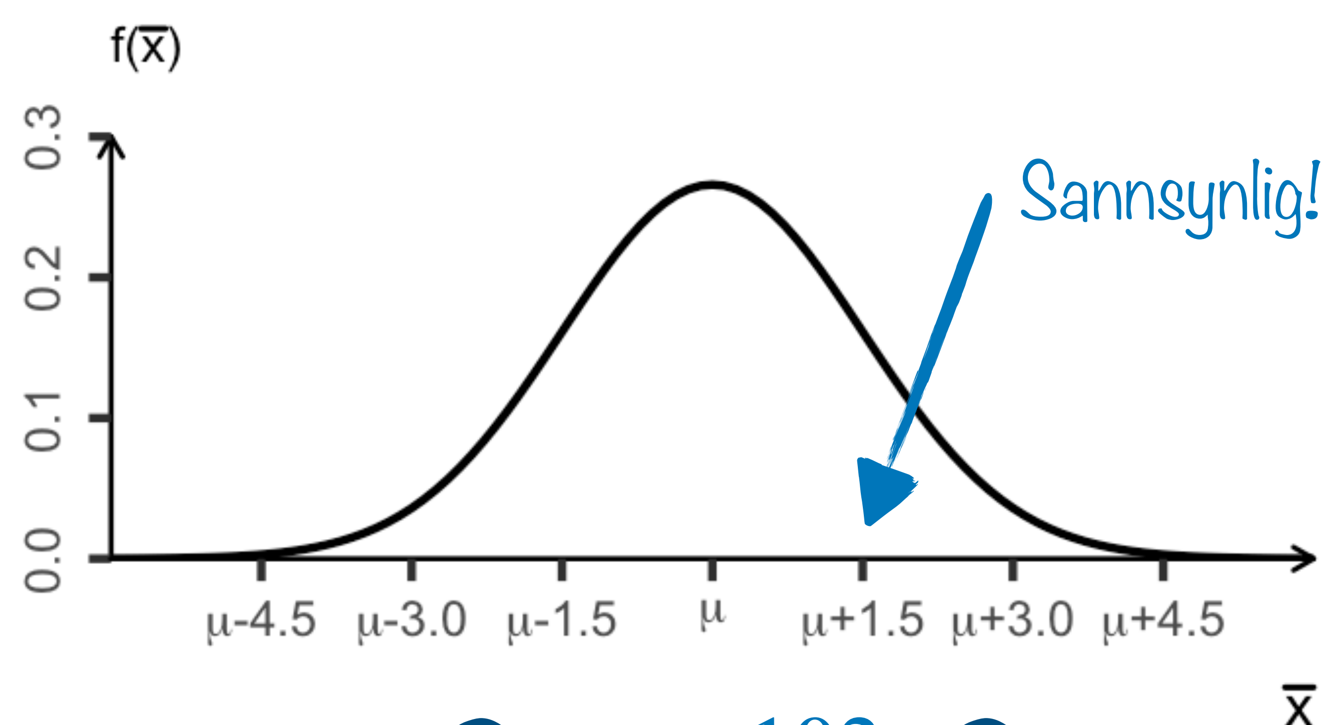
Estimator → $$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100} \qquad \bar{X} \sim N\left(\mu, 1.5\right)$$

2. Vi gjennomfører IQ-tester på n = 100 studenter og regner ut et estimat

$$x_1, x_2, \ldots, x_{100}$$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_{100}}{100}$$

$$\bar{x} = 103.5$$



Hva tror vi egentlig om verdien til parameteren μ?

? $\mu = 102$?

