

Svar till Göran Nilsson

Först en terminologisk fråga: Vid hypotestestning har man att göra med två hypoteser. Dels den vetenskapligt uppställda hypotesen, t ex: regelbundet intag av denna medicin minskar frekvensen av hjärtinfarkt. Dels nollhypotesen, i detta fall en hypotes enligt vilken medicinen i fråga inte har någon sådan effekt. Med fraser som "den testade hypotesen" menar man i gängse språkbruk den vetenskapligt uppställda hypotesen. I den vetenskapsfilosofiska diskussionen om verifiering och falsifiering av hypoteser använder man också detta språkbruk. Den diskussionen handlar alltså om verifiering och falsifiering av vetenskapligt uppställda hypoteser, inte av nollhypoteser. I min artikel anslöt jag mig till detta allmänna språkbruk.

Det finns emellertid också ett mera tekniskt språkbruk, där man med "den testade hypotesen" menar nollhypotesen. Detta språkbruk är ändamålsenligt i statistiska sammanhang, och jag har inga invändningar mot det. Göran Nilsson använder detta språkbruk, vilket förvisso är helt i sin ordning. Däremot blir det förvirrande när han tolkar mig som om jag också gjorde det.

Han har kommentarer dels om mitt tärningsexempel, dels om vart och ett av de fyra missförstånd jag tog upp.

*Tärningsexemplet:* Inledningsvis använde jag ett exempel där den uppställda hypotesen är att en tärning alltid ger en sexa. Nollhypotesen är i detta exempel uppenbart att frekvensen sexor är  $1/6$ . Nilsson tolkar mig emellertid som om jag skulle mena att nollhypotesen är att tärningen alltid ger sexa. Detta beror på en trivial feltolkning av det jag kallade "vår hypotes" som skulle testas. Jag följde som sagt här det allmänna, inte det tekniska språkbruket.

Jag använde tärningsexemplet bl a för att visa hur absurda resultat man kan få om man uppställer sin hypotes först efter att testet har genomförts. Det är svårt att se hur Nilsson kan få intrycket att "Hansson väljer det kritiska området sedan han sett resultatet". Jag gav exempel på felaktiga resonemang där detta misstag görs, och framhöll att de är felaktiga.

*Missförstånd 1:* Det Nilsson kallar "en vänlig tolkning" är såvitt jag kan se den enda tolkning som är förenlig med vad jag skrev.

*Missförstånd 2:* Nilsson talar om vad han kallar min "märkliga uppfattning att resultat som stämmer exakt med den uppställda hypotesen är signifikanta". Något sådant har jag inte sagt och det är obegripligt för mig hur denna misstolkning kan uppstå. Det finns gott om exempel på utfall som stämmer med den uppställda hypotesen, men där ett test gentemot en adekvat nollhypotes inte ger signifikant utfall. Jag har inte lyckats lista ut vad han åsyftar med "förvirrat resonemang", men kanske är det återigen den ovannämnda termologiska förbistringen som spelat ett spratt.

*Missförstånd 3:* Möjligen finns här en åsiktsskillnad mellan Nilsson och mig. Antag att ett medium säger sig ha förmågan fastställa könet på en person bakom ett skynke i minst 75 % av fallen. Enligt min mening bör ett test utformas så att mediet faktiskt måste lyckas i minst 75 % av fallen för att testet ska anses ge stöd åt hypotesen att hon har denna förmåga. Härigenom möjliggörs också, som jag skrev i artikeln, att man använder låga kritiska  $p$ -värden gentemot en nollhypotes om rent slumpmässigt utfall. I sin kommentar har Lars Ångquist rätt ut detta betydligt mera grundligt än vad jag gjorde i min ursprungliga artikel.

Mitt förslag innebär att en person som säger sig ha en sådan förmåga avkrävs ett besked om förmågans storlek, innan den kan testas. Alternativet är att räkna alla signifikanta avvikelser (åt "rätt" håll) från nollhypotesen (dvs hypotesen att försökspersonen "enbart gissar") som stöd för hypotesen att förmågan finns. Detta leder enligt min mening fel eftersom gängse praxis för valet av kritiska  $p$ -värden (såsom 0,05) förutsätter hypoteser med avsevärt högre vetenskaplig trovärdighet.

I en tänkt värld där man kan göra experiment helt utan felkällor vore det intressant om försökspersonen överträffar slumpen med en träffsannolikhet som är statistiskt signifikant skild från nollhypotesen, även om skillnaden är liten. Men i praktiken har vi alltid felkällor, inte minst i parapsykologiska experiment. En trolig förklaring till små avvikelser från nollhypotesen är därför att försöket påverkats av (kända eller okända) felkällor. Om däremot försökspersonen före experimentet uppger sig kunna uppnå 75 % andel korrekta träffar, och klarar av att demonstrera den förmågan på ett statistiskt övertygande sätt, då vore det utfallet mycket intressant och i stort behov av ytterligare forskning och förklaringsförsök. Som alltid: Man måste se både till effektens storlek och till den statistiska signifikansen.

*Missförstånd 4:* Jag nämnde här som exempel hypotesen att en ond ande som bor i en tärning påverkar utfallet så att den ger en sexa endast i en åttondel av kasten. Detta är förvisso en hypotes (i ordets allmänna bemärkelse). Dessutom är det en hypotes som kan testas. Om vi kastar tärningen 8000 gånger och får 1330 sexor har vi starka skäl att förkasta hypotesen. Men, som jag skrev i artikeln, om vi får 1000 sexor av 8000 har vi ändå inte fått nämnvärt stöd för ande-hypotesen eftersom det finns ett alternativ och mycket trovärdigare hypotes (skev massfördelning) som också får stöd av detta test.

"En ande förorsakar att det blir sexa i en åttondel av kasten" och "skev massfördelning förorsakar att det blir sexa i en åttondel av kasten" är två olika hypoteser som emellertid inte kan särskiljas genom test där man kastar tärningen upprepade gånger.

Nilsson föredrar ett annat språkbruk enligt vilket ingen av dessa utsagor är hypoteser, utan det är bara "det blir sexa i en åttondel av kasten" som är en hypotes. Detta uttrycker han så att jag "blandar ihop" en statistiskt formulerad hypotes med en förklaring till hypotesen. Men jag vill varna mot det språkbruk som han här förordar. I det allmänvetenskapliga språket är en hypotes ett "antagande som göres

för att förklara visst faktum eller vissa fakta" (SAOB). Steget från en vetenskaplig hypotes till en precisering som fungerar i statistisk hypotesprövning är grannlaga och ofta betydligt svårare än i det här exemplet. Men inget är vunnet med att inskränka innebörden av ordet "hypotes" så att det inte omfattar vad man i forskningen i allmänhet menar med ordet.

Sven Ove Hansson