Statistik-
främjandet

FINNS SLUMPEN? FRÅN ORDNING TILL OORDNING OCH VV

SIDORNA 2-6

PRISTAGARE/ MINNEN OCH MENINGAR

SIDORNA 20-22

Qvintensen

NR 2 2017

Certifiering på gång

SIDAN 12

Hög tid att förbereda inför GDPR

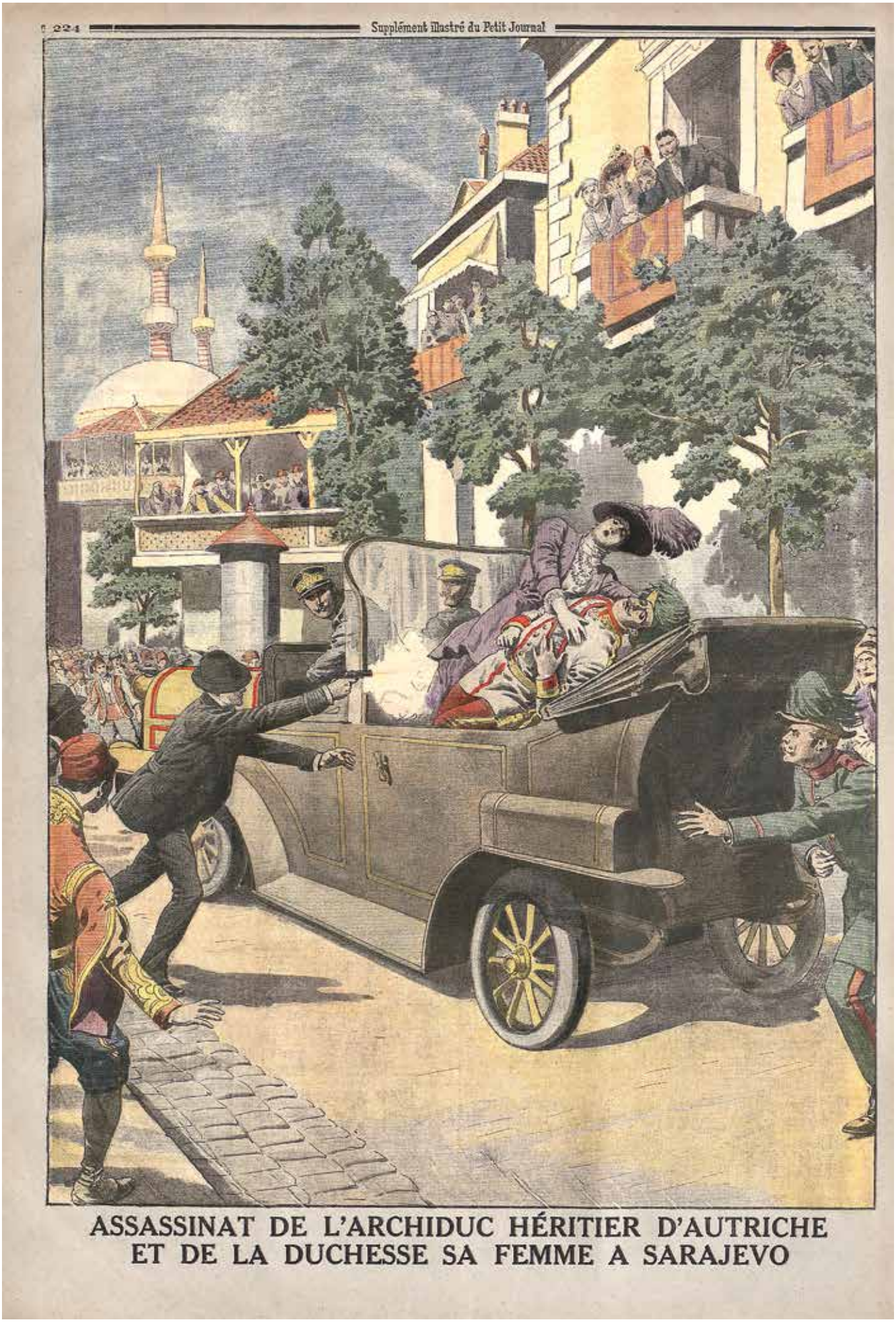
SIDORNA 10-12

Matematisk statistik som universitets-ämne – vart är vi på väg?

SIDORNA 25-27

Fakta eller fiktion –så rapporterar media om opinions-mätningar

SIDORNA 13-14



Historievetenskapen vimlar av slumpmässiga händelser – med stundom ödesdigra följder. Det klassiska exemplet är skottet i Sarajevo den 28 juni 1914, som ledde till Hitler och Stalin och det europeiska 1900-talets brutala historia. Bilden visar tecknaren Achille Beltrames bild av dådet, publicerad på förstasidan av Domenica del Corriere, en italiensk tidning.

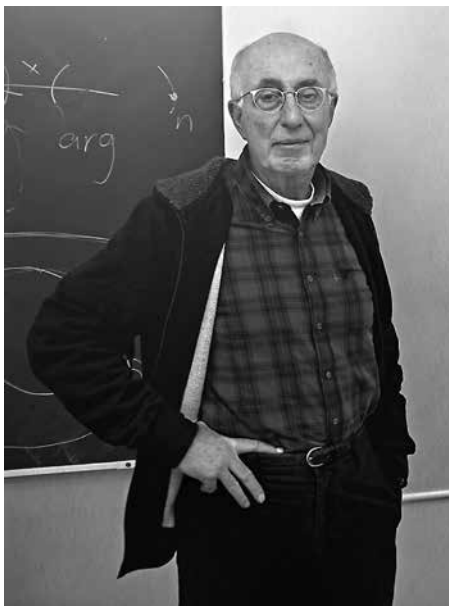
Finns slumpen? – Från ordning till oordning och tvärtom

Slumpen är nästan påträngande närvarande i vår tid och kultur. Postkodlotteriets käcka uppmaningar dimper ner i brevlådan. Internetsurfandet avbryts med uppmaningar att spela än det ena och än det andra. ICA-butiken om hörnet hävdar att just den sålde den stora vinstlotten. Tipset lever och frodas, liksom den klassiska vadslagningen kring trav och galopp. Visst finns slumpen i spel och dobbel, som i sin tur tycks finnas överallt.

Men finns slumpen i allvarligare sammanhang? Och vad är den i så fall?

Den klassiska fysiken hade ingen plats för slumpen. Eller med markisen Simon de Laplaces ord: "För en högre intelligens, som i en och samma formel skulle kunna sammanfatta rörelserna av universums största kroppar och den lättaste atom, för dess ögon skulle framtiden såväl som det förflutna vara ständigt närvarande." Slumpen var snarast vetenskapens antites, en uppgiven axelryckning inför det man inte kunde förklara, sedan gud hade svikit.

Det är först på 1900-talet, med den kvantmekaniska revolutionen, som osäkerhet och slump bryter in i fysiken. Heisenbergs osäkerhetsrelation uttryckte detta genom att säga att produkten av standarddeviationerna för en partikels rörelsemoment och för position inte kunde underskrida ett visst positivt



Peter Jagers.

värde, givet av Plancks konstant. Alternativt kan man tala om en osäkerhet i energi och tid. Inom kvantmekaniken är alltså slumpens närvaro principiell. Noggrannast möjliga beskrivning av en partikels energi vid en viss tidpunkt måste innehålla osäkerhet.

Men slumpen kan också vara av annat slag. Utfallet av ett tärningskast faller inom den klassiska mekanikens domäner. Vet du hur jag håller tärningen, känner avståndet till golvet, inblandade materialegenskaper osv., kan du i princip räkna ut hur många prickar tärningen kommer att visa efter kastet. Inom meteorologin förefaller situationen att vara densamma: Mekaniken är känd, men alldeles för invecklad för mer långsiktiga prognoser. Vi vet alla hur väderrapportens tillförlitlighet avtar från god de närmaste dagarna till mycket tvivelaktig tio dagar framöver. Här verkar det vara komplexiteten i ett i sak välbestämt förlopp som ger upphov till oförutsägbarhet. Man kan notera att det är just oförutsägbarhet som ofta anges som själva essensen i slumpmässighet, t.ex. i den svenska nationalencyklopedin.

Det kan tyckas att denna komplexitetsgrundade slumpmässighet är av ett mindre fundamentalt slag än vad vi ovan kallade principiell slumpmässighet. Men här får man vara försiktig. Vad skall man säga om beräkningen av en bana blir så komplicerad att den omöjligen kan utföras under den tid som förloppet tar? Om alltså t.ex. själva den detaljerade uträkningen av vädret söndag eftermiddag om en vecka tar hela två veckor i datortid? På snabbast möjliga utrustning, förstås. Det kan noteras att när A. N. Kolmogorov ville definiera begreppet slumpmässig talföljd, så utnyttjade han just komplexiteten hos den kortaste algoritm som kunde generera följden.

En tredje typ av osäkerhet handlar om vår »»»

Visst finns slumpen i spel och dobbel, som i sin tur tycks finnas överallt.

»»» FINNS SLUMPEN? – FRÅN ORDNING TILL OORDNING OCH TVÄRTOM

okunskap. Om blindtarmsinflammation säger exempelvis landstingens vårdinformation att orsaken oftast är okänd ”men det troliga i många fall är att något täpper till mynningen till blindtarmsbihanget vilket gör att trycket i bihanget ökar. Det i sin tur kan leda till en virus- eller bakterieinfektion i den tilltäppta delen, vilket gör att den svullnar upp och att var bildas. Ibland upptäcks i samband med operation något som har fastnat vid bihangets öppning. Det kan vara en hård matbit eller en avföringsklump. Någon enstaka gång kan en ficka, sammanväxning, polyp eller tumör täppa till bihanget. I de flesta fall hittar inte kirurgen vid operationen något som täpper till bihanget utan bara ett förtjockat och inflammerat blindtarmsbihang.” Man kan tala om okunskapsgrundad slump.

Slumpen i fysiken

Givetvis förekommer kombinationer av ovanstående faktorer. Vi kan vara okunniga om såväl enkla som komplexa förlopp och ett komplext förlopp kan tänkas ha sin grund i en kvanthändelse.

I denna anda har man spekulerat om universums uppkomst. Skedde den genom en kvantfluktuation – ren slump alltså? Enligt de Heisenbergska osäkerhetsrelationerna är ju tiden och energin inte helt bestämda. Alltså kan energin öka för ett ögonblick, som kan tänkas vara mycket kort. Då kan energin bli mycket stor, och dessutom behövs kanske inte så mycket.

Man lånar liksom, säger astrofysikern Bengt Gustavsson, lite tid av framtiden och kan också låna lite energi, och så kan man sätta igång universum, på lånad tid och energi. Det är ungefär som att gå till banken, låna pengar, göra en investering och lämna tillbaka pengarna innan bankfolket bråkar. ”Big Bang” skulle alltså ha varit så liten att den först inte märktes.

Samma kvantslump menar man – nu mindre spekulativt – ligger bakom galaxernas uppkomst, en kort tidsrymd senare, 10^{-40} – 10^{-30} sekunder talar astronomerna om. Galaxerna blåstes sedan upp till jätteformat genom universums tidiga oerhört snabba expansion (inflationen). Evidensen för detta står att finna i storleksfördelningen av fluktuationerna i den kosmiska mikrovågsbakgrunden, som är nära den som förväntas av tidiga kvantfluktuationer.

