



Estimatorar

Torstein Fjeldstad

Institutt for matematiske fag, NTNU

11.10.2018

I dag



- Repetisjon
- Er dataa mine normalfordelt?
- Estimatorar
- Eigenskapar til S^2
- Kahoot



Repetisjon

Observator

Ein **observator** (eng: statistic) er ein funksjon av stokastiske variablar i eit tilfeldig utval.

Utvalsfordeling

Fordelinga til ein observator kallast utvalsfordelinga (eng: sampling distribution).

Ofte brukte observatorar

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

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

Sentralgrenseteoremet (utvalgsfordelinga til \bar{X})

Teorem

La X_i vere uavhengig identisk fordelte (u.i.f.) stokastiske variablar med $E(X_i) = \mu_X$ og $Var(X_i) = \sigma_X^2 < \infty$ for $i = 1, \dots, n$. La $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Då vil fordelinga til

$$Z = \frac{\bar{X}_n - E(\bar{X}_n)}{SD(\bar{X}_n)} = \frac{\bar{X}_n - \mu_X}{\sigma_X / \sqrt{n}}$$

gå mot ei standard normalfordeling når $n \rightarrow \infty$.

Sentralgrenseteoremet (utvalgsfordelinga til \bar{X})

Teorem

La X_i vere uavhengig identisk fordelte (u.i.f.) stokastiske variablar med $E(X_i) = \mu_X$ og $Var(X_i) = \sigma_X^2 < \infty$ for $i = 1, \dots, n$. La $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Då vil fordelinga til

$$Z = \frac{\bar{X}_n - E(\bar{X}_n)}{SD(\bar{X}_n)} = \frac{\bar{X}_n - \mu_X}{\sigma_X / \sqrt{n}}$$

gå mot ei standard normalfordeling når $n \rightarrow \infty$.

Dette tilsvarar

$$\bar{X}_n \approx n \left(x; \mu_X, \frac{\sigma_X}{\sqrt{n}} \right).$$



Er dataa mine normalfordelt?

Fysisk kondisjon ved maksimalt oksygenopptak



- Helseundersøkinga i Nord-Trøndelag (HUNT) er ei av verdas største folkehelseundersøkingar. HUNT 3 pågjekk i perioden 2006-2008
- Me skal sjå på data frå 1471 menn frå *Kondisprosjektet i HUNT 3*
- Prosjektet ville kartlegge samanhengen mellom kondisjon, fysisk aktivitet, smerte, karfunksjon osv.

Maksimalt oksygenopptak (VO_{2max}) måles ved å springe på ei tredemølle med oksygenmaske til ein ikkje klarer meir.

