

Elementær statistik

Lektion 3

Peter Tibert Stoltze
stat@peterstoltze.dk

22. februar 2010

Dagens program

- ▶ Kapitel 3: Centraltendens og spredning
 - ▶ Kort opsamling
 - ▶ Opgave 2
- ▶ Kapitel 4: Sandsynlighed og statistiske modeller
 - ▶ Gennemgang (fortsat)
 - ▶ Opgave 4 og 5
- ▶ Kapitel 5: Estimation
 - ▶ Gennemgang
 - ▶ Opgave 6 og 7

Del I

Kapitel 3: Central tendens og spredning

Oversigt

Opsamling

Opgave 2

Opsamling

- ▶ Vi regner på en simpelt tilfældigt udvalgt **stikprøve** fra en **population**
- ▶ **Centraltendens** kan beskrives ved modus, stikprøvegennemsnit og median
- ▶ **Spredning** kan beskrives ved variationsområde, stikprøvestandardafvigelse og interkvartilafstand
- ▶ Den samlede fordeling kan beskrives med den **empiriske fordelingsfunktion**, dvs. fraktilværdien afsat som funktion af observationsværdien.

Opgave 2

Fælles gennemgang af opgave 2 på tavlen...

Del II

Kapitel 4: Sandsynlighed og statistiske modeller

Oversigt

Fortsat gennemgang

Opgaver

Fortsat gennemgang

Se slides fra lektion 2...

Opgaver

- ▶ Opgave 5
 - ▶ Regnes fælles på tavlen nu
- ▶ Opgave 4
 - ▶ Regnes individuelt eller i grupper efter opgave 5
 - ▶ Fælles opsamling på tavlen

Del III

Kapitel 5: Estimation

Oversigt

Indledning

Stikprøvefejl

Konfidensintervaller

Opgaver

Indledning

- ▶ Stikprøvegennemsnittet er et **estimat** for middelværdien i en population
- ▶ Stikprøvegennemsnittet kan ikke forventes at være eksakt lig populationens middelværdi
- ▶ Men vi kan bestemme et konfidensinterval for estimatet og så (typisk) sige, at *middelværdien ligger med en sandsynlighed på 95% indenfor dette interval*
- ▶ Husk at elementerne i stikprøven skal være udtrukket simpelt tilfældigt

