

Hauge, Lilland

Å telle de usynlige

Hvordan lager man statistikker av titusener av rapporter om arrestasjoner, savnede eller henrettede? Hvordan settes ulike statistikker sammen med vitneutsagn slik at mønstre trer frem? Statistikk og maskinlæring (se under) har vist seg å være effektive våpen i kampen mot brudd på menneskerettigheter. Human Rights Data Analysis Group (HRDAG) er en gruppe med statistikere og databehandlere som jobber for menneskerettigheter, og har hatt en lang rekke prosjekter. Blant annet var de vitner i rettssaken mot tidligere president i Serbia Slobodan Milošević, og de har avdekket hemmelige gravsteder i Mexico og systematisk tortur i Tsjad. Et motto de har, er å telle de usynlige; å sette lys på de henrettelser som ikke er dokumentert, og som de sier: «We show that statistics have human consequences.»¹ For tiden jobber de med et litt annerledes prosjekt, nemlig om politivold i USA. For deres iherdige og verdifulle arbeid vant HRDAG Raftoprisen i 2021.

Kjellrun Hiis Hauge

Høgskulen på Vestlandet
khh@hvl.no

Inger Elin Lilland

Høgskulen på Vestlandet
iel@hvl.no

Human Rights Data Analysis Group (HRDAG)
Statisticians for human rights

<https://hrdag.org/>

HRDAG ble opprettet i 1991 av Patrick Bell, der han utviklet en database for El Salvador der militære offiserers handlinger og brudd på menneskerettigheter ble kartlagt i forbindelse med landets fredsprosess. Siden har HRDAG hatt prosjekter i 15 land fordelt på fem verdensdeler. HRDAG består av en liten kjerne av statistikere og databehandlere, men får hjelp fra en lang rekke frivillige med nøkkelkompetanser. Alle prosjektene foregår i samarbeid med andre, for eksempel sannhetskommisjoner og menneskerettighetsorganisasjoner.

Raftoprisen

<https://www.rafto.no/>

Raftoprisen deles ut årlig til kandidater som kjemper for menneskerettigheter. HRDAG vant prisen i 2021. Første pris ble delt ut i 1987, og det er Raftostiftelsen som står bak tildelingen etter en vurdering av innkomne forslag. Raftostiftelsen ble stiftet i 1986 for å minnes professor i økonomihistorie Thorolf Rafto og hans arbeid for menneskerettigheter og demokrati. Hovedkontoret ligger i Raftohuset, Bergen.

Forfatterne av denne artikkelen var så heldige å bli invitert til å treffe prisvinnerne i for-

bindelse med prisutdelingen, fordi vi holder et årlig kurs for lærere i regi av Raftostiftelsen. Kursets fokus er rettet mot hvordan matematikk- og norskfaget kan brukes til kritisk demokratisk danning, og da særlig knyttet til arbeid med kontroversielle temaer i klasserommet. I samtalen med Patrick Bell, grunnleggeren av HRDAG, var vi selvsagt interessert i arbeidet til HRDAG. Vi var i tillegg interessert i å høre om de gjennom sine erfaringer fra arbeidet i HRDAG hadde gjort seg noen tanker om matematikkundervisningen, og da særlig emnet statistikk i skolen.

I det følgende presenterer vi to av deres prosjekter, litt om metodene de bruker, og hva de etterlyser av allmennekunnskaper i statistikk. Avslutningsvis presenterer vi noen tanker om hvordan matematikkfaget i skolen kan bidra til at den enkelte i større grad kan bli i stand til å verdsette, vurdere og stille spørsmål ved tallbasert argumentasjon knyttet til samfunnsaktuelle temaer, for eksempel menneskerettigheter.

Ofrene i borgerkrigen i Colombia^{2,3}

Etter flere tiår med borgerkrig ville myndighetene i Colombia ha svar på hvem som ble drept, kidnappet eller som forsvant, og hvem som utførte de voldelige handlingene. En sannhetskommisjon ble opprettet, og 30 ulike organisasjoner bidro med hvert sitt datasett bestående av totalt 200 filer med til sammen rundt 20 millioner registreringer. HRDAG veiledet samarbeidspartnere og lærte dem opp i aktuelle metoder. De involverte samarbeidspartnerne lærte å vaske data (klargjøre data for analyser), blant annet hvordan man velger felles format, finner overlappende data og vurderer kvaliteten på dataene, og samarbeidspartnerne lærte om ulike analyseverktøy. Slik hjelper HRDAG til med å finne «manglende data» som skal få frem ofre som ikke er registrert. For eksempel ble det estimert at 40 % av forsvinningene under borgerkrigen ikke var rapportert. Arbeidet krever at institusjoner, organisasjoner og frivillige jobber sammen om data.

