

Statistikens vänner och kritiker

Olle Häggström

Ibland drabbas man som akademiker och forskare av att någon utanför den egna ämnesspecialiteten har synpunkter på ens ämne. Ett sätt att då reagera är med taggarna utåt: vad ger dig, som väl knappt ens läst mitt ämne på elementäraste grundkursnivå, rätt att yttra dig? Ett annat är att glädjas åt att det finns andra än de närmast sörjande som intresserar sig för ämnet, och ta del av synpunkterna med öppet och konstruktivt sinnelag.

Läsaren kanske redan anar vilket av dessa förhållningssätt jag föredrar. Frågan aktualiseras av två uppmärksammade och relativt färska böcker där icke-statistiker har kritiska synpunkter på statistikens teori och (framför allt) praktik: *The Black Swan* av finansmäklaren och filosofen Nassim Nicholas Taleb, och *The Cult of Statistical Significance: How the Standard Error Costs us Jobs, Justice and Lives* av nationalekonomerna Stephen Ziliak och Deirdre McCloskey. Ingen av böckerna är invändningsfria, men båda har mycket intressant att säga. De förtjänar att läsas av var och en som intresserar sig för statistikens tillämpningar och för förhållandet mellan statistisk modell och verklighet. Jag skall här fokusera på den senare boken.¹

Det finns en rad olika problem med bruket av signifikanstest, och Ziliak och McCloskey tar upp flera av dem, men lägger störst krut på det fenomen de kallar *sizeless science*. Antag exempelvis att en ny blodtryckssänkande medicin skall utprovas. Antag vidare att det faktiska förhållandet är att medicinen visserligen har en positiv effekt jämfört med placebo, men att skillnaden är så liten att den inte har någon praktisk betydelse för patientens hälsa eller välbefinnande. Om försöket omfattar tillräckligt många patienter, så kommer likväl denna lilla skillnad med stor säkerhet att detekteras, och statistisk signifikans erhållas. Lärdomen av detta är att det i en medicinsk studie inte räcker att uppnå statistisk signifikans för att ge stöd åt det nya läkemedlet – vad som också skall till är att den observerade skillnaden är *medicinskt signifikant*. På motsvarande sätt behöver en nationalekonom (psykolog, geolog, etc) utöver statistisk signifikans också beakta ekonomisk (psykologisk, geologisk, etc) signifikans. Underlåtelse att göra det är vad författarna kallar *sizeless science*.

Bruket av *sizeless science* inom bl.a. medicin och psykologi diskuteras i boken, men största utrymmet ägnar Ziliak och McCloskey naturligt nog nationalekonomin. I en imponerande litteraturstudie har de gått igenom samtliga de 369 artiklar som under

¹ Vad gäller Talebs bok så recenserade jag den i Axess 6/2007; se <http://www.math.chalmers.se/~olleh/Taleb.pdf>

1980- och 1990-talet publicerats i tidskriften *American Economic Review* och som inbegriper statistisk analys. Inte mindre än 70% av artiklarna under 80-talet visade prov på sizeless science, en siffra som för 90-talet förvärrats ytterligare till 79%. En rad andra slags missbruk av statistisk metodik studeras i samma undersökning, och resultaten är nästan genomgående nedslående.

Ett särskilt vanligt felgrepp bland dem författarna diskuterar, och som varje lärare i matematisk statistik suckar tungt igenkännande åt, är förväxlandet mellan å ena sidan sannolikheten för erhållna data givet nollhypotesen, och å andra sidan sannolikheten för nollhypotesen givet data. (Den senare sannolikheten är som bekant överhuvudtaget inte åtkomligt med mindre än att Bayesiansk modellering tillgrips, något som sällan förekommer i de empiriska vetenskaper som här diskuteras.)

Boken har en utpräglat polemisk stil, vilket leder till roande läsning, som t.ex. i den minst sagt svartvita historiska skildringen av förhållandet mellan ängeln William Gossett och djävulen Ronald Fisher. Men emellanåt går polemiken över styr, som när författarna på s 240 frågar sig: "Om nu signifikanstest är så idiotiskt som vi och dess övriga kritiker länge hävdade, hur kan det då ha överlevt?". Jag är helt enig med författarna om att det ensidiga användandet av signifikanstest är ett oskick; icke desto mindre är den statistiska signifikansen ett avgörande komplement till medicinsk/nationalekonomisk/etc dito, ty utan den har vi inte grund för att säga att de mönster vi tycker oss se är verkliga mönster och inte bara brus.

Vad bör då göras åt den rådande situationen? Någon mirakelkur kan Ziliak och McCloskey inte erbjuda, men de är tydliga med att förespråka en högre grad av pluralism bland statistiska metoder – dock utan att beröra risken att större flexibilitet i valet av statistisk metod ökar risken att metod väljs i syfte att ge önskat resultat för just de data som råkat erhållas. Många av författarnas kommentarer tycks implicera ett ställningstagande för Bayesiansk statistik, men frågan är om de ens är klara över den saken; hur som helst tar de aldrig explicit klivet ut ur den Bayesianska garderoben.

Sammanfattningsvis kan om Ziliaks och McCloskeys bok (liksom för den delen även Talebs) sägas att det inte är svårt att hitta passager där författarnas brist expertkunskap i själva statistikämnet skiner igenom, men att boken likväl erbjuder lärorika perspektiv på statistiska inferensproblem. I fall som detta gör vi nog klokt i att betrakta statistikens kritiker också som våra vänner.