Estimere forventningsverdien μ i normalfordelingen

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

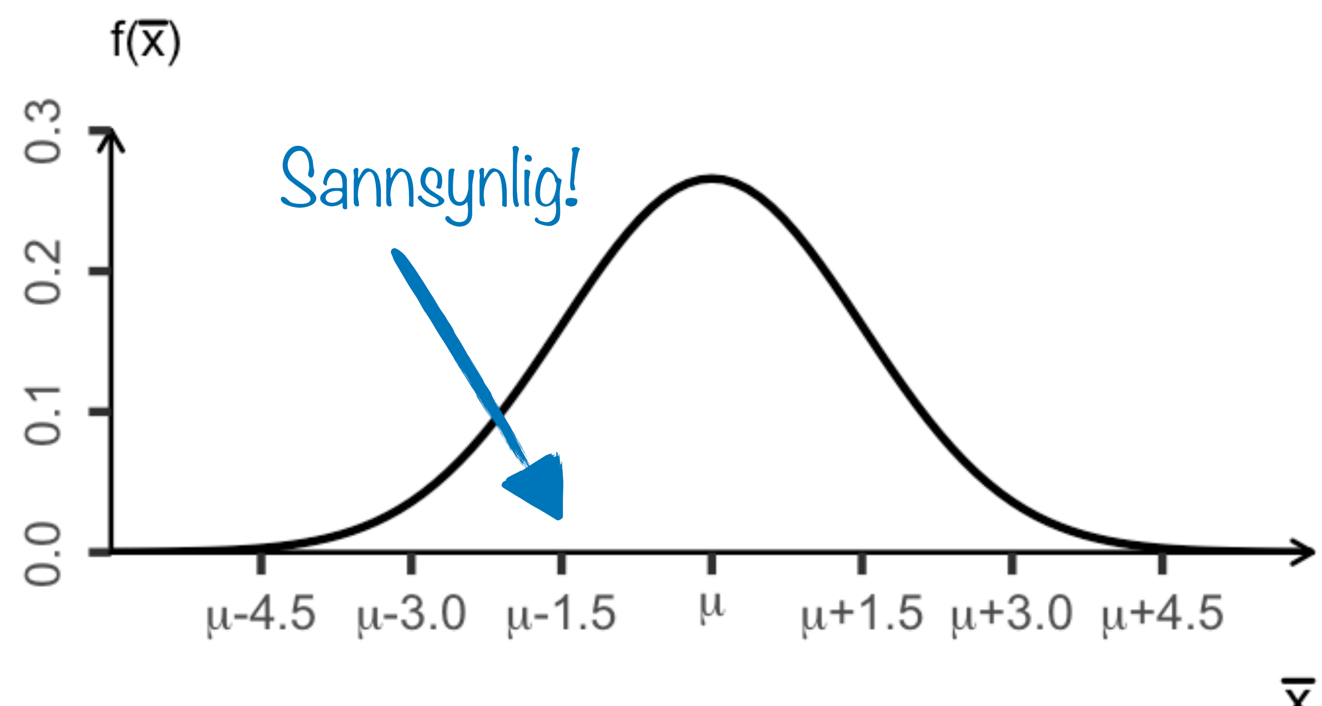
Estimator → $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_{100}}{100}$ $\qquad \bar{X} \sim N\left(\mu, 1.5\right)$

2. Vi gjennomfører IQ-tester på n = 100 studenter og regner ut et estimat

$$x_1, x_2, \ldots, x_{100}$$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_{100}}{100}$$

$$\bar{x} = 103.5$$



Hva tror vi egentlig om verdien til parameteren μ?

? $\mu = 105$?

Vi bør bruke observasjonene våre til å lage et **intervall** av rimelige verdier for den ukjente forventningsverdien