Politivold og rasisme i USA⁴

Siden 2015 har HRDAG jobbet med udokumenterte ofre for politivold og rasisme i USA. Analysemetodene er de samme som de har brukt i andre saker. Funnene er urovekkende. USA har programvare som skal hjelpe politi og dommere til å treffe fordomsfrie valg. Etter å ha analysert store mengder med data konkluderer HRDAG med at valgene likevel er rasistiske. Ikke rart, kanskje, da det blant annet viser seg at rasisme er innbakt i den såkalte fordomsfrie programvaren, fordi den bygger på tidligere (rasistiske) erfaringer. Dette gjelder både programvaren som skal hjelpe politiet med å ta beslutninger, og programvaren som hjelper dommerne i å avsi passende dom; et varsku om at kvaliteten på data er vesentlig i statistiske analyser, samt at slike analyser slett ikke trenger å være objektive. Videre i denne saken fant HRDAG at tiltalte som det ble kausjonert for, hadde større sannsynlighet for å bli dømt skyldig. I tillegg fant de at av alle drap i USA begått av fremmede for den drepte var en tredjedel begått av politiet. HRDAG har også laget statistikker over pågripelser med militære våpen i hjem med barn til stede, og de har sett på seksuell trakassering ved kroppsvisitering. Basert på politirapporter, offisielle klager og vitneutsagn fant de en rekke grove tilfeller av seksuelle overgrep bak rapporteringsordlyden «Improper Search of Person».

Metodene

HRDAG arbeider aldri alene fordi de trenger samarbeidspartnere som har den kontekstuelle kunnskapen. Deres egen kompetanse er innenfor analyse av data, dvs. statistiske metoder og maskinlæring. Maskinlæring er en form for kunstig intelligens og består av å gjøre datamaskiner i stand til å gjenkjenne komplekse mønstre basert på data. HRDAG trenger maskinlæring når de får et utall dokumenter og det ikke er mulig å få lest gjennom alle. Når det er snakk om tusenvis eller millioner av dokumenter, sier det seg selv at det er nødvendig med datahjelp. I politivoldsaken der de skulle kartlegge pågripel-

ser med militære våpen i hjem der det var barn til stede, for eksempel, var det flere måter barns tilstedeværelse ble rapportert på. Noen rapporter hadde rubrikker for slik informasjon, men de var merket på ulikt vis, og i andre tilfeller kom det frem i tekst som beskrev hendelsen. Dette betyr at dataprogrammet må finne rapporter som inneholder pågripelser i hjem der militært utstyr er brukt, de må avdekke ulike ordsammensetninger som viser at barn var til stede, og informasjon som identifiserer rapportene slik at de kan hentes frem. Når papirdokumenter er skannet til en lang sammenhengende fil, trengs det dessuten maskinlæring for å finne ut hvor hver enkelt rapport begynner.

For å få tilstrekkelig informasjon til å belyse en sak må man kvalitetssikre dataene, og det trengs flere datakilder som må overføres til et felles format. Hvis det, for eksempel, er flere uavhengige oversikter over drepte eller forsvunne, er det vesentlig å finne overlapp mellom kildene. Deretter kan det totale antallet drepte estimeres, inkludert de som ikke er registrert. I Guatemalas langvarige borgerkrig ble utallige mennesker drept. I spørsmålet om drepte i Ixil-regionen i 1982–83, sammenliknet de antall overlappende registreringer fra to uavhengige kilder med enkeltregistreringer. Ved hjelp av sannsynlighetsberegninger, nærmere bestemt multiple systems estimation (MSE⁵), estimerte de antall ikke-registrerte drepte til å være en fjerdedel av totalt antall drepte.

