

Kursus 02403

Introduktion til Statistik

Birgitte Brøndum

Praktisk Information



- Undervisning: Hver dag de første 13 dage fra 8.30-12.30
 - ◊ 2 timer forelæsning (dagens pensum)
 - ◊ 2 timer øvelser
- Selvstudie
 - ◊ Hjemmeopgaver
 - ◊ Læse forud
- Eksamen: 4 timer multiple choice (mandag 25/6)

02403 Introduktion til Statistik



- Praktisk information om kurset
- Introduktion til statistik
 - ◊ Anvendelsesområder
- Beskrivende statistik:
 - ◊ Kapitel 1: Statistik for ingeniører
 - ◊ Kapitel 2: Beskrivende statistik

- Lærebog: Richard A. Johnson, Miller and Freund's Probability and Statistics for Engineers, 7th Edition Pearson Prentice Hall (2004), ISBN: 0-13-127840-1
- Kan købes i polyteknisk boghandel



Praktisk Information



- Hjemmeside: www.imm.dtu.dk/courses/02403
 - ◊ Pensumliste
 - ◊ Oversigt over forelæsninger og dagens opgaver
 - ◊ Opgaver & løsninger
 - ◊ Slides
- Campusnet: www.campusnet.dtu.dk
 - ◊ Meddelelser, konference og chat

Statistik og Ingeniører



- Statistik er et vigtigt værktøj i problemløsning
 - ◊ Analyse af data
 - ◊ Kvalitetforbedring
 - ◊ Forsøgsplanlægning
 - ◊ Forudsigelse af fremtidige værdier
 - ◊ .. og meget mere!

Introduktion til Statistik

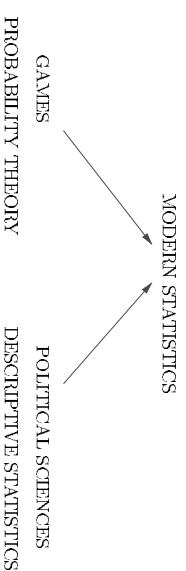


- Hvordan behandle (eller analysere) data?
- Hvad er tilfældig variation?
- Statistik er et værktøj til at træffe beslutninger:
 - ◊ Hvor mange computere har vi solgt det sidste år?
 - ◊ Hvad er forventet pris af en aktie?
 - ◊ Er maskine A mere effektiv end maskine B ?
- Statistik er et metodefag, der kan anvendes inden for de fleste fagområder, og er derfor et meget vigtigt værktøj

Statistik



- Moderne statistik har baggrund i sandsynlighedsregning og beskrivende statistik





- Statistik handler ofte om at analysere en *stikprøve*, der er taget fra en *population*
- Baseret på stikprøven, prøver vi at generalisere (eller udtale os) om populationen
- Det er derfor vigtigt, at stikprøven er *repræsentativ* for populationen

Kapitel 2: Nøgletal



Vi anvender en række *nøgletal* for at opsummere og beskrive data (stokastiske variable)

- Gennemsnit $\bar{x}$
- Median
- Varians s^2
- Standardafvigelse s
- Fraktiler

Bogens forslag til statistisk fremgangsmåde



- Definer formål for din undersøgelse
- Bestut hvilke data du vil indsamle og hvordan dette skal ske
- Anvend en passende statistisk metode til at trække information ud af data
- Fortolk dine resultater og drag en konklusion

Gennemsnit



Gennemsnittet er et *nøgletal*, der angiver tyngdepunkt eller centrering af data

- Gennemsnit:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Vi siger, at $\bar{x}$ er et *estimat* af middelværdien

Median



Medianen er et også nøgletal, der angiver tyngdepunkt eller centering af data. I nogle tilfælde, f.eks. hvis man har ekstreme værdier, er medianen at foretrække frem for middelværdien

- Median:

Den midterste observation (i den sorterede rækkefølge)

Variationskoefficient



Standardafvigelse og variansen er nøgletal for den absolutte variation. Hvis man gerne vil sammenligne variationen mellem forskellige datasæt, er det en god idé at anvende et relativt nøgletal, nemlig variationskoefficienten:

$$V = \frac{s}{\bar{x}} \cdot 100$$

Varians og standardafvigelse



Variansen (eller standardafvigelsen) siger noget om hvor meget data spredt:

- Varians

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- Standardafvigelse (spredning)

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Fraktiler



Medianen beregnes som det punkt, der deler data ind i to halvdele. Man kan naturligvis finde andre punkter, der deler data ind i andre dele, og det man kalder fraktiler.

Oftest beregner man fraktilerne

- ◊ 0, 25, 50, 75, 100 % fraktiler og/eller
- ◊ 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 % fraktiler

Bemærk: 50% fraktilen svarer til medianen

Grupperede data

Når man ikke har de enkelte observationer, men kun de grupperede data, kan man bruge følgende formler til at beregne gennemsnit og varians (af de grupperede data):

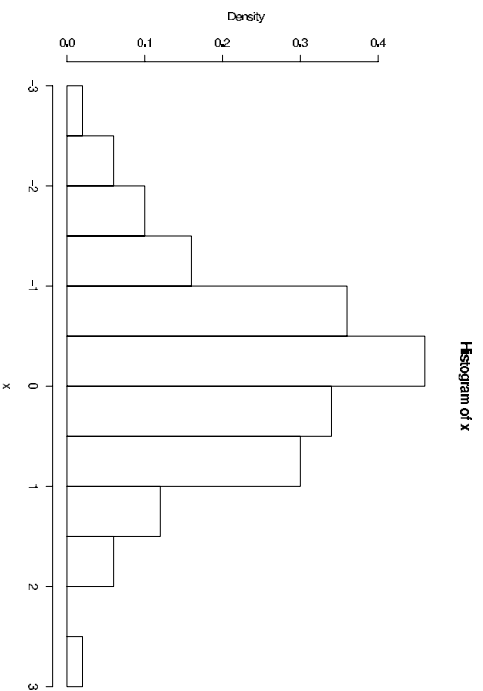
$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

$$s^2 = \frac{n \cdot \sum_{i=1}^n x_i^2 f_i - \left(\sum_{i=1}^n x_i f_i\right)^2}{n(n-1)}$$

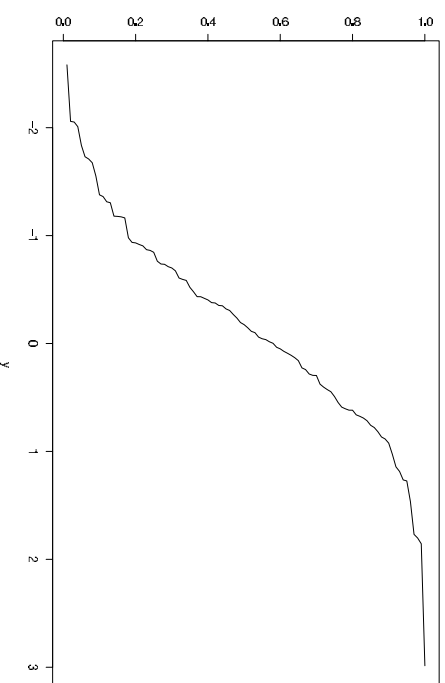
Figurer

- Scatter plot (xy plot)
- Histogram
- Kumulativ fordeling
- Boxplots
- Bar charts (pareto diagram)
- Pie charts

