
Kari og Ola finnes ikke

MEDISIN OG TALL

JO RØISLIEN

jo@joroislien.no

Jo Røislien er professor i medisinsk statistikk ved Det helsevitenskapelige fakultet, Universitetet i Stavanger, og profilert fagformidler og programleder på TV.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

KATHRINE FREY FRØSLIE

Kathrine Frey Frøslie er statistiker ved Nasjonal kompetansetjeneste for kvinnehelse, Oslo universitetssykehus, og driver den populærvitenskapelige strikkebloggen Statistikk.no.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Gjennomsnittsmennesket er et spøkelse fra fortiden.

En viktig bruk av statistikk er å oppsummere mange enkeltobservasjoner med noen få oppsummerende tall, såkalt deskriptiv eller beskrivende statistikk. Det finnes etter hvert noen prinsipper for hva som er gode måter å gjøre det på, men underveis har det oppstått noen seiglivede feilslutninger. Gjennomsnittsmennesket er en slik.

En statistikkvind

På midten av 1800-tallet blåste en statistikkvind over verden. Naturvitenskapene hadde i lang tid hatt stor suksess med å bruke tall for å beskrive naturen, og andre vitenskaper kom etter. Som en samfunnsmessig parallell til «physics» ville man lage et «social physics», det som nå heter sosiologi, der målet var å beskrive samfunnet ved hjelp av tall. Nasjonale statistikkbyråer ble opprettet i land etter land, og med store tallmengder om samfunnet og dets innbyggere kom også behovet for å oppsummere og presentere disse tallene på meningsfylt og forståelig vis.

En av pionerene i dette arbeidet var belgiske Adolphe Quetelet (1796–1874), med sin banebrytende innsats innen statistikk og kvantitativ analyse. I tillegg til kroppsmasseindeks (KMI) for å beskrive kroppssammensetning var en av Quetelets store vitenskapelige oppfinnelser «l’homme moyenne» – gjennomsnittsmannen: et menneske med gjennomsnittlige egenskaper (1). Ideen var besnærende. I stedet for rader og kolonner med tallinformasjon kunne man formelig se oppsummeringstallene foran seg, manifestert som et faktisk menneske. Ideen spredte seg, og etter hvert ga mange land til og med navn til gjennomsnittsmennesket sitt. I Norge har vi Kari og Ola Nordmann, svenskene har sin Medel-Svensson, mens britene har Jane og John Doe.

Gjennomsnittsmannen møter motstand

Men Quetelets gjennomsnittsmenneske møtte motstand. Det var vanskelig å se for seg at et menneske med gjennomsnittlige egenskaper faktisk fantes. Menneskekroppen er et finjustert system der alt henger sammen med alt, og gjennomsnittstall for ulike menneskelige egenskaper kan ikke uten videre limes sammen til et nytt menneske. Som prinsipp ga gjennomsnittsmennesket rett og slett ikke mening.

Men gjennomsnittsmannen klamret seg til livet. Da som nå ble det å formidle vitenskapelige resultater på forståelig vis ansett som viktig, og matematikeren Joseph Diaz Gergonne (1771–1859) introduserte i 1825 «mannen i gata» som mål på om man evnet å forklare en teori (2). Fra 1900 og utover ble «mannen i gata» popularisert av den tyske matematikeren David Hilbert (1862–1943), og «mannen i gata» fant også veien til den amerikanske politikken, i betydning «den alminnelige velger». Til Norge kom «mannen i gata» for fullt i forbindelse med folkeavstemningen om norsk EU-medlemskap i 1972. Ideen om ett menneske som representant for de mange slo an nok en gang.

Suffisiensprinsippet

Det Quetelet og hans samtidige pirket i, er et av de grunnleggende spørsmålene i statistikk: Hvordan oppsummere en mengde med tall uten at essensiell informasjon går tapt i prosessen? Det kan virke som et u håndterlig spørsmål, men når det formuleres matematisk blir det overraskende håndterbart. Svaret er det som i matematisk statistikk kalles «suffisiensprinsippet» – tilstrekkelighetsprinsippet (3). Suffisiensprinsippet gir oss verdifull innsikt i hva som faktisk må til for å oppsummere data på skikkelig vis.

Gjennomsnittet er ikke nok

Det er flere måter å finne midten i et datasett på. Medianen, som er det midterste tallet, og gjennomsnittet, som er en matematisk beregning av det geometriske tyngdepunktet i tallene, er de mest brukte. Statistikere elsker gjennomsnittet på grunn av dets matematiske egenskaper. Det er lettere å regne med enn medianen, og dukker opp av seg selv flere steder. Men der medianen ofte er en faktisk observasjon, så er ikke

gjennomsnittet et tall som finnes blant målingene. Gjennomsnittet er ikke typisk, det er ikke vanlig, det er ikke normalt. Gjennomsnittet er faktisk så unormalt at det ikke engang finnes blant observasjonene.

Men det er et større, mer fundamentalt problem med å la ett tall representere de mange, uavhengig av om dette tallet er en matematisk konstruksjon eller en smart valgt observasjon: Det er ikke nok.

Bare nok er nok

I et vanlig datasett, passelig normalfordelt og uten uteliggere (outliers), viser suffisiensprinsippet at vi trenger to oppsummeringstall: gjennomsnitt og standardavvik. Rapporterer du bare det ene, tilbakeholder du essensiell informasjon om datasettet. Ett gjennomsnitt, to gjennomsnitt, hundre gjennomsnitt, limt sammen til et helt gjennomsnittsmenneske: Det er fortsatt ikke nok. For å gi en fullgod beskrivelse av en populasjon er variasjonen i tallmaterialet beviselig like viktig som tyngdepunktet.

Det er ikke så rart folk ikke kjenner seg igjen i gjennomsnittsmennesket, eller at ingen føler at de selv er «mannen i gata». Det er fordi Kari og Ola Nordmann ikke representerer dem. Ved å fokusere på gjennomsnittsmennesket har essensiell informasjon om gruppen gått tapt. Samtidig innsnevres forståelsen av hva som er typisk, vanlig og normalt. For normalfordelte data vil ca. 95 % av observasjonene ligge i intervallet $\text{gjennomsnitt} \pm \text{to standardavvik}$. Det er et intervall som ofte gir rause rammer for hva som er vanlig.

Populær

Quetelets gjennomsnittsmenneske er fortsatt populært. Et Google-søk på «average person» gir i skrivende stund nærmere 15 millioner treff. Det er mer enn både Whitney Houston, Marlon Brando og Sir Isaac Newton kan skilte med. Vitenskapen avlivet gjennomsnittsmannen for over 100 år siden fordi han ikke ga mening. Men samfunnet han var ment å skulle beskrive, har holdt ham kunstig i live siden.

LITTERATUR

1. Faerstein E, Winkelstein W. Adolphe quetelet: statistician and more. *Epidemiology* 2012; 23: 762 - 3. [PubMed][CrossRef]
2. Universitetet i Agder. Mannen i gata er 200 år. *Forskning.no*. <https://forskning.no/2016/12/har-funnet-mannen-i-gata/produsert-og-finansiert-av/universitetet-i-agder> (8.11.2018)
3. Birnbaum A. On the foundations of statistical inference. *J Am Stat Assoc* 1962; 57: 269 - 306. [CrossRef]

Publisert: 1. januar 2019. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.18.0824