HRDAG bruker som regel bayesiansk statistikk i sine analyser i kombinasjon med matematisk modellering. Bayesiansk statistikk skiller seg fra frekventistisk statistikk blant annet ved at det tas utgangspunkt i *a priori* sannsynlighetsfordelinger på ukjente parametre, altså at de behandles som stokastiske variabler (Heuch, 1997). Med støtte i innsamlede data og informasjon blir så den opprinnelige sannsynlighetsfordelingen justert i analysen. Slik kan bayesianerne håndtere visse usikre situasjoner bedre enn frekventistene, men det krever representative data som grunnlag for *a priori*-fordelingen.

HRDAG tar i bruk og videreutvikler statistiske metoder kontinuerlig. Fordi de har metodene til det, jobber de stort sett med forsvinninger eller dødsfall som ikke er dokumenterte. Enten står de selv for analysene, eller de lærer opp samarbeidspartnerne. HRDAG er avhengig av tillit og troverdighet, så de må være sikre på at påstandene bygger på solide analyser. Dette krever tålmodighet og langsiktighet. I noen saker tar det 10 til 20 år før dataene og analysene kan brukes i kampen mot brudd på menneskerettigheter.

Et av mottoene til HRDAG er at «hvert menneske som har blitt myrdet, skal bli husket.» Dette betyr ikke at de avdekker identiteten til hvert av ofrene. «Hvis vi ikke kan navngi hvert offer, kan vi i det minste telle dem,» som Megan Price, HRDAGs daglige leder, sa i takketalen under prisutdelingen. Men de er forsiktige med å oppgi *ett* tall for antall ofre, og opererer i større grad med intervaller bestemt ved hjelp av sannsynlighetsvurderinger og estimering når de skal si noe om omfanget av forsvinninger og drap. Det må være solid dekning for konklusjonene, ellers kan anklagede bli felt på feil grunnlag, og HRDAG kan miste troverdighet.

I tilfeller der dataene ikke er gode nok til at man kan anslå antall menneskerettighetsbrudd godt nok, eller til at man kan anslå et intervall, kan mønstre i dataene være verdifulle. For eksempel avslørte analyser i prosjektet om politivold at det kunne ligge grove seksuelle overgrep gjemt bak «Improper Search of Person». Det gir grunn til å mene at det må gjøres noe med rapporteringen. Andre eksempler er mønstre som avdekker skjulte gravsteder, og mønstre som avdekker hvem som har gitt ordre som fører til brudd på menneskerettigheter.

Betydning for matematikk- og statistikkundervisning

Problematikken som HRDAG arbeider med, sammen med fortellingen om hvordan de arbeider med å «telle de usynlige», er viktig i

seg selv å få formidlet til elever. Det gir eksempel på hvordan statistikk og tall kan bidra til å gi innsikt i grunnlaget for sosiopolitiske beslutningsprosesser i samfunnet, men også eksempel på hvordan tall kan brukes for å legitimere eller dekke over mer eller mindre kriminelle handlinger utført av et lands myndighetspersoner eller andre beslutningstakere. Det handler om å formidle betydningen statistikk har som redskap for å få innsikt i, men også når det gjelder å kunne skjule, grunnlaget for sosiopolitiske beslutningsprosesser i samfunnet. Samtidig kan fortellingen fungere som inspirasjon til utforskningsprosjekter i matematikkfaget alene eller som del av et tverrfaglig prosjekt i skolen. Gjennom det tverrfaglige temaet demokrati og medborgerskap understreker Kunnskapsløftet viktigheten av å gi elevene kompetanse i å utforske, analysere og vurdere funn fra reelle datasett fra hverdagsliv og samfunnsliv, samt gjøre elevene bevisste på «føresetnader og premisser for matematiske modeller som ligg til grunn for avgjerder i deira eige liv og i samfunnet».