Slumpen i biologin

Går vi från universum till livet, kan man fundera över slumpens roll i livets själva uppkomst, i molekylers irrande och möten i en dypöl för länge sedan. Redan Darwin var medveten om slumpens betydelse i det senare stadium som evolutionen utgör, och efter den moderna syntesen mellan genetik och evolutionsbiologi har det märkliga samspelet mellan det slumpmässiga och det nödvändiga i evolutionen kunnat beskrivas precist matematiskt. ”Le hasard et la nécessité” var titeln på en berömd essä av biologen Jacques Monod. Mutationer är slumpmässiga felkopieringar. I sällsynta undantagsfall leder de till uppkomsten av nya livskraftiga arter, som eventuellt (!) etablerar sig i en ny nisch eller genom att slå ut sämre rustade föregångare. Etablerandet består då av två skeden, varav det första är slumpmässigt eftersom populationen är så liten att tillfälligheter kan slå ut den. Det senare skedet är stabilt om bara de initiala farorna överlevts och inte slumpen slår till utifrån. (Detta är ju vad man tror skedde för 65 miljoner år sedan då en rymdkollision ledde till dinosauriernas utdöende.)

Slumpen i samhällsvetenskapen och historien

Låt oss gå vidare till samhällsvetenskapen, där de mest eklatanta exemplen står att finna i den gren som har de största pretentionerna på exakthet, nationalekonomin. I Riksbankens prognosunderlag till beslutet att höja räntan till 4,5 % den 3 september 2008 hette det:

"Sveriges riksbank använder sedan något år tillbaka en ny makroekonomisk allmän jämviktsmodell av svensk ekonomi som går under namnet RAMSES ... Modellen har anpassats för att beskriva utvecklingen av ett antal centrala makroekonomiska tidsserier med hjälp av de bästa tillgängliga ekonometriska metoderna, och metodmässigt ligger modellen i den vetenskapliga frontlinjen."

Vad man förväntade sig behöva hantera var vad man kallade ”en varaktigt hög prisnivå”. Tolv dagar senare gick Lehman Brothers i konkurs.

Inflationen föll till noll och svensk BNP med ca 5 procent 2009.

Detta skeende hör nästan till historien nu, och historievetenskapen vimlar förstås av händelser som svårligen kan beskrivas som annat än slumpmässiga – med stundom ödesdigra följder. Det klassiska exemplet är skottet i Sarajevo den 28 juni 1914, som ledde till Hitler och Stalin och det europeiska 1900-talets brutala historia. Enligt de gängse beskrivningarna hade en tidning publicerat fel kortegerutt, som Gavrilo Princip tog del av. Där han ställde sig fanns således ingen skyddande polis och militär. Men ärkehertigparets chaufför hade läst samma felaktiga vägbeskrivning.

Historikern och författaren Peter Englund har återkommande intresserat sig för slumpen i historien och bland annat gett en färgstark beskrivning av dess roll i Cortés krossande av aztekernas rike:

Den 22 april 1519 dyker spanjorerna upp utanför den mexikanska kusten, och – sig själva ovetande – seglar rakt in genom ett knappnålsstort hål i den aztekiska kalendern. En av de många berättelserna i aztekernas rika och förstämda mytologi, handlade om överguden Quetzalcoatl, som någon gång i tidernas gryning fördrivits men som hotat att återvända och hemsöka dem på nytt. Profetian sade bland annat att detta skulle ske på året "ett säv" och på dagen "nio vind". Utseendet då?

Tja, alla bilder visade honom iförd svart dräkt.

Översatt till europeisk kalender blir "ett säv" och "nio vind" den 22 april 1519. Den dagen råkade Cortés stiga i land i Mexiko, och då det också råkade vara långfredagen bar han därför – svarta kläder.

(Dagens Nyheter, 2002-03-31)

Och så kan man fortsätta: Kaiser Wilhelm hade kunnat bli skjuten av prickskytten Annie Oakley i Buffalo Bills Wild West show i Berlin på 1890-talet om hon inte varit så träffsäker – han gick upp scenen och höll upp ett spelkort som hon prickade. Eller med den sovjetiske hjälten S. E. Petrov som inte förde ett felaktigt larm om missilangrepp vidare den 26 september 1983 och därmed troligen räddade världen från ett fullskaligt kärnvapenkrig. Historien är mer av en slumpvandring än vad deterministiska historiker, materialister eller hegelianska dialektiker tycks kunna föreställa sig.

»Historien är mer av en slumpvandring än vad deterministiska historiker, materialister eller hegelianska dialektiker tycks kunna föreställa sig.»

Slumpen i juridiken

Även i juridiken är sannolikhetsresonemangen legio. Vad menar man när man talar om att någon är ”på sannolika skäl” misstänkt för ett brott? Envar som har blivit fälld för att ha stått för länge på en parkeringsplats har på ett trivialt sätt blivit utsatt för detta. För hur gör parkeringsvakten, när hon lappar en bil? Jo, hon kommer tillbaka efter säg två timmar och kontrollerar vänster framhjuls ventilställning – åtminstone var detta ett klassiskt förfarande. Pekar ventilen på samma klockslag (i timmar) nu som förra gången, så drar hon den slutsatsen att bilen inte har varit på någon utflykt. Men vad händer om du i själva verket har varit ute och kört, återvänder och finner din gamla plats ledig? Nja, åtminstone om din utflykt varit kort, borde man kunna räkna hjulvarven och nog är det möjligt att ventilen hamnar på samma sätt igen. I praktiken låter detta sig förstås inte göras och det är rimligt att resonera som när vi kastar tärning: alla tolv ventilställningar borde vara lika tänkbara. Sannolikheten att jag helt regelvidrigt fälls för att ha parkerat för länge blir då 1/12, det vill säga över åtta procent. Är 92 procent ”sannolika skäl” för att jag ska vara skyldig? Mindre triviala frågor av ett likartat slag möter i DNA-bevisningens problematik.

För några år sedan aktualiserades sannolikhetsteorin i ett känt svenskt rättsfall. Om sannolikheten att en person under vissa omständigheter är skyldig till ett brott är p_1, under andra omständigheter p_2 och ytterligare andra omständigheter p_3, vad kan man då säga om den totala sannolikheten för personens skuld? (Samma resonemang kan förstås föras om antalet betingade situationer är fler eller färre än tre.) I boken Quickologi, s. 296-299, menar justitierådet Göran Lambertz att svaret ges genom addition av de inblandade betingade sannolikheterna, i vårt fall alltså av $p_1+p_2+p_3$. Eftersom Lambertz delar upp situationen i skuldbörda under en rad olika omständigheter, är det inte att förundra sig över att han hamnar långt över visshet (sannolikheten 1, det vill säga 100 %). I det fall han diskuterar blir Thomas Quicks skuldsannolikhet inte mindre än 183 %. (Detta fall har diskuterats av statistikern Olle Häggström i en rad blogginlägg.)

Politiska opinionsundersökningar

Låt oss se på ett sista exempel, de politiska opinionsundersökningarna mellan samhällsvetenskap och journalistik. Länge trodde man att säkerheten i en sådan låg i undersökningens omfattning. Vändpunkten kom 1936 då tidskriften Literary Digest gjorde en undersökning baserad på två miljoner läsare, men missade att förutsäga valet av Franklin Roosevelt till USAs president. George Gallup, däremot, hamnade rätt, med ett slumpmässigt urval av 50.000 respondenter. Detta ledde till en snabb utveckling av och dominans för slumpmässiga urval, som ju har en väldefinierad felmarginal, ett intervall som är sådant att det korrekta värdet med en känd sannolikhet, ofta 95 %, håller sig inom marginalen. Särskilt i vårt land, med dess stora möjligheter att enkelt få fram slumpmässiga urval ur befolkningsregister, blev detta en metod som fått bred användning.

Men en ny tid möjliggör nya lösningar, inte alltid bättre. Dagens opinionsundersökare frestas att göra sina urval ur egna datapaneler dit deltagarna själva anmält sig. Felmarginalen blir då en marginal inte till vad som är relevant, nämligen sympatierna i hela valmanskåren, utan en möjlig missvisning i förhållande till panelen. Hur denna i sin tur förhåller sig till populationen som helhet vet man sedan mindre om. Åtminstone ett av de svenska opinionsinstituten med brett utbasunerade undersökningar har använt sig av sådana självrekryterade paneler och också rapporterat resultat som avviker från vad andra »»»

institut trorsig se. På senare tid har en vis tillnyktring kanske sett; man säger sig ha bättre koll på panelerna.

Vad är slumpmässighet?

Slumpen kommer tydligen in i många helt olika sammanhang, och sannolikhetsteoretiska och statistiska fenomen spelar en roll i snart sagt all vetenskap. Lite slagordsmässigt kan man skilja på vägen in i slumpen och vägen ut ur densamma. Bägge vägarna kännetecknas av en växande komplexitet.

Vetenskapligt, i klassisk sannolikhetsteori eller kvantfysik, kan slumputfall ses som ett primärt begrepp, beskrivet genom den fördelning som anger sannolikheten av olika utfall. Jag har dock redan nämnt att 30- och 40-talens försök att precisera matematisk beräkningsbarhet ledde till Kolmogorovs algoritmiska definition av slumpmässiga följder av nollor och ettor. Denna uttrycker i en precis form just de tankar som ligger bakom kaosbegreppet och kanske också det vardagliga slumpbegreppet. När orsakskedjorna blir riktigt trassliga, då ger vi upp, rycker på axlarna och säger: ”Det var nog mest slumpen.”

Startar vi från slumpmässiga storheter, vad nu detta är, som kan beskrivas med sannolikhetsfördelningar – till exempel positioner hos irrande partiklar, storleken av observerade enheter, behandlingseffekter eller resultaten av tärningskast – så kan vi ställa frågan vad som sker när de observerade fenomenen sätts samman i allt större system. Vad händer när tärningar kastas många gånger, eller när nyföddas vikt i en stor stad studeras? Eller när olika fenomen orsakar ett observationsfel? Jo, när systemen blir allt större och påverkas av allt fler oberoende eller svagt beroende komponenter, där inga enskilda dominerar och mätskalan anpassas, så återkommer stabiliteten. Fysikern kan då tala om makroskopisk determinism kontra mikroskopisk slump, sannolikhetsteoretikern om de stora talens lag, eller den berömda centrala gränsvärdessatsen som beskriver fluktuationer i de stora talens lag.

Men det finns många sådana stabiliseringsfenomen. Själv har jag forskat i hur strukturen stabiliseras i en växande population, åldersfördelningen, släktskapsrelationer, och därmed den fylogeni, som spelar en viktig roll i modern populationsbiologi. Ett område som har väckt uppmärksamhet är förbindelsestrukturen i stora nätverk – vi är alla sex handslag från Einstein, som det kan heta. Typiskt för denna forskning är att den inspireras av speciella empiriska mönster från naturvetenskap eller samhälle. Resultaten blir emellertid renodlat matematiska och har således bred giltighet. Idag är vi vana vid detta, men man kan föreställa sig Gauss fascination när han upptäckte samma normalfördelning i felen vid sina astronomiska observationer som för vikterna hos nyfödda barn i dåtidens Breslau, dagens Wróclaw.

