

10 av 9 blir lurt av statistikk

I 1991 ble britten Dennis Adams dømt for voldtekt, hovedsakelig på grunnlag av en DNA-profil. Alle de øvrige indisiene lot til å peke bort fra Adams, ikke minst et vitneutsagn fra offeret, som sa at det ikke var ham. Likevel lot juryen seg overbevise av at Adams' DNA-profil tilsvarte gjerningsmannens.

Årsaken er først og fremst å finne i statistikk. Den statistiske nøyaktigheten på DNA-profilen var nemlig så høy at bare én av en million menn ville ha en tilsvarende profil. Mye taler for at juryen oppfattet det som om det bare var én sjanse av en million for at Adams var uskyldig. Men i virkeligheten er dette en

feilslutning. Forholdet én til en million betyr her bare at for hver million menn vil det være én med en DNA-profil som tilsvare gjerningsmannens.

I Storbritannia er det omtrent 20 millioner voksne menn. Det vil si at det foruten Adams er 19 andre menn som også har en DNA-profil som er lik gjerningsmannens. Enhver av disse 19 mennene kunne i prinsippet like godt være den skyldige. En DNA-profil er altså bare noe verd hvis den blir sett i sammenheng med andre bevis og fakta.

Eksempelet viser hvordan statistikk blir misforstått og i verste fall misbrukt. Historien om Dennis Adams har sammen med andre saker fått en britisk professor i statistikk, Philip Dawid ved University College i London, til å foreslå at dommere og advokater sendes på kurs for å bli flinkere til å tolke statistikk.

Hivrisiko ble halvert

Det kunne kanskje mange andre også ha utbytte av. For det er nemlig ikke bare i rettssalene at statistikk kan bety forskjellen mellom liv og død.

Tyske forskere ved Max Planck-instituttet stilte i et eksperiment 20 behandlere på rådgivningsklinikker som utfører hivtester, overfor følgende situasjon:

En ung mann som ikke tilhører noen spesiell risikogruppe, kommer inn for å bli testet. Men først har han et spørsmål:

Hver gang vi åpner en avis eller slår på fjernsynet, blir vi bombardert med statistiske fakta om alt. Men statistikk er vanskeligere å forstå enn vi tror. Derfor kan den misbrukes – og det blir den ofte.

Hvis testen skulle være positiv, hvor stor er i så fall sannsynligheten for at han faktisk er smittet med hiv?

Behandlerne, som hovedsakelig var leger, fikk oppgitt at for unge menn uten spesiell risikofaktor ville 1 av 10 000 statistisk sett være smittet. Hivtesten var så god at hvis en person var smittet, ville den alltid avsløre det. Hvis han derimot ikke var smittet, ville testen likevel være positiv i 1 av 10 000 tilfeller.

Av de 20 behandlerne erklærte 10 at testen var absolutt sikker, mens ytterligere fire mente at sikkerheten var 99,9 prosent. En enkelt trodde at sannsynligheten for at den unge mannen faktisk var smittet, var 90 prosent. Dette tallet var det laveste som overhodet ble foreslått.

Det korrekte svaret er at hvis mannens test er positiv, vil han i halvparten av tilfellene *ikke* være smittet. Årsaken er

