

Statistisk signifikans

Den sjunde artikeln i vår serie om vetenskapens kunskapssyn handlar om statistisk signifikans.

En vanlig frågeställning i vetenskaplig hypotesprövning är om ett experimentellt resultat verkligen skall tillmätas någon betydelse eller om det "bara beror på slumpen". Det vanligaste hjälpmedlet för att avgöra detta är den *statistiska signifikansen*. För att klargöra vad detta innebär kan vi ta följande exempel:

En forskare vill undersöka om saltsyraångor ger upphov till magsår hos råttor. Han utsätter 30 råttor (försöksgruppen) för saltsyraångor under ett års tid. En kontrollgrupp om 30 råttor slipper saltsyraångorna

men får i övrigt samma levnadsvillkor som försöksgruppen. Efter ett år, när råttorna obduceras, visar det sig att 7 av försöksgruppens råttor har klara tecken på magsår mot 5 i kontrollgruppen.

Ett sådant resultat skulle inte ge något stöd för hypotesen att saltsyra ger upphov till magsår. Skillnaden är ju så liten att den lika väl kan bero på slumpen. Om däremot 20 av försöksgruppens råttor hade fått magsår mot endast 5 i kontrollgruppen, hade försöket gett stöd åt hypotesen att saltsyraångor ger magsår. Detta förefaller uppenbart.

Men hur skulle man ha resonerat om 10 av försöksgruppens råttor hade fått magsår? Eller 15? Uppenbarligen räcker inte vår statistiska intuition här. Det

behövs någon form av regler för att avgöra om slumpvariationer är en rimlig förklaring till ett resultat.

Det gängse synsättet är att man skall ställa upp en *nollhypotes*. Därmed avses hypotesen att skillnaden mellan betingelserna för de båda grupperna (här saltsyran) inte har någon betydelse för det man mäter (här förekomsten av magsår). Man räknar sedan ut hur stor sannolikheten skulle vara, om nollhypotesen vore sann, att få minst så stora skillnader

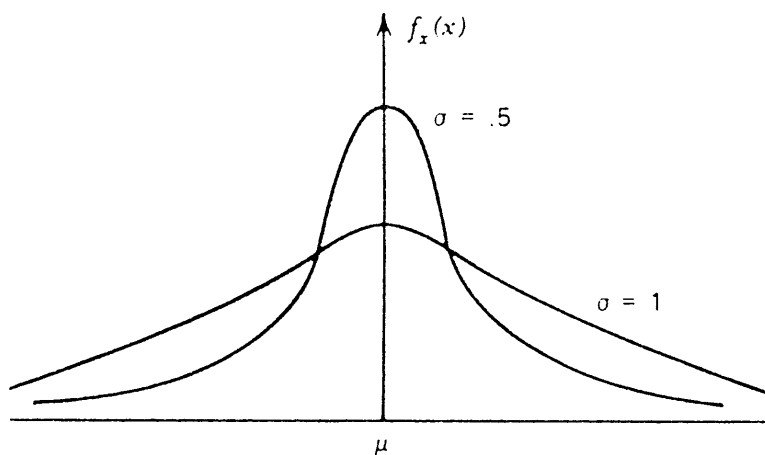
mellan grupperna som man iakttagit.

I vårt exempel med de två grupperna om 30 djur är sannolikheten 37 % att få minst en sådan skillnad som den mellan 5 och 7; förutsatt att nollhypotesen är sann, dvs att saltsyran inte har

ingen effekt på uppkomsten av magsår. Den är 12 % för att få minst en sådan skillnad som mellan 5 och 10, 1 % att få minst en sådan skillnad som mellan 5 och 15 och 0,01 % att få minst en sådan skillnad som mellan 5 och 20.

Dessa sannolikheter (sannolikheten för effekten, givet nollhypotesen) anger den statistiska signifikansen i de olika fallen. Det är praxis att dra en gräns vid 5 %. Om sannolikheten för effekten, givet nollhypotesen, är mindre än 5 % så är resultatet statistiskt signifikant, annars inte. I vårt fall betyder detta att ett resultat om 7 eller 10 magsårsfall i försöksgruppen inte är signifikant, medan 15 eller 20 skulle vara ett signifikant resultat.

Gränsen vid 5 % är i viss mening godtycklig. Om man sänker denna gräns, ökar man risken att bortse från verkliga effekt-



er. Om man höjer gränsen, ökar man i stället risken att tolka slumpeffekter som regelbundenheter hos naturen. Signifikansgränsen kan därför ses som en avvägning mellan nackdelarna med felaktig kunskap och nackdelarna med ingen kunskap alls.

Självfallet är denna avvägning i högsta grad värderingsberoende. Det framkommer särskilt tydligt när bedömningarna skall användas som underlag för samhällsbeslut. Om man t ex vill vara på den säkra sidan vid bedömningen av kemiska risker, kanske man är betydligt mera angelägen att inte missa verkliga effekter än att inte misstolka slumpeffekter som verkliga. Då kan 5 %-gränsen framstå som alltför låg. Dessutom kan helt andra statistiska mått bli aktuella vid en sådan bedömning.

Även för de rent inomvetenskapliga bedömningarna gäller att signifikanstal och andra statistiska kalkyler bara kan ge en mycket grov vägledning. Först och främst säger signifikansen inget om hur pass väl experimentet är utfört. En aldrig så stark statistisk signifikans kan aldrig kompensera brister i det praktiska utförandet. Ur ett dåligt utfört experiment kan man inte dra några säkra slutsatser, oavsett vilken signifikans som uppnås.

Därtill kommer att signifikansmättet inte tar någon hänsyn till att man måste bedöma teorier mot bakgrund av tidigare gjorda observationer. Om ett enstaka experiment skulle ge signifikanta resultat som strider mot tyngdlagen, räcker detta inte för att kullkasta tyngdlagen, eftersom så mycken annan, tidigare insamlad bevisning har bekräftat tyngdlagen.

I regel finns många olika sätt att bearbeta experimentella data. Om man provar tillräckligt många olika sätt att signifikans-testa sitt material, så hittar man i regel åtminstone något samband som är statist-

iskt signifikant. Möjligheterna att leta efter sådana samband har ökat kraftigt de senaste åren genom tillkomsten av behändiga data-program för statistikkalkyler.

Men signifikanser som tillkommit på det sättet har mycket ringa vetenskapligt värde. Signifikansprövning är en metod för att testa i förväg uppställda hypoteser som har ett stöd i befintliga teorier eller i tidigare gjorda observationer. Den måste genomföras efter i förväg fastställda riktlinjer för den statistiska bearbetningen. Det är endast i ett sådant sammanhang som den 5 %-iga signifikansgränsen är rimlig och användbar. Att tillämpa den på hypoteser som ställts upp i efterhand är ett farligt missbruk som lätt kan leda in forskaren på fel vägar.

Sven Ove Hansson