

Introduktion til  
PRAKTISK STATISTIK



# Introduktion til PRAKTISK STATISTIK

2. udgave

*Sanne Lund Clement & Jan Holm Ingemann*

Metodeserie for Social- og Sundhedsvidenskaberne, bind 4

SYDDANSK UNIVERSITETSFORLAG 2016

*Introduktion til praktisk statistik*

© Forfatterne og Syddansk Universitetsforlag 2011, 2015, 2016

Sats og tryk: Specialtrykkeriet Viborg

Omslag: Donald Jensen, UniSats ApS

2. udgave, 1. oplag

ISBN 978-87-7674-981-1

*University of Southern Denmark Studies in History and Social Sciences vol. 530*

Mekanisk, fotografisk, elektronisk eller anden mangfoldiggørelse af denne bog er kun tilladt med forlagets tilladelse eller ifølge overenskomst med Copydan

Syddansk Universitetsforlag

Campusvej 55

5230 Odense M

[www.universitypress.dk](http://www.universitypress.dk)

# Indhold

<b>REDAKTØRENS FORORD</b> .....	9
<b>FORFATTERNES FORORD</b> .....	11
<b>FORORD TIL 2. UDGAVE</b> .....	13
<b>DEL 1: STATISTIKKENS ROLLE I UNDERSØGELSE SARBEJDET</b>	15
<b>KAPITEL 1</b>	
<b>FRA PROBLEM TIL STATISTIK</b> .....	17
Statistik og faglig proces .....	17
Fra samfundsmæssigt til videnskabeligt problem. ....	19
Tilrettelæggelse af arbejdet .....	21
Indsamling af data .....	24
Behandling af data .....	25
Fortolkning .....	26
Konklusion .....	27
Opsamling .....	28
<b>KAPITEL 2</b>	
<b>DATA OG MÅLING</b> .....	29
Validitet og reliabilitet .....	29
Sekundær statistik .....	31
Hvor finder vi de sekundære data? .....	33
Faldgruber samt brug og misbrug .....	34
Måleproblemer .....	36
Stikprøveproblemer .....	38
Opsamling .....	39
<b>KAPITEL 3</b>	
<b>STATISTISKE REDSKABER OG PROBLEMSTILLINGER</b> .....	41
Hvad er statistik? .....	41
Analytiske elementer: måling, sammenligning og kontrol .....	43
Variable, enheder og værdier .....	46
Opsamling .....	51

<b>DEL 2: DESKRIPTIV STATISTIK</b> .....	53
<b>KAPITEL 4</b>	
<b>LÆS OG FORSTÅ STATISTIKKEN - ÉN VARIABEL</b> .....	55
Teorier, begreber og indikatorer .....	55
At læse og fortolke den univariate analyse .....	59
Indeksering og tidsrækker .....	63
Centraltrendensen - middelværdi, median og modus .....	67
Spredning - afvigelse, varians og standardafvigelse .....	71
Opsamling .....	74
<b>KAPITEL 5</b>	
<b>LÆS OG FORSTÅ STATISTIKKEN</b>	
<b>- TO ELLER FLERE VARIABLE</b> .....	75
Afhængige og uafhængige variable .....	75
Om kausalitet .....	75
Typer af sammenhænge .....	77
Krydstabeller .....	81
Hvad er en krydstabel? .....	81
Hvordan læser og fortolker man den bivariate analyse? .....	83
Hvornår kan vi ikke anvende krydstabeller? .....	84
Opsamling .....	87
<b>DEL 3: TEORETISK STATISTIK</b> .....	89
<b>KAPITEL 6</b>	
<b>SANDSYNLIGHED</b> .....	91
Usikkerhed og ufuldstændig viden .....	91
Sandsynlighed, risiko og beregning .....	92
Sandsynlighedsfordeling .....	93
Opsamling .....	96
<b>KAPITEL 7</b>	
<b>STIKPRØVEN</b> .....	97
Stikprøven som estimat .....	97
Sikkerhedsinterval for kontinuerte variable .....	99
Sikkerhedsinterval for relative frekvenser .....	102
Sikkerhedsniveau .....	102
Den lille stikprøve .....	104
Opsamling .....	105

**KAPITEL 8****TEORETISK STATISTIK - REGRESSION OG LINEÆR**

<b>SAMMENHÆNG MELLEML TO VARIABLE</b> . . . . .	107
Afhængig og uafhængig variabel . . . . .	107
Korrelation . . . . .	108
Estimering . . . . .	109
Hvor god er sammenhængen? . . . . .	112
Opsamling . . . . .	118

**KAPITEL 9**

<b>HYPOTSETEST</b> . . . . .	119
Sammenligning og sammenhæng . . . . .	119
Sikkerhedsintervallet . . . . .	120
Signifikanstest for afhængige andele . . . . .	122
Signifikanstest ved uafhængige andele . . . . .	123
Sammenhænge mellem variable - Chi <sup>2</sup> -testen . . . . .	127
Et eksempel - sammenhæng mellem køn og forventet stemme ved folkeafstemning om EU . . . . .	129
Opsamling . . . . .	132

**DEL 4: FORTOLKNING OG KONKLUSION** . . . . . 133**KAPITEL 10**

<b>FORTOLKNING</b> . . . . .	135
Den samlede proces . . . . .	135
Analysens inputside . . . . .	136
Tolkningsproblemer . . . . .	137
Analysens outputside: Konklusion . . . . .	142
Formidling . . . . .	143
Opsummering . . . . .	146

**KAPITEL 11****OM USIKKERHED OG SANDSYNLIGHED**

<b>- HVAD VED VI, OG HVAD VED VI IKKE?</b> . . . . .	149
Om at søge, hvor der er lys . . . . .	149
Sikker viden og usikkerhed . . . . .	151
Viden, vægtning og objektiv sandsynlighed . . . . .	153
Opsamling og afslutning . . . . .	155

<b>INDEKS</b> .....	157
<b>LITTERATUR</b> .....	159
<b>APPENDIKS</b> .....	161



# Redaktørens forord

Det er som regel en af følgende to reaktioner, man mødes med, når man over for studerende nævner betegnelsen 'statistik'. Enten ser man panikangstens utvetydige udtryk malet i ansigtet på dem, eller også oplever man et øjeblikkeligt anfald af træthed kombineret med en totalt manglende interesse samt ikke mindst et ønske om hurtigst muligt at tale om noget andet. Det er da også sandt, at statistik ikke er et fagområde, som almindeligvis forbindes med morskab eller tøjsløs fantasi. For mange studerende (og måske særligt i samfundsvidenskabelige sammenhænge) anses det som et nødvendigt onde, som man nødvendigvis skal igennem, fordi det er en integreret del af mange uddannelsesprofiler og fagkompetencer. Det er således forholdsvis sjældent - medmindre man skal læse til økonom eller aktuar - at statistik er noget, der omfavnes eller opsøges frivilligt af de studerende.

Mine egne erfaringer fra min studentertid er da også en kombination af de to ovenstående reaktioner - panikangsten over den forestående statistikeksamamen (hvor en af denne bogs forfattere faktisk var min eksaminator) og en udbredt træthed over at tærskte et tungt pensum igennem i teoretisk statistik i en bog på hen ved 1000 sider. Statistik var også dengang noget, man kom - eller kæmpede sig - igennem, og ikke noget, man i almindelighed så frem til med stor faglig appetit. Dog var det sådan, at man til sin dødsdag huskede det, man havde lært, fordi man havde terpet så grundigt af frygt for ikke at bestå faget og skulle til reeksamen.

Trods disse ofte tvivlsomme eller ligefrem negative følelser, man som studerende knytter til statistikfaget, er statistik imidlertid et område, der er uendeligt vigtigt. En stor del af vores viden om verden stammer faktisk fra statistik. Vi lever således i dag i et samfund med uendelige mængder af statistisk materiale, som man kan analysere på - lige fra statistikker over forandringer i fertilitet via vælgeradfærd, uddannelsesniveauer og forbrugsmønstre over til sygdoms- og dødsårsager. Store dele af nutidens menneske- og samfundsliv er således opgjort statistisk. Statistik er da ej heller hverken hokusfokus eller nødvendigvis nørdet talknuseri. Det er en seriøs interesse for at indfange, forstå og fortolke verden gennem tilgængeligt talmateriale, som gemmer sig i forskellige typer af registre, arkiver og databaser. Selvom statistik - og arbejdet hermed - ofte på misforstået vis forbindes med sociologiens store prygelknabe i disse år, positivismen, så er det dog en noget fejlagtig sammenkædning, ligesom det, at man sætter lighedstegn mellem kvantitativ samfundsforskning og statistik, også er det. Man kan snildt arbejde med og analysere på statistisk materiale uden at være positivist (faktisk

kan man både betragte sig som hermeneutiker, fænomenolog eller socialkonstruktivist og ikke desto mindre drage nytte af statistisk materiale i sine undersøgelser), ligesom man ikke behøver at identificere sig som kvantitativ samfundsforsker for at snuse til eller uddrage nyttig viden fra statistikker over sociale fænomener (kvalitative samfundsforskere kan således med fordel inddrage statistikker for at understøtte, nuancere eller modbevise deres egne kvalitativt funderede teser). Statistisk materiale – uanset karakteren eller omfanget heraf – kan således berige en hvilken som helst forskningsrapport, analyse eller argumentationskæde. Statistik er dermed noget, som – selvom det sjældent kan stå alene som forklaringsgrundlag – er værdifuldt at inkorporere i sine samfundsanalyser og -diagnoser.

Denne bog tilbyder en indføring i praktisk statistik for social- og sundhedsvidenskaberne. Det er således ikke teoretisk statistik, der udgør omdrejningspunktet for bogen (om end det berøres i et kapitel), men derimod den måde, hvorpå en viden om at udarbejde og analysere på statistik kan berige vores forståelser af forskellige dele af samfundslivet. Jan Holm Inge-mann og Sanne Lund Clement har skrevet en særdeles læseværdig, interessant og anvendelig lille bog, der med stor fordel kan finde vej ind i undervisning og projektarbejde inden for en bred vifte af samfundsvidenskabelige discipliner og fagområder. Jeg ønsker god læselyst med bogen med forhåbningen og forventningen om, at den kan bidrage til at gøre undervisningen i og indlæringen af statistik knap så angstfremkaldende og kedsommelig, som mine egne ovenstående erindringer og erfaringer ellers tilsiger, at faget ofte er behæftet med.

*Michael Hviid Jacobsen*

Aalborg Universitet, sommeren 2011

# Forfatternes forord

Brugen af statistisk materiale har i både samfunds- som sundhedsvidenskaben vundet stadig større indpas som pålidelig kilde til dokumentation af påstande, holdninger, betydninger og effekter. Praktikere på områderne pålægges i stigende omfang at frembringe sammenlignelig og evidensbaseret information, ofte baseret på statistisk materiale. Studerende på uddannelserne forventes at kunne forholde sig konstruktivt og kritisk til andres brug af statistisk materiale, ligesom de selv skal kunne udarbejde mere eller mindre avancerede statistiske analyser. Vores erfaringer som undervisere i praktisk statistik er imidlertid, at mange studerende har en i udgangspunktet lettere nervøs tilgang til faget. Det indebærer, at specielt de studerende, der ikke føler sig stærke rent matematisk, så at sige giver op på forhånd, og de opnår derfor aldrig den viden og den erkendelse, at statistik i praksis er et særdeles nyttigt redskab i rigtig mange tilfælde.

Denne bog er derfor et forsøg på at udarbejde en anvendelsesorienteret lærebog i statistik, der i særlig grad fokuserer på at koble anvendelsen af statistik til de problemorienterede og projektor organiserede studiemetoder og at gøre materialet praksisorienteret, således at det alene sigter mod at sætte læserne i stand til at anvende det til at kvalificere arbejdet med problemformulering og simpel analyse.

Hermed har vi også sagt, at praktisk statistik som sådan ikke kun handler om tal. Naturligvis er en stor del af statistikken udtrykt i tal, og derfor spiller håndteringen af tal også en væsentlig rolle i denne bog. Men bogen lægger samtidig primært vægt på netop at være anvendelsesorienteret, og hensigten er således også at vise, hvordan man relativt enkelt dels kan blive bedre til at forstå andres brug af statistisk materiale, dels kan blive i stand til selv at udarbejde simpel statistisk dokumentation og analyse.

Det er således heller ikke tilfældigt, at det allerede i bogens titel pointeres, at det er 'praktisk statistik', vi har med at gøre. For i denne sammenhæng, og i brugen af statistik i det problemorienterede projektarbejde, er statistik meget mere end matematiske formler - det er et praktisk redskab, der skal hjælpe os med at afprøve påstande, belyse udviklingstendenser, illustrere pointer, sandsynliggøre sammenhænge og meget andet.

Men hvor lærebøger i statistik ofte lægger vægt på at lære en række matematiske formler og beregninger, lægger denne lille introduktionsbog om praktisk statistik mere vægt på at være anvendelsesorienteret. Hensigten er således at vise, dels hvor man kan finde materiale i form af sekundær statistik, som kan anvendes i projektarbejdet såvel i relation til problemformule-

ring som analyser af mere deskriptiv karakter, dels at give en introduktion til den statistiske bearbejdning og formidling af dette datamateriale.

Således er det også indirekte understreget, at vi ikke gør meget ud af at redegøre for de matematiske beviser eller logikker bag de statistiske metoder og redskaber. I stedet sigter vi mod at give en mere intuitiv forståelse af, hvorledes man kritisk kan anvende dem. Selvom teksten er skrevet i et forholdsvist almindeligt sprog med meget få matematiske symboler, så ændrer det ikke ved, at statistik er et redskabsfag med særlige regler og logikker, og teksten i denne bog kan følgelig ikke blot læses, som var den en avisartikel. Forudsætningen for et højt fagligt udbytte er, at man tager sig tid til grundigt at arbejde med teksten og ikke mindst arbejde med eksempler. Derfor opfordres læserne til at have papir og blyant ved hånden, så de kan arbejde med små eksempler undervejs og til tilsvarende at arbejde med eksempler under anvendelse af et regnearksprogram. I den forbindelse kan læseren drage nytte af at indtaste bogens eksempler og så arbejde med dem.

Selvom vi har begrænset vores anvendelse af matematiske formler og symboler, så er de dog ikke helt udeladt, idet de for nogle studerende giver en bedre forståelse af det statistiske arbejde, og fordi matematisk sprog og logik reelt er en integreret del af det statistiske fag. Vi skal understrege, at den matematiske notation kan variere fra lærebog til lærebog, og at vi i denne bog baserer os på Wonnacot & Wonnacot (1972) samt Kvanli m.fl. (1997).

Endelig bør det nævnes, at forfatterne begge er samfundsvidenskabsfolk, og at eksemplerne derfor langt hen ad vejen vil være samfundsvidenskabelige, men at vi naturligvis forsøger at inddrage sundhedsvidenskaben, i det omfang vores indsigt rækker til det.

Afslutningsvis ønsker vi at takke lektor Henrik Lolle, der beredvilligt har stillet upubliceret materiale til rådighed. Også tak til vores nationale og internationale kollegaer i vores netværk inden for metodeforskning for inspiration og kritik. Sidst, men ikke mindst, også en stor tak til de mange studerende, der i årenes løb kritisk og konstruktivt har givet os inspiration til såvel forsknings- som formidlingsmæssigt at arbejde med metode.

*Sanne Lund Clement & Jan Holm Ingemann*  
Aalborg Universitet, maj 2011

# Forord til 2. udgave

Bogen er blevet anvendt som lærebog for tusindvis af studerende siden udgivelsen af den første udgave. Det har givet anledning til mange erfaringer, der har dannet afsæt for denne nye udgave, hvor bogen er blevet væsentligt omarbejdet, og der er tilføjet fire helt nye kapitler. Hermed kommer vi nu omkring et større pensum, hvor udvidelsen især vedrører den teoretiske statistik. Samtidig er bogen disponeret om sådan, at den nu tæt følger de kronologiske faser fra et umiddelbart problem over fremskaffelse af data og analyse frem til konklusion.

Manuskriptet har været igennem nyt review i den forbindelse, og vi takker for gode og konstruktive kommentarer og forslag. Også tak til Majbritt Christine Severin for en fin indsats med indeks, samling af formler og anden hjælp til håndtering af manuskriptet.

Vi håber, at brugerne vil værdsætte den nye udgave.

*Sanne Lund Clement & Jan Holm Ingemann*  
Aalborg Universitet august 2016