Ett nöjsamt, men mindre känt exempel på en sådan stabilisering, som dessutom var och en enkelt kan kontrollera, ges av Benfords lag. Denna säger att i stora siffermaterial, som taxeringslängder, telefonkataloger, listor över fysikaliska data eller företagsräkenskaper, så tenderar förstasiffrorna att förekomma på följande sätt. Sannolikheten för en etta är log(1+1/1) = log 2 = 0,301. Sannolikheten för en tvåa är log(1+1/2) = log 1,5 = 0,176. En trea uppträder med sannolikheten log(1+1/3) och så vidare, allt sammanfattat i figur 1.

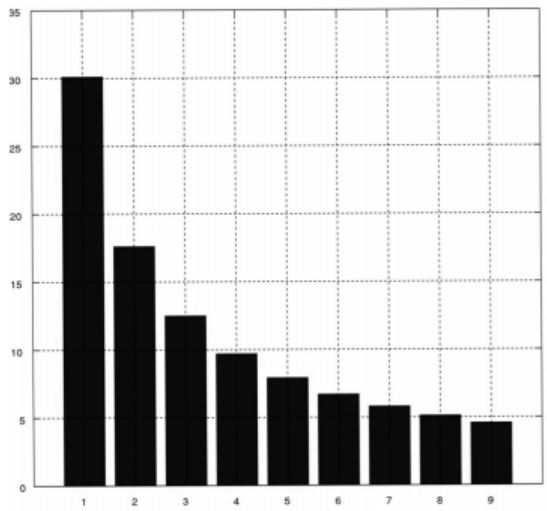


Fig. 1. Antalsfördelningen i procent hos första heltalet i en stor population av flersiffriga tal – Benfords lag.

Beviset av Benfords lag avser situationen när materialets storlek går mot oändligheten. Många sådana resultat är sedan tillämpliga på ganska små material, och frågan om konvergenshastigheten i stabiliseringsresultat är ett viktigt forskningsområde. Men resultaten får hanteras med en viss försiktighet – och en dos förnuft. Googlesökning på talen 165, 265, 365, 465, 565, 665, 765, 865 och 965 ger (i skrivande stund) resultaten: 1 220, 722, 955, 400, 507, 297, 286, 247, 288 miljoner träffar. Tre avvikelser sticker ut: 365, som är årets dagar, 565 som är en palindrom, och 965, som jag inte kan förklara. Så fick vi en – om än oavsiktlig – påminnelse om att användningen av statistik och sannolikhetsteori är en mötesplats för precis matematik, sakvetenskap, sunt förnuft och vanskliga tillfälligheter.

(Denna artikel baserar sig på föredrag inför Kungl. Vetenskaps- och Vitterhetssamhället i Göteborg samt Sällskapet Heimdall i Malmö.)

PETER JAGERS
MATEMATISKA VETENSKAPER
CHALMERS OCH GÖTEBORGS UNIVERSITET

Qvintensen

Ansvarig utgivare
John Öhrvik

Redaktör
Jan Wretman, 070-781 78 35

Redaktion
Alf Fyhrlund
Rolf Larsson
Marie Linder
Ingemar Sjöström
Marika Wenemark

E-post wretman.jan@gmail.com

Produktion
Form och redigering: Mezzo Media AB
Tryckeri: Trydells Tryckeri AB

Annonser
Annonser i Qvintensen bokas med redaktören.
Annonsutskick per e-post bokas med Statistikfrämjandets sekreterare. Priset är 6 000 kronor för institutionsmedlem eller motsvarande, och
8 500 kronor för övriga annonsörer. På alla priser tillkommer 25 % moms.

Statistik-
främjandet

SVENSKA STATISTIKFRÄMJANDETS STYRELSE
Ordförande
John Öhrvik, 070-666 08 08,
john.ohrvik@regionvastmanland.se

Vice ordförande Patrik Rydén

Skattmästare Maria Juhlin

Sekreterare Linnea Johansson Kreuger

Hemsidesansvarig Alf Fyhrlund

Klubbmästare Sabri Danesh

Surveyföreningen
Åke Wissing

FMS
Anna Lindgren

Industriell statistik
Hans Alberg

Cramérsällskapet
Rebecka Jörnsten

Övriga ledamöter
Magnus Ohlsson, Silke Tindrebäck

Adress
Svenska statistikfrämjandet
Sekreterare: Linnea Johansson Kreuger

E-post sekrfram@gmail.com

Webbplats www.statistikframjandet.se

Situationen i Sverige med traditionellt stabila väljarsympatier har förbytts i stor rörlighet i väljarkåren, vilket visas hos samtliga institut som mäter detta.

KARIN NELSSON SIDORNA 15-16

Qvintensen
STATISTIKFRÄMJANDET
NR 2 2017

Innehåll

FINNS SLUMPEN?
– Från ordning till oordning och tvärtom 2-6

REDAKTÖRENS RUTA 8-9

WEBBVERSION AV QVINTENSEN 9

GDPR: Hög tid att förbereda sig 10-12

CERTIFIERING PÅ GÅNG 12

OPINIONSUNDERSÖKNINGAR/
Fakta eller fiktion? 13-14
Före och efter Trump 15-16

OFFICIELL STATISTIKPRODUKTION 17-19

MINNEN OCH MENINGAR 20-22

MINNESORD/Jan Hoem 22

MATEMATISK STATISTIK SOM UNIVERSITETSÄMNE 25-27

Våra föreningar

SURVEYFÖRENINGEN
Kvalitetsseminarium 2018 23
Efterlysta svenskar 23

FÖRENINGEN FÖR MEDICINSK STATISTIK
FMS 30 år 24
Marie Eriksson – ny i styrelsen 24

ISSN 2000-1819
Författare och andra som utan ersättning bidrar med material för publicering behåller upphovsrätten till sina verk. Svenska statistikfrämjandet äger rätten att publicera dessa verk i valfri form och upplaga samt att behålla alla eventuella intäkter från sådan publicering, om inget annat avtalas. Författare och andra som utan ersättning bidrar med material för publicering garanteras rätten att själva få sprida sitt verk under förutsättning att Qvintensen anges som originalkälla med angivande av nummer och år. Svenska statistikfrämjandet äger upphovsrätten till publicerat material då det är framställt av redaktionsmedlem, styrelsemedlem, någon till Främjandet knuten förenings styrelse eller person som erhållit arvode för sitt bidrag. © Svenska statistikfrämjandet 2016.

Det här handlar inte om statistik i allmänhet, utan bara om sådan statistik som ska beskriva hur det ser ut i samhället. Dit hör både officiell statistik och sådan statistik som produceras i vetenskapligt eller kommersiellt syfte (till exempel opinionsundersökningar).

Hur ska det gå med samhällsstatistiken?

Ibland känner jag oro över vissa tendenser. Det är tre frågor jag tänker på:

- Väljer växande grupper av människor att inte lita på statistik?
- Kan ”big data” innebära risk för missbruk av data?
- Kommer vi att återgå till icke-sannolikhetsurval?

Jag tar nu upp dessa frågor mycket kortfattat, i tur och ordning.

Inte lita på statistik? Att samhällsstatistiken ljuger, eller att man kan bevisa vad som helst med statistik, är fraser som florerat i mer än hundra år. Men vettiga människor har i regel insett att det viktiga är att statistik inte ska missbrukas. Dock undrar jag, i stunder av dysterhet, om vi nu håller på att förlora kontroll över utvecklingen. Jag tänker på den ökande förekomsten av det som med en aktuell term kallas kunskapsresistens.

I The Guardian kan man hitta en intressant och lätt provokativ artikel av den engelske sociologen William Davies med titeln ”How statistics lost their power – and why we should fear what comes next” (https://www.theguardian.com/politics/2017/jan/19/crisis-of-statistics-big-data-democracy). Davies beskriver tendenser i England och USA i samband med Brexitvalet och det senaste amerikanska presidentvalet. Tendenserna har kanske inte nått Sverige ännu, men en grogrund finns nog även här.

Det handlar om den ökade förekomsten i västerländska demokratier av populistiska (och ofta främlingsfientliga) rörelser, som bland annat misstror samhällets ”experter”. Dessa experter antas vilja manipulera folket. Ibland ser man officiell statistik som lögnaktig. Man föredrar att istället lita på anekdotisk evidens, och ”hur det känns”. Det här gäller bland annat politiskt känsliga statistikområden som arbetslöshet, invandring, kriminalitet och miljöfrågor.

Här är två exempel från Davies artikel. President Donald Trump figurerar i båda.

(1) Vid en presskonferens sade Vita husets talesman Sean Spicer till journalisterna: ”The president, he’s not focused on statistics as much as he is on whether or not the American people are doing better as a whole.” Hur bär han sig åt för att få kunskap om detta?

(2) Under presidentvalskampanjen uttalade sig Trump om arbetslöshetssiffror från Bureau of Labor Statistics: ”The 5% figure is one of the biggest hoaxes in American modern politics.” Och lite senare i samma tal: ”The number’s probably 28, 29, as high as 35. In fact, I even heard recently 42%.” Av vem hade han hört detta?

”Big data”? Med ”big data” menas digitalt lagrade datamängder av enorm storlek, ofta skapade i realtid. Sådana datamängder skapas inom dagligvaruhandel, sociala media, kommunikationstjänster och satellitövervakning, för att bara ta några exempel. Förhoppningar (ibland kanske överdrivna) har uttryckts om att utnyttja den här typen av data för statistikproduktion. En balanserad artikel, med ett SCB-perspektiv, är Lili Japecs ”Så kan vi göra statistik av big data” (https://www.scb.se/sv_/Hitta-statistik/Artiklar/Sa-kan-vi-gora-statistik-av-big-data/). Jag föreställer mig att ”big data” inte passar som underlag till den typ av traditionell samhällsstatistik, som vill ge svar på frågor av typen ”Hur stor andel av personerna i vissa noggrant definierade befolkningsgrupper är det som har den ena eller andra egenskapen?” Men det kommer kanske att uppstå helt nya statistikgrenar.

ILLUSTRATION: THIERRY GREGORIUS

"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here"

Kan "big data" göra psykologiska profiler av befolkningen? Ett amerikanskt företag tycker sig kunna identifiera ett antal människotyper med olika psykologisk profil genom att på individnivå analysera stora datamängder från sociala media.

Det finns ett annat problem, av mer etiskt slag, som har att göra med användning av ”big data” i syfte att omärkligt påverka politiska opinioner. Ett ofta nämnt namn i detta sammanhang är det amerikanska företaget Cambridge Analytica. Genom att på individnivå analysera stora datamängder från sociala media har företaget tyckt sig kunna identifiera ett antal människotyper med olika psykologisk profil. Detta har skett utan att personerna informerats. I samband med Donald Trumps presidentvalskampanj lär man så ha skickat ut diversifierad information till olika grupper av människor, speciellt anpassad

till gruppens psykologiska profil, i syfte att få dem att rösta på önskat sätt.