I samtalen vi hadde med Patrick Bell, uttrykte han at «It's not necessary for people to understand statistics, but to appreciate statistics.» Med det mente han at det ikke er nødvendig at oppdragsgivere og samarbeidspartnere forstår de kompliserte statistiske metodene, men at de verdsetter styrken ved statistiske metoder og har en bevissthet om hvor begrensningene ligger. Han fortalte videre at myndigheter og personer de ellers kom i kontakt med, ofte hadde begrensede allmennkunnskaper i statistikk. For eksempel var det ikke rent sjelden at oppdragsgivere hadde forventninger om objektive og entydige svar straks det forelå statistikker. Blant annet ble HRDAG møtt med liten forståelse for betydningen av at gode analyser krever data med god kvalitet. En vanlig oppfatning er at tallfestede opplysninger, og som dessuten er behandlet ved hjelp av en datamaskin, gir klare og objektive svar. Den nye læreplanens etterspørsel etter aktiviteter med reelle data og bevisstgjøring av tallmateriale og modellers

begrensninger imøtekommer HRDAGs ønske om at folk skal kunne anerkjenne hva analysene krever av datakvalitet.

Fenomenet med overdreven tillit til kvantifiseringer er noe kritisk matematikkdiraktikk har vært opptatt av. Blant annet har Skovsmose (2005, 2006) påpekt at matematikkundervisning kan skape en form for overdreven tillit til tall om undervisningen domineres av et oppgaveparadigme der hver oppgave har kun ett gyldig svar. Det kan støtte opp under en ubegrunnet tro på at tallbeskrivelser er objektive. Videre peker Skovsmose på at anvendelse av matematikk og statistikk kan legitimere tvilsomme beslutninger, kan gi autoritet til fremstillinger og kan forskyve og filtrere etisk ansvarlighet. Når det foretas statistiske analyser, er også konteksten sentral. Patrick Bell understreket betydningen av å samarbeide med folk som hadde god kjennskap til konteksten til tallmaterialet. Uten denne innsikten kunne ikke HRDAG ha utført gode nok analyser. Dette er en erkjennelse som oppdragsgivere ikke har nok bevissthet rundt, sukket Patrick Bell. I tillegg fant han det problematisk at folk ikke vet forskjell på et estimat og en variabel, og han var opptatt av at her hadde skolen en jobb å gjøre når det gjaldt å legge til rette for at elever lærte seg denne forskjellen. Så, hvorfor er dette så viktig for HRDAG? Mens et estimat bare er et anslag på en størrelse, og beheftet med usikkerhet, kan en variabel representere flere verdier.

Datakvalitet kan være påvirket av så mangt. I tillegg til for få data, inkonsistente data og irrelevante data kan mindre opplagte forhold påvirke datakvaliteten, for eksempel hvordan data er samlet inn, hvordan undersøkelsen kan påvirke svarene, og hvordan konteksten og designet av undersøkelsen kan være verdiladd (Weiland, 2017). Hvordan seksuell trakassering ikke ble rapportert i saken beskrevet over, og dermed ikke kom frem i statistiske oversikter, er et eksempel på hvordan valg av rubrikker i et skjema muliggjør at grov trakassering blir gjemt eller dekket over. I mange av sakene HRDAG

jobber med, er det i overgripernes interesse at drap ikke blir rapportert, eller at den ansvarlige ikke blir navngitt.

Hvordan kan matematikkundervisning i skolen øke bevisstheten om godt anvendt statistikk? Selvsagt må grunnleggende begreper være på plass, men i denne teksten konsentrerer vi oss om de kritiske perspektivene. Vi understreker betydningen av å inkludere samfunnsaktuelle og samfunnskritiske perspektiver der det er mulig, parallelt med begrepslæring. Uansett om statistikk er knyttet til brudd på menneskerettigheter eller til langt mindre dramatiske samfunnsspørsmål, er det å kunne stille kritiske spørsmål til data en viktig del av å være en aktiv medborger og verdifullt for demokratiet. Det krever imidlertid erfaring med ulike typer svake data og hvordan det påvirker kvaliteten i resultatet. I tillegg krever det erfaring med hva man faktisk kan finne ut av med gode data, og hva de kan bidra til å belyse, slik som for eksempel HRDAG demonstrerer. Dette er et ansvar skolen kan ta og gi elever på ulike trinn erfaringer med, for eksempel ved at elevene selv planlegger, gjennomfører og kritisk analyserer egne og andres undersøkelser foretatt i klasse eller i nærmiljø.

Når det gjelder spørsmål knyttet til kontroversielle temaer som for eksempel nedleggelse av grendeskoler, flyktnings- og innvandringsproblematikk, koronatiltak osv., kan elever få i oppgave å undersøke påstander eller spørsmål som presenteres og diskuteres i mediene. Til dette finnes datasett og tallmateriale til utforskning lett tilgjengelig på sidene til Statistisk sentralbyrå (særlig relevant er Kostra⁶), Flykninghjelpen, FN-sambandet og Folkehelseinstituttet, for å nevne noen. Når det gjelder å få innsikt i forskjellen mellom et estimat og en variabel, foreslår Lysø (2014) at elever i en klasse gruppevis kan få i oppdrag å finne middelveien i en populasjon hvor en er ute av stand til å undersøke alle verdiene i populasjonen. Deretter kan de undersøke hvordan gruppens beregnede middelveier varierer mellom gruppene, og

hva dette kan ha å si for hvordan vi kan snakke om middelveien i populasjonen.

Kritisk vurdering av kontekst er et annet område som HRDAG er opptatt av. Data kan være samlet inn for et visst formål, i en viss kontekst, men det kan være fristende å bruke de samme dataene til andre formål som dataene ikke støtter i samme grad. Trolig er det dette HRDAG ønsker større forståelse for når de blir presentert for mulige brudd på menneskerettigheter samt oversikter og tallmateriale som det var tenkt skulle støtte slike påstander. Relevansen av informasjon og tilnærming for problemstillingen er sentralt både i statistiske analyser og for matematisk modellering, som igjen kan være del av eller bygge på statistiske analyser. For eksempel er metavalidering et begrep som er brukt i forbindelse med matematisk modellering. Dette innebærer å ha et kritisk blikk på data, dataenes relevans og hvordan antagelser og begrensninger påvirker resultatet (Niss, 2015).

Å øve opp elevenes evne til kritisk tenkning står sentralt i Kunnskapsløftet, hvor det understrekes at et hovedmål for skolen er å stimulere elevene til å bli aktive medborgere og gi dem kompetanse til å delta i videreutviklingen av demokratiet i Norge. Det handler om å gi elevene kunnskaper og ferdigheter til å møte utfordringer i tråd med demokratiske prinsipper, øve opp evnen til å tenke kritisk og lære seg å håndtere meningsbrytninger og respektere uenighet.

En kritisk tilnærming til statistikk og matematisk modellering i skolen og til å være kritisk kan bety å være vurderende og evaluerende, for eksempel ved å spørre om noe er feil. Men det å være kritisk kan også forstås som en prosess mot det å forstå og kanskje endre egne oppfatninger om noe (Weiland, 2017). Dette betyr at å jobbe med statistikk og matematisk modellering kan avsløre misbruk, men også utvide ens egen horisont. En kritisk tilnærming handler om å erfare både styrker og svakheter ved bruk av statistikk og matematisk modellering.

Tangenten: tidsskrift for matematikkundervisning

Noter

- 1 <https://hrdag.org/>
- 2 <https://hrdag.org/2021/10/28/colombia-truth-reconciliation/>
- 3 <https://hrdag.org/casanare-colombia/>
- 4 <https://hrdag.org/usa/>
- 5 <https://hrdag.org/2013/03/11/mse-the-basics/>
- 6 <https://www.ssb.no/offentlig-sektor/kostra/>

Referanser

- Heuch, I. (1997). Striden mellom bayesianere og frekventister om idégrunnlaget for statistiske slutninger. I R. Strand & G.A. Bristow (Red.) *Naturvitere filosofer* (s. 14–28). Megaloceros forlag.
- Lysø, K. O. (2014). *Sannsynlighetsregning og statistisk metodelære* (2. utg.). Caspar Forlag.
- Niss, M. (2015). Prescriptive modelling – challenges and opportunities. I G.A. Stillman (Red.), *Mathematical modelling in education research and practice. Cultural, Social and Cognitive Influences* (s. 67–79). Springer International Publishing.
- Skovsmose, O. (2005). *Travelling Through Education. Uncertainty, Mathematics, Responsibility*. Sense Publisher.
- Skovsmose, O. (2006). Kritisk matematikkundervisning – for fremtiden. I O. Skovsmose & M. Blomhøj (Red.), *Kunne det tænkes? – om matematiklæring* (s. 273–295). Malling Beck.
- Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics*, 96, 33–47.