

Jan Erik Kristiansen



Jan Erik Kristiansen er sosiolog og seniorrådgiver i Formidlingsavdelingen i Statistisk sentralbyrå. Han har lang erfaring i å presentere statistikk på en brukervennlig måte og har holdt en rekke kurs i statistikkforståelse og bruk av statistikk.

Han utga i 2007 boken «Tall kan temmes!» (U-forlaget), som delvis danner grunnlaget for denne spalten. (jan.erik.kristiansen@ssb.no)

Statistikk er en sosial konstruksjon

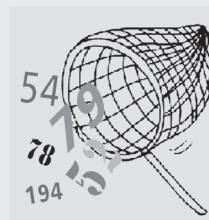
Det snakkes ofte om datainnsamling som det viktigste og avgjørende punktet eller momentet i en undersøkelse. Og selvfølgelig er datainnsamlingen viktig: uten data – ingen statistikk. Men det kan likevel være grunn til å minne om de mange andre elementene som inngår i produksjonskjeden.

Innsamlingen av data omtales i dag ofte som datafangst, og man får lett et inntrykk av at data er noe som finnes «der ute» og lar seg «fange», omtrent som sommerfugler eller fisk. Slik er det naturligvis ikke: I hovedsak er statistikk en sosial konstruksjon. Og statistikkproduksjon består (på samme måte som andre undersøkelser) av en rekke operasjoner eller trinn, som alle innebærer vurderinger og avgrensninger. For eksempel:

Enheter/populasjon: Et første hovedpunkt er ofte å avgjøre hvem statistikken skal omfatte: Hva er populasjonen? Skal for eksempel en undersøkelse om helse omfatte alle personer i landet, bare voksne personer eller bare personer i yrkesaktiv alder? Et annet eksempel: Om vi skal lage statistikk om innvandreres levekår, må vi først avgrense eller definere hvem undersøkelsen skal omfatte: bare førstegenerasjons innvandrere eller også annengenerasjon, det vil si deres etterkommere? Eller også andre med «innvandringsbakgrunn»? Definisjoner og begreper er helt avgjørende for hva statistikken viser, og hvordan den skal tolkes.

Avhengig av temaet for undersøkelsen/statistikken må man så definere det man vil undersøke: Hva ønsker vi å få vite om de enhetene vi undersøker? I en undersøkelse om arbeidsmiljø må første skritt være å bestemme hvilke aspekter av arbeidsmiljøet som er viktig, deretter må disse variablene operasjonaliseres; det vil si at man må konkretisere hvordan de

kan måles. I tillegg må man bestemme hvilke bakgrunnsvariabler som skal innhentes.



I neste omgang må man bestemme hvordan undersøkelsen skal gjennomføres. Skal den baseres på et utvalg av enheter? Hvor stort skal utvalget være? Eller skal alle telles? Dette får igjen betydning for graden av usikkerhet som

knytter seg til resultatene.

Når dataene er samlet inn, må de kodes for videre databearbeiding, de må kontrolleres og eventuelt rettes. Deretter må de ofte klassifiseres; noe som også innebærer definering – for eksempel av «eldre» og «unge» eller «lav», «middels» og «høy» inntekt.

I tillegg vil man ofte foreta visse korrekasjoner i forhold til ulike feilkilder som kan forekomme i datainnsamlingen, for eksempel sesongjusteringer eller justeringer for frafall i spesielle grupper.

Til slutt vil man ofte foreta en rekke ulike beregninger for å gjøre tallene mer sammenliknbare, som beregning av prosenter eller rater, gjennomsnitt eller andre sammenfattende mål, som for eksempel indekser.

Deretter kan man endelig begynne med det som egentlig er hovedformålet med statistikk: å sammenlikne tallene. Alt i alt innebærer dette at statistikk – enten det er i form av såkalt offisiell statistikk eller andre undersøkelser – slett ikke er uberørt av menneskehender. Snarere



tvært imot: Statistikk er *et sosialt produkt*, og er som sådan et resultat av de mange valg og beslutninger som inngår i innsamlings-, produksjons- og bearbeidingsprosessen. Derfor er det hele tiden viktig å spørre: Hvordan er denne statistikken framkommet? Hvordan er begrepene definert?