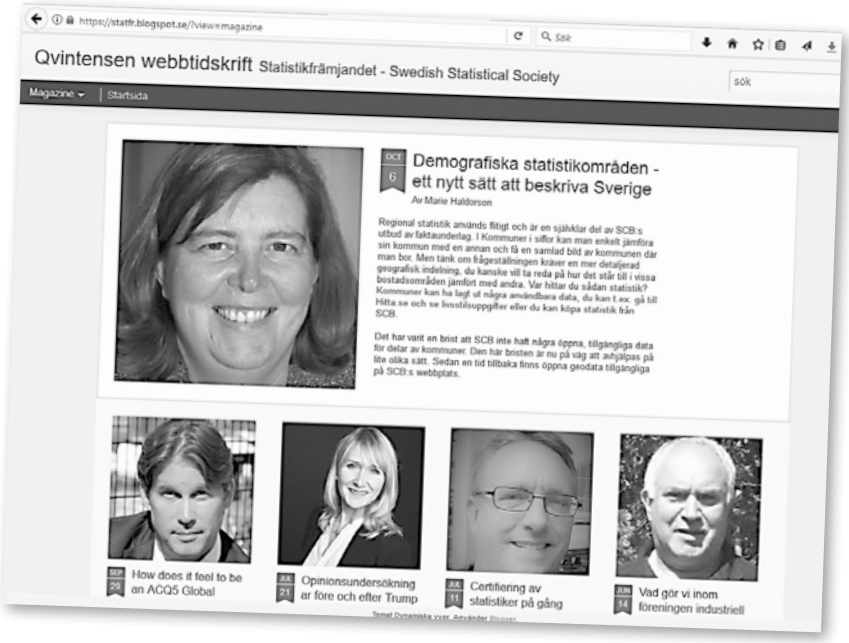
Icke-sannolikhetsurval? Införandet av sannolikhetsurval i samhällsvetenskapliga undersökningar från 1920-talet och framåt kan förknippas med namn som Bowley, Neyman och Hansen-Hurwitz-Madow. Det innebar ett stort steg framåt för samhällsstatistiken. Det blev bland annat möjligt att diskutera osäkerhet hos statistiska skattningar med hjälp av statistisk teori. En av förutsättningarna för att teorin skulle fungera var att bortfallet var inget eller bara obetydligt.

Under de första decennierna var också bortfallssiffrorna låga i samhällsvetenskapliga undersökningar. I arbetskraftsundersökningarna på 1960-talet var bortfallet runt 2 procent, enligt SCB:s jubileumsskrift Summa summarum. Men situationen har förändrats. Det är nu vanligt att undersökningar har ett bortfall på mer (ibland betydligt mer) än 50 procent. Nu gäller inte längre den gamla teorin för beräkning av skattningars osäkerhet, och någon ny teori finns inte. Det hjälper inte om det ursprungliga urvalet gjorts som ett sannolikhetsurval. Många opinionsinstitut struntar nu i att över huvud taget använda slumpmässiga urval, utan använder i stället urval konstruerade enligt andra principer, ibland till och med självrekryterade urval.

Antag att det i framtiden blir så att sannolikhetsurval inom samhällsstatistiken kommer att helt ersättas av icke-sannolikhetsurval. Det här låter för mig som att utvecklingen går bakåt, tillbaks till 1800-talet. Kommer det att uppstå någon ”vetenskap och beprövad erfarenhet” att stödja sig på? Jag vet inte. Man blir tvungen att avstå från sannolikhetsteoretiskt grundade resonemang om osäkerhet, felmarginaler och liknande. För att få ”beprövad erfarenhet” skulle man behöva skaffa erfarenhet av hur pass bra erhållna undersökningsresultat överensstämmer med de resultat som skulle ha erhållits vid en totalundersökning. Några studier av detta slag känner jag inte till. (Man skulle möjligen kunna säga att när politiska partisympatiundersökningar utförs omedelbart före ett val, så kan man få visst underlag för bedömning av hur bra man lyckats. Men hur människor säger att de ska rösta och hur de sedan faktiskt röstar är egentligen två olika saker.)

Avslutningsvis: Kommentarer och debattinlägg om dessa och andra saker är som vanligt välkomna.

JAN WRETMAN



Webbversion av Qvintensen

■ Qvintensen finns på försök i en webbversion, en så kallad webbtidskrift. Innehållet är inte helt identiskt med innehållet i pappersversionen. Till skillnad från pappersstidskriften utkommer den inte i form av ett antal distinkta nummer, utan i stället i form av en så kallad blogg, som uppdateras fortlöpande.

Qvintensens webbversion kan läsas på mobil och läsplatta (med så kallad responsiv design). Det är enkelt att skriva ut enskilda artiklar och att dela dem i olika sociala media.

Du hittar Qvintensens webbversion här: https://statfr.blogspot.se/. Den finns även länkad på Statistikfrämjandets hemsida: http://statistikframjandet.se/.

Redaktör för webbversionen är Alf Fyhrlund (alf.fyhrlund@gmail.com).

Har du flyttat?

■ Meddela din nya adress genom att skicka ett e-brev till Linnea Johansson Kreuger på sekrfram@gmail.com.

EU:S NYA DATASKYDDSREFORM (GDPR)

Hög tid för undersökningsbr att förbereda sig

Presentation av artikeln

■ EU:s nya dataskyddsförordning (General Data Protection Regulation, GDPR) börjar gälla den 25 maj 2018. Förordningen ställer krav på företag och organisationer att på ett dokumenterat sätt visa att personuppgiftshanteringen uppfyller förordningens omfattande krav – i annat fall kan dryga sanktionsavgifter på upp till 20 miljoner euro (!) utdömas.

Det nya regelverket kommer att i Sverige ersätta den nu gällande personuppgiftslagen (PuL). Det här är viktigt bland annat för dem som producerar statistik baserad på individdata, t.ex. SCB och opinionsinstitut, samt för dem som bedriver vetenskaplig forskning baserad på registerdata. Förutom i följande artikel finns också information att hämta på Datainspektionens hemsida.

Artikelförfattarna, advokaterna Daniel Tornberg och Alexander Jute, båda vid advokatfirman MarLaw, har mångårig erfarenhet av rådgivning kring företags och organisationers personuppgiftshantering och är flitigt anlitade föreläsare på området. I denna artikel belyser de några av de utmaningar som just undersökningsbranschen står inför, där det som bekant inte är ovanligt med blandningar av olika datakällor med personuppgifter, såsom existerande register, stickprovsdata, eller uppgifter insamlade genom olika typer av specialanpassade undersökningar.

Som en följd av den alltmer omfattande personuppgiftshanteringen – framdriven av gränslös teknisk utveckling – har EU beslutat att ersätta nuvarande direktiv 95/46/EG om skydd för personuppgifter med en dataskyddsförordning ((EU) 2016/679) för att stärka individens integritet och samtidigt skapa bättre möjligheter för fri handel på EU:s inre (digitala) marknad. Skillnaden mot tidigare är att individens rättigheter tydliggörs, samtidigt som kraven på den som behandlar personuppgifter – själv eller för annans räkning – skärps betydligt. I kombination med att sanktioner (som redan nämnts) och tillsyn ska intensifieras, räknar EU med att företag, organisationer och myndigheter i betydligt högre grad än tidigare faktiskt ska efterleva personuppgiftslagstiftningen.

Personuppgifter och behandling

Personuppgifter definieras i förordningen som varje upplysning som avser en identifierad eller identifierbar fysisk person (nedan kallad en registrerad), varvid en identifierbar fysisk person är en person som direkt eller indirekt kan identifieras särskilt med hänvisning till en identifierare som ett namn, ett identifikationsnummer, en lokaliseringsuppgift, onlineidentifikatorer eller en eller flera faktorer som är specifika för den fysiska personens fysiska, fysiologiska, genetiska, psykiska, ekonomiska, kulturella eller sociala identitet (Art. 4.1)

Behandling av personuppgifter definieras i förordningen som en åtgärd eller kombination av åtgärder beträffande personuppgifter eller uppsättningar av personuppgifter, oberoende av om de utförs automatiserat eller ej, såsom insamling, registrering, organisering, strukturering, lagring, bearbetning eller ändring, framtagning, läsning, användning, utlämning genom överföring, spridning eller tillhandahållande på annat sätt, justering eller sammanförande, begränsning, radering eller förstöring (Art. 4.2).

Det inses därför snabbt att förordningens tillämpningsområde är synnerligen brett och träffar de flesta företags, organisationers och myndigheters verksamhet.

Laglig grund en förutsättning för behandling

För att personuppgiftshantering ska vara tillåten krävs att behandlingen bygger på en laglig grund. Behandling kan endast vara laglig om och i den mån som åtminstone ett av följande villkor är uppfyllt: samtycke från registrerad, nödvändig för att fullgöra avtal, nödvändig för att fullgöra en rättslig förpliktelse, nödvändig för att skydda intressen som är av grundläggande betydelse för den registrerade eller för en annan fysisk person, nödvändig för att utföra en uppgift av allmänt intresse eller som ett led i den personuppgiftsansvariges myndighetsutövning, eller nödvändig för ändamål som rör den personuppgiftsansvariges eller en tredje parts berättigade intressen, om inte den registrerades intressen eller grundläggande rättigheter och friheter väger tyngre och kräver skydd av personuppgifter, särskilt när den registrerade är ett barn.

I de flesta fall kan den insatte oftast ganska snabbt konstatera på vilken laglig grund som uppgifterna behandlas. Problemet är emellertid att detta förutsätter att alla personuppgifter som hanteras måste inventeras och kategoriseras – ett arbete som kan vara minst sagt betungande. I undersökningsbranschen är det vanligt förekommande att den lagliga grunden utgörs av samtycke. Ett problem med många samtycken är att de skrevs innan förordningen tillkommit och således inte uppfyller förordningens krav. Det kan (beroende på inventeringen) alltså bli nödvändigt att inhämta nya samtycken för att personuppgifterna ska kunna behandlas. Detsamma gäller naturligtvis externa register eller databaser som tillhandahållits av tredje part.

Med förordningen följer även omfattande informationskrav för den som behandlar per-

anschen

sonuppgifter (art. 12-14). Exempelvis ska den behandlade ges information om den behandlades identitet, kontaktuppgifter, ändamål och rättslig grund för behandling, berättigade intressen (om intresseavvägning som grund), mottagare av personuppgifter om överföring sker till annat land etc. etc.

Det är således ett omfattande arbete som måste nedläggas för att inventera och kategorisera de personuppgifter som hanteras, i kombination med att samtyckestexter, samarbetsavtal, personuppgiftspolicies och dylika dokument i många fall måste skrivas om för att förordningen ska efterlevas. Arbetet inför dataskyddsförordningen visar sig inte sällan vara större än förväntat, och ett fungerande projektarbete involverar alltså också fler aktörer än vad man kan anta vid en första anblick. Självklart kan inte hela organisationen släppa allt och jobba med personuppgiftsfrågan – upprättandet av en projektgrupp bör dock göras med vetskapen om att samspel med övriga avdelningar sannolikt kommer att krävas. Även för större organisationer som har möjlighet att ta in hjälp utifrån (projektledningsresurser, IT-verktyg, juridisk expertis), så kommer det i princip oundvikligen ändå att krävas insatser från berörda grupper inom den egna personalen. Detaljkunskapen om hur personuppgifter faktiskt behandlas i den egna verksamheten finns ju just där – inom den egna organisationen – och ingen annanstans. Att etablera en medvetenhet inom organisationen redan från början om att ingen avdelning ”slipper undan” personuppgiftsarbetet kan alltså underlätta.

