



# Bootstrapping – å løfte seg selv etter håret?

---

## MEDISIN OG TALL

EVA SKOVLUND

E-post: [eva.skovlund@ntnu.no](mailto:eva.skovlund@ntnu.no)

Eva Skovlund er professor i medisinsk statistikk ved Institutt for samfunnsmedisin og sykepleie, NTNU, og seniorforsker ved Folkehelseinstituttet.

Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

---

I en tid der alle går rundt med stor regnekraft så å si i lomma, er det egentlig ikke nødvendig å basere statistiske analyser på fordelingsantagelser. I stedet kan vi utnytte simuleringer til å estimere usikkerhet som grunnlag for både konfidensintervall og hypotesetester.

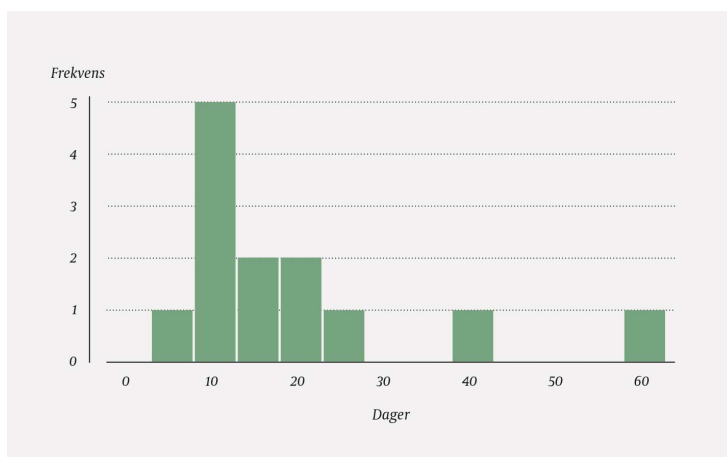
Tradisjonelle parametriske analyser baserer seg på å gjøre antagelser om fordelingen observasjonene i et datasett stammer fra, for eksempel at den er normalfordelt. Dette fungerer stort sett greit så lenge datasettet er rimelig stort. I situasjoner der man ikke uten videre kan bruke standardmetoder, er såkalt bootstrapping et aktuelt alternativ. Metoden baserer seg på å estimere den ukjente underliggende fordelingen ved hjelp av gjentatte tilfeldige utvalg fra datasettet og deretter trekke slutninger basert på denne.

## Idéen

Idéen ble publisert av den amerikanske statistikeren Bradley Efron i 1979 (1, 2). Han foreslo å benytte simuleringer til å konstruere et estimat av fordelingen data kommer fra, og deretter fortsette analysen som om denne fordelingen er den korrekte. Man trekker  $n$  observasjoner med tilbakelegging fra det reelle datasettet bestående av  $n$  observasjoner, og estimerer den parameteren man er interessert i. Dette gjentas  $B$  ganger, og fordelingen av disse  $B$  bootstrap-estimatene gir robuste estimater av usikkerhet, i form av for eksempel standardfeil og konfidensintervall for parameteren. Bootstrapping kan også brukes til å konstruere hypotesetester.

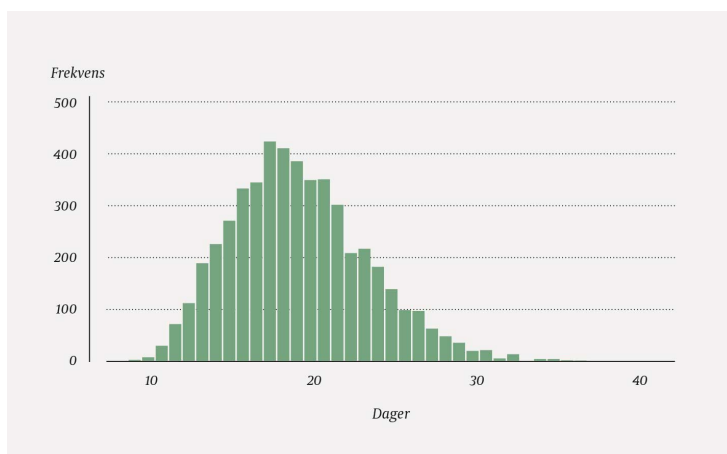
## Et enkelt eksempel

Anta at vi har registrert antall dager på sykehus for 13 pasienter med en gitt diagnose (hhv. 3, 9, 10, 10, 10, 12, 13, 14, 18, 21, 27, 38 og 62 dager). Hvis vi ønsker å estimere kostnad eller behov for personell, er gjennomsnittet en mer interessant parameter enn medianen. Fordelingen er vist i figur 1. Den er klart høyreskjev, og gjennomsnittet er 19 dager. Vi kan enkelt estimere et parametrisk 95 %-konfidensintervall for gjennomsnittet (9,46 til 28,54), men en tradisjonell fremgangsmåte antar normalfordelte observasjoner. Selv om prosedyren er robust mot små avvik, er det ikke opplagt at tilnærmingen er god når vi har så få observasjoner som her.



**Figur 1** Fordeling av antall liggedager for 13 tenkte pasienter.

Hvis vi i stedet benytter bootstrapping, trekker vi 13 observasjoner, med tilbakelegging, fra de 13. Et mulig ordnet utfall er 3, 9, 10, 10, 13, 13, 14, 14, 14, 21, 27, 27, 27. Legg merke til at samme observasjon kan gjentas flere ganger, mens andre ikke blir trukket ut. I akkurat dette utvalget blir gjennomsnittet 15,54. I alt trekker man  $B$  utvalg. I figur 2 vises fordelingen til  $B = 5\ 000$  gjennomsnitt, hvert av dem basert på  $n = 13$  uttrukne observasjoner. Gjennomsnittet av disse 5 000 gjennomsnittene er 19,05. Skjevheten ( $19,05 - 19,00 = 0,05$ ) er altså liten.



**Figur 2** Fordeling av 5 000 bootstrap-estimer av gjennomsnittlig liggetid for de samme 13 tenkte pasientene som i figur 1.

Standardavviket til de 5 000 gjennomsnittene er 4,24. Dette er et ikke-parametrisk bootstrap-estimat av standardfeilen. Mange statistikkpakker benytter dette til å estimere et 95 %-konfidensintervall for gjennomsnittet på vanlig måte ( $\bar{x} \pm z_{0,025} \cdot SE = 19 \pm 1,96 \cdot 4,24$ ). Resultatet blir 10,69 til 27,31 – et mer robust estimat enn det enkle, parametriske estimatet vi fikk over. Merk at en ny simulering ville gitt et litt annet resultat. En ulempe med bootstrapping er usikkerheten i grensene i konfidensintervallet. Denne unngås når man gjør en fordelingsantagelse.

## Prosentiler

Et alternativ til å estimere et 95 %-konfidensintervall basert på den simulerte standardfeilen slik som over er å benytte prosentilene i fordelingen til de  $B$  bootstrap-estimatene. Øvre og nedre 2,5-prosentil i eksempelet blir 11,92 og 28,31. Dette er et 95 %-ikke-parametrisk prosentilintervall for gjennomsnittet. Det blir forskjellig fra estimatet basert på simulert standardfeil fordi fordelingen i figur 2 er litt skjev.

## Egenskaper

De tilnærmede verdiene vi får ved bootstrapping, vil være avhengige av hvor godt fordelingen som genererer bootstrap-utvalgene passer med den sanne, underliggende

fordelingen. Dermed kan man spørre hva man egentlig vinner ved bootstrapping. En viktig fordel er at bootstrap-metoder er robuste mot avvik fra fordelingsantagelser, noe som er spesielt nyttig når vi har lite data og begrenset informasjon om modellen data er generert fra. Enda viktigere er det kanskje at man kan unngå kompliserte analytiske beregninger. Bootstrapping er fremfor alt et nyttig verktøy til å estimere usikkerhet i komplekse situasjoner.

Metoden har ikke noe norsk navn (3). Et uttrykk som *gjentatte tilfeldige utvalg* forklarer hva man faktisk gjør, men ordet *bootstrapping* har vært i bruk så lenge at det neppe er hensiktsmessig å endre den. Å løfte seg selv etter støvlehempene gjør man ikke på norsk, men man kan kanskje utnytte data så godt at man (nesten) løfter seg selv etter håret?

---

#### LITTERATUR:

1. Efron B. Bootstrap methods: Another look at the jackknife. *Ann Stat* 1979; 7: 1–26. [CrossRef]
2. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC, 1993.
3. Haugen TB. Bootstrapping-kjær metode uten norsk navn. *Tidsskr Nor Legeforen* 2012; 132: 1497. [PubMed][CrossRef]

---

Publisert: 18. november 2019. *Tidsskr Nor Legeforen*. DOI: 10.4045/tidsskr.19.0413  
© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no