

Jan Erik Kristiansen



Jan Erik Kristiansen er sosiolog og seniorrådgiver i Statistisk sentralbyrå, Formidlingsavdelingen. 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!» (IJ-forlaget), som delvis danner grunnlaget for denne spalten.

(jan.erik.kristiansen@ssb.no)

Nøyaktighetens tyranni: Store tall og desimaler

Man kan ofte få inntrykk av at statistikk er en svært nøyaktig disiplin. For eksempel var det 4 681 134 innbyggere i Norge i 2007, det ble født 58 459 barn, brutto nasjonalprodukt (BNP) per innbygger var 483 725 kroner, det ble anmeldt 271 712 forbrytelser, det var 2 027 894 personbiler i landet, gjennomsnittlig månedslønn for heltidsansatte var 32 257 kroner, samlet fruktbarhetstall var 1,901, og forventet levealder for menn og kvinner var henholdsvis 78,24 år og 82,66.

Alt sammen svært nøyaktige tall. Og på en måte kan man si at statistikkens formål nettopp er å bidra til å gi presise beskrivelser av ulike samfunnsfenomener, slik at vi kan erstatte generelle utsegn – som for eksempel at «sykefraværet er høyt» – med mer nøyaktig informasjon om *hvor høyt* det er.

På den annen side innebærer all statistikk også en form for forenkling: Statistikk innebærer på en måte å se samfunnet i fugleperspektiv. Undersøkelser gir opplysninger om utvalgte egenskaper eller variabler ved bestemte enheter, og resultatet er i stor grad avhengig av hvordan variablene er definert. I tillegg er resultatet ofte forbundet med en relativt stor grad av usikkerhet. På denne bakgrunn må mye av statistikkens tilsynelatende presisjon betraktes som illusorisk.

Avrunding

Svært få av tallene ovenfor er helt korrekte. Tallet for BNP per innbygger er foreløpig og vil bli revidert både én og to ganger. Og om man oppgir befolkningen i Norge i 2007 til 4 681 134 personer, bør man også opplyse om tidspunktet for denne tallfestingen (1. januar), siden befolkningen øker med vel 100 personer per dag. Allerede 1. april 2007 var

antallet bosatte økt til 4 691 502. (Antall personbiler, derimot, blir oppgitt ved årets utgang).

I tillegg vil de mange sifrene og desimalene som oftest være langt mer forstyrrende enn opplysende, særlig når de presenteres i løpende tekst. Les for eksempel følgende setning (hypotetisk eksempel): «Fra 1990 til 2001 økte antall kvinnelige studenter fra 33 764 til 65 756, mens antall mannlige studenter økte fra 28 435 til 43 567».

Les så setningen en gang til: «Fra 1990 til 2001 økte antall kvinnelige studenter fra 33 800 til 65 800, mens antall mannlige studenter økte fra 28 400 til 43 600». De fleste vil synes at det er langt lettere å lese og oppfatte hovedpoenget i den siste setningen, fordi tallene her er avrundet. Vi ser da lettere at antallet kvinnelige studenter økte med nesten 100 prosent, mens antallet mannlige studenter økte med om lag 50 prosent. Og for langt de fleste brukere og formål er de avrundede tallene nøyaktige nok. Presisjon øker ikke alltid tilliten til tallene, og i alle fall ikke forståelsen av dem – snarere tvert imot.

Avrundede tall er både lettere å huske, og de er lettere å forstå og å sammenlikne: De fleste kan foreta hoderegning med tall med to siffer – blir det flere, får vi problemer. Prøv for eksempel å sammenlikne tallene 633 og 214. Om vi runder av til 630 og 210, ser vi fort at det ene er tre ganger så stort som det andre. Og når vi skal kjøpe bil, tenker vi neppe på prisen som for eksempel 149 670 eller 164 730 kroner, men som 150 000 eller 165 000.