Den 25 maj gäller GDRP – EU-förordningen som ställer krav på företag och organisationer att på ett dokumenterat sätt visa att personuppgiftshanteringen uppfyller vissa krav – i annat fall hotar dryga sanktionsavgifter.

Undersökningsbranschens utmaningar

I undersökningsbranschen bygger ofta själva produkten på en analys av personuppgifter. Branschens överlevnad är således beroende av att personuppgifter fortsatt får behandlas, och därför vilar ett alldeles särskilt tungt ansvar på aktörerna att efterleva förordningen. Branschorganisationerna ESOMAR (European Society for Opinion and Marketing Research) och EFAMRO (European Federation of Associations of Market Research Organisations) publicerade så sent som juni 2017 en guide över hantering av dataskyddsförordningen i undersökningsbranschen: ”General Data Protection Regulation (GDPR) Guidance Note for the Research Sector: Appropriate use of different legal bases under the GDPR” . [1] Guiden som visserligen är ett levande dokument innehåller en hel del nyttig information likt den ovan beskrivna, men även mer uppmanande förslag till den som är aktiv i undersökningsbranschen. Dit hör exempelvis uppmaningen att i de fall det är möjligt anonymisera (avsnitt 3.2) personuppgifter så att de inte kan kopplas till en viss individ – vilket naturligtvis löser problematiken. I de fall detta inte är möjligt genomgås istället vilka rättigheter som individen faktiskt har, och hur information om dessa rättigheter måste ges av den som behandlar personuppgifter.

I förordningens art. 89 finns ett undantag som anger att *behandling för arkivändamål av allmänt intresse, vetenskapliga eller historiska forskningsändamål eller statistiska ändamål ska omfattas av lämpliga skyddsåtgärder i enlighet med denna förordning för den registrerades rättigheter och friheter. Skyddsåtgärderna ska säkerställa att tekniska och organisatoriska åtgärder har införts för att se till att särskilt principen om uppgiftsminimering iakttas.*

Som guiden anger (se avsnitt 4) kan art. 89 tolkas som att under dessa syften kan behandling av personuppgifter vara tillåten utan föregående särskilt samtycke. Det ska understrykas att det »»»

»»» HÖG TID FÖR UNDERSÖKNINGS-BRANSCHEN ....

råder viss oklarhet om hur art. 89 ska tolkas, och om den exempelvis kan omfatta kommersiella undersökningar. Här hänvisas till den så kallade artikel 29 gruppen, som är en rådgivande och oberoende organisation, som ska se till att direktivet tillämpas enhetligt i medlemsstaterna. Gruppen ska också yttra sig till EU-kommissionen om skyddsnivån i gemenskapen och i tredje land, samt ge råd till kommissionen om förslag till ändringar av direktivet. Dessutom ska gruppen avge yttranden om branschöverenskommelser som utarbetas inom EU[2]. Ännu finns inga yttranden som specifikt rör undersökningsbranschen. Det ska även nämnas att den svenska regeringen den 7 juli 2016 beslutade om ett kommittédirektiv i syfte att bland annat analysera vilken reglering av personuppgiftsbehandling för forskningsändamål som är möjlig och kan behövas utöver den generella reglering som Dataskyddsutredningen kommer att föreslå med anledning av EU:s nya dataskyddsförordning. Utredningen ska redovisas den 27 april 2018.

Slutsatser

Som framgått medför den nya dataskyddsförordningen en stor omställning för oss alla. Det är ett betydande arbete som måste nedläggas för att det på ett korrekt sätt ska vara möjligt att efterleva förordningen. Undersökningsbranschen har kanske alldeles särskilda utmaningar eftersom datakällorna ofta är många och av olika karaktär. Förhoppningsvis kan guider från EU, kommande praxis eller uttalanden från våra svenska myndigheter leda till praktiska lättnader i personuppgiftsarbetet för just undersökningsverksamhet. Där är vi dock inte idag, utan det är fortsatt förordningskraven som gäller – även om det bör vara av särskilt intresse att följa utvecklingen på området den kommande tiden…

DANIEL TORNBERG OCH
ALEXANDER JUTE,
MARLAW

Fotnot

1. Se https://www.esomar.org/uploads/public/government-affairs/position-papers/EFAMRO-ESOMAR_GDPR-Guidance-Note_Legal-Choice.pdf

2. Se http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=50083

Besök certstat.se för fler detaljer.

Certifiering av statistiker på gång

Sedan förra hösten deltar jag i en projektgrupp inom FenStatS för ackreditering/certifiering av statistiker (FenStatS = The Federation of European National Statistical Societies). Projektgruppens namn är Committee on Professional Ackreditation. Vi arbetar utifrån att skapa ett system som är jämbördigt med de som utarbetats av Royal Statistical Society (RSS) och American Statistical Association (AMSTAT), och som ska innebära jämförbarhet inom EU.

Diskussion om certifiering av statistiker har pågått under flera år. I vissa andra länder har sådana system redan införts, t.ex. i USA där benämningen ”Ackredited Professional Statistician (PStat)” innebär en personlig certifiering som handhas av AMSTAT, och i Storbritannien där benämningen ”Chartered Statistician” är en personlig certifiering som handhas av RSS. Senast infördes ett liknande system i Hong-Kong med den amerikanska certifieringen som direkt förebild.

Förslaget vi utarbetar för FenStatS följer i stort sett det amerikanska systemet. Basen är att det är en ”portfölj” av meriter som certifieras, dvs. det ska inte genomföras speciella test eller liknande. Certifieringen ska kunna ges till alla statistiker oavsett bransch och är därmed även tänkt att kunna göra det lättare att flytta mellan olika arbetsområden.

Följande kriterier används av AMSTAT (ungefär desamma gäller för RSS):

1. Examen i statistik på nivån MSc eller högre
2. Fem års yrkesverksamhet
3. Kommunikation och pedagogik
4. Professionell utveckling
5. Följa etiska standarder
6. Medlem i nationell statistikerorganisation

AMSTAT har en procedur för hur dessa kriterier ska dokumenteras och rapporteras vid ansökan. Proceduren har också visat sig fungera. I huvudsak ska man lämna en detaljerad förteckning över meriter, lämna två exempelrapporter (eller liknande), samt ange två referenter som stöder ansökan.

För detaljer om argumenten för certifiering hänvisar jag till tidigare artiklar i Qvintensen (bl.a. 2013:4, 2014:1 och 2014:4) samt på Certstats hemsida (www.certstat.se), (Certstat = Föreningen för Certifiering av Statistiker).

Under hösten kommer vi att lägga fram ett förslag som bygger på att FenStatS står som huvudman för en standard för certifiering som sedan tillämpas av FenStatS medlemmar. Tanken är att ansökan sänds in nationellt, bl.a. för att undvika språkbarriärer. FenStatS lägger ribban för vad som räknas som tillräcklig nivå och för även ett register över ansökningarna. Med lite tur kan de första ansökningarna tas emot under 2018.

MAGNUS PETTERSSON,
ACKREDITERAD PSTAT
STATISTIKKONSULTERNA, GÖTEBORG

Magnus Pettersson är sammankallande i FenStatS ”Committee on professional accreditation”. Kommittén består dessutom av följande personer: Arne Bathke (Österrike), Luigi Pieri (Italien), Sebestien Marque (Frankrike), Paul Koopman (Nederländerna), Dominik Rozkrut (Polen), Gerardo Sanz Sáiz (Spanien) och Feridum Turkman (Portugal).

Jesper Strömbäck och Per Oleskog Tryggvason undersöker hur journalister förhåller sig till statistiska osäkerheter när de rapporterar om partisympatiundersökningar.

Fakta eller fiktion

– OM HUR MEDIER RAPPORTERAR OM OPINIONSUNDERSÖKNINGAR

I teorin finns det mycket som förenar vetenskapen med journalistiken. Båda syftar exempelvis till att ta fram och tillhandahålla information som är verifierad och sann, som är trovärdig, och som bidrar till att människor kan forma välgrundade uppfattningar om sin omvärld. Samtidigt är det också uppenbart att det finns en inneboende konflikt mellan journalistiska normer och arbetssätt och de inom akademin. Inom journalistiken är en av de största synderna att missa en stor story, medan en av kardinalsynderna inom akademin är att dra för stora växlar som inte stöds av data. Det finns alltså diametralt skilda syner på huruvida det är värst att begå ett så kallat typ-I-fel (*falskt positivt*) eller ett typ-II-fel (*falskt negativt*).

Det finns antagligen ingen journalistisk genre där denna normkonflikt är så uppenbar som när journalister skriver om resultat från opinionsundersökningar. Att medierna har en fäbless för opinionsundersökningar har visats av många studier och hänger samman med rådande nyhetsvärderingar. Opinionsmätningar handlar om politiska eliter, är lätta att visualisera och passar samtidigt in i nyhetsmedielogiken. De ger också journalister möjligheter att inta en kvasi-neutral roll i den politiska rapporteringen genom att vara baserade på undersökningar, samtidigt som resultaten skapar utrymme för journalistiska tolkningar. Parallellt har medierna genom att använda sig av opinionsundersökningar gett sig själva möjligheten att uttala sig om folkets röst i siffror och framställa sig som ombud för folkets röst.

Dessvärre har journalistiken inte varit lika villig att förhålla sig till de förbehåll som kommer med den metodologi som gjort allt detta möjligt. Även om det är värt att poängtera att det finns en rad möjliga felkällor vid urvalsundersökningar (exempelvis ofullständig urvalsram, fallande svarsfrekvenser, undermåliga frågeinstrument) så finns det dessutom något som alla undersökningar med slumpurval måste förhålla sig till: den felmarginal som beror på att man gjort ett urval i stället för en totalundersökning.

Det grundläggande faktumet att vi förväntar oss en viss grad av variation när vi nöjer oss med att fråga ett urval av hela populationen är något som är självklart för alla som läst någon statistisk grundkurs. Men i praktiken förvandlas dessa förväntade förändringar ofta till förstasidesnyheter – i stil med opinionsras – inte sällan ackompanjerat med förklaringar till varför förändringen skett.

Att medierna ofta brister när de rapporterar om statistiska undersökningar är något som noterats och påpekats i samhällsdebatten otaliga gånger, men det har fram till rätt nyligen inte studerats systematiskt. Sedan ett par år finns det dock en växande internationell litteratur som visar att journalister gång på gång rapporterar om förändringar i opinionsmätningar som inte har täckning i datamaterialet.

Till denna litteratur hör en studie som vi nyligen publicerade i den vetenskapliga tidskriften *Journalism Studies*, där vi undersöker hur journalister förhåller sig till statistiska osäkerheter när de rapporterar om partisympatiundersökningar. Mer specifikt kartlägger och analyserar vi hur de tidningsredaktioner som har etablerade samarbeten med opinionsinstitut som genomför partisympatiundersökningar med slumpmässigt urval[1] behandlar statistiska osäkerheter när de kommenterar skillnader inom och mellan undersökningar samt när de ger förklaringar till varför ett parti gått upp eller ner i opinionen. Vår studie baseras på samtliga huvudartiklar där Dagens Nyheter (DN), Svenska Dagbladet (SvD) och Expressen redovisar sina regelbundna partisympatiundersökningar under 2010-2011 samt under 2014-2015.