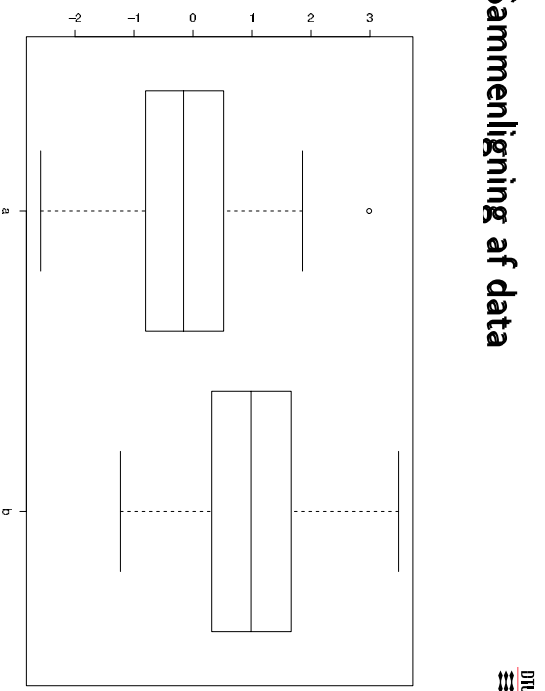
Histogram



Empirisk kumulativ fordeling



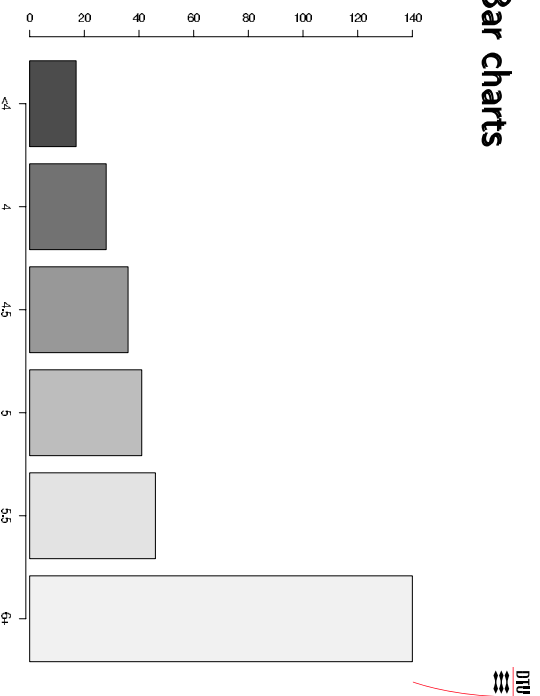
Sammenligning af data



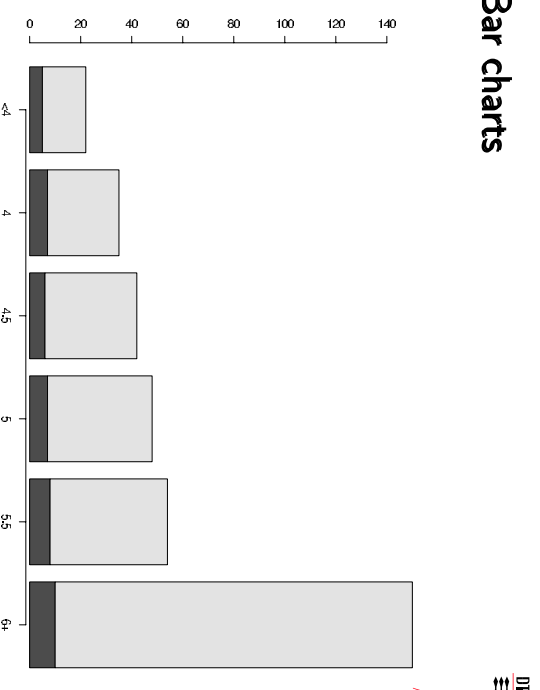
Tabeller

- Grupperede data kan beskrives ved tabeller
 - ◊ Grupperede data kan også beskrives ved Bar charts (pareto diagram) eller Pie charts

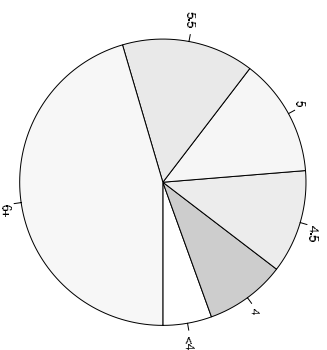
Bar charts



Bar charts



Pie charts



I morgen:

- Diskrete fordelinger - kapitel 4

