

AVTAGANDE KORRELATIONER

En av grundprinciperna inom vetenskaplig forskning är att ett forskningsrön håller måttet om experimenten det baserar sig på kan upprepas med samma resultat av flera oberoende forskare. Kim Berts diskuterar ett problematiskt fenomen inom statistiska undersökningar, där de ursprungliga resultaten i och för sig bekräftas i senare undersökningar, men där sambanden blir allt svagare ju mer man undersöker fenomenet.

Text: Kim Berts

För några år sedan råkade jag läsa, antagligen i en dagstidning eller något populärvetenskapligt magasin, att forskare noterat att honor hos olika djurarter föredrar hanar med symmetriska drag. Det kunde till exempel röra sig om fågelhonor som föredrog hanar med symmetriska stjärtfjädrar. Man tänkte att detta var ett sätt för honorna att urskilja vilka hanar som hade goda gener och därför lönade sig att para sig med. Det är naturligtvis en annan femma att kunna förklara i detalj varför djuren uppvisar detta beteende, men att man med statistiska metoder lyckats påvisa en korrelation slog mig inte som problematiskt i sig.

Enligt artikeln "The Truth Wears Off" (*The New Yorker* 13.12.2010), skriven av **Jonah Lehrer**, har sambandet mellan symmetri och framgång hos honor vid senare undersökningar avklingat. Den ursprungliga undersökningen handlade om ladasvalors stjärtfjädrar och publicerades i början av 1990-talet. Den inspirerade många andra liknande studier vilka fann samma koppling mellan symmetri och framgång vid fortplantning. Men, redan 1997 kunde bara en tredjedel av undersökningarna uppvisa någon korrelation och det samband som kunde påvisas i denna tredjedel var dessutom svagare än i de första studierna.

Detta fenomen har också drabbat många andra statistiska undersökningar, exempelvis effektiviteten hos olika typer av medicinering. Det samband man först fann, och som ofta bekräftats i många oberoende studier, har visat sig vara svagare i senare undersökningar, om det ens har syns alls. Det märkliga är att det inte rör sig om uppenbara misstag i den första undersökningen – dess resultat har som sagt kunnat upprepas i andra undersökningar – problemet är att själva sambandet förefaller avklinga ju mer det undersöks. Fenomenet är nytt och har ännu inte något

etablerat namn, skriver Lehrer, men han använder benämningen *the decline effect*. Benämningen brukar användas om den uteblivna bekräftelse som är typisk när man försöker upprepa en undersökning som till synes ger stöd åt någon form av pseudovetenskapligt fenomen, exempelvis förmågan att hitta vatten med hjälp av slagruta.

Lehrers artikel har väckt stor uppmärksamhet och detta hänger säkert delvis ihop med hans antydning om bristande vetenskaplighet i de studier som drabbas. Ett kriterium för att en vetenskaplig upptäckt ska anses hålla måttet är att en motsvarande undersökning vid ett annat tillfälle eller plats ska

ge samma resultat ifall förhållandena i övrigt är jämförbara. Det problematiska med denna typ av avtagande är att just möjligheten att upprepa undersökningen med samma resultat inte uppfylls. Eller, den uppfylls kanske till en början för att sedan avta, vilket gör det svårt att veta hur man ska ställa sig till upptäckterna.

Betyder detta att gränsen mellan pseudovetenskap och vetenskap suddas ut? Lehrer skriver att fenomenet visar hur svårt det är att bevisa något. Det är

Det märkliga är att det inte rör sig om uppenbara misstag i den första undersökningen – dess resultat har som sagt kunnat upprepas i andra undersökningar – problemet är att själva sambandet förefaller avklinga ju mer det undersöks.

kanske lätt att drabbas av cynism och tänka att vi inte kan veta något med säkerhet, eller att detta kastar tvivel över all vetenskaplig kunskap eller något liknande. Men måste man det? Fenomenet förefaller vara kopplat till statistiska undersökningar i huvudsak. Dessa tar fasta på statistiskt signifikanta avvikelser inom ett sampel eller på korrelationer mellan två variabler, och säger som bekant inget om vad som orsakar avvikelserna eller om att det ena fenomenet orsakar det andra. Eftersom fenomenet med avtagande bara kan drabba statistiska under-

sökningar går den vetenskapliga kunskap, som härrör ur andra typer av undersökningar, fri från fenomenet.