Fornøyelsesparker er rene dødsfeller

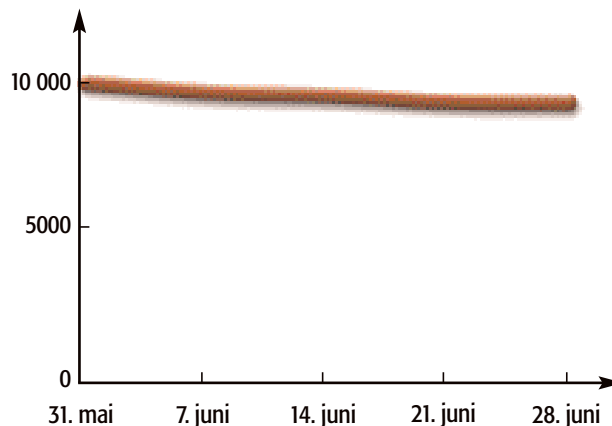
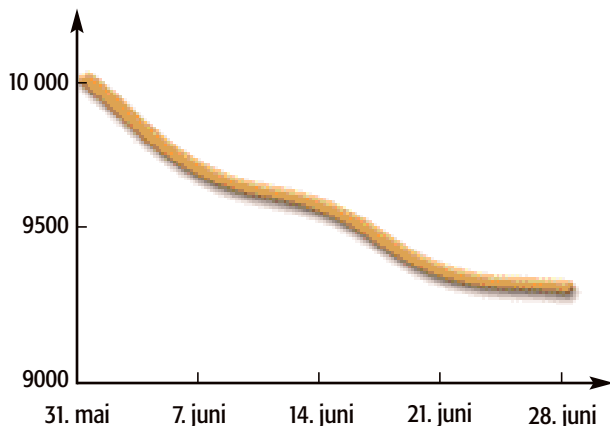
I 2001 gikk antallet ulykker i amerikanske fornyelsesparker kraftig opp. Men som all amerikansk ulykkesstatistikk var undersøkelsen basert på tall fra 100 tilfeldig utvalgte sykehus. Nærmere granskning avslørte at det økte antallet kom fra ett sykehus, etter all sannsynlighet fordi en større fornyelsespark nylig var åpnet nær dette sykehuset.

Aksjekursen stupdykket ...

eller gjorde den?

Den amerikanske aksjeindeksen Dow Jones falt kraftig i juni 2002 – eller gjorde den?
Til venstre ser vi utviklingen i indeksen der bare et utsnitt av vertikalaksen er vist.
Til høyre er den samme kurven tegnet inn, men her er hele vertikalaksen tatt med.

CLAUS LUNAU



denne: Av 10 000 menn som testes, vil det i gjennomsnitt være én smittet. Men på grunn av testens usikkerhet vil det blant de samme 10 000 også være én som tester positivt uten å være smittet.

Er mannens test positiv, kan han altså like godt være den som virkelig er smittet, som den som har fått et feilaktig positivt testresultat. Sannsynligheten for at han faktisk er smittet, er bare 50 prosent.

Selv om eksempelet er tenkt, ligger tallene ikke langt fra virkelighetens verden. Forståelse for statistikk er derfor ikke bare et spørsmål om vitenskapelige spissfindigheter. Det er eksempler på at folk har levd i angst i årevis etter at de har fått vite om en feilaktig positiv hivtest. Enkelte har til og med tatt livet av seg.

Et eldorado for juksemakere

Når selv høyt utdannede leger kan gå seg så vill i tallene, kan det neppe være vanskelig å villede oss andre med statistikk.

Det er da også en kjensgjerning at statistikk egner seg ualminnelig godt til juks. Det er tallrike eksempler på at politikere, journalister, advokater og til og med vitenskapsmenn mer eller mindre bevisst har utnyttet folks manglende innsikt i tallenes kronglete verden.

For å avsløre misbruk av statistikk må man være oppmerksom på at statistisk arbeid består av tre prosesser: innsamling av data, bearbeiding av data og presen-

tasjon av data. Det er mulig å jukse i alle tre ledd, og for å oppnå upålitelige resultater er det bare nødvendig å manipulere ett av leddene. Samtidig blir jukset betydelig vanskeligere å avsløre hvis man er uetisk bare det ene stedet, men redelig andre steder. Man kan for eksempel gjøre mye ut av å presentere sine resultater på en troverdig og overbevisende måte, uten å komme nærmere inn på hvordan materialet er blitt samlet inn.

Lag grafen som du vil ha den

Den enkleste måten å manipulere statistikk på er i forbindelse med presentasjonen av resultatene, som i mange tilfeller

foregår ved grafisk fremstilling. Et klassisk eksempel på manipulasjon av grafer er å bruke koordinatsystemer uten noe nullpunkt på vertikalaksen. Ved å vise bare den øverste delen av aksene kan man få dramatiske utslag i stigninger og fall på en kurve. Hvis den samme grafen ble tegnet i et koordinatsystem med et nullpunkt, ville utviklingen kanskje virke ganske ubetydelig. Og det blir umulig å gjennomskue en graf hvis man fullstendig unnlater å angi enheter på aksene.