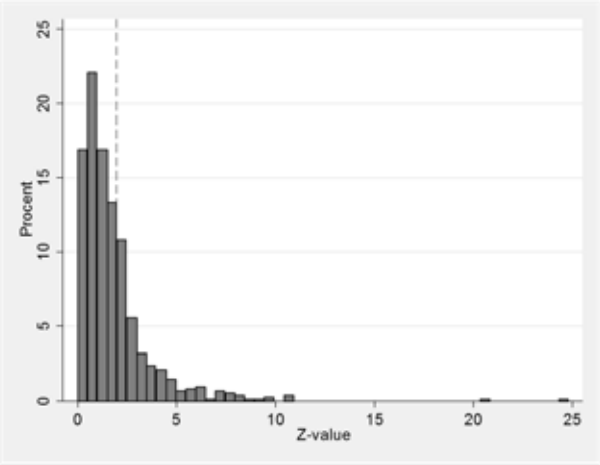
Totalt består vårt material av 142 artiklar. I dessa gör journalister totalt 757 jämförelser mellan resultaten i den senaste partisympatiundersökningen och föregående undersökning, föregående riksdagsval eller hur det ligger till i förhållande till fyraprocentsspärren. I samma urval återfinns 121 tillfällen när journalisten eller någon extern expert (vanligtvis en statsvetare eller opinionsanalytiker) försöker förklara varför ett parti har förändrat sitt opinionsstöd jäm-

»»»

Fotnot

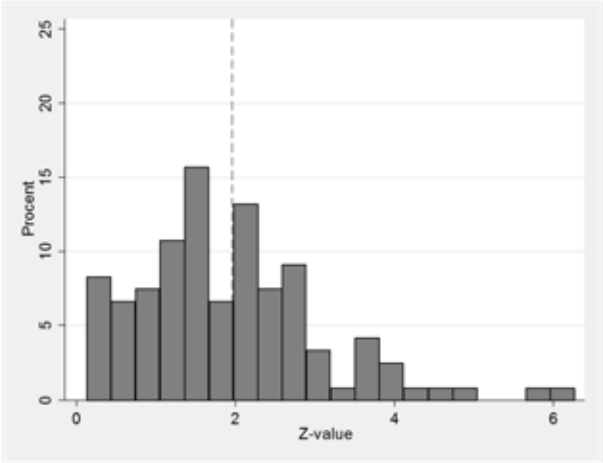
1. Det finns en pågående diskussion huruvida de urvalsförfaranden som opinionsinstituten har när de gör sina partisympatiundersökningar kan betraktas som slumpmässiga i en strikt statistisk mening. Vår utgångspunkt är dock att behandla undersökningarna som slumpmässiga i och med att det är så instituten själva beskrev och behandlade sina undersökningar vid tidpunkten för studien.

Antal observationer n = 757. Den streckade linjen representerar z = 1,96.



Figur 1. z-värdenas fördelning i de fall där ett partis sympatistöd jämförts med a) skattningen i den senaste mätningen b) partiets röststöd i det senaste valet eller c) gentemot fyraprocentsspärren.

Antal observationer n = 121. Den streckade linjen representerar z = 1,96.



Figur 2. z-värdenas fördelning i de fall där journalisten eller annan expert gett en förklaring till att partiets sympatistöd förändrats jämfört med föregående mätning eller jämfört med föregående riksdagsval.

Tabell 1. Illustration av andelen förändringar som är statistiskt signifikanta och vilken information som ges i relation till förändringen i artikeln (total procent).

	Ej signifikant	Signifikant
Nämner ingenting om statistiska osäkerheter i texten	Typ I Fel 41 %	Korrekt 22 %
Nämner att förändringen inte är statistisk säkerställd	Korrekt 26 %	Typ II Fel 2 %
Nämner att förändringen är statistisk säkerställd	Typ I Fel 1 %	Korrekt 8 %

Tabellen bygger på antalet observationer då ett partis opinionsresultat jämförts med resultatet i föregående mätning, föregående riksdagsval eller gentemot fyraprocentspärren (n = 757).

fört med föregående mätning eller jämfört med föregående riksdagsval. För att undersöka i vilken utsträckning som dessa jämförelser och förklaringar av opinionsförändringar förhåller sig till konventionen om statistisk säkerhet på 95-procentsnivån har vi använt opinionsdata från samtliga undersökningar som ligger till grund för artiklarna. Därefter har vi analyserat i vilken utsträckning som journalisterna rapporterar om materialet på ett korrekt sätt.

Av figur 1 framgår fördelningen av z-värden för samtliga 757 tillfällen där ett partis opinionsstöd beskrivits som skilt från föregående mätning, riksdagsval eller där man jämfört opinionsstödet med fyraprocentsspärren. Mer specifikt visar den att 68 procent av alla jämförelser har ett z-värde under 1,96, vilket är gränsen för förändring inom säkerhetsmarginalen (med 95 procents säkerhet).

I figur 2 visas z-värdenas fördelning för de gånger då man i artikeln gett en förklaring till varför ett partis opinionssiffror har gått upp eller ner. Av de 121 gånger där journalisten presenterar en förklaring till varför ett parti har gått upp eller ner så har 55 procent ett z-värde under gränsvärdet 1,96. Histogrammet i figur 2 skvallrar om att det inte är ovanligt att man lägger fram förklaringar även till oerhört små förändringar.

Från ett journalistiskt perspektiv kan man möjligen argumentera för att hållningen i figur 1 är alltför konservativ och oförenlig med journalistiska normer att förmedla information. Att ge förklaringar till förändringar som befinner sig inom felmarginalen och som ur ett statistiskt perspektiv utgör brus förtjänar dock hård kritik: om en förändring inte är statistiskt säkerställd kan den heller inte förklaras.

Om man ser till helhetsbilden och hur ofta de undersökta tidningarna förmedlar en rättvisande bild till läsarna visar våra resultat att tidningarna i 44 (= 41+2+1) procent av de undersökta fallen ger undermålig eller felaktig information. Det framgår av tabell 1. Den vanligaste typen av fel är att man rapporterar om förändringar som inte är signifikanta utan att nämna något om statistisk osäkerhet.

Slutsatser

Att journalister brister när de rapporterar om partisympatiundersökningar är kanske föga förvånande givet journalistikens nyhetsvärderingar och nyhetsmedielogiken, men det är ändå bekymmersamt. Ytterst är journalistikens syfte att tillhandahålla sådan information som människor behöver för att fritt och självständigt kunna ta ställning i samhällsfrågor, och då är ett grundkrav att den information som förmedlas är verifierad och korrekt. När journalister väljer att skapa nyheter baserade på uppgifter som mycket väl kan vara osanna – vilket det handlar om när en förändring inte är statistiskt säkerställd – utan att poängtera att förändringen är så pass liten att den kan uppstå av ren slump brister journalistiken vad gäller kraven på såväl korrekthet som källkritik och transparens. Då handlar det inte om att rapportera om fakta. Då handlar det om att rapportera om något som mycket väl kan vara fiktion.

PER OLESKOG TRYGGVASON,
DOKTORAND VID INSTITUTIONEN FÖR JOURNALISTIK, MEDIER OCH KOMMUNIKATION, GÖTEBORGS UNIVERSITET

JESPER STRÖMBÄCK,
PROFESSOR I JOURNALISTIK OCH POLITISK KOMMUNIKATION VID INSTITUTIONEN FÖR JOURNALISTIK, MEDIER OCH KOMMUNIKATION, GÖTEBORGS UNIVERSITET

Källa:

Oleskog Tryggvason, Per & Strömbäck, Jesper (2017). Fact or Fiction? Investigating the Quality of Opinion Poll Coverage and its Antecedents. *Journalism Studies*, online early.

Donald Trump under talet inför USA:s samlade kongress den 28 februari 2017. I bakgrunden syns vicepresident Mike Pence och talman Paul Ryan.

OPINIONSUNDERSÖKNINGAR

Före och efter Trump

Opinionsundersökningar är viktiga i en demokrati. De visar vad medborgarna tycker och ger beslutsfattare viktiga insikter och kunskaper om attityder och åsikter hos allmänheten. Ett område som är viktigt att undersöka är politiken. Går de folkvalda i takt med väljarna? I flera fall de senaste åren uppfattas väljarundersökningar i andra länder ha misslyckats med att förutspå valutgångar. Det gäller såväl det brittiska valet 2015, Brexit och presidentvalet i USA hösten 2016. I andra länder och val har mätningarna fungerat mycket bra, men hur som helst har vissa media uttalat kris för opinionsundersökningar och vi själva har i varierande grad rannsakat oss själva. Men är det kris och vad kan vi i så fall göra åt det?

»I flera fall de senaste åren uppfattas väljarundersökningar i andra länder ha misslyckats med att förutspå valutgångar. »

Undersökningsbranschen befinner sig i ett paradigmskifte och i USA förklarade välrenommerade företag som Gallup och Pew Research tidigt att det var för svårt att mäta presidentvalet förra hösten med hänsyn till sjunkande svarsfrekvenser och behov av nya metoder, och man valde därför att inte genomföra några mätningar om kandidaternas opinionsläge. Men vad gör vi medan vi väntar på ”de nya metoderna”? De kommer sannolikt att växa fram och utvecklas över tiden på samma sätt som de traditionella sannolikhetsurvalen – när man gick ifrån totalundersökningar till stickprovsundersökningar med en inledningsvis stor skepsis. Stickprovsteorin utvecklades till en seriös och helt accepterad vetenskap.

De flesta andra institut gjorde undersökningar, och med något undantag förutspåddes Clinton att bli nästa president. Efter Trumps överraskande seger intensifierades arbetet med metodutveckling både inom akademien och privata företag, där branschorganisationen AAPOR (American Association of Public Opinion Research) är drivande. AAPOR publicerade nyligen rapporten "An Evaluation of 2016 Election Polls in the U.S.", som man får tillgång till genom AAPOR:s hemsida. »»»

»»» OPINIONSUNDERSÖKNINGAR FÖRE OCH EFTER TRUMP

I Storbritannien valde man att tillsätta en kommission efter det misslyckade brittiska valet 2015, ledd av Patrick Sturgis, professor vid University of Southampton och Director of the NCRM (National Center for Research Methods). Kommissionen publicerade i mars 2016 rapporten ”Report of the Inquiry into the 2015 British general election opinion polls”, som man får tillgång till genom NCRM:s hemsida. I denna konstaterades att det är en utmaning för opinionsundersökarna att kunna skapa en bra sammansättning på urvalet så att vissa undergrupper blir representerade. Även valdeltagandet utgör en osäkerhetsfaktor. Om valdeltagandet sammantaget sjunker något kan det samtidigt vara så att i vissa grupper ökar valdeltagandet medan det i andra grupper är svårt att mobilisera och få folk att gå till valurnorna.

Karin Nelsson.