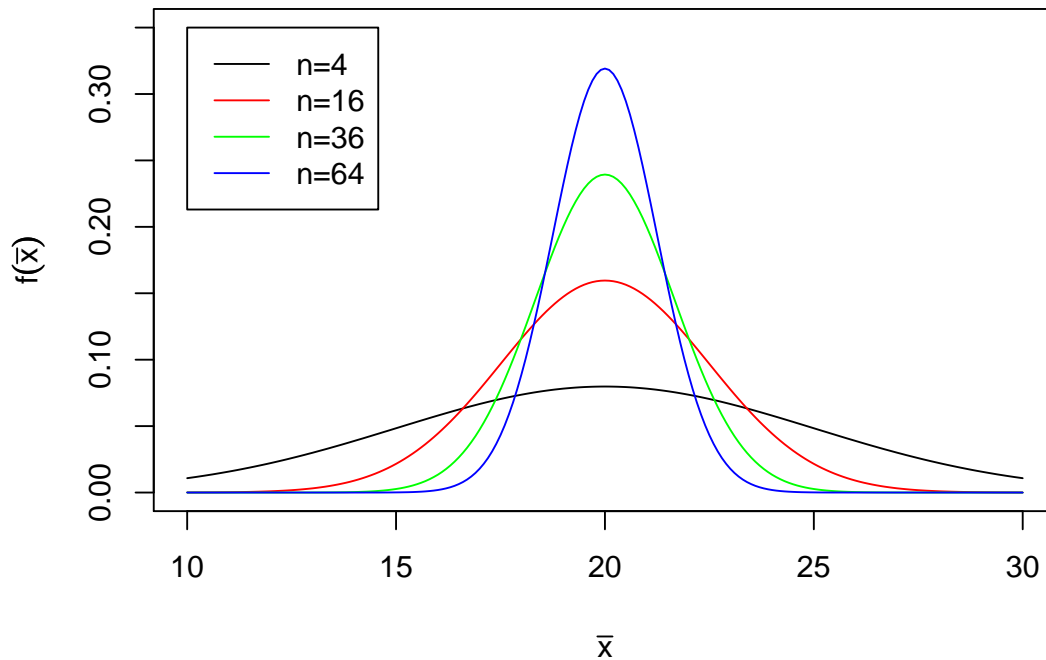
Fordeling af stikprøvegennemsnit

- ▶ Hvis vi udtager et antal forskellige stikprøver af samme størrelse n fra samme population, så vil vi opnå forskellige stikprøvegennemsnit
- ▶ Jo større n , jo mere vil stikprøvegennemsnittene alt andet lige ligne hinanden
- ▶ Spredningen på stikprøvegennemsnittet kaldes også standardfejlen og den beregnes som

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Fordeling af stikprøvegennemsnit

Tæthed idet $\bar{x} = 20$ og $s = 10$



Konfidensinterval

- ▶ Konfidensinterval for middelværdien baseres på stikprøvegennemsnittet og spredningen på dette (standardfejlen)
- ▶ Ved store stikprøver ($n > 30$) følger stikprøvegennemsnittet noget, der ligner en normalfordeling
- ▶ I eksemplet med vokalvarighed var $n = 40$ og gennemsnittet 208,9 og standardafvigelsen 9,79
- ▶ Vi vil nu bestemme et 95% konfidensinterval for middelværdien

Eksempel

- ▶ Først beregnes standardfejlen

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{9,79}{\sqrt{40}} = 1,548$$

- ▶ Dernæst findes i tabellen de z-værdier, der afgrænser de nederste og øverste 2,5% af fordelingen, nemlig -1,96 og +1,96
- ▶ Nu kan konfidensintervallet beregnes som