Før vi teller, må vi definere

Statistikk dreier seg selvfølgelig om tall. Men før vi kan begynne å telle, måle og beregne, må vi definere. Og svært ofte er statistikkens resultater mer avhengig av hvordan ulike variabler eller indikatorer er definert enn av datainnsamlingen og datakvaliteten. I denne forstand kan vi si at statistikk til en viss grad er elastisk, det vil si at ved å «strekke» definisjonene, får vi svært ulike resultater.

Et svært enkelt eksempel på hvor stor betydning definisjoner har, er de såkalte ferieundersøkelsene i Statistisk sentralbyrå, hvor en «ferietur» defineres som minimum fire dager (overnattinger) hjemmefra (operasjonell definisjon). Dermed fanger ikke disse undersøkelsene opp de mange helgeturene som er blitt så vanlig de siste årene; disse går gjerne til storbyer og varer ofte bare 2-3 dager. Ifølge ferieundersøkelsen reiser nordmenn nå på omtrent 3,3 millioner ferieturer til utlandet i året. I andre undersøkelser – de såkalte reiseundersøkelsene – registreres også kortere ferieturer, og når disse tas med, øker antallet ferieturer til utlandet med nesten 2 millioner, eller to tredjedeler. Feriestatistikkens resultater er derfor langt mer avhengig av hvordan vi definerer ferie, enn av ulike feilkilder og usikkerhet i data.

Et annet eksempel på definisjonens avgjørende betydning kan være begrepet «innvandrere»: Antallet innvandrere i Norge er i stor grad avhengig av hvor vid eller trang vi gjør definisjonen: Tar vi for eksempel med bare dem som er født i utlandet med to utenlandsfødte foreldre, blir antallet 342 000. Tar vi også med personer født i Norge med to foreldre som er innvandrere (etterkommere/

«annengenerasjonsinnvandrere», SSBs definisjon av innvandrerbefolkningen), blir tallet 415 000. Hvis vi i tillegg også tar med dem «med innvandringsbakgrunn», som for eksempel utenlandsadopterte og personer med én utenlandsfødt forelder, blir antallet omtrent dobbelt så høyt.

Arbeidsledige eller -søkende?

Et tredje eksempel på definisjoners betydning kan være hvordan vi måler arbeidsledighet. Ifølge Statistisk sentralbyrås arbeidskraftundersøkelse (AKU) var det 54 000 arbeidsledige i fjerde kvartal 2007. Samtidig meldte NAV om 39 000 arbeidsledige. Når SSB og NAV publiserer tall som spriker såpass mye, skyldes dette rett og slett at de opererer med litt ulike definisjoner av arbeidsløshet.

Hos NAV registreres de som melder seg ledige på arbeidskontorene. SSBs arbeidskraftundersøkelse – som er en intervjuundersøkelse – inkluderer også arbeidsledige som ikke registrerer seg ved arbeidsformidlingen. Dette gjelder blant annet et relativt stort (men ukjent) antall elever og studenter som ønsker seg en deltidsjobb ved siden av studiene. Alle elever/studenter som ønsker en jobb av minst én times varighet, regnes dermed som arbeidsledige. Dette omfatter dermed også en gruppe heltidselever og -studenter, som mange kanskje snarere vil betegne som arbeidssøkende enn som arbeidsløse. Legg også merke til at AKU omfatter personer i alderen 67-74 år, selv om den generelle pensjonsalderen i Norge er 67 år.

Også mange andre personer som ønsker arbeid, faller utenfor begge disse offisielle definisjonene av arbeidsløshet. Dette kan være personer som ikke har søkt arbeid, som ikke har søkt de siste fire ukene, eller som ikke kan begynne i arbeid umiddelbart. Mange deltidsysselsatte ønsker også mer arbeid. Om alle disse ble definert som arbeidsledige, ville ledigheten sannsynligvis nærme seg det doblete.

Tabell 1. Antall innvandrere, etter ulike definisjoner. 1. januar 2007

	Antall
Førstegenerasjonsinnvandrere	342 000
Første- og annengenerasjonsinnvandrere	415 000
«Med innvandringsbakgrunn»	665 000

Kilde: Innvandrerbefolkningen, Statistisk sentralbyrå.