Sturgis säger i rapporten ”*Det har förekommit många teorier och spekulation om vad som gick fel 2015, men efter att ha övervägt de tillgängliga bevisen, har kommissionen kommit till slutsatsen att den främsta felkällan till att undersökningsinstituten missade att förutspå rätt resultat är de sätt på vilka urvalen är konstruerade.*”

»Många av de metoder som används idag av instituten skapades för över 50 år sedan, och det kanske är dags att revidera och utveckla dessa? »

Väljarnas rörlighet är ett annat problem. Det närmaste exemplet på detta ser vi i det nyligen genomförda valet i Frankrike. På ett år har Macrons parti En Marche! gått från att vara ett nystartat parti till att bli det överlägset största franska partiet.

Situationen i Sverige med traditionellt stabila väljarsympatier har förbytts i stor rörlighet i väljarkåren, vilket visas hos samtliga institut som mäter detta. Det gäller i synnerhet denna sista mandatperiod. Det har också blivit vanligt att rösta på olika partier i de lokala och nationella valen. Gruppen ”kärnväljare” minskar. I skrivande stund har hittills 16 nya partier registrerat sig inför nästa års val. Kan något av dessa vara ett nytt En Marche!?

Internet och digitaliseringen kan vara en förklaring. För vissa väljare tycks det inte mycket annorlunda att uttrycka partisympati än att ”gilla” något på Facebook.

Många av de metoder som används idag av instituten skapades för över 50 år sedan, och det kanske är dags att revidera och utveckla dessa? Ingen ska tvivla på att det är branschens yttersta vilja att kunna göra bra och tillförlitliga mätningar. Andra typer av aktörer provar också att mäta väljarsympatier. Det är allt från vadslagningsfirmor, sociala medier som Facebook och Twitter till den senaste snackisen Cambridge Analytica, som kombinerar psykografiska metoder med aktivitet på sociala medier. Hittills har framgången varit begränsad men uppfattas ha stor utvecklingspotential. (Beror naturligtvis på vem som uttalar sig.)

I Sverige är förutsättningarna för undersökningar om partisympatier sammantaget bättre än i många andra länder. Vi har trots allt högt valdeltagande, god officiell statistik och vårt valsystem gör det lättare att mäta än i andra länder.

Undersökningsbranschen befinner sig alltså i en spännande utveckling. Behovet av och intresset för opinionsundersökningar ökar. Det faktum att samtidigt med denna utveckling allt färre personer vill svara på undersökningar kräver fortsatt utveckling kring urvalsmetoder och sätt att justera skevheter.

Det är en grannlaga uppgift att göra opinionsundersökningar, speciellt som nu i en situation där vi som bransch ifrågasätts. Det är dock för drastiskt att säga att undersökningsbranschen är i kris i Sverige. Det är viktigare än någonsin att vi samarbetar kring metodutveckling och ökad transparens för att stärka vår ställning och vårt förtroende bland medier och allmänhet. Medlemmarna i Statistikfrämjandet och Surveyföreningen har en självklar och viktig uppgift att verka för att öka förtroendet för vår bransch.

KARIN NELSSON
SENIOR PARTNER, INIZIO

(Något bearbetad version av artikel tidigare publicerad i Qvintensen webb på Statistikfrämjandets hemsida.)

OFFICIELL STATISTIKPRODUKTION

Statistiska undersökningar och statistisk inferens

Vi artikelförfattare anser att statistisk inferens bör få en tydligare plats än nu i produktion och presentation av (officiell) statistik. Sådana synsätt och resonemang kan underlätta val av undersökningsdesign, skattningsförfaranden samt uppskattningar och bedömningar av osäkerheter. Den här artikeln har sin utgångspunkt i ett arbete på SCB om design av statistiska undersökningar.

1. Inledning: statistikproduktionsprocessen

Det var förmodligen Nya Zealand som 2006 var först med en så kallad processkarta för statistikproduktion. SCB tog snabbt efter idén 2007. Nu använder många upplägget, se UNECE (2013) för version 5 av den *Generic Statistical Business Process Model* (GSBPM) som har arbetats fram internationellt.

SCB:s modell har på den översta nivån följande åtta processer:

1. Fastställ behov
2. Designa och planera
3. Skapa och testa
4. Samla in
5. Bearbeta
6. Analysera
7. Redovisa och kommunicera
8. Utvärdera och återkoppla

Arbetet är inte så linjärt som listan antyder. Modellen är ändå en praktisk sammanhållande beskrivning. Processerna 1–3 utgör förberedelser där bl.a. användarbehov, undersökningsdesign och produktionssystem utreds och bestäms. Processerna 4–7 beskriver genomförandet från insamling till redovisning. Process 8 är en utvärdering och återkoppling för framtiden.

Den resulterande statistiken – som avser exempelvis befolkning, bebyggelse, priser, handel och nationalräkenskaper – är främst deskriptiv. Den presenteras ofta i form av en tabellplan med tabeller som, direkt eller indirekt, avser populationer av t.ex. individer, fastigheter, företag eller geografiska områden. Varje tabellcell i de i förväg planerade tabellerna tänks innehålla ett okänt tal, en ”parameter”, vars värde ska skattas; sanna parametervärden antas existera.

Avsnitten 2–4 ger exempel från statistikproduktion och tillhörande statistiska synsätt. Avsnitten 5–7 diskuterar statistisk inferens och hur denna kan illustreras. Benämningen statistikproducent används för SCB och andra som tar fram, eller ansvarar för, officiell statistik.

2. Data och datainsamling

En statistisk undersökning kan exempelvis samla in data direkt från individer eller företag *eller* hämta administrativa eller andra befintliga data från ett register *eller* hämta makrodata som t.ex. statistik på fin nivå för vidarebearbetningar. En kombination av uppgiftskällor är vanligt förekommande.

För en statistiker kan det vara naturligt att se data som genererade på något sätt sådant att sannolikheter ingår. Det kan gälla

- en beskrivning av ett urvalsförfarande
- en beskrivning av erhållna mätvärden jämfört med sanna värden eller av svarsbenägenheter i en enkätundersökning
- en modell med slumpmässiga (stokastiska) inslag som beskriver hur vissa variabelvärden genererats utifrån andra.

Med ett sannolikhetsurval i en urvalsundersökning finns en slumpmekanism som statistikproducenten rår över. Bortfallsmekanismen styrs inte av statistikproducenten, men med en förhandskunskap om den mekanismen kan motverkande förhandsåtgärder och bortfallsuppföljning planeras med hänsyn till kvalitet och kostnader. Motsvarande resonemang kan föras för andra osäkerhetskällor, t.ex. mätning, med hjälp av teori för svarsprocesser och förhandstest.

3. Urval och estimation

Runt urvalsundersökningar och urvalsförfaranden finns mycket teori och flera statistiska skolor. Ett vanligt förekommande synsätt på skattning kallas **designbaserat**. Då är det den slumpmässiga urvalsdragningen med urvalssannolikheter som dominerar utformning av skattningar och beräkning av tillförlitlighet. Det finns en utvidgning där modeller tas till hjälp utan att frångå grundprincipen. Den metoden kallas **modellassisterad** estimation. Kalibrering är ett typiskt exempel; en eller flera hjälpvariabler från register inkluderas i skattningen. En ursprunglig tanke med förfarandet var att minska variansen.

I dag används dock förfarandet även i syfte att minska risken för att skattningar snedvrids på grund av bortfall. Hjälpvariablerna är, enkelt uttryckt, kraftfulla om de ”fångar” variationer i undersökningsvariabeln och i svarsbenägenheten. Detta förfarande kan i allt väsentligt sägas vara **modellbaserat**, eftersom de skattningar som produceras baseras på (implicita) modellantaganden. Andra situationer där modellbaserade inslag ofta förekommer, åtminstone delvis, är vid imputering, skattning för små redovisningsgrupper samt vid tidiga skattningar.

Modeller och modellantaganden kan avse verklighetens oftast ändliga population eller ha en mer generell karaktär. Detta gäller ofta även för modeller som används i modellassisterade och modellbaserade skattningar. Den ändliga populationen kan t.ex. tänkas vara genererad stokastiskt från vad som brukar kallas en superpopulation. Målet för skattningarna kan vara karaktäristika hos den ändliga populationen. Målet kan alternativt vara parametrar i en modell, t.ex. parametrar som beskriver superpopulationen.

4. Registerbaserad statistik

Det är även vid totalundersökningar och registerbaserad statistikproduktion viktigt att tydligt uttala målet för parameterskattningar, t.ex. vilken population som skattningarna avser. Det kan finnas täckningsbrister, saknade data och andra osäkerhetskällor att hantera i skattningar och tillförlitlighetsbedömningar. Modellering och omräkningar/justeringar kan behövas, t.ex. för att härleda eller modellera statistiska variabler utifrån de administrativa variabeldefinitionerna och för att räkna om från administrativa objekt till de objekt som definieras för statistiken.

Administrativa data som har samlats in av någon annan än statistikproducenten kan vara svårare att hantera och kvalitetsbedöma än direktinsamlade data som statistikproducenten mer styr och kontrollerar själv.

5. Statistisk inferens

Statistisk inferens är enligt Nationalencyklopedin en induktiv vetenskap där man drar slutsatser ur empiriska data under en osäkerhet orsakad av slumpmässighet i data. Denna definition framhåller erfarenhet, osäkerhet och resonemang; det är inte en säker härledning (deduktion).

Att den officiella statistiken främst är deskriptiv hindrar inte att det är en statistisk inferens – mer eller mindre explicit – som sker: från empiriska data till mer allomfattande parametrar som skattas med hänsyn tagen till osäkerheter. Det är i första hand statistikanvändarna som gör jämförelser och analyser. De vill kunna dra slutsatser, t.ex. om likheter och olikheter vid jämförande studier mellan grupper och över tid. För att statistiken ska kunna tolkas och användningen bli meningsfull behöver information finnas om statistikens kvalitet, t.ex. en beskrivning av dess innehåll (de parametrar som skattas), skattningarnas tillförlitlighet och möjligheter att använda skattningar tillsammans, särskilt från flera undersökningar.

För ändliga populationer finns det (ännu) inte en allmänt vedertagen princip för inferens i situationer med data som är ”totalmaterial” eller näst intill – det vore önskvärt. Zhang (2012) ger ett förslag till teoribildning.

6. Inferens och statistisk teori – en illustration

Statistisk teori har allmänt två huvuddelar, nämligen sannolikhetslära och statistisk inferensteori (med skattning och test). Detta kan ses som statistiska beräkningar åt ”två håll”. Det ena bedömer och beskriver i förväg hur det kan tänkas bli, och det andra går från observerade data till något mer allmängiltigt.

