

Kære bruger

Denne pdf-fil er downloadet fra Illustreret Videnskabs website ([www.illvid.dk](http://www.illvid.dk)) og må ikke videregives til tredjepart.

Af hensyn til copyright indeholder den ingen fotos.

Mvh

Redaktionen

# 10 ud af 9 bliver snydt af statistik

I 1991 blev englænderen Dennis Adams dømt for voldtægt, primært på basis af en dna-profil. Alle de øvrige indier syntes at pege bort fra Adams, ikke mindst et vidneudsagn fra offeret, der sagde, at det ikke var ham. Alligevel lod juryen sig overbevise af, at Adams' dna-profil svarede til gerningsmandens.

Årsagen skulle først og fremmest findes i statistik. Den statistiske nøjagtighed på dna-profilen var nemlig så høj, at kun én ud af en million mænd ville have en tilsvarende profil. Meget taler for, at juryen opfattede det, som om der kun var én chance ud af en million for, at Adams var uskyldig. Men i virkeligheden er det en

fejlslutning. Når forholdet én til en million nævnes, betyder det blot, at for hver million mænd vil der være én, hvis dna-profil svarer til gerningsmandens.

I Storbritannien er der omkring 20 millioner voksne mænd. Det vil sige, at der foruden Adams er 19 andre mænd, som også har en dna-profil, der er magen til gerningsmandens. Enhver af disse 19 mænd kunne i princippet lige så godt være den skyldige. En dna-profil er altså kun noget værd, hvis den bliver sammenholdt med andre beviser og fakta.

Eksemplet viser, hvordan statistik bliver misforstået og i værste fald misbrugt. Historien om Dennis Adams har sammen med andre sager fået den engelske professor i statistik Philip Dawid fra University College i London til at foreslå, at dommere og advokater sendes på kursus for at blive dygtigere til at afkode statistik.

## Hiv-risiko blev halveret

Det kunne mange andre måske også have gavn af. For det er nemlig ikke kun i retssalene, at statistik kan være forskellen imellem liv og død.

Tyske forskere fra Max Planck-instituttet stillede i et eksperiment 20 behandlere på rådgivningsklinikker, der udfører hiv-test, over for følgende situation:

En ung mand, der ikke tilhører nogen speciel risikogruppe, kommer ind for at blive testet. Men først har han et spørgs-

Hver gang vi åbner en avis eller tænder fjernsynet, bliver vi bombarderet med statistiske fakta om alt. Men statistik er sværere at forstå, end vi tror. Derfor kan det misbruges – og bliver det hyppigt.

mål: Hvis testen er positiv, hvor stor er sandsynligheden så for, at han rent faktisk er smittet med hiv?

Behandlerne, der primært var læger, fik opgivet, at for unge mænd uden særlig risikoadfærd ville 1 ud af 10.000 statistisk set være smittet. Hiv-testen var så god, at hvis en person var smittet, ville den altid afsløre det. Hvis han derimod ikke var smittet, ville den alligevel være positiv i 1 ud af 10.000 tilfælde.

Ud af de 20 behandlere erklærede 10, at testen var absolut sikker, mens yderligere fire mente, at sikkerheden var 99,9 procent. En enkelt troede, at sandsynligheden for, at den unge mand faktisk var smittet, var 90 procent. Det tal var det laveste, der blev foreslået.

Det korrekte svar er, at hvis mandens test er positiv, vil han i halvdelen af tilfældene *ikke* være smittet. Årsagen er den-

## Forlystelsesparker er rene dødsfælder

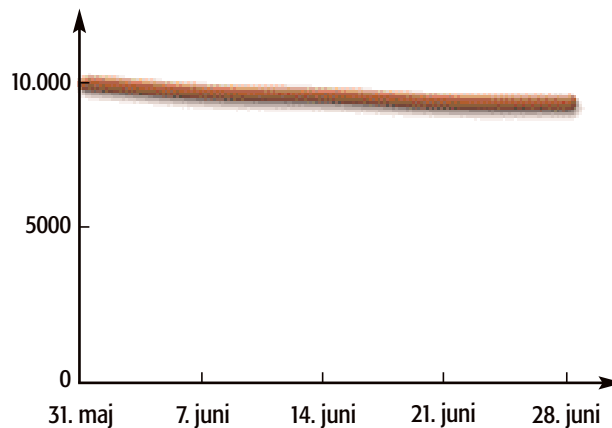
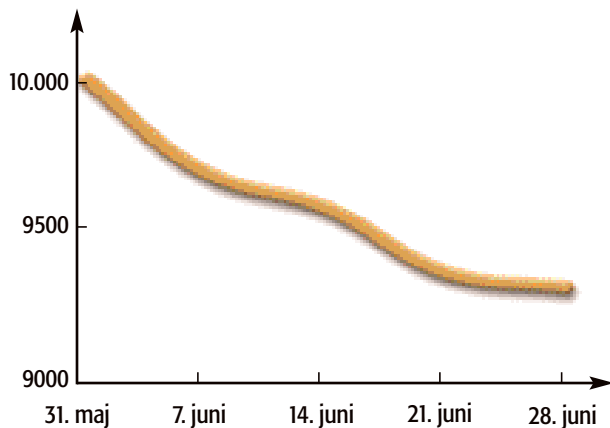
I 2001 gik antallet af ulykker i amerikanske forlystelsesparker kraftigt op. Men som al amerikansk ulykkesstatistik var undersøgelsen baseret på tal fra 100 tilfældigt udvalgte skadestuer. Nærmere granskning afslørede, at hele stigningen kom fra ét hospital. Efter al sandsynlighed fordi en større forlystelsespark netop var åbnet tæt på dette hospital.

## Aktiekursen styrtdykkede ...

## eller gjorde den?

Det amerikanske aktieindeks Dow Jones faldt kraftigt i juni 2002 – eller gjorde det? Til venstre ses udviklingen i indekset, hvor kun et udsnit af den lodrette akse er vist. Til højre er den samme kurve tegnet ind, men her er hele den lodrette akse med.

CLAUS LUNAU



ne: Ud af 10.000 testede mænd vil der i gennemsnit være én smittet. Men på grund af testens usikkerhed vil der blandt de samme 10.000 mænd være én, der får en positiv test uden at være smittet.

Er mandens test positiv, kan han altså lige så godt være ham, der virkelig er smittet, som ham, der har fået et fejlagtigt positivt resultat. Sandsynligheden for, at han faktisk er smittet, er kun 50 procent.

Selv om eksemplet er tænkt, ligger tallene ikke langt fra virkelighedens verden. Så forståelse for statistik er ikke kun et spørgsmål om videnskabelige spidsfindigheder. Der er eksempler på folk, som har levet i angst i årevis efter at have modtaget en fejlagtig positiv hiv-test. Enkelte har endda taget deres eget liv.

### Et eldorado for fuskere

Når selv højtuddannede læger kan gå så galt i byen, burde det være en smal sag at vildlede os andre med statistik.

Det er da også en kendsgerning, at statistik er ualmindeligt velegnet til fusk. Der er talrige eksempler på, at politikere, journalister, advokater og sågar videnskabsmænd mere eller mindre bevidst har udnyttet folks manglende indsigt i tallenes kringlede verden.

For at afsløre misbrug af statistik skal man være opmærksom på, at statistisk arbejde består af tre processer: indsamling af data, bearbejdning af data og præ-

sentation af data. Det er muligt at fuske i alle tre led, og for at opnå upålidelige resultater er det kun nødvendigt at manipulere med ét af leddene. Samtidig bliver snyderiet sværere at afsløre, hvis man kun er uetisk det ene sted, men redelig de andre. Man kan for eksempel gøre meget ud af at præsentere sine resultater på en troværdig og overbevisende måde uden at komme nærmere ind på, hvordan man har indsamlet sit materiale.

### Få din graf, som du vil

Den nemmeste måde at manipulere med statistik på er i forbindelse med præsentationen af resultaterne, der i mange

tilfælde foregår i form af grafer. Et klassisk eksempel på manipulation med grafer er at bruge koordinatsystemer, der ikke har et nulpunkt på den lodrette akse. Ved kun at vise den øverste del af akserne kan man få stigninger og fald på en kurve til at vise dramatiske udsving. Blev den samme graf lavet i et koordinatsystem med et nulpunkt, ville udviklingen måske se ret ubetydelig ud. Helt umulig bliver en graf at gennemskue, hvis man fuldstændig undlader at angive enheder på akserne.

Man kan også beskære en graf, så den kun viser det, man vil have den til. Analyserer man for eksempel en bestand af lemninger, viser det sig, at antallet vokser og falder i en cyklus, der strækker sig over fire år. Sådant er det nu engang med lemninger, og det er der ikke noget unormalt i. Men beskærer man sin graf, kan kurven over bestandens udvikling tegne et dramatisk billede af, at bestanden af lemninger de seneste to år er i voldsom tilbagegang – eller dramatisk fremgang, alt efter hvad ens formål er.

### Vi misforstår procenter

Statistiske resultater bliver ofte fremlagt ved at bruge procenter, da det er en nem måde at beskrive en udvikling på. Men man skal aldrig glemme, at procenter kun giver mening, hvis man ved, hvilke tal der ligger bag. Det lyder for eksempel katastrofalt, når man får at vide, at ►

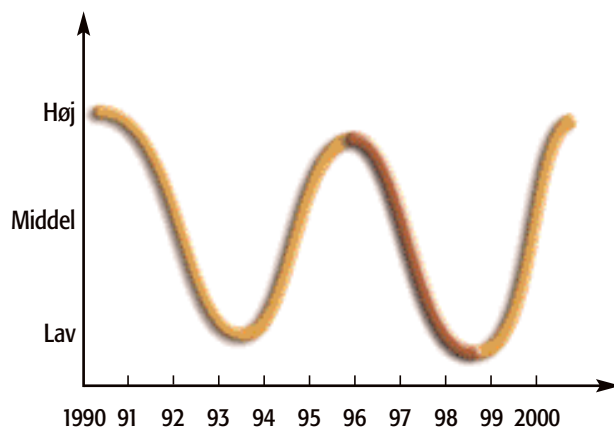
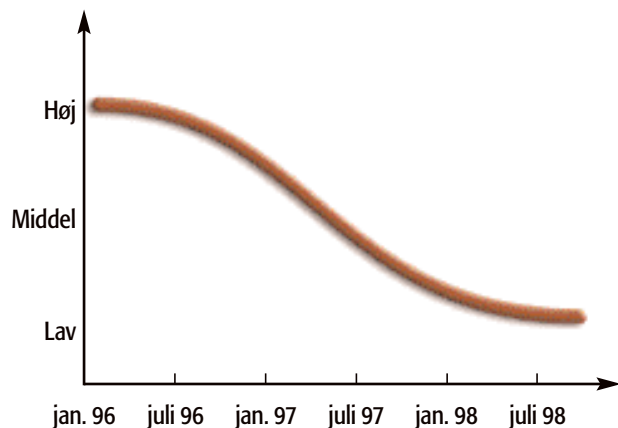
**Asiatiske kvinder bryder sig ikke om hvide mænd**  
En amerikansk undersøgelse viser, at kun 1,9 procent af hvide mænd har asiatiske kvinder som partnere. Det fik forskerne til at konkludere, at asiatiske kvinder går uden om hvide mænd. Nej! Der er blot mange flere hvide i USA end asiater. Faktisk deler hver anden asiatiske kvinde bord og seng med en hvid mand.

## Lemmingebestanden truet ...

## hvert fjerde år

Verden rundt svinger bestanden af lemninger i perioder på cirka fire år som vist her. Men en uærlig statistiker kan benytte sig af de regelmæssige udsving og nøjes med at vise en begrænset del af kurven. Pludselig er lemningerne i dramatisk tilbagegang.

CLAUS LUNAU



► hyppigheden af en dødelig sygdom er vokset med 150 procent på verdensplan i løbet af et år. Tallet blegner lidt, hvis det viser sig, at der på hele kloden er tale om en stigning fra fire til ti tilfælde.

Samtidig har overraskende mange mennesker svært ved at forstå procenter. For eksempel viser en tysk undersøgelse, at en tredjedel af alle tyskere ikke har tjekket på almindelige procenter. I undersøgelsen blev folk blandt andet bedt om at forklare, hvad "40 procent" ville sige. Nogle mente, at det betød en fjerdedel, mens andre skød på, at det var én ud af 40.

Man kan også få en fornemmelse af procenters upålidelighed via dette eksempel: En virksomhed lancerer sig selv som et progressivt firma, der lægger vægt på ligestilling. Virksomheden dokumenterer sin indstilling med en rask lille statistik:

Virksomheden har to afdelinger, som den kalder A og B. I løbet af det seneste år er andelen af kvinder i begge afdelinger vokset, i afdeling A fra 10 til 20 procent og i afdeling B fra 50 til 60 procent. Det lyder godt og ser pænt ud i virksomhedens præsentationsmateriale.

Men en emsig statistiker går nu tallene efter i sømmene og finder ud af følgende: Afdeling A havde sidste år 10 medarbejdere, heraf 1 kvinde (10 procent). Afdeling B havde 40 medarbejdere, heriblandt 20 kvinder (50 procent). I løbet af det seneste år har virksomheden haft fremgang, og der er blevet ansat flere folk.

Afdeling A har nu 100 medarbejdere, heraf 20 kvinder (20 procent), mens afdeling B har 50 medarbejdere, heriblandt 30 kvinder (60 procent). Kvoten af kvinder er ganske rigtigt steget i begge afdelinger. Alligevel viser det sig, at den samlede andel af kvinder i virksomheden er faldet! Fra 42 procent det første år (21 ud af 50 medarbejdere) til 33 procent det andet år (50 ud af 150 medarbejdere). Pludselig er der tale om en gammeldags og kvindefjendsk virksomhed.

Årsagen til dette overraskende resultat er, at afdeling A er vokset uforholdsmæssigt meget i forhold til afdeling B. Selv om der også er kommet flere kvinder i afdeling A, er der ansat så mange ekstra

mænd her, at det forrykker det samlede billede. Den ene kvinde, der var her i forvejen, talte så at sige for meget.

Det har ligefrem fået sit eget navn, når procenter opfører sig på denne særlige måde. Fænomenet kaldes Simpson-paradokset opkaldt efter en britisk statistiker.

### Spis is, og bliv rask

Ligesom der kan snydes med procenter, kan statistik også bruges til at påvise sammenhænge, der er det rene vrøvl. For eksempel kan det bevises rent statistisk, at man er mindre forkølet, hvis man spiser meget is. Tallene fortæller nemlig, at de fleste mennesker er mindre forkølede om sommeren, og ligeledes er forbruget af is højere om sommeren end om vinteren. Voila: Is er sundt!

At der ikke er nogen sammenhæng mellem is og snue, er nemt nok at gennemskue. Men det er ikke altid tilfældet. Det måtte en gruppe amerikanske forskere erkende, efter at de i 1999 bekendtgjorde, at børn, der sov med lyset tændt, var tilbøjelige til at blive nærsynede. Forskningsresultatet blev offentliggjort i det ansete videnskabelige magasin Nature, men historien holdt ikke. Det viste sig, at forældre, der selv er nærsynede, ofte får nærsynede børn. Og for at undgå at falde over legetøj, når de skulle kigge til deres afkom om aftenen, var de nærsynede forældre mere tilbøjelige end andre til at lade lyset brænde i barnekammeret.

### Flyvning er verdens sikreste transportform

Det er langt sikrere at flyve end at køre i bil, hører man hyppigt. Måske. Ser man på antallet af passagerkilometer, er der ganske rigtigt kun 0,3 dødsfald pr. milliard kilometer i fly, hvor der i bil er 4. Men kigger man i stedet på antallet af ture, er der pludselig 55 døde pr. milliard flyveture mod kun 45 pr. milliard udflugter i bil.

Derfor var der oftere tændt lys i nærsy- nede børns soveværelser end i deres kammeraters. Men det var ikke lyset, som gjorde deres syn ringere.

### Skotland vrimler med sorte får

Også i indsamlingen af data kan der gå kludder i tingene, ligesom der er rige muligheder for løgnagtighed. Specielt skal man være opmærksom på, om der er et rimeligt forhold mellem mængden af data og den præcision, man udtaler sig med.

Problemet illustreres tydeligt af en gammel anekdote:

En britisk politiker er i forbindelse med en valgkamp kommet til Skotland. Mens han sidder i toget, ser han nogle sorte får på en mark. Politikerens konstaterer, at alle får i Skotland må være sorte, og glæder sig over, at denne detalje kan bruges i en tale. Hans rådgiver, der er statistiker, bemærker, at det eneste, man kan sige med sikkerhed, er, at nogle får i Skotland er sorte på den ene side.

Det er ikke altid til at sige, hvor stor en datamængde der kræves. Men skal man udtale sig med sikkerhed, kræver det meget grundige observationer. Eksempelvis er der altid risiko for, at man har overset nogle få får i en fjern afkrog.

### Flere ben end gennemsnittet

Ligeledes skal man passe på med ren pseudostatistik. Det er oplysninger, der præsenteres som statistik, men i virkeligheden ikke er det. Det kan blandt andet forekomme i forbindelse med gennemsnit. For eksempel er det et faktum, at langt størstedelen af befolkningen her i landet har mere end det gennemsnitlige antal ben. Nogle få etbenede trækker snittet ned, og så har man et pudsig, men temmelig meningsløst udsagn.

Ligeledes blev det præsenteret som en statistisk kendsgerning, at Danmark var på vej mod at vinde VM i sommeren 2002, da det danske landshold havde slået de forsvarende franske mestre 2-0. Baggrunden var, at de sidste tre gange, Danmark og Frankrig havde mødt hinanden ved en slutrunde (EM 1992, VM 1998 og EM 2000), var det vindende hold endt som mestre. Som bekendt holdt den statistik ikke vand. Brasilien vandt VM.

Baggrunden er selvfølgelig, at resultatet af de indbyrdes opgør ikke har nogen – eller kun meget ringe – betydning for det endelige udfald. Selv om Danmark og

Billedet er udeladt a.h.t. copyright

E. DORNBACH/SCANPIX

**O.J. Simpsons forsvarer, Robert Shapiro (til venstre), anvendte statistik i sit vellykkede forsøg på at få O.J. frifundet for mord.**

## Statistik satte O.J. på fri fod

I den meget omtalte sag mod den amerikanske footballstjerne O.J. Simpson spillede statistik en stor rolle. Selv om ingen med sikkerhed kan sige, hvorfor juryen frifandt O.J. for anklagen om mord på sin kone, har kritikere påpeget den måde, som forsvareren benyttede statistik på. O.J.s forsvarer fortalte nævningene, at ud af 1000 mænd, der slår deres koner, er der kun én, der går så vidt, at han dræber hende. Det skulle give juryen det indtryk, at

der kun var én chance ud af 1000 for, at O.J. var skyldig.

Denne statistik blev ikke modgået af anklageren, men er på ingen måde et indicium på O.J.s uskyld. Omvendt kunne anklageren have påpeget, at når en kvinde, der har været udsat for hustruvold, bliver myrdet, er det i hvert andet tilfælde hendes mand, der er morderen. Vendt på denne måde var der altså 50 procents sandsynlighed for, at O.J. var gerningsmanden.

Frankrig havde mødt hinanden nok så mange gange med et bestemt resultat, ville det ikke sige meget om finalevinderen.

Det samme gør sig gældende for de systemer, der hævdes at kunne forudsige lottotal. Man kan i dag få computerprogrammer, der er konstrueret til at generere de mest sandsynlige vindertal i ugens lottospil, når de fodres med tidligere vindende kombinationer. Program-

merne kigger så på, hvilke tal der ikke er kommet ud længe, hvilke tal der ofte optræder sammen og så videre.

Problemet er blot, at den maskine, der trækker lottotalerne ikke ved noget om alt dette. Det er blot en maskine, der helt tilfældigt udvælger nogle små bolde. Så sandsynligheden for at vinde i lotto er stadig mindre end for at blive ramt af lynet. Og det er statistisk bevist! □