

NyS

Forfatter:	Laura Winther Balling & Rikke Slot Johnsen	
Anmeldt værk:	Carsten Elbro & Mads Poulsen: <i>Hold i virkeligheden. Statistik og evidens i uddannelse</i>	
Kilde:	<i>NyS – Nydanske Sprogstudier 48. Sprogtilegnelse</i> , 2015, s. 174-181	
Udgivet af:	NyS i samarbejde med Dansk Sprognævn	
URL:	www.nys.dk	

© NyS og artiklens forfattere, 2015

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre NyS-numre (NyS 1-36) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Anmeldelse

Carsten Elbro & Mads Poulsen: *Hold i virkeligheden. Statistik og evidens i uddannelse*. Hans Reitzel, 2015.

LAURA WINTHER BALLING & RIKKE SLOT JOHNSEN

INDLEDNING

Vi har haft den fornøjelse at læse Elbro og Poulsens bog, i vores egen- skab af hhv. underviser i statistiske metoder for humanister og konsu- lent i en kommunal børne- og ungeforvaltning. Vi prøver på den bag- grund at vurdere bogen ud fra både ekspert- og brugerperspektiv. Men først en præsentation af bogen og dens sigte, efterfulgt af et resumé kapitel for kapitel.

BOGENS SIGTE OG OPBYGNING

Som titlen siger, handler bogen om statistik og evidens i uddannelse; bagsideteksten beskriver den nærmere som en hjælp til at forstå ud- dannelsesforskning og, i mindre grad, en håndbog for dem der selv vil forsøge sig med sådan forskning. Den er rettet mod ”læreren, skole- ledelsen og konsulenterne [der] har brug for at have hold i virkelighe- den” (s. 10). I forordet beskriver forfatterne dog bogens udspring som undervisning i statistik for audiologopæder, som forekommer at være en noget smallere, mere specialiseret gruppe.

Bogen består af et forord, syv kapitler og en oversigt over statistiske tests. Hvert kapitel starter med en kort liste over kapitlets sigte, hvad man kunne kalde kapitlets læringsmål, og slutter med en opsummering og mulige svar på de aktiviteter der optræder som hjælp til refleksion og forståelse i løbet af kapitlet. Bogen bruger også jævnlige lyserøde tekstbokse, men det præcise formål med dem er lidt uklart: Nogle gange beskriver de scenarier fra virkeligheden, men sådanne scenarier beskrives også i selve brødteksten; andre gange indeholder de korte

definitioner af de termer der er blevet introduceret i den umiddelbart forudgående del af brødteksten.

RESUMÉ

Kapitel 1 definerer titlens og med den bogens centrale begreber evidens og statistik, sat i forhold til det valgte praksisfelt, altså undervisning: Evidens bygger til enhver tid på iagttagelser, der kan komme fra lærerens daglige praksis, fra mere eller mindre strukturerede interviews eller fra tests. I forbindelse med tests understreger forfatterne at brugen af tests som evidens ikke nødvendiggør et menneskesyn hvor elever reduceres til testresultater, en besværgelse der nok er nødvendig i forhold til den herskende uddannelsespolitik. Statistik beskrives som et redskab til at skabe overblik over en større samling iagttagelser, det man i statistikerjargon kalder beskrivende eller deskriptiv statistik, og til at skelne mellem systematiske og tilfældige tendenser, altså inferentiell statistik. Disse tekniske begreber bruges dog ikke.

Derudover introduceres bogens dobbelte formål i kapitel 1: at gøre læseren i stand til at forstå evidensbaseret uddannelsesforskning og i nogen grad selv at designe empiriske undersøgelser af uddannelsespraksis og analysere resultaterne. Forfatterne understreger, ganske rigtigt, at bogen i forhold til det sidstnævnte formål ikke kan stå alene, men kræver øvelse og vejledning, ligesom teoriundervisningen til køreprøven ikke sætter en i stand til at køre bil. Denne sammenligning bruges flere gange i bogen og virker fint til at vise muligheder og begrænsninger. Til sidst i kapitlet præsenteres bogens heltinde: Rikkens lærer som underviser Rikke og andre voksenkursister i dansk. Eksemplerne i de efterfølgende kapitler er alle baseret på situationer hvor Rikkens lærer undersøger sit holds færdigheder.

Kapitel 2 handler om de forskellige mål og begreber der kan bruges til at beskrive et individs, en gruppes eller et segments præstation, dvs. deskriptiv statistik der bruges til at beskrive det man har undersøgt, i bogens eksempler elever. Grundpræmissen i kapitlet er at enhver vurdering og rådgivning er baseret på implicit eller eksplicit sammen-

ligning; målet med kapitlet er at præsentere redskaber til at foretage sammenligning på systematisk vis. Med udgangspunkt i Rikkens hold af voksenkursister i dansk og deres resultater i semantisk ordforrådsprøve og staveprøve viser forfatterne hvordan gennemsnit og standardafvigelse (dvs. deltagernes gennemsnitlige afvigelse fra gennemsnittet) kan bruges til at omsætte de faktiske scorer i testene til z-scorer på standardskalaen, hvor elevernes gennemsnit i testen har værdien 0, og en standardafvigelse svarer til 1 på skalaen. Derudover introduceres percentiler, primært medianen som er den værdi som 50 % af holdet scorer over, og 50 % scorer under. Endelig diskuteres gulv- og loftseffekter, altså det fænomen at der i nogle tests, her eksemplificeret med staveprøven, er uforholdsmæssigt mange der får den højeste score, en loftseffekt, mens der i andre tests er uforholdsmæssigt mange der får den laveste score, en gulveffekt. Lofts- og gulveffekter har forskellige implikationer for en fortolkningen af testscorerne, bl.a. betyder loftseffekter at testresultater ikke kan bruges til at skelne mellem de dygtigste og de allerdygtigste, og at fremskridt kan blive undervurderet.

I kapitel 3 flyttes fokus fra deskriptiv til inferentiell statistik, altså den form for statistik der bruges til at skelne mellem tilfældige og systematiske effekter. Rikkens lærer vil gerne vide om den fremgang der er på Rikkens hold fra en test taget før til en test taget efter kursusforløbet, er tilfældig eller pålidelig, og tilsvarende for forskellen mellem Rikkens hold og det hold læreren havde sidste år. Til det formål bruger hun t-tests, hvis vigtigste output er en p-værdi: Sandsynligheden for at den forskel der er observeret, er opstået tilfældigt. Hvis den sandsynlighed er lav, typisk defineret som under 5 % eller 0,05, antages det at forskellen er systematisk; hvis ikke, at den er tilfældig. Forfatterne understreger at t-testen ikke kan afsløre om resultaterne skyldes en systematisk fejl; her må man selv vurdere testen og designet i øvrigt. Det er en vigtig og heldigvis gennemgående pointe i bogen, også i kapitel 4 om variansanalyse og i kapitel 5 om sammenhænge.

Den grundlæggende logik i statistiske tests beskrives meget fint og forståeligt, dels i forklaringen af p-værdien, dels i forklaringen af hvad der skal til for at man kan være sikker på at en forskel er systematisk snarere end tilfældig. Her gælder det at jo større en forskel er, jo mindre

en forskel varierer fra elev til elev, og jo flere elever der har deltaget i undersøgelsen, jo mere sandsynligt er det at der er tale om en systematisk forskel. De data t-testen anvendes på, opfylder ikke helt de krav der almindeligvis stilles til dens brug (kontinuerte data eller over 30 observationer), men den pædagogiske pointe med at tage udgangspunkt i den meget anvendte og relativt letforståelige t-test kan nok siges at opveje dette problem. Kapitlet beskriver hvilke betingelser der skal være opfyldt for at en t-test giver mening, men dækker kun delvis spørgsmålet om type af data. Endelig forklares konfidensintervaller som beskriver det interval som man kan være 95 % sikker på at det faktiske gennemsnit falder indenfor.

Mens kapitel 3 altså handler om situationer hvor man vil undersøge virkningen af en enkelt uafhængig variabel, hold, på en afhængig variabel, præstation i stave- og andre prøver, handler kapitel 4 om de mere komplicerede situationer hvor man vil undersøge virkningen af flere uafhængige variable, f.eks. både hold og tidspunkt (før/efter kursus) ved hjælp af variansanalyse. Fokus i kapitlet er på sammenligning af fremgange mellem forskellige grupper, hvor Rikkens og det tidligere hold sammenlignes i forhold til hvor meget de er gået frem i læseforståelse fra en test taget før til en test taget efter deres kursusforløb. Dette undersøges først med en variansanalyse med en faktor hvor den afhængige variabel er fremgangen i læseforståelse, og den uafhængige er faktoren hold. Dernæst forklares den mere informative og bredt anvendelige variansanalyse med to faktorer, hvor den afhængige variabel er scoren i læseforståelsestesten, og de uafhængige variable er test før/efter kurset og hold. Her kan analysen vise en hovedeffekt, hvis en variabel har en virkning uanset værdien af den anden variabel, f.eks. hvis test før/efter kurset har den samme effekt for begge hold. Hvis der derimod er forskel på holdenes udvikling i løbet af kurset, vil variansanalysen vise en interaktion hvor virkningen af den ene variabel afhænger af den anden variabel, dvs. hvor effekten af før/efter afhænger af holdet. Sådant en interaktion kan give grundlag for at konkludere at undervisningen på det ene hold virker bedre end på det andet, forudsat naturligvis at der ikke er indbyggede forskelle og systematiske fejl i designet, og at der tages passende forbehold i forhold til undersøgelsens

størrelse. Udover det centrale eksempel med udvikling over tid indeholder kapitlet flere gode eksempler hvor variansanalysen kunne være relevant, både i selve teksten og i aktiviteterne.

Et forbehold overfor kapitel 4 fra et statistisk synspunkt er at variansanalyserne kræver at de uafhængige variable består af kategorier; det giver mening i tilfældet med Rikkens hold kontra det tidligere hold, men er mere kunstigt, og statistisk problematisk, hvis det drejer sig om variable der i virkeligheden består af tal, som f.eks. indkomst eller sengetid (der begge bruges som eksempler). Som det var tilfældet med t-testen, kan der være udmærkede pædagogiske grunde til fokuset på variansanalysen, men man kan dog stadig savne at variansanalyserne i kapitel 4 og de variabeltyper de kræver, sættes i forhold til korrelationsanalyserne i kapitel 5.

Kapitel 5 behandler sammenhænge mellem variable, f.eks. mellem scorer i tekstlæsning og ordkendskab. De analytiske redskaber til at forstå sådanne sammenhænge er korrelation og regression. Korrelationsanalysen giver et mål, r , der viser hvor tæt sammenhængen er mellem to variable; mere forståelig er måske R^2 , der fortæller hvor stor en procentdel af variationen i de to variable der er fælles. Udover at vise hvor stærk sammenhængen er, kan korrelationstests etablere om den er signifikant, dvs. systematisk, ved hjælp af en p -værdi svarende til den der findes i t-test og variansanalyser. Endelig kan regressionsanalyser bruges til at undersøge sammenhænge og til at forudsige testscorer (og andre afhængige variable); her kan især multipel regression være nyttig fordi mere end én uafhængig variabel kan inkluderes. Dette eksemplificeres med en undersøgelse af hvordan testscoren i tekstforståelse afhænger af både ordkendskab og læseomfang.

De analyser der forklares i kapitel 5, beskrives som værende brugbare når man vil lede efter sammenhænge og lade sig inspirere, snarere end til ”systematiske udviklingsarbejder” (s. 110), og det er da også i overensstemmelse med traditionen indenfor bogens praksisfelt og i øvrigt mange andre. Der er dog i nyere forskning en tendens til at bruge regressionstilgangene også til mere planlagte undersøgelser, som måske nok kunne nævnes.

I kapitel 6 dækkes kort de metoder der kan bruges når ens mål er kategorier, snarere end tal, dvs. primært ja/nej-data (vil du anbefale

kurset til en ven?) og smiley-skalaer til at udtrykke (u)tilfredshed (hvor tilfreds er du med din egen indsats?). Her bruges hhv. chi i anden-test og Mann-Whitney-test, der forklares kort. Desuden kommer forfatterne ind på Spearmans korrelationstest for sammenhænge mellem rangordnede data. De tre tests repræsenterer et passende udvalg af tests, men man kunne igen godt ønske sig at skalaerne var lidt mere nuanceret forklaret, da det faktisk ikke er tilstrækkelig forudsætning for at bruge de tests der præsenteres i kapitel 3-5, at man har numeriske, ikke-kategoriske data.

Bogen slutter med kapitel 7 der ikke primært handler om statistiske metoder, men om de grundlæggende overvejelser der er relevante for fortolkningen af og tiltroen til de statistiske tests, med fokus på tre aspekter: gyldighed, pålidelighed og følsomhed. Gyldighed eller validitet har at gøre med om en test faktisk måler det den er konstrueret til at måle. Pålidelighed har at gøre med hvor følsom testen er i forhold til tilfældig variation; her er sammenligning af gentagne målinger, f.eks. før og efter et kursus, særligt udfordret fordi der er tilfældighed i målingen både før og efter. Endelig bør man overveje følsomheden af en test: Giver den et tilstrækkelig klart billede af hvad deltagerne faktisk kan? Hvis ikke, kan spørgsmål eller svarmuligheder omformuleres, evt. gøres mere kvalitative, eller der kan tilføjes flere spørgsmål. Bogens allersidste afsnit handler om en overvejelse der vel egentlig kan siges at være indledende for alt forskningsarbejde, nemlig hvorvidt det man er interesseret i, kan måles og hvordan. Spørgsmålet er om en kvalitativ undersøgelse er det mest relevante redskab, enten i sig selv eller som grundlag for en videre udforskning ved hjælp af kvantitative metoder.

VURDERING

Hold i virkeligheden er en nyttig bog der dækker mange vigtige aspekter med forbilledligt tilgængelige forklaringer. Det er interessant og relevant læsning, med en behagelig afvæbnende tone og passende sværhedsgrad. Selvom målgruppen er bredere end hvad vi repræsenterer, er det vores vurdering at bogen vil være ideel for den målgruppe den retter sig mod. Her er det en vigtig pointe at det kan være overordentlig brugbart at have en grundlæggende forståelse af statistik, også hvis

man ikke selv skal foretage egentlige statistiske analyser. For det første kan bogen udgøre en støtte for novicer indenfor statistik i samarbejdet med statistikere og andre mere erfarne statistikbrugere. For det andet kan den hjælpe til at de indsats der udvikles indenfor uddannelses-sektoren (og i øvrigt også bredere i offentlig forvaltning), fra starten bliver opbygget så de egner sig til at blive evalueret statistisk

Et vigtigt, pædagogisk begrundet træk ved bogen er at den hele tiden tager udgangspunkt i det samme hold af voksenkursister i dansk, Rikkes hold. Det er der åbenlyst gode grunde til: Rikkes hold er i den grad til at forholde sig til, og man skal ikke ustandseligt forstå nye scenarier og forskningsspørgsmål. Det giver imidlertid også nogle problemer: Det ensidige fokus gør at tests og eksempler kan blive svære at holde ude fra hinanden, og fordi alle eksempler tager samme udgangspunkt, er dækningen i forhold til uddannelseslitteraturen nok også dårligere end den kunne have været med flere eksempler – der er simpelthen færre virkelige scenarier bogen kan præsentere paralleller til når udgangspunktet altid er det samme. Derudover har forfatterne også valgt at undgå ord som population og stikprøve, hvilket måske gør eksemplerne mere tilgængelige, men det betyder også at læseren ikke lærer dem at kende, og de kan være relevante for at forstå den uddannelsesforskning som forfatterne sigter til at læseren skal kunne forstå. Så selvom fokuset på de håndgribelige størrelser – Rikke, Rikkes hold, voksenhold i dansk generelt – er fornuftigt, burde de nok relateres til de abstrakte størrelser individ, stikprøve og population.

Fra et statistisk synspunkt tages der lige lovlig let på forudsætningerne for de tests der forklares, primært i forhold til de afhængige variablers skala, som vi har nævnt ovenfor i forbindelse med et par af testene. Dette er mestendels et problem hvis læseren bruger bogen som vejledning til selv at lave undersøgelser. Det kunne måske også indvendes at størrelsen på Rikkes hold er lige lovlig lille, men det er et mindre problem.

Bogens struktur er gennemgående vellykket: Definitionsbokse og indeks er gode støtter til at lære og huske de centrale begreber, og aktiviteterne provokerer læseren til at tænke sig om uden at bringe hende for langt væk fra det kapitel hun er i gang med. Progressionen fra kapitel til kapitel fungerer også fint. Rent layoutmæssigt er den blandede

funktion af de lyserøde bokse lidt forvirrende, og forsidebilledet rent ud sagt gådefuldt – hvor kommer det mon fra, og hvad er meningen med det? – men det er kun små skønhedsfejl som vi mest nævner som forslag til den 2. udgave som vi håber og forventer vil blive relevant for denne glimrende bog.

Laura Winther Balling
Dept. of International Business Communication
Copenhagen Business School
lwb.ibr@cbs.dk

Rikke Slot Johnsen
Pædagogik og Integration, Børn og Unge
Aarhus Kommune
rikke.sj@gmail.com