$$208,9 - 1,96 \cdot 1,548 = 205,9$$

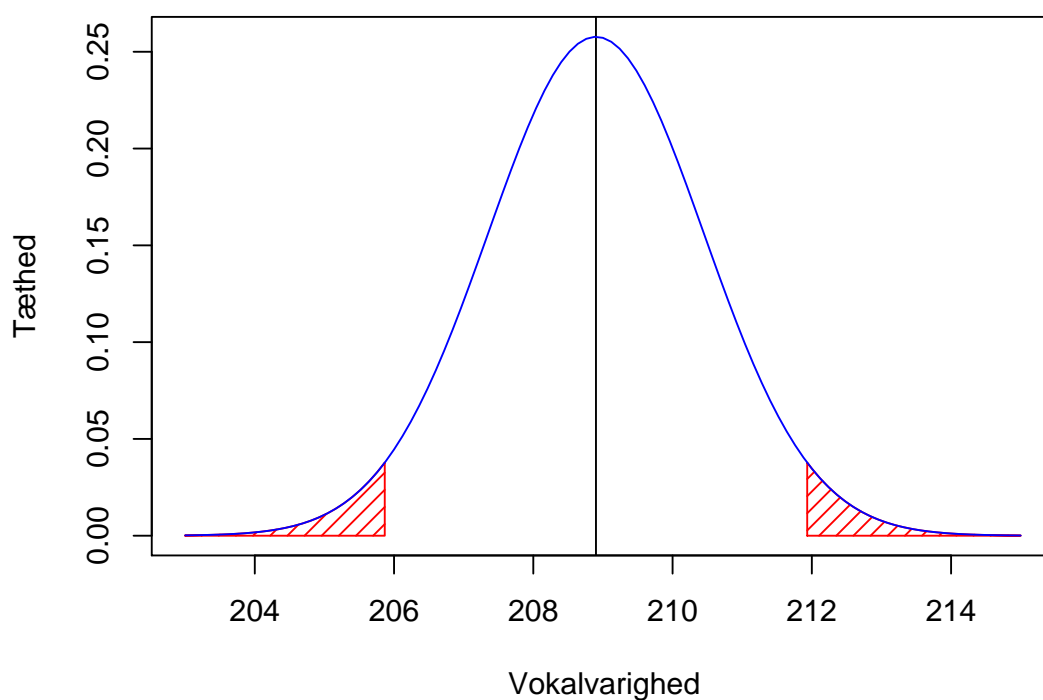
$$208,9 + 1,96 \cdot 1,548 = 211,9$$

- ▶ På baggrund af stikprøven kan vi altså sige, at med en sikkerhed på 95% ligger den sande middelværdi i intervallet fra 205,9 til 211,9
- ▶ Vi kan også beregne konfidensintervaller med andre konfidensgrader $1 - \alpha$ efter opskriften

$$\bar{x} \pm u_{1-\alpha/2} \cdot s_{\bar{x}}$$

hvor $u_{1-\alpha/2}$ er en fraktil fra normalfordelingen

Konfidensinterval grafisk



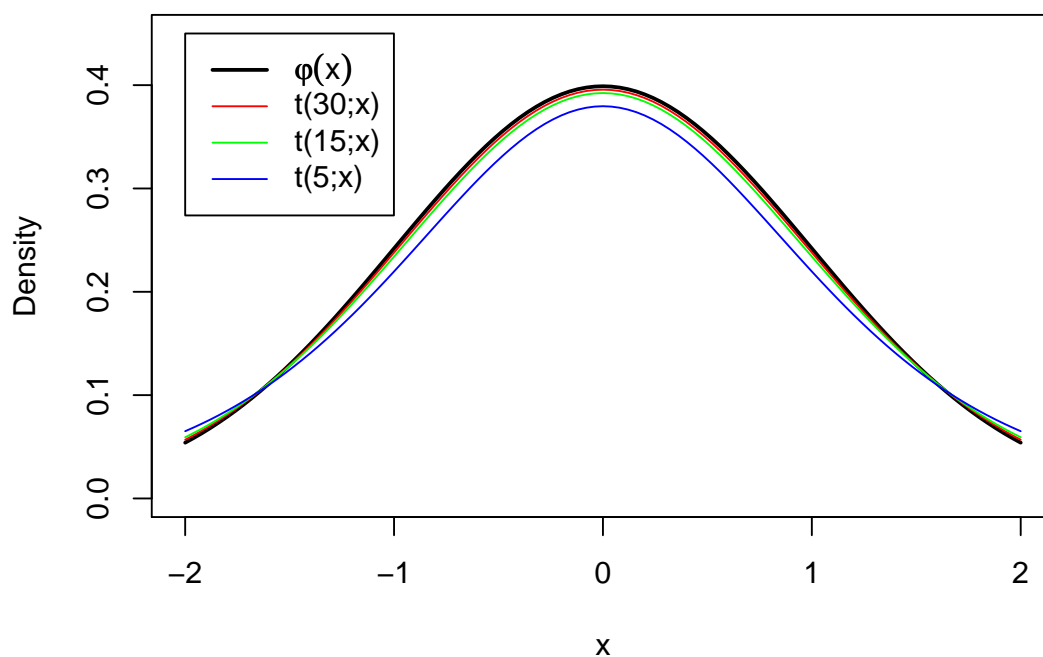
Konfidensintervaller med t -fordelingen

- ▶ Hvis $n < 30$ er konfidensintervaller baseret på normalfordelingen for optimistiske
- ▶ Vi benytter stadig samme opskrift, men nu bruger vi en fraktil fra en t -fordeling med $n - 1$ frihedsgrader i stedet for:

$$\bar{x} \pm t_{n-1;1-\alpha/2} \cdot s_{\bar{x}}$$

- ▶ Selvom $n > 30$ er dette den rigtige måde at regne på, såfremt spredningen s er estimeret

t -fordelingen



Opgaver

- ▶ Opgave 7
 - ▶ Regnes fælles på tavlen
- ▶ Opgave 6
 - ▶ Regnes individuelt eller i grupper
 - ▶ Fremlæggelse næste gang...