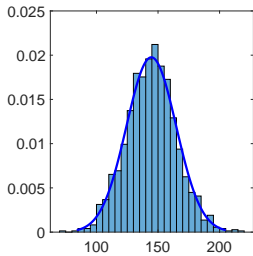


Figure: Histogram



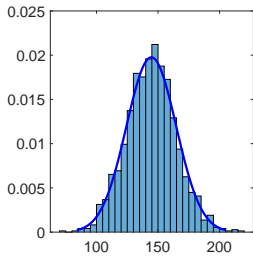


Figure: Histogram

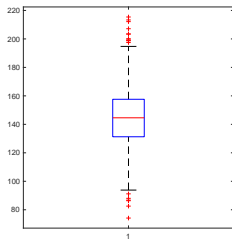


Figure: Boksploott

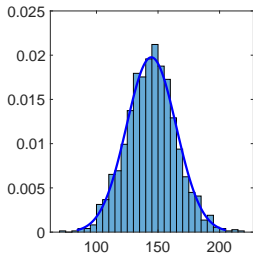


Figure: Histogram

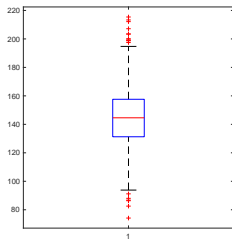


Figure: Bokslott

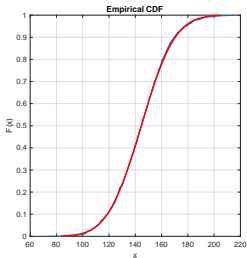


Figure: Kumulativ fordeling

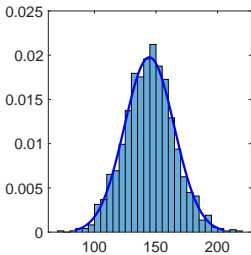


Figure: Histogram

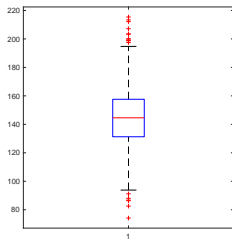


Figure: Boksploott

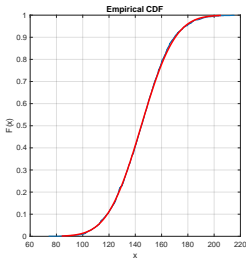


Figure: Kumulativ fordeling

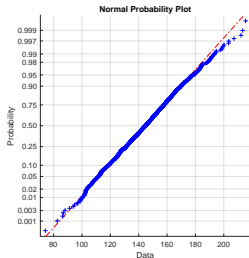


Figure: Normalplott





Estimatorar

Definisjon

Anta at me har eit tilfeldig utval X_1, X_2, \dots, X_n frå $f(x; \theta)$ -populasjonen, der verdien til parameteren θ er ukjend. Ein estimator er ein observator som nyttast til å anslå verdien til θ .

Definisjon

Anta at me har eit tilfeldig utval X_1, X_2, \dots, X_n frå $f(x; \theta)$ -populasjonen, der verdien til parameteren θ er ukjend. Ein estimator er ein observator som nyttast til å anslå verdien til θ .

Forventningsrett estimator

Ein observator $\hat{\theta}$ seies å verre ein forventningsrett (eng: unbiased) estimator for parameteren θ dersom

$$E(\hat{\theta}) = \theta.$$

Definisjon

Anta at me har eit tilfeldig utval X_1, X_2, \dots, X_n frå $f(x; \theta)$ -populasjonen, der verdien til parameteren θ er ukjend. Ein estimator er ein observator som nyttast til å anslå verdien til θ .

Forventningsrett estimator

Ein observator $\hat{\theta}$ seies å verre ein forventningsrett (eng: unbiased) estimator for parameteren θ dersom

$$E(\hat{\theta}) = \theta.$$

Effisient estimator

Av fleire forventningsrette estimatorar for θ seier me at den med minst varians er den mest effisiente.

Me føretrekk den mest effisiente estimatoren.

Illustrasjon ved simulering

— Situasjon:

- X_1, X_2, \dots, X_n tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til μ er ukjent.
- Skal estimere (anslå) verdien til μ .

— Naturlige estimatorer:

$$\hat{\mu}_1 = \bar{X} \text{ (gjennomsnitt)} \quad \text{og} \quad \hat{\mu}_2 = \tilde{X} \text{ (empirisk median)}$$

Illustrasjon ved simulering

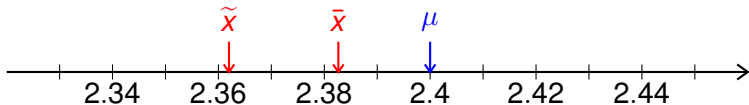
— Situasjon:

- X_1, X_2, \dots, X_9 tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til μ er ukjent.
- Skal estimere (anslå) verdien til μ .

— Generer observasjoner $X_i \sim n(x_i; 2.4, 0.1)$.

2.480, 2.262, 2.560, 2.281, 2.503, 2.362, 2.258, 2.282, 2.456

$$\bar{x} = 2.3827 \quad \text{og} \quad \tilde{x} = 2.362$$



Illustrasjon ved simulering

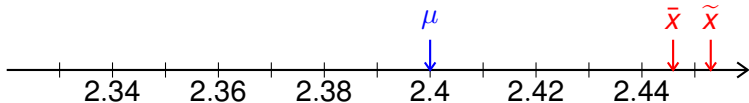
— Situasjon:

- X_1, X_2, \dots, X_9 tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til μ er ukjent.
- Skal estimere (anslå) verdien til μ .

— Generer observasjoner $X_i \sim n(x_i; 2.4, 0.1)$.

2.482, 2.407, 2.415, 2.406, 2.480, 2.453, 2.512, 2.331, 2.527

$$\bar{x} = 2.4459 \quad \text{og} \quad \tilde{x} = 2.453$$

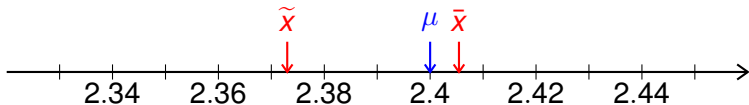


Illustrasjon ved simulering

- Situasjon:
 - X_1, X_2, \dots, X_9 tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen.
 - Verdien til μ er ukjent.
 - Skal estimere (anslå) verdien til μ .
- Generer observasjoner $X_i \sim n(x_i; 2.4, 0.1)$.

2.464, 2.277, 2.547, 2.393, 2.373, 2.369, 2.321, 2.369, 2.536

$$\bar{x} = 2.4054 \quad \text{og} \quad \tilde{x} = 2.373$$



Illustrasjon ved simulering

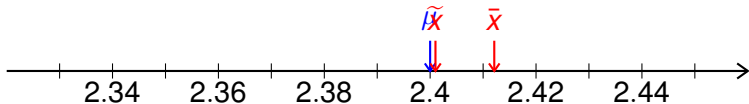
— Situasjon:

- X_1, X_2, \dots, X_9 tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til μ er ukjent.
- Skal estimere (anslå) verdien til μ .

— Generer observasjoner $X_i \sim n(x_i; 2.4, 0.1)$.

2.544, 2.401, 2.465, 2.430, 2.389, 2.461, 2.330, 2.325, 2.364

$$\bar{x} = 2.4121 \quad \text{og} \quad \tilde{x} = 2.401$$

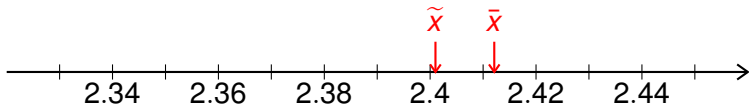


Illustrasjon ved simulering

- Situasjon:
 - X_1, X_2, \dots, X_9 tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen.
 - Verdien til μ er ukjent.
 - Skal estimere (anslå) verdien til μ .
- Observerer verdier for X_1, X_2, \dots, X_9 .

2.544, 2.401, 2.465, 2.430, 2.389, 2.461, 2.330, 2.325, 2.364

$$\bar{x} = 2.4121 \quad \text{og} \quad \tilde{x} = 2.401$$



Illustrasjon ved simulering

- Situasjon:
 - X_1, X_2, \dots, X_9 tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen.
 - Verdien til μ er ukjent.
 - Skal estimere (anslå) verdien til μ .
- Gjentar forsøket 10 000 ganger.
 - Lager histogram over de 10 000 verdiene av \bar{x} og \tilde{x} .



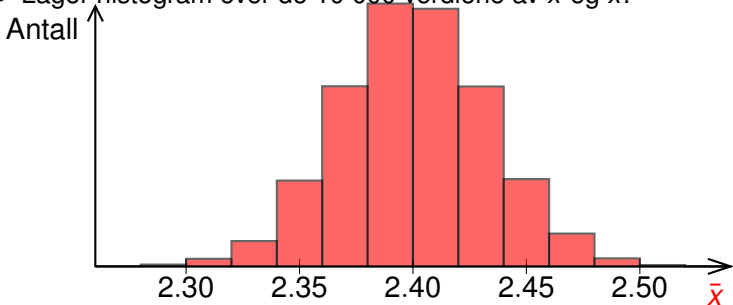
Illustrasjon ved simulering

— Situasjon:

- X_1, X_2, \dots, X_9 tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til μ er ukjent.
- Skal estimere (anslå) verdien til μ .

— Gjentar forsøket 10 000 ganger.

- Lager histogram over de 10 000 verdiene av \bar{x} og \tilde{x} .



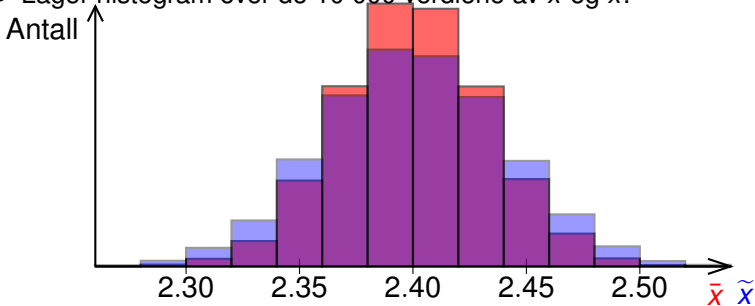
Illustrasjon ved simulering

— Situasjon:

- X_1, X_2, \dots, X_9 tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen.
- Verdien til μ er ukjent.
- Skal estimere (anslå) verdien til μ .

— Gjentar forsøket 10 000 ganger.

- Lager histogram over de 10 000 verdiene av \bar{x} og \tilde{x} .





Egenskaper til S^2

Eigenskapar til S^2



Teorem

Anta X_1, X_2, \dots, X_n u.i.f. frå $n(x; \mu, \sigma)$ -populasjonen. Då er $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ein forventningsrett estimator for σ^2 , det vil seie

$$E(S_n^2) = \sigma^2.$$



Kahoot

Neste veke



- Sannsynsmaksimeringsestimator
- Konfidensintervall