Man kan også beskjære en graf slik at den bare viser det man vil. Analyserer man for eksempel en bestand av lemen, viser det seg at antallet øker og minker i en syklus som strekker seg over fire år. Slik er det nå engang med lemen, og det er det ikke noe unormalt i. Men hvis man beskjærer grafen, kan kurven over bestandens utvikling tegne et dramatisk bilde av at bestanden av lemen i de siste to årene er i voldsom tilbakegang – eller i dramatisk fremgang, alt etter formålet.

Vi misforstår presenter

Statistiske resultater blir ofte fremlagt ved å bruke presenter fordi det er en enkel måte å beskrive en utvikling på. Men man må aldri glemme at presenter gir mening bare hvis man vet hvilke tall som ligger bak dem. Det lyder for eksempel katastrofalt når vi får vite at hyppigheten av en dødelig sykdom er ►

Asiatiske kvinner bryr seg ikke om hvite menn

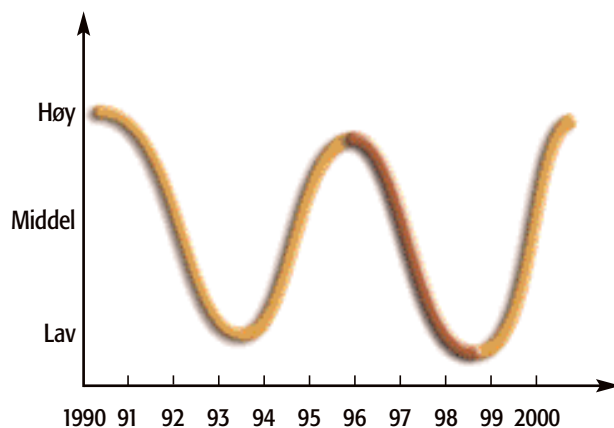
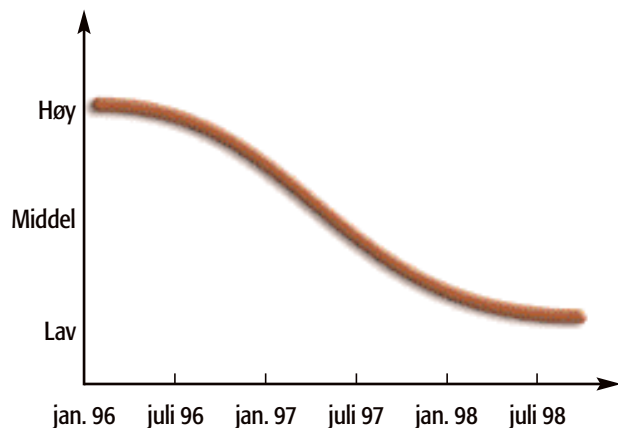
En amerikansk undersøkelse viser at bare 1,9 prosent hvite menn har asiatiske kvinner som partnere. Det fikk forskerne til å konkludere med at asiatiske kvinner ikke velger hvite menn. Nei! Det er bare mange flere hvite i USA enn asiater. Faktisk deler annenhver asiatiske kvinne bord og seng med en hvit mann.

Lemenbestanden er truet ...

hvert fjerde år

Verden rundt svinger bestanden av lemen i perioder på cirka fire år som vist på grafen. Men en uærlig statistiker kan benytte seg av de regelmessige svingningene og nøye seg med å vise en begrenset del av kurven. Plutselig er lemenbestanden i dramatisk tilbakegang.

CLAUS LUNAU



► økt med 150 prosent på verdensbasis i løpet av et år. Tallet blir atskillig mindre imponerende hvis det skulle vise seg at det dreier seg om en økning fra fire til ti tilfeller på hele kloden.

Samtidig har overraskende mange mennesker problemer med å forstå prosenter. For eksempel viste en tysk undersøkelse at en tredel av alle tyskere ikke har noe begrep om vanlig prosentregning. I undersøkelsen ble folk blant annet bedt om å forklare hva 40 prosent betydde. Noen mente at det betydde en firedel, mens andre trodde at det var én av 40.

Vi kan også få en følelse av prosentenes upålitelighet gjennom dette eksempelet: En bedrift lanserer seg selv som et progressivt firma som legger vekt på likestilling. Bedriften dokumenterer sin innstilling med en rask liten statistikk:

Vedkommende bedrift har to avdelinger, som den kaller A og B. I løpet av det siste året er kvinneandelen i begge avdelingene økt, i avdeling A fra 10 til 20 prosent og i avdeling B fra 50 til 60 prosent. Dette lyder bra og tar seg pent ut i bedriftens presentasjonsmateriale.

Men en skeptisk statistiker går nå tallene etter i sømmene og finner ut følgende: Avdeling A hadde 10 medarbeidere i fjor, av dem 1 kvinne (10 prosent). Avdeling B hadde 40 medarbeidere, av dem 20 kvinner (50 prosent). I løpet av det siste året har bedriften hatt fremgang, og det er blitt ansatt flere folk. Avdeling A har nå

100 medarbeidere, av dem 20 kvinner (20 prosent), mens avdeling B har 50 medarbeidere, av dem 30 kvinner (60 prosent).

Tallene viser derfor at kvinnekvoten ganske riktig har økt i begge avdelingene. Likevel viser det seg at den samlede kvinneandelen i bedriften har minket! Den har falt fra 42 prosent det første året (21 av 50 medarbeidere) til 33 prosent året etter (50 av 150 medarbeidere). Plutselig viser det seg at det dreier seg om en gammeldags og kvinnefiendtlig bedrift.

Årsaken til dette overraskende resultatet er at avdeling A har vokst uforholdsmessig mye i forhold til avdeling B. Selv om det også er kommet flere kvinner i avdeling A, er det ansatt så mange

ekstra menn der at det forrykker totalbildet. Den ene kvinnen som var ansatt der i forveien, telte så å si for mye.

Det har faktisk fått sitt eget navn når prosenter oppfører seg på denne måten. Fenomenet kalles Simpson-paradokset, oppkalt etter en britisk statistiker.

Spis is og bli frisk

Akkurat som det kan jukses med prosenter, kan statistikk også brukes til å hevde sammenhenger som er det rene vrøvl. For eksempel kan det bevises rent statistisk at man er mindre forkjølet hvis man spiser mye iskrem. Tallene forteller nemlig at de fleste mennesker er mindre forkjølet om sommeren, samtidig som forbruket av is er mye høyere om sommeren enn om vinteren. Voilå: Is er sunt!

Det er lett nok å gjennomskue at det ikke er noen sammenheng mellom iskrem og snue. Men slik er det ikke alltid i andre tilfeller. Det måtte en gruppe amerikanske forskere innrømme etter at de i 1999 meddelte at barn som sov med lyset tent, var tilbøyelige til å bli nærsynte. Forskningsresultatet ble offentliggjort i det kjente vitenskapelige tidsskriftet Nature, men historien holdt ikke. Det viste seg at foreldre som selv er nærsynte, ofte får nærsynte barn. Og for å unngå å snuble over leker når de skulle se til avkommet om kvelden, var de nærsynte foreldrene mer tilbøyelige enn andre til å la lyset brenne i barneværelset. Derfor

Flyging er verdens sikreste transportform

Det er mye sikrere å fly enn å kjøre bil, hører vi ofte. Men er det slik? Hvis vi ser på antallet passasjerkilometer, er det ganske riktig bare 0,3 dødsfall per milliard kilometer i fly, mens det er 4 i bil. Men ser vi i stedet på antallet turer, er det plutselig 55 døde per milliard flyturer mot bare 45 omkomne per milliard utflukter i bil.

skjedde det oftere at lys var tent i nærsynte barns soveværelser enn i jevnaldrende kameraters. Men det var ikke lyset som ga dem dårligere syn.

Skottland er fullt av svarte sauer

Også i innsamlingen av data kan det bli surr i tingene, og det er rike muligheter for løgnaktighet. Spesielt skal man være oppmerksom på om det er et rimelig forhold mellom mengden av data og den presisjonen man uttaler seg med.

Dette problemet illustreres tydelig av en gammel anekdote:

En britisk politiker reiser til Skottland i forbindelse med en valgkamp. Mens han sitter på toget, ser han noen svarte sauer på et jorde. Politikerens konstaterer at alle sauene i Skottland må være svarte, og gleder seg over at denne detaljen kan brukes i en tale. Hans rådgiver, som er statistiker, bemerker at det eneste man kan si med sikkerhet, er at noen sauer i Skottland er svarte på den ene siden.

Det er ikke alltid lett å si hvor stor datamengde som kreves. Men vil man uttale seg med sikkerhet, kreves det svært grundige observasjoner. For eksempel er det alltid en viss risiko for at man har oversett noen få sauer i en fjern avkrok.

Flere ben enn gjennomsnittet

Videre må man passe seg for ren pseudo-statistikk. Dette er opplysninger som presenteres som statistikk, men i virkeligheten ikke er det, og kan blant annet forekomme i forbindelse med gjennomsnitt. For eksempel er det et faktum at et overveiende flertall av befolkningen her i landet har mer enn det gjennomsnittlige antallet ben. Et fåtall ettbente personer trekker snittet ned, og så får vi et pussig, men temmelig meningsløst utsagn.

På samme måte ble det presentert som en statistisk kjensgjerning at Danmark var på vei til å vinne fotball-VM sommeren 2002, da det danske landslaget hadde slått de regjerende franske mestrene 2-0. Årsaken var at de siste tre gangene Danmark og Frankrike hadde møtt hverandre i en sluttrunde (EM 1992, VM 1998 og EM 2000), ble det vinnende laget mestere til slutt. Som kjent holdt ikke statistikken vann. Brasil vant VM.

Bakgrunnen er selvsagt at resultatet av de innbyrdes oppgjørene ikke har noen – eller bare svært liten – betydning for det endelige utfallet. Selv om Dan-

Bildet er utelatt
pga. copyright

O.J. Simpsons forsvarer, Robert Shapiro (til venstre), brukte statistikk i sitt vellykkede forsøk på å få O.J. frifunnet for mord.

Statistikk satte O.J. på frifot

I den meget omtalte saken mot den amerikanske rugbystjernen O.J. Simpson spilte statistikk en stor rolle. Selv om ingen med sikkerhet kan si hvorfor juryen frifant O.J. for anklagen om mord på sin kone, har kritikere påpekt den måten som forsvarer benyttet statistikk på. O.J.s forsvarer fortalte jurymedlemmene at av 1000 menn som slår sin kone, er det bare én som går så langt at han dreper henne. Det skulle gi juryen et inntrykk

at det bare var én sjanse av 1000 for at O.J. var skyldig.

Denne statistikken ble ikke imøtegått av anklageren, men er på ingen måte et indisium på O.J.s uskyld. Omvendt kunne anklageren ha påpekt at når en kvinne som er blitt mishandlet av ektemannen blir myrdet, er det i annethvert tilfelle hennes mann som er morderen. Med et slikt perspektiv var det altså 50 prosent sannsynlighet for at O.J. var gjerningsmannen.

mark og Frankrike hadde møtt hverandre utallige ganger, ville ikke resultatet si mye om hvem som kom til å vinne VM.

Det samme gjør seg gjeldende for systemer som påstås å kunne forutsi lottotall. Man kan i dag få kjøpt dataprogrammer som er utviklet for å frembringe de mest sannsynlige vinnertallene i ukens lottospill når de blir føret med tidligere vinnende kombinasjoner. Programmene

analyserer hvilke tall som ikke er kommet ut på lenge, hvilke tall som ofte forekommer sammen, og så videre.

Problemet er bare at maskinen som trekker ut lottotalle, ikke vet noe om alt dette. Den er bare en maskin som helt tilfeldig plukker ut noen små baller. Så sannsynligheten for å vinne i lotto er fremdeles mindre enn for å bli rammet av lynet. Og det er statistisk bevist! □