Man brukar ibland säga att man kan bevisa vad som helst med statistik och kanske har vi i det här fenomenet ett exempel på bedrägligheten hos statistiska resultat. Att resultat av statistiska undersökningar kan vara svåra att tolka och därför ibland är missvisande är välkänt och syns bland annat när dagstidningar lyfter fram ett statistiskt samband på ett sätt som låter påskina att man upptäckt ett orsakssamband. En i detta fall viktigare fråga är om avtagandet kan förklaras med att undersökningarna på något oförutsett vis var bristfälliga – eller om också en undersökning utförd enligt god praxis, t.ex. beträffande urval av sampel, är opålitlig i sig. Annorlunda uttryckt: är statistik bedrägligt på grund av att det är ovanligt

Annorlunda uttryckt: är statistik bedrägligt på grund av att det är ovanligt lätt att göra fel när man planerar och genomför undersökningar eller på grund att resultaten är mindre tillförlitliga än andra typer av resultat?

lätt att göra fel när man planerar och genomför undersökningar eller på grund att resultaten är mindre tillförlitliga än andra typer av resultat? Det kan ju i varje fall knappast vara så djur år 1991 föredrog symmetri hos sina partners men att detta hann gå ur mode mot slutet av 1990-talet. (En tänkbar möjlighet är dock att denna typ av preferens bara gäller vissa djurarter och kroppsdelar men inte andra och att liknande studier på andra djurarter eller kroppsdelar därför inte finner samma koppling.)

Ett exempel på en fallgröp i genomförandet av statistiska undersökningar som Lehrer nämner och som delvis kunde förklara avtagandet är att de fenomen man undersöker kan vara svåra att kvantifiera korrekt. I synnerhet när det gäller medicinens verkan kan det vara svårt att jämföra vad som har bättre effekt. Denna typ av snedvridning av ett resultat är något som blivande forskare varnas för redan i början av sina studier, men problemet tycks ändå svårt att rå på, särskilt i de fall där det finns ekonomiska eller andra intressen i att få ett visst resultat.

Ett annat problem som också kan ha sin andel är att forskare tenderar att publicera rapporter om funna korrelationer och avvikelser, men inte om uteblivna. Det samma gäller också tidskrifternas redaktioner när de väljer vilka artiklar

som trycks. Detta kan ge upphov till snedvridet intryck trots att undersökningarna är korrekta utförda.

Kan det till och med vara så att vissa fynd är rent slumpmässiga, att det råkar vara så att ett visst sampel uppvisar någon egenskap i långt högre grad än vad som förekommer i verkligheten? Det kan tyckas lustigt att statistik som handlar om att försöka utesluta slumpens makt över det enskilda fallet genom att koncentrera sig på ett stort antal fall, ändå är utsatt för sådant. Man måste då komma ihåg att ifall man i en studie finner en statistiskt signifikant avvikelse från det förutspådda så betyder det inte att man funnit något som ofelbart gäller i allmänhet utan enbart att sannolikheten för att avvikelsen ska vara ett resultat av slumpen är mindre än en den givna konfidensen. Ett förslag till förklaring av resultatens avtagande är just att de "funna" fenomen som drabbats bara var slumpmässiga egenheter i samplen och att de därför avtar för att till sist avspeglar verkligheten. Mot detta talar förstås att upptäckterna faktiskt upprepats av andra vilket vore mycket osannolikt om det bara handlade om en slumpmässig avvikelse.

Lehrer hävdar i varje fall att inga av dessa problem ensamma kan förklara fenomenet. Det kan förvisso stämma att inget av de nämnda problemen som föreslagits ensamt kan förklara fenomenet med avtagande resultat, men kanske de sammantaget kan göra det? Lehrer tenderar att falla in i den cynism som nämndes ovan och se själva den vetenskapliga metoden som bristfällig. Dessutom avslutar han med att misstänkliggöra empiriska undersökningar i allmänhet och skriver att experiment inte nödvändigtvis leder oss till sanningen. Men eftersom det fenomen han diskuterar hänger ihop med statistiska undersökningar förefaller denna

generella skepticism angående empirisk vetenskap malplacerad. Ifall däremot resultatens avtagande inte är en följd av brister hos den vetenskapliga metoden utan tvärtom en följd av brister i undersökningarna, i till exempel sättet som vetenskap offentliggörs, så vore ett annat förhållningssätt till *the decline effect* naturligare. Man kunde se den inte som ett problem utan som ett sundhetstecken: de resultat som inte håller måttet klingar av och principen om möjligheten att upprepa undersökningen och få samma resultat gör sitt jobb.

Det positiva med artikeln är att den visar att det finns utrymme för en kritisk blick trots att ett resultat har blivit bekräftat av oberoende undersökningar och passerat det vetenskapliga samfundets fackgranskning (peer-review).



Kim Berts är doktorand i filosofi vid Åbo Akademi