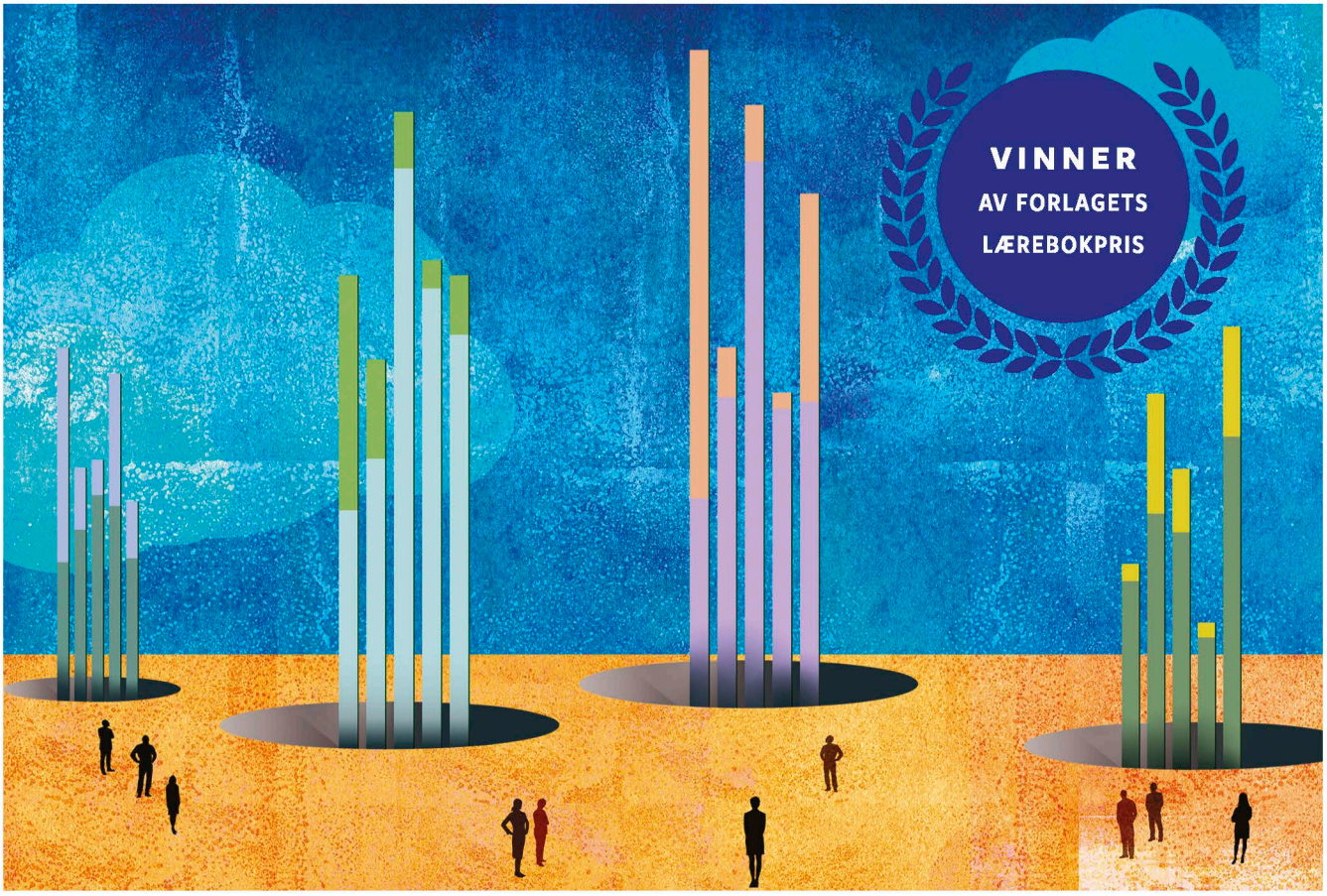

FOLDNES | GRØNNEBERG | HERMANSEN | WELLÉN

STATISTIKK OG DATAANALYSE

En moderne innføring

2. UTGAVE

CAPPELEN DAMM AKADEMISK

Statistikk og dataanalyse

Njål Foldnes,
Steffen Grønneberg,
Gudmund Horn Hermansen og
Einar Christopher Wellén

Statistikk og dataanalyse

En moderne innføring

2. utgave

CAPPELEN DAMM AKADEMISK

Forord

Statistikk som et verktøy til å forstå verden

Velkommen til statistikkfaget! Her skal du lære om hvordan du kan analysere data og bruke det til å ta bedre avgjørelser.

Vi får stadig tilgang til mer og mer data, og etterspørselen etter folk som kan tolke kvantitativ informasjon, er økende.

Denne boken er en introduksjon til hvordan man kan bruke tall og data til å finne ut hvordan verden er skrudd sammen.

Verden blir mer og mer kvantitativ – det samles inn mer og mer digitalisert informasjon. Ta for eksempel Wal-Mart, en amerikansk kjøpesenterkjede med en årlig omsetning på størrelse med Norges brutto nasjonalprodukt. Wal-Mart lagrer informasjon om mer enn én million handler hver eneste time! Alle slags organisasjoner lagrer data om sin virksomhet, og de beste bruker dette for alt det er verdt for å maksimere verdiskapingen sin. Her hjemme samler Statistisk sentralbyrå inn store mengder kvantitative data fra alle samfunnsområder. Denne informasjonen blir brukt til å analysere sammenhenger og trender. Dette er viktig informasjon til politikere og andre beslutningstakere.

Kvantitativ Informasjon: Informasjon som omhandler tall og målinger. I motsetning til kvalitativ informasjon. Alt som kan måles og telles er kvantitativ informasjon. Kvantitative metoder er altså matematiske og statistiske teknikker som hjelper oss å forstå tallmateriale.

Med moderne datateknologi kan du selv gjøre kvantitativt arbeid og oppdage nye sammenhenger. Noen få (men smarte) tastevalg i regneark som Excel eller statistisk programvare som JMP, R eller SPSS kan gi ny informasjon som kan hjelpe din bedrift. Etterspørselen etter folk som kan tolke og «knuse» tall kommer garantert til å vokse også i årene framover.

Vi som underviser i statistikk vet godt at mange studenter har et anstrengt forhold til tall og matematikk. Mange studenter i samfunnsfagene føler seg distansert fra tall og regning, og mange har dessverre negative erfaringer fra skolen. Om dette skulle gjelde deg, også – prøv likevel å være åpen! I denne boken kreves det ingen ekstraordinære matematikkferdigheter. Likevel viker vi på ingen måte unna dette viktige faget, for målet vil alltid være å forstå dataene, og da trenger vi noe matematikk, rett og slett.

I denne boken legger vi vekt på tenkning og forståelse. Du vil lære å tolke resultatene av statistisk analyse i den sammenhengen du vil bruke dem i.

Formler og formelle statistiske prosedyrer vil være en del av kurset, men til syvende og sist er det din forståelse av analysen som skal være grunnlaget for avgjørelser du tar.

Enkelte overskrifter i boken er merket med (*). Dette betyr at teksten som følger er fordypningsstoff, og ikke behøver å leses grundig ved første gjennomgang av boken.

Slik er boken organisert

Boken har fire hovedbolker. Det kan være greit å få en oversikt over disse bolkene før man begynner å lese. De første tre bolkene fokuserer på statistisk analyse av én variabel mens den siste og fjerde bolken handler om analyse av to eller flere variabler.

Vi starter i **bolk 1** med å introdusere hovedideene i en statistisk analyse. Hovedmålet med boken er å få deg til å utføre slike analyser på egenhånd. Vi presenterer en skjematisk oversikt over det vi kaller de fire elementene i en statistisk analyse. Dette gir et godt utgangspunkt for å lese boken. Du vil også bli gjort kjent med de mest grunnleggende begrepene i statistikk og herunder bli kjent med de ulike variabeltypene.

Vi tar deg deretter gjennom de tre første stegene i en statistisk analyse. Det første steget er å definere problemstillingen du ønsker å løse. Deretter lærer vi deg å trekke et representativt utvalg (andre steget i en statistisk analyse). Til slutt tar vi deg gjennom det tredje steget i en statistisk analyse, som handler om å beskrive data fra utvalget med nøkkeltall (f.eks. gjennomsnitt) og grafer (f.eks. histogram).

Det siste steget er inferens, som betyr hvordan vi kan gå fra informasjon om utvalget til å si noe om hele populasjonen utvalget er trukket fra. Dette steget kommer først i bolk 3, og er det mest sofistikerte steget. For å kunne gjøre og forstå inferens må vi først studere sannsynlighet. Dette gjøres i bolk 2 av boken.

Bolk 2 tar for seg sannsynlighetsregning og består av regneregler og bruk av matematiske teknikker. Her er det altså mange formler. Vi lærer regler om sannsynligheten av hendelser, og vi introduserer begrepet tilfeldig variabel, og dens sannsynlighetsfordeling. Kunnskap om sannsynlighetsregning er nødvendig for å kunne gå fra steg 3 (beskrivende statistikk) til steg 4 (inferens) i en statistisk analyse, men vi lærer også sannsynlighetsregning fordi vi ofte trenger å bruke sannsynlighetsregning direkte på problemstillinger innen økonomi, finans, markedsføring og andre fag.

Bolk 3 handler om inferens: hvordan kan vi gå fra informasjon om utvalget til å si noe om hele populasjonen som utvalget er trukket fra. Inferens er steg 4 og det siste steget i den statistiske analysen. Når vi utfører inferens svarer vi på problemstillingen fra steg 1 i den statistiske analysen. For å kunne utføre inferens kreves kjennskap til sannsynlighetsregning (broen mellom steg 3 og steg 4 i en statistisk analyse) og bruk av nøkkeltall fra den beskrivende analysen (steg 3 i en statistisk analyse). Inferens gjøres ved å beregne konfidensintervaller og utføre hypotesetester.

I **bolk 4** lærer vi å analysere sammenhengen mellom to variable. Vi studerer samvariasjon mellom to variable via simultane sannsynlighetsfordelinger, regresjon og khikvadrattester. Kan du regresjon, har du et godt grunnlag for å studere mer avanserte statistiske modeller i videregående kurs.

Kapitteloversikt

Innhold

KAPITTEL 22

Enkel regresjon 536

KAPITTEL 23

Samvariasjon for to kategoriske variable 570