En vanlig tommelfingerregel når det gjelder avrunding, er å runde av til to eller tre tellende (betydningsfulle) sifre,

for eksempel at 56 887 rundes til 57 000 eller 56 900. En spesiell form for avrundning er å skrive for eksempel «seks av ti», «hver fjerde» eller «en tredjedel» i stedet for å oppgi nøyaktige prosentandeler. Denne formen for brøkgregning er ofte uttrykk for et ønske om å variere språket og å angi et grovt mål på størrelsen. Ved sammenlikninger bør man imidlertid være forsiktig: «To tredjedeler» og «en fjerdedel» er ikke lettere å sammenlikne enn 67 og 25 prosent.

Store tall

Store tall kan altså med fordel avrundes. Men også avrundede tall kan bli for store. Dagbladet hadde 3. august 2005 et oppslag om norske husholdningers samlede gjeld, som da hadde nådd et rekordnivå. I avisen ble beløpet presentert slik:

Vi skylder 1 300 000 000 000 kr*
1300 milliarder!

Beløpet var altså avrundet. Men avisen skjønte nok at et beløp med elleve nuller kan bli i meste laget og forklarte derfor i en slags fotnote at dette er 1 300 milliarder. Men hvor mye er egentlig dette? Da journalisten forsøkte å illustrere beløpet ved å si at det tilsvarer 2 890 utgaver av det nye Bislett stadion til 450 millioner per stykk, er det vel et åpent spørsmål om dette blir så mye mer meningsfylt eller forståelig. Det å illustrere et stort tall ved hjelp av en slik analogi (flere store tall) er ikke alltid like vellykket.

En bedre måte å alminneliggjøre store tall på er ofte å beregne hva de betyr per innbygger. I dette tilfellet blir gjelden per innbygger 281 000 kroner, et tall som er langt mer forståelig og meningsfylt. Eller man kan – som Dagbladet også gjør – beregne gjelden per innbygger i alderen 18-60 år (de yngste og de eldste har sjelden gjeld); den blir da 498 000 kroner, eller en halv million kroner per innbygger. Det aller beste her hadde kanskje vært å beregne gjelden per husholdning. Dermed får man et tall som leserne kan sammenlikne, for eksempel med sin egen situasjon.

Desimaler

Av og til kan en få inntrykk av at desimaler er statistikkens adelsmerke. Men desimaler gir ofte inntrykk av en presisjon eller nøyaktighet som det ikke alltid er grunnlag for. Ett av de mange månedlige partibarometrene kan for eksempel oppgi Arbeiderpartiets oppslutning til 26,8 prosent. Men når vi vet at slike undersøkelser vanligvis baserer seg på et utvalg av befolkningen på omtrent 1 000 personer, og at usikkerheten for dette tallet er i størrelsesorden plus/minus 2 prosentpoeng, blir desimalen uttrykk for en falsk nøyaktighet (eksaktifisering). Hvis vi derimot rapporterer valgresultatet, blir saken en annen: Her er det helt naturlig å bruke én desimal: 26,8 prosent.

Selv om regneark og kalkulatorer er svært rundhåndet med antall desimaler, kan vi si at desimaler vanligvis er unødvendig når vi presenterer prosentandeler på grunnlag av utvalgsundersøkelser. Unntak er undersøkelser basert på svært store utvalg (som for eksempel SSBs arbeidskraftsundersøkelse – AKU). Når vi presenterer statistikk basert på tellinger (som folke- og boligtellinger) eller registerdata (jamfør VGs arbeidsledighets-tall), er én desimal på sin plass. Men heller ikke flere.

Når man omtaler prosentvise endringer, er det vanlig å bruke bare én (eller ingen) desimal. Men det finnes unntak, spesielt når endringene er små: Årlig folketilvekst oppgis vanligvis med to desimaler, som for eksempel 0,56 prosent. Også enkelte små prosentandeler oppgis med to desimaler: for eksempel utviklingshjelp som prosent av bruttonasjonalinntekten = 0,87.

Avrundning av desimaler

Vær oppmerksom på at desimaltall svært ofte allerede er avrundet: Hvis vi så sløyfer en desimal, innebærer dette at vi runder av en gang til, noe som kan få





relativt store konsekvenser når tallene er små: 2,5 kan være en avrunding av 2,47. Hvis vi så sløyfer en desimal, kan 2,5 i neste omgang bli avrundet til 3, til tross for at det opprinnelige tallet altså var nærmere 2. Et regneark, derimot, vil huske også det opprinnelige tallet og foreta en korrekt avrunding til 2. Om vi senere multipliserer to avrundede tall med hverandre, kan resultatet bli svært forskjellig: $2,47 * 5,42 = 13,4$, mens $2 * 5 = 10$. Avrundingen bør derfor skje helt til slutt.

Det vanlige er altså å runde av til nærmeste hele tall. Men det finnes ett unntak: Når det gjelder *alder*, foretar vi nesten alltid en avrunding nedover. En person regnes således som 32 år helt til han eller hun fyller 33. Dette innebærer derfor at vi må legge til et halvt år når vi beregner gjennomsnittsalderen i et utvalg eller en populasjon.

Eksempel I: Forventet levealder

Forventet levealder beregnes og publiseres ofte med to desimaler: For Norge er levealderen i 2007 beregnet til 78,24 år for menn og 82,66 år for kvinner. Dette kan betraktes som en slags prognose for dem som ble født dette året, under forutsetning av at dødeligheten på ulike alderstrinn også i framtiden vil være den samme som i 2007.

Selv om vi nå et øyeblikk aksepterer denne forutsetningen, signaliser likevel to desimaler en nøyaktighet som det neppe verken er grunnlag for eller behov for. Nå er imidlertid forutsetningen om en framtidig stabil aldersspesifikk dødelighet vanskelig å akseptere: De siste 50 årene har således den forventede levealder ved fødselen økt med mellom seks og sju år.

Nå vet vi selvfølgelig ikke om denne økningen vil fortsette, men foreløpig er det ingen ting som tyder på noen snarlig stagnasjon. Dette betyr at de publiserte tallene med stor sannsynlighet undervurderer forventet levealder med flere år for dem som fødes i dag, kanskje tre til seks år. På denne bakgrunn blir det å oppgi

den beregnede alderen med to desimaler meningsløst; selv én desimal er sannsynligvis én for mye.

Eksempel II: Flypassasjerer

Statistikken over flypassasjerer som publiseres av Avinor (og SSB), oppgir at det i 2007 var 26 455 387 innenlandske passasjerer på landets flyplasser. Altså et svært nøyaktig tall. Og med de rutiner og sikkerhetskontroller som i dag finnes på alle flyplasser, er det vel all grunn til å anta at dette tallet er riktig. Likevel ville det nok for de fleste formål være nøyaktig nok å oppgi tallet som 23,5 millioner.



Særlig fordi dette tilsynelatende presise tallet dekker over en mer grunnleggende usikkerhet: Hva er en flypassasjer? En flypassasjer er her nemlig ikke det samme som en person eller reisende: Flypassasjerer telles to ganger: både ved avreise og ankomst. Dermed kan antallet passasjerer egentlig halveres. I tillegg kommer at noen passasjerer telles flere ganger: En person som reiser fra Oslo til Bodø med flybytte i Trondheim, utgjør fire «passasjerer». På denne bakgrunn blir det oppgitte, nøyaktige passasjerantallet litt meningsløst. Særlig fordi antallet innenlandske flypassasjerer i en annen statistikk (innenlandske transportytelser) oppgis til 10 millioner passasjerer.

Oppsummert: Siden mange sosiale fenomener er svært avhengige av hvordan de defineres og måles – i tillegg til at de ofte også er beheftet med en betydelig usikkerhet – gir det vanligvis liten mening å presentere veldig nøyaktige tall: For de fleste formål er det tilstrekkelig med avrundede tall, selvfølgelig avhengig av formål og datakvalitet. *Det er svært ofte bedre å presentere et omtrentlig, men tilnærmet riktig tall, enn et helt nøyaktig, men galt tall.*