Kommentarer till Sven Ove Hanssons artikel *Signifikansmissbruket i pseudovetenskap* i nr 3 av Qvintensen 2010

Göran Nilsson

Sven Ove Hansson använder sig av tärningskast som exempel. Han antar först att hypotesen vi ska testa, nollhypotesen, är att tärningen alltid ger en sexa, dvs. att sannolikheten för sexa i ett kast är 1. Ett försök med tre kast, som alla gett en sexa, betraktar Hansson nu som signifikant eftersom sannolikheten för detta utfall bara är ungefär 0,005 under antagande av att alla sidor är lika sannolika, men detta var ju inte den hypotes som skulle testas. Om nollhypotesen är sann är ju tre sexor just vad vi kan förvänta oss. Det kritiska utfallet, oberoende av antalet kast som ingår i försöket, är naturligtvis minst ett kast, som inte ger en sexa. Men som Hansson påpekar verkar tre kast vara i minsta laget. För att bedöma hur många kast som är rimliga att göra måste vi beakta sannolikheten för fel av andra slaget, dvs. sannolikheten att acceptera en felaktig hypotes. Denna sannolikhet beror på hur fel hypotesen är. Hansson har beräknat denna sannolikhet till ungefär 0,005 (approximation av $1/6^3$) då sannolikheten för sexa är 1/6. Man kan naturligtvis beräkna sannolikheten för fel av andra slaget för alla tänkbara värden på sannolikheten att få en sexa och kan då ange sannolikheten att acceptera hypotesen som funktion av det sanna värdet att få en sexa. Vanligen anger man i stället sannolikheten att förkasta hypotesen som funktion av sanna värdet. Denna funktion kallas styrkefunktionen för testet och kan användas då man vill bestämma ett lämpligt antal kast i försöket. Antag att vi vill ha en sannolikhet på minst 0,9 att förkasta hypotesen om den verkliga sannolikheten för en sexa är 0,8. Vi ska då bestämma antalet kast n så att $0{,}8^n < 0{,}1$, vilket ger $n = 11$.

Om Hansson i stället menar att nollhypotesen är att alla sidor på tärningen är lika sannolika borde de kritiska utfallen rimligen även ha innefattat åtminstone 3 ettor, 3 tvåor osv. Med denna nollhypotes får man intrycket av att Hansson väljer det kritiska området sedan han sett resultatet. Om vi inte förutsätter att resultatet i varje kast är helt slumpmässigt, oberoende av tidigare utfall, utan vill testa även detta, blir det knepigt att bestämma det kritiska utfallsområdet. Vi måste då bestämma när resultaten i en sekvens av kast ska betraktas som slumpmässiga och det är ett problem som brytt många skarpa hjärnor.

Hansson tar sedan upp vad han kallar fem vanliga missförstånd om statistisk signifikans.

Missförstånd 1: Ett uttalade om signifikans behöver inte hänvisa till en i förväg gjord förutsägelse.

En vänlig tolkning av denna formulering är att man inte ska bestämma kritiska utfall först och sedan formulera en nollhypotes som gör just dessa utfall osannolika. Man kan naturligtvis inte låta det kritiska utfallsområdet vara just det utfall man fått. Om detta är vad Hansson menar så har han rätt.

Missförstånd 2: Statistisk signifikans mäter sannolikheten att det man iakttar beror på slumpen.

Hansson fortsätter med tärningsexemplet och sin märkliga uppfattning att resultat som stämmer exakt med den uppställda hypotesen är signifikanta. Efter ett förvirrat resonemang påpekar han i alla fall att signifikansen är sannolikheten att förkasta en sann hypotes.

Missförstånd 3: Signifikant avvikelse från nollhypotesen är ett adekvat mått på graden av vetenskaplig bevisning.

Här tar Hansson ett exempel där det gäller förmågan att avgöra könet på en person som sitter bakom ett skynke och argumenterar helt riktigt för att man bör ställa en hypotes om träffprocenten före ett försök men resonemanget är ur statistisk synpunkt lika förvirrande som tidigare. Om en försöksperson påstår sig kunna förutsäga kön på en person bakom ett skynke med en träffprocent på i genomsnitt 75 % rätt tycker Hansson att det är juste att ha just 75 % rätt som beslutsgräns då man genomför ett visst antal försök. Sannolikheten att förkasta en nollhypotes om en viss träffsannolikhet (oberoende av vilken) då den verkliga träffsannolikheten är just 75 % är naturligtvis då i runda slängar 50 % och det verkar väl inte särskilt juste (skulle träffsannolikheten vara 75 % kan en nollhypotes lika gärna testas genom slantsingling). Det vanliga statistiska synsättet är väl snarast följande: Vår nollhypotes är att det inte finns någon förmåga att avgöra könet utan att det sker helt slumpmässigt, dvs. sannolikheten för rätt könsangivelse är 50 %. För ett givet antal försök kan vi beräkna en kritisk gräns för den erhållna träffprocenten så att vi får t ex högst 5 % sannolikhet att få ett större värde om hypotesen är sann (ensidigt test). Om försökspersonen tror sig ha en träffsannolikhet på 75 % är det rimligt att bestämma antalet försök så att sannolikheten att förkasta hypotesen (styrkan) vid detta värde är relativt stor, t ex 90 %. Den kritiska gränsen kommer då att vara ett värde mellan 50 och 75 %. Bara för att statistiska metoder används för att avslöja humbug finns det inget skäl till att tillämpa andra synsätt.

Hanssons resonemang kring förmågan att avgöra kön tycker jag inte har så stor koppling till formuleringen av missförståndet. Jag vill i stället belysa det på följande sätt: För det erhållna resultatet beräknar man sannolikheten att man skulle få så stor avvikelse eller större från nollhypotesen och kallar detta för erhållet *p*-värde. Vi har nu två resultat att beakta: den erhållna träffprocenten och det erhållna *p*-värdet. Naturligtvis är det så att man känner sig säkrare på att man gör rätt när man förkastar nollhypotesen ju mindre *p* är. I den meningen skulle man kunna säga att *p*-värdet är ett mått på styrkan i den vetenskapliga bevisningen. Enbart *p*-värdet säger dock inget om betydelsen (storleken) av avvikelsen från nollhypotesen. Ibland verkar det dock som om man var mer intresserad av *p*-värdet än den faktiska storleken på avvikelsen.

Missförstånd 4: Ett höggradigt signifikant resultat ger ett starkt stöd åt den testade hypotesen.

Ett höggradigt signifikant resultat ger snarast ett starkt stöd för att förkasta den testade hypotesen men Hansson anser att påståendet är felaktigt eftersom vi sällan direkt testar en hypotes. Naturligtvis testar man en hypotes vid statistisk hypotesprövning, men hypotesen måste vara formulerad i statistiska termer såsom sannolikhet för ett visst utfall eller medelvärde i en population. Hansson blandar ihop en statistiskt formulerad hypotes med mer eller mindre spektakulära förklaringar till att hypotesen skulle vara sann, t.ex. en ond ande som bor i en tärning. Sådana förklaringar kan naturligtvis inte testas statistiskt. Olika förklaringar, som ger samma statistiskt formulerade hypotes, kan självklart inte särskiljas med statistisk hypotesprövning.