**«Ved å “strekke”
definisjonene
får vi svært
ulike resultater»**



«Statistikkens definisjoner er med på å bestemme hvordan vi oppfatter samfunnet.»

Fattigdom til hodebry

Et siste eksempel: Fattigdom kan defineres og måles på en rekke ulike måter. I Norge brukes vanligvis en såkalt relativ fattigdomsgrense, det vil si at fattigdom (eller lavinntekt) defineres relativt til andres inntekt, snarere enn i forhold til en absolutt minimumsgrense, for eksempel minstepensjon. Men også en slik relativ grense kan fastsettes på forskjellige måter.

Tabellen nedenfor viser resultatet av to ulike, men vanlige beregningsmåter eller definisjoner. Felles for begge beregningsmåtene er at de tar utgangspunkt i medianinntekten, det vil si den inntekten som deler fordelingen i to like store deler, slik at halvparten tjener mer enn dette og halvparten mindre. Selv om de to beregningsmåtene har samme utgangspunkt, definerer de fattigdomsgrensen som henholdsvis 50 og 60 prosent av medianen. Disse to grenseverdiene er selvfølgelig i en viss forstand tilfeldige, man bruker runde tall og ikke for eksempel 51 eller 58 prosent av medianen. Valget av 50 og 60 prosent skyldes også muligheten for internasjonale sammenlikninger. I tillegg justeres det også for størrelsen på husholdningen, fordi store husholdninger antas å ha visse «stordriftsfordeler». Her finnes det to såkalte ekvivalensskalaer, som er mye brukt: en EU-skala og en OECD-skala.

Når vi kombinerer disse to tilnærmingene, får vi fire forskjellige fattigdomsgrenser, som i neste omgang resulterer i fire ulike anslag for hvor mange fattige (eller personer med lav inntekt) vi har i Norge.

Tabell 2. **Andel personer med lav inntekt. 2005. Prosent**

Lavinntekts- grense	Ekvivalensskala			
	OECD-skala		EU-skala	
	50 prosent	60 prosent	50 prosent	60 prosent
Andel med lav inntekt	5,6	10,0	6,5	11,4

Kilde: Inntekts- og formuesundersøkelsen for husholdninger, Statistisk sentralbyrå.

Som vi ser, varierer anslagene fra 6 til 11 prosent. I tillegg kan tilsvarende andeler også beregnes uten studenter, siden studielån ikke registreres som inntekt i inntektsstatistikken, og fordi studenters lave inntekt betraktes som forbigående. Dermed får man ytterligere fire (noe lavere) anslag for andelen med lav inntekt. Endelig kan man også beregne andelen som er fattige over lengre tid, for eksempel en treårsperiode, såkalt vedvarende lavinntekt – en beregningsmåte som gir enda lavere andeler.

På denne måten kan den enkelte bruker selv velge «fattigdomsnivå», noe som nok en gang illustrerer at statistikk i stor grad er et spørsmål om definisjoner. Igjen ser vi hvordan utenomstatistiske vurderinger og valg får stor betydning for resultatet av slike sammenlikninger. Valg av ekvivalensskala kan selvfølgelig forankres i antakelser eller empiriske data om for eksempel stordriftsfordeler. Når det gjelder valget mellom 50 eller 60 prosent av medianinntekten, kan dette begrunnes med internasjonal sammenliknbarhet. Men i siste omgang er valget ofte nettopp det – et valg.

Definisjonsmakt og ansvar

Poenget her er at definisjoner av for eksempel innvandrere, fattigdom/lav inntekt eller arbeidsledighet ikke er gitt – de er sosialt konstruert eller bestemt. Det er i seg selv ikke noe galt i dette, men det er viktig å være klar over at det er slik – og at det må være slik. Statistikk innebærer altså en betydelig definisjonsmakt, i den forstand at statistikkens definisjoner er med på å bestemme hvordan vi oppfatter samfunnet.

Disse eksemplene viser også hvor viktig det er å oppgi kilde når man bruker statistikk, slik at leseren har mulighet til å kontrollere opplysningene og finne ut for eksempel hvilke definisjoner som ligger til grunn for statistikken, hvordan dataene er samlet inn, og så videre.