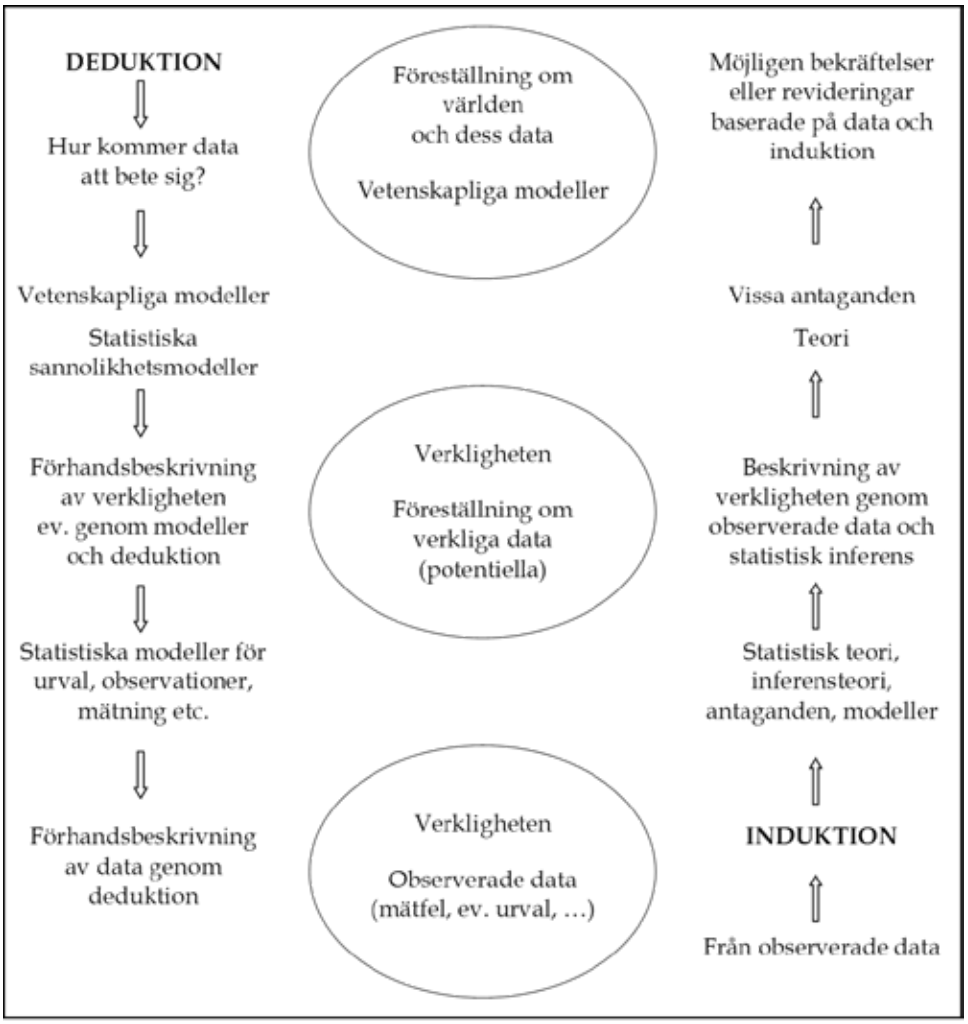
Figur 1 nedan är inspirerad av figurer om inferens i Barnett (1999; sidan 8), Makar och Rubin (2009; sidan 85) samt Kass (2011; sidorna 5 och 6). De två typerna av statistiska beräkningar går ”nedåt” respektive ”uppåt”, och de avser tre tillstånd som åskådliggörs genom ovaler.

Den mellersta och den nedersta ovalen avser data i verkligheten om t.ex. befolkning, bebyggelse och priser. Den mellersta ovalen ”innehåller” faktiska data som helhet utan praktiska begränsningar. Exempelvis finns variabler där, oavsett om de är observerbara eller inte. Den nedersta ovalen innehåller observerade data. Den översta delen av figur 1 är mer abstrakt. Den ovalen är övergripande eller vetenskaplig, och den avser en abstrakt tänkbar värld. Denna kan vara intressant i sig eller vara ett hjälpmedel för att beskriva verkligheten i figurens mitt. Exempelvis kan en tänkt slumpmekanism ingå som en beskrivningsmodell för verkligheten.

Till **vänster** i figur 1 illustreras ”förflyttningar” med två huvudsakliga steg. Den övre går från den abstrakta tänkta verkligheten till förhandsbeskrivningen av den faktiska verkligheten, medan den nedre går från denna faktiska verklighet till observationer. I den nedre förflyttningen ingår praktiska förfaranden som ramframställning (för att nå de objekt som på ett eller annat sätt ska bidra med variabler till datainsamlingen), urvalsdragning (när sådan görs), mätförfarande (t.ex. genom ett formulär som fylls i) etc. Denna förflyttning kan ses som en form av härledning (deduktion) i förväg av data som kommer att observeras med hänsyn tagen till de samband och osäkerheter som finns. Här ingår t.ex. sannolikhetsteori som används i förhandskalkyler av resursallokering, bortfall, mätosäkerhet eller ett urvalsförfarande. Syftet är att i förväg bedöma hur den statistiska undersökningen bör utformas och hur tillförlitlig statistiken kan väntas bli. Den övre huvudförflyttningen kan vara ett hjälpmedel vid utformningen av den nedre.

Till **höger** i figur 1 illustreras omvända förflyttningar, först från observerade data till verklighetens helhet, typiskt med utsedda statistiska parametrar som ska skattas. Det kan sedan finnas ytterligare en förflyttning upp till den tänkta världen. Induktionen ”räknar i motsatt riktning” jämfört med vad som sker på figurens vänstra sida. Den beaktar således de osäkerheter som finns och söker skattningar av de utsedda parametrarna, skattningar med goda egenskaper som ingen eller liten bias (skevhet, systematisk avvikelse) och liten varians. För att kunna prata om bias och varians måste det finnas en beskrivning av systematiska och slumpmässiga avvikelser. Parametrarna och den statistiska inferensen kan avse den mellersta ovalen eller den översta ovalen, beroende på statistikanvändares intressen och statistikproducentens val. Den officiella statistiken utgörs i allmänhet av skattningar avseende den mellersta ovalen.

När statistikanvändare gör jämförelser och andra analyser kan de vara inriktade



Figur 1. Statistisk inferens, inklusive deduktion och induktion.

på populationer i verkligheten och på förhållanden i dessa ofta ändliga populationer. Ofta är de intresserade av svar på frågan *hur många?* men det kan även gälla mer komplexa förhållanden som samband av olika slag. Ibland använder de modeller i och för analyser. Det kan t.ex. vara en ekonometrisk modell eller en regressionsmodell med ett slumpmässigt (stokastiskt) inslag i modellen. Sådana modeller kan ha en vidare karaktär än just den ändliga population som undersökningen avser. Inte sällan gäller det undersökningar som genomförs i syfte att studera underliggande fenomen snarare än förhållanden vid en viss tidpunkt – parametrarna i en regressionsmodell kan t.ex. avse en superpopulation snarare än den ändliga populationen. Ofta, men inte alltid, är frågan *varför?* av intresse; se t.ex. Deming (1953) eller Kalton (2002, avsnitt 2.5). De två frågorna kan även särskiljas genom att använda orden *beskrivande* respektive *förklarande*. Det kan i den förklarande situationen vara naturligt att arbeta med hela figuren, inklusive den översta abstrakta delen.

7. Avrundande diskussion

Vi författare anser att statistisk inferens bör få en tydligare plats än nu i produktion och presentation av (officiell) statistik. Det gäller särskilt i en tid när många statistikproducenter är under stort tryck att anamma nya arbetssätt, datakällor och metoder i syfte att producera mer och bättre användaranpassad statistik till lägre kostnad.

I arbetet med att lyfta fram och tydligare betona statistisk inferens kan den modell som beskrivs i figur 1 i föregående avsnitt tjäna oss väl. Det som finns på den vänstra sidan av figuren fyller en viktig funktion i designen av en statistisk undersökning. Ju mer förhandskunskap, desto bättre möjligheter att göra en god utformning med hänsyn till tillförlitlighet och risker för avvikelser. Även skattningsförfarandet – den ”återvändande” högersidan – vinner på att beakta osäkerhetskällorna och dessas troliga systematiska och slumpmässiga effekter.

De syn- och arbetssätt som visas i figur 1 kan således underlätta val av undersökningsdesign, skattningsförfaranden samt uppskattningar och bedömningar av osäkerheter. Utöver att vara användbara för statistikproducenter, kan de även, direkt eller indirekt, underlätta kommunikationen med statistikanvändare.

EVA ELVERS OCH MARTIN AXELSON
STATISTISKA CENTRALBYRÅN (SCB)

Referenser

Barnett, V. (1999). *Comparative Statistical Inference.* Third Edition. Wiley.

Deming, W.E. (1953). On the Distinction between Enumerative and Analytic Surveys. *Journal of the American Statistical Society*, 48, 244–255.

Kalton, G. (2002). Models in the Practice of Survey Sampling (Revisited). *Journal of Official Statistics*, 18, 129–154.

Kass, R.E. (2011). Statistical Inference: The Big Picture. *Statistical Science*, 26, 1–9.

Makar, K. och Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8, 82–105.

UNECE (2013). Generic Statistical Business Process Model. Version 5.0 – December 2013. The United Nations Economic Commission for Europe (UNECE).

Zhang, L.-C. (2012). Topics of statistical theory for register-based statistics and data integration. *Statistics Neerlandica*, 66, 41–63.

Rolf Sundberg utsågs av Statistikfrämjandet till årets statistikfrämjare. Här berättar han om hur han blev statistiker och sedan fortsatt att jobba inom detta område. Dessutom synpunkter på kurser, forskning, teori/tillämpningar och bredd/djup i forskningen.

Minnen och meningar

Hur mycket betyder slumpen för oss? För mig som statistiker är idén om något slags slumpmässighet i data förstås fundamental och nödvändig, men här tänker jag på de tillfälligheter som knuffar våra liv i olika riktningar. Var vi råkar hamna, vilka vi råkar träffa, och vilka mer eller mindre välgrundade beslut vi tar.

Det första viktiga beslutet för min utbildning var att som 10-åring lämna den kända världen i Fredhälls folkskola för att börja i realskolan i Kungsholmens läroverk. Jag ville/vågade egentligen inte, som jag minns det, men övertalades av mina föräldrar, som själva aldrig hade fått chansen. Gymnasietiden där ledde nog vetenskapligt framåt, fast jag inte förstod det då. Klassen var ett bra kamratgäng, som fortfarande träffas var femte år eller så. Jag har senare förvånat konstaterat att fyra ur min klass blev professorer. Jag hade ett par utmärkta, stimulerande och mycket betydelsefulla matte-fysik-lektorer, och när jag lämnade skolan var jag inställd på att bli som dem – så långt sträckte sig mina vyer. Teknisk fysik på KTH valde jag bort, i tron att man då skulle vara mer praktiskt-tekniskt lagd – så var det inte, insåg jag senare. Felbeslut? Kanske. Kamratinflytandet var dock också viktigt och bidrog till att jag började läsa matte vid SU hösten 1962, den naturliga inledningen till fysikstudier.

»Jag ville ha något mer verklighetsnära, med matematiken som verktyg i stället för självändamål.»

Jag var inte ensam nybörjare det året, vi var väl 500 eller så, och antalet nybörjare växte ytterligare de närmaste åren. Det förde med sig att institutionen behövde många lektionsledare/övningsassistenter, och jag blev tillfrågad och rekryterad redan efter första året, och blev kvar vid matematiken. Det var ju smickrande, kul, lärorikt, och gav vänner och lön, men det gick inte ihop med mina tidigare planer på fysikstudier. Det relativt nya ämnet numerisk analys (inkl. programmering) valde jag också bort, för det sades vara så jobbigt och tidskrävande. Det kanske var dumt, men i stället blev matematