

STATISTIK

Samhällets naturlagar?

Några nedslag i statistikens historia

Det började som ett försök att övervaka befolkningens storlek, slavarnas antal och mängden elfenben i kolonierna. Men med tiden har den förändrat både vår syn på samhället och vetenskapen. **Henrik Serup Christensen** gör några nedslag i historien om statistikens inträde i den moderna världsbilden.

Statistik är som
en gatlykta:
Inte särskilt upplysande,
men **bra** att
stöda sig på.

Robert Storm Petersen
(Storm P, 1882-1949).



tära makt, detta på grund av införseln av värnplikt i flera länder. Som en följd etablerade de flesta länder nationella statistikkontor för att på detta sätt hålla reda på befolkningstrenderna. Så viktiga var dessa uppgifter att de länge var förbjudna att publicera. Först på 1800-talet blev det vanligt med offentligt tillgängliga statistiska register.

Samhälleliga lagar i siffror

Benämningen "statistik" om denna typ av data användes första gången av tysken **Gottfried Achenwall** (1719-1772). Men han syftade inte på sådana kvantitativa data som vi idag förbinder med begreppet utan på mera kvalitativa och beskrivande data. Som många andra på den tiden – däribland **Auguste Comte**, positivismens och sociologins fader – var Achenwall skeptisk till kvantifiering.

Trots motståndet mot kvantifiering blev de insamlade data i allt högre grad kvantitativa, precisa mätmeter var den tidens ideal inom vetenskapen. Samtidigt växte datamängden explosivt, vilket ledde till ett behov av probabilistiska verktyg för att kunna tolka de insamlade data. Då det insamlade datamaterialet på 1800-talet blev offentligt tillgängligt ökade intresset för att använda de stora datamängderna till att förklara och förstå samhället med hjälp av statistiska lagar. Enligt **Ian Hacking**, i sitt verk *The Taming of Chance* om statistikens idéhistoria,

Statistiken är ett relativt nytt fenomen som har utvecklats under några få århundraden. Även om statistikens historia som helhet bäst beskrivs som en långsträckt utveckling snarare än en avgränsad mängd avgörande genombrott är det ändå möjligt att peka ut några väsentliga händelser och namn.

Redan tidigare fanns statistiska register över väsentliga samhällsfenomen, men det var i förbindelse med den europeiska moderniteten från slutet av 1600-talet och framåt som insamlingen av statistiska data raffinerades. De europeiska stormakterna fick då behov av att hålla reda på sina kolonier och därför började staterna noggrant samla in data för att hålla reda på viktiga faktorer såsom befolkningsmängd, antal slavar och mängden elfenben.

Detta blev sedan också vanligt i kolonialmakterna, där insamlingen av data fick en väsentlig roll i statens försök att övervaka och registrera befolkningen för att kunna insamla skatt. Samtidigt började befolkningsstorleken uppfattas som en viktig indikator för en nations mili-

skedde vid denna tid ett skifte i synen på hur samhället fungerar och hur vi kan förklara och förstå omvärlden.

Vetenskapen kvantifierades

Idag kan det kännas självklart att det är möjligt att beskriva världen i probabilistiska termer, men det har inte alltid varit självklart att uppfatta världen på detta sätt. Först runt år 1840 blev det vanligt att uppfatta numeriska regelbundenheter som lagar för samhället, ett sätt att tänka på som var okänt några få decennier tidigare. Denna förändring sammanföll med en kvantifiering av vetenskapen, då precisa kvantitativa mått också började spela en central roll inom fysiken och kemien.

Data som insamlats många år i rad på

kortare än sina föräldrar. Galton insåg att förhållandet mellan föräldrarnas och barnens egenskaper var mera komplicerat än han först antagit.

Mot medelmåttighet

Detta ledde till den idé som i dag benämns "regression mot medelvärdet", i Galtons termer "regression mot medelmåttighet", som han utvecklade i slutet av 1800-talet. En konsekvens av detta var också idén om korrelationsbegreppet och därmed förutsättningen för regressionsanalys, som gör det möjligt att analysera samband mellan variabler. Denna metod hör ännu idag till de mest använda statistiska metoderna, även om den med tiden har raffinerats och vidareutvecklats.

ser. Men vissa fenomen är svåra att mäta precis, vilket medför att de antingen utesluts från analyser eller inkluderas, men mäts på ett otillfredsställande sätt. Statistiska analyser ger alltså inte objektiva svar på frågor, utan är alltid till en viss del beroende av forskningens kytimer och forskarens förförståelse. Dessa förbehåll lyfts sällan fram när resultaten publiceras i vetenskapliga tidskrifter, och inte alls när de populariseras i medierna, där probabilistiska samband behäftade med en massa om och men presenteras som kausala mekanismer.

Att det finns utrymme för skepsis innebär inte nödvändigtvis att de statistiska analyserna ger felaktiga konklusioner, men innebär en högre grad av osäkerhet. Det är värt att komma ihåg att statistik enbart är ett redskap i försöket på att uppnå klarhet: det kan brukas och missbrukas. Statistik ger inte alla svar, men kan vara till hjälp i försöket på att uppnå dem.



Henrik Serup Christensen
är doktorand i statsvetenskap
vid Åbo Akademi

.....

”Idag kan det kännas självklart att det är möjligt att beskriva världen i probablilistiska termer, men det har inte alltid varit självklart att uppfatta världen på detta sätt.”

.....

enskilda fenomen hade uppvisat en förvånande nivå av regularitet, och som en följd började regelbundenheterna uppfattas som konstanta samhälleliga lagar. Det fanns plötsligt en "naturlig" nivå av sjukdom, död och kriminalitet som gjorde det möjligt att beskriva skillnader mellan nationer i termer av statistiska skillnader i till exempel den genomsnittliga livslängden.

Den genomsnittliga mannen

Adolphe Quetelet började på 1830-talet använda det som senare har blivit känt som normalfördelningen för att beskriva sociala fenomen med sin konstruktion av "den genomsnittliga mannen" baserad på en rad beskrivningar av de kännetecken som utmärker befolkningen i genomsnitt. Även om hans metoder var bristfälliga, blev den genomsnittliga mannen för Quetelet ett ideal för hur människan borde vara – avvikelser från det normala var onaturliga och måste botas. Detta medan andra var mera intresserade av ytterligheterna och hur exceptionella egenskaper kunde nedärvas.

Francis Galton (1822-1911) – halvkusin till **Charles Darwin** – vidareutvecklade Quetelets idéer. Galton var intresserad av biologisk ärftlighet och förknippas ofta med uppkomsten av tanken om rashygiene, och han var övertygad om att de flesta mänskliga egenskaperna var ärftliga. Han undersökte därför sambandet mellan föräldrars och barns egenskaper med avseende på en rad olika faktorer. Men även om det fanns ett visst samband fanns det också till exempel en tendens till att barnen till extremt högväxta individer närmade sig medelvärdet, eftersom de var

De statistiska metoderna utvecklades i ett raskt tempo på 1900-talet och har gjort det möjligt att använda statistiken till att belysa en rad olika fenomen. De statistiska teknikerna vidareutvecklades sedan av bland annat Galtons lärjunge **Karl Pearson**, som innehade den första professuren i matematisk statistik och utvecklade väsentliga statistiska verktyg som den *multivariata statistiska analysen* och *chi2-testet*.

Han bidrog därmed också till att utveckla möjligheten för statistisk inferens, när sannolikheter används till att beskriva okända enheter utgående från de data som finns tillgängliga, med andra ord att på basis av ett urval kunna uttala sig om en population. Här har normalfördelningen också en central roll, men den har blivit kompletterad med andra statistiska fördelningar, var dataobservationerna inte grupperar sig jämnt runt medelvärdet.

Statistik och sund skepsis

I dag får vi dagligen ta del av nya "statistiska upplysningar" i dagstidningen. Datorns frammarsch har gjort det lättare att producera och sprida statistik, vilket medverkar till en lavin av statistiska upplysningar, i vilken dagens sanningar imorgon motsägs av helt nya forskningsrön.

Det finns fog för en viss skepsis gentemot de många undersökningar som på statistiska grunder hävdar nya forskningsrön som motsäger tidigare resultat. Resultaten av alla undersökningar baserade på statistiska uträkningar är beroende av vilka variabler som inkluderas, hur de mäts, hur sammanhang mellan variablerna modelleras, och vilka antagningar som görs om deras frekven-

LITTERATUR

Ian Hacking *The Taming of Chance* ger en bra introduktion i statistikens historia. För en historisk översikt över de väsentligaste matematiska utvecklingarna inom statistiken, se **Anders Halds** *A History of Mathematical Statistics From 1750 to 1930*. För en mera populärvetenskaplig introduktion, se **Philip Balls** *Critical Mass*, som trots vissa problematiska slutsatser är underhållande.

ORDET: Normalfördelningen

- En central statistisk idé är att en rad sociala fenomen i tillräckligt stora antal tenderar att gruppera sig runt medelvärdet, medan betydliga avvikelser från detta är sällsynta.
- Normalfördelningen, som denna idé kallas, involverade flera av tidens framträdande vetenskapsmän såsom den tyska matematikern **Johann Carl Friedrich Gauss** (1777-1855), och fransmannen **Pierre Simon de Laplace** (1749-1827).
- Illustreras ofta med exemplet att singla slant. Om du gör detta tio gånger kan det hända att det blir 9 kronor och 1 klave, men om du gör det 1000 gånger blir fördelningen mellan de två jämnare, relativt sett. Med andra ord kan man uppnå en viss regelbundenhet genom att upprepa ett försök tillräckligt många gånger.