Konfidensintervall for forventningsverdien μ

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

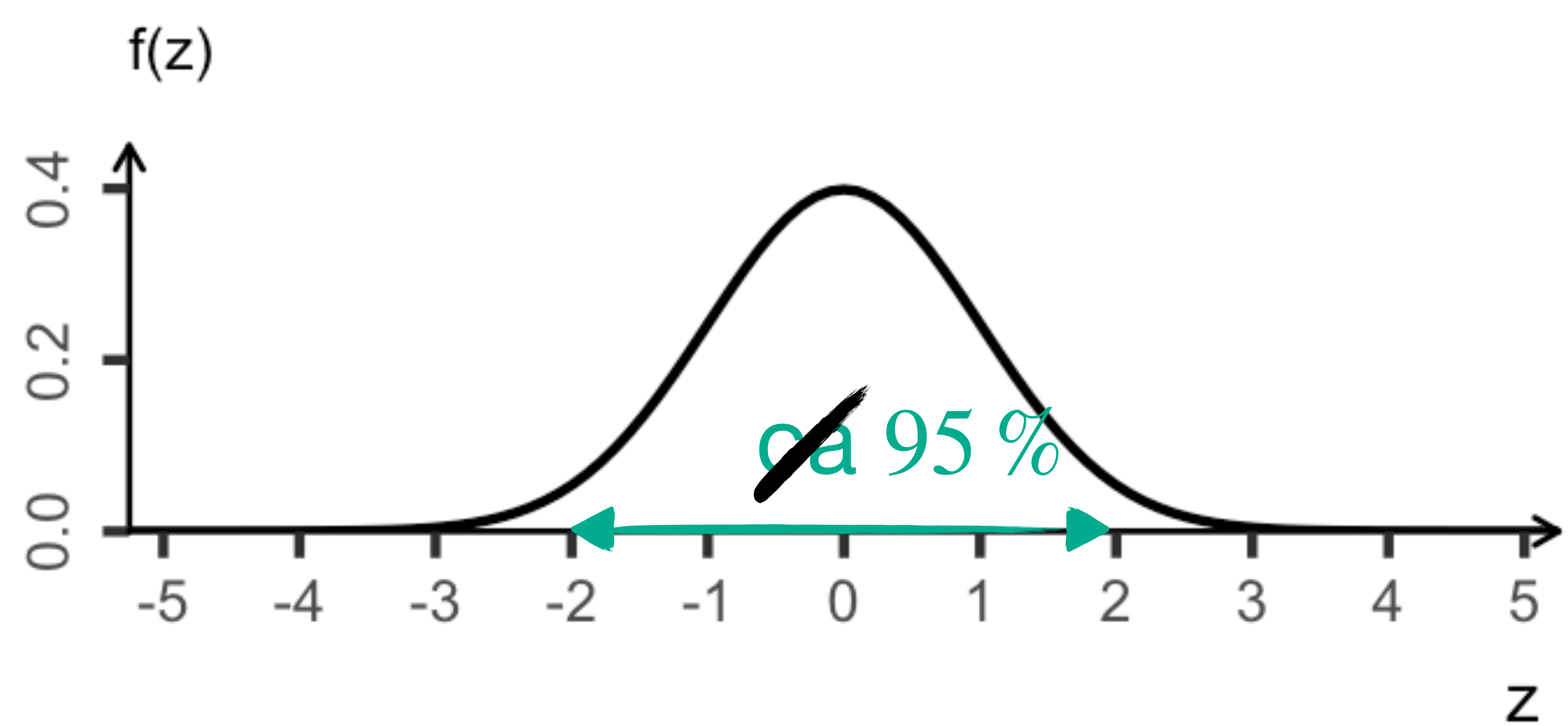
$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

Estimator → $$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100} \qquad \bar{X} \sim N\left(\mu, 1.5\right)$$

Standard normalfordelt variabel

$$Z = \frac{\bar{X} - \mu}{1.5}$$

$$Z \sim N(0, 1)$$



Konfidensintervall for forventningsverdien μ

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

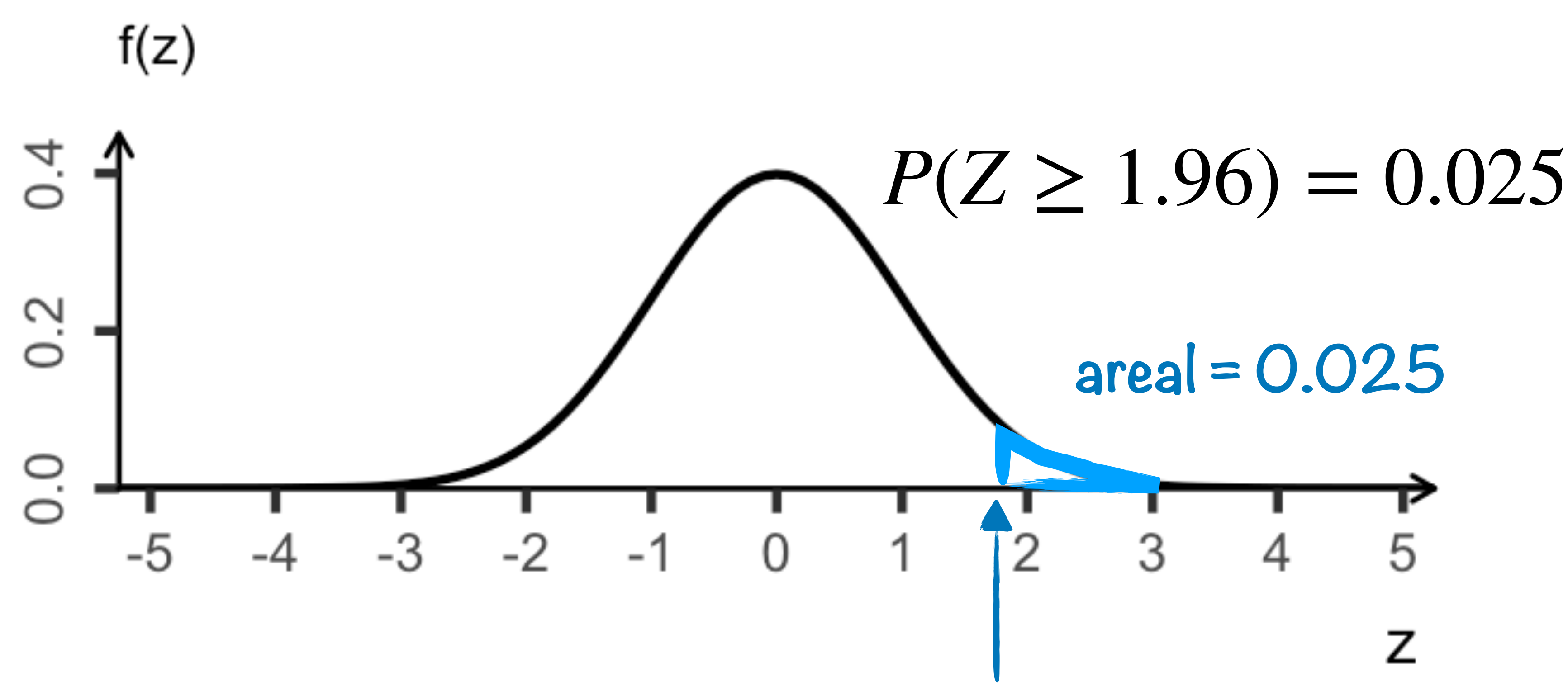
$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

Estimator → $$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100} \qquad \bar{X} \sim N\left(\mu, 1.5\right)$$

Standard normalfordelt variabel

$$Z = \frac{\bar{X} - \mu}{1.5}$$

$$Z \sim N(0, 1)$$



Kvantil i standard normalfordelingen: $z_{0.025} = 1.96$

Konfidensintervall for forventningsverdien μ

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

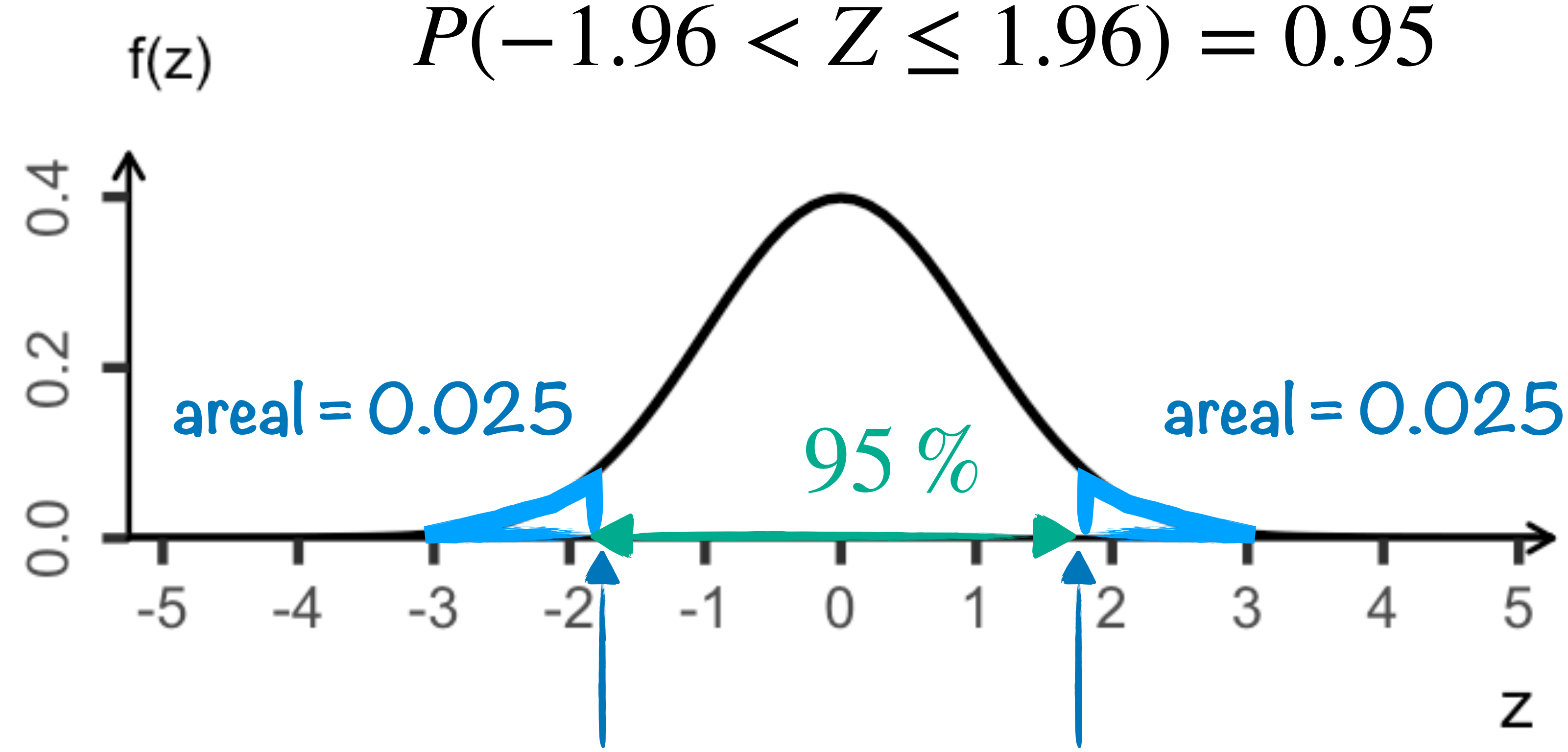
Estimator $\rightarrow$ $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_{100}}{100}$ $\qquad \bar{X} \sim N\left(\mu, 1.5\right)$

Standard normalfordelt variabel

$$Z = \frac{\bar{X} - \mu}{1.5}$$

$$Z \sim N(0, 1)$$

$$P(-1.96 < Z \leq 1.96) = 0.95$$



$$-z_{0.025} = -1.96 \qquad z_{0.025} = 1.96$$

Konfidensintervall for forventningsverdien μ

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

Estimator → $$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100} \qquad \bar{X} \sim N\left(\mu, 1.5\right)$$

Standard normalfordelt variabel

$$Z = \frac{\bar{X} - \mu}{1.5} \qquad P(-1.96 < Z \leq 1.96) = 0.95$$

$$P\left(-1.96 < \frac{\bar{X} - \mu}{1.5} \leq 1.96\right) = 0.95$$

$$Z \sim N(0, 1)$$

⋮

Øvre grense

$$P(\bar{X} - 1.96 \cdot 1.5 < \mu \leq \bar{X} + 1.96 \cdot 1.5) = 0.95$$

Nedre grense

$$\left[\bar{X} - 1.96 \cdot 1.5, \quad \bar{X} + 1.96 \cdot 1.5\right]$$

Konfidensintervall for forventningsverdien μ

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

Estimator → $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_{100}}{100}$ $\qquad \bar{X} \sim N\left(\mu, 1.5\right)$

95% konfidensintervall

Standard normalfordelt variabel

$$Z = \frac{\bar{X} - \mu}{1.5} \qquad P(-1.96 < Z \leq 1.96) = 0.95$$

$$P(-1.96 < \frac{\bar{X} - \mu}{1.5} \leq 1.96) = 0.95$$

$$Z \sim N(0, 1)$$

Øvre grense

$$P(\bar{X} - 1.96 \cdot 1.5 < \mu \leq \bar{X} + 1.96 \cdot 1.5) = 0.95$$

Nedre grense

$$\left[\bar{X} - 1.96 \cdot 1.5, \quad \bar{X} + 1.96 \cdot 1.5\right]$$

Konfidensintervall for forventningsverdien μ

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

Estimator → $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_{100}}{100} \qquad \bar{X} \sim N\left(\mu, 1.5\right)$

95% konfidensintervall for μ: $\left[\bar{X} - 1.96 \cdot 1.5, \quad \bar{X} + 1.96 \cdot 1.5\right]$

2. Vi gjennomfører IQ-tester på n = 100 studenter og regner ut et punktestimat og et intervall

$$x_1, x_2, \ldots, x_{100}$$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_{100}}{100} \qquad [\bar{x} - 1.96 \cdot 1.5, \quad \bar{x} + 1.96 \cdot 1.5]$$

$$\bar{x} = 103.5 \qquad [100.6, \quad 106.4]$$

Konfidensintervall for forventningsverdien μ

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

Estimator → $$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100} \qquad \bar{X} \sim N\left(\mu, 1.5\right)$$

95% konfidensintervall for μ: $$\left[\bar{X} - 1.96 \cdot 1.5, \quad \bar{X} + 1.96 \cdot 1.5\right]$$

2. Vi gjennomfører IQ-tester på n = 100 studenter og regner ut et punktestimat og et intervall

Hvis vi gjør steg 2 veldig mange ganger....

$$x_1, x_2, \ldots, x_{100}$$

Ligger parameterverdien μ inne i dette intervallet?

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_{100}}{100}$$

... så vil 95% av alle intervallene inneholde parameterverdien μ

$$\bar{x} = 103.5 \qquad [100.6, \quad 106.4]$$

... men 5% vil bomme

Konfidensintervall for forventningsverdien μ

1. Vi ser på et tilfeldig utvalg av n = 100 studenter

$$X_1, X_2, \ldots, X_{100} \qquad X_i \sim N(\mu, 15) \qquad i = 1,\ldots,100$$

Estimator → $$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100} \qquad \bar{X} \sim N\left(\mu, 1.5\right)$$

$$z_{0.005} = 2.576$$

95% konfidensintervall for μ: $$\left[\bar{X} - \cancel{1.96} \cdot 1.5, \quad \bar{X} + \cancel{1.96} \cdot 1.5\right]$$

99%

2. Vi gjennomfører IQ-tester på n = 100 studenter og regner ut et punktestimat og et intervall

$$x_1, x_2, \ldots, x_{100}$$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_{100}}{100}$$

$$\bar{x} = 103.5$$

Ligger parameterverdien μ inne i dette intervallet?

$$[99.6, \quad 107.4]$$

$$\cancel{[100.6, \quad 106.4]}$$

Hvis vi gjør steg 2 veldig mange ganger....

99%

... så vil 95% av alle intervallene inneholde parameterverdien μ

1%

... men 5% vil bomme

Resultat:

Konfidensintervall for forventningsverdien μ
når standardavviket σ er kjent

$X_1, X_2, \ldots, X_n \qquad X_i \sim N(\mu, \sigma) \qquad i = 1, \ldots, n$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

"Standardfeilen til estimatoren"

$SE(\bar{X})$

95% konfidensintervall for μ: $\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$

Resultat:

Konfidensintervall for forventningsverdien μ når standardavviket σ er kjent

$$X_1, X_2, \ldots, X_n \qquad X_i \sim N(\mu, \sigma) \qquad i = 1, \ldots, n$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

"Standardfeilen til estimatoren"

$$SE(\bar{X})$$

$$z_{0.025} = 1.96$$

95% konfidensintervall for μ: $\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$

$$P(Z \leq -1.96) = 0.025 \qquad P(Z \geq 1.96) = 0.025$$

Resultat:

Konfidensintervall for forventningsverdien μ når standardavviket σ er kjent

$$X_1,\ X_2,\ \ldots,\ X_n \qquad X_i \sim N(\mu,\sigma) \qquad i = 1,\ldots,n$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

95% konfidensintervall for μ: $\left[\bar{X} - z_{0.025}\cdot\frac{\sigma}{\sqrt{n}},\quad \bar{X} + z_{0.025}\cdot\frac{\sigma}{\sqrt{n}}\right]$

(1-0.05) 100%

$-z_{0.05/2}$ $\qquad$ $z_{0.05/2}$

Resultat:

Konfidensintervall for forventningsverdien μ
når standardavviket σ er kjent

$$X_1, X_2, \ldots, X_n \qquad X_i \sim N(\mu, \sigma) \qquad i = 1, \ldots, n$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

95% konfidensintervall for μ: $\left[\bar{X} - z_{0.05/2} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{0.05/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$

$\alpha \qquad \alpha$

(1-0.05)100%

α

Resultat:

Konfidensintervall for forventningsverdien μ
når standardavviket σ er kjent

$$X_1,\, X_2, \ldots, X_n \qquad X_i \sim N(\mu, \sigma) \qquad i = 1, \ldots, n$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

(1-α)100% konfidensintervall for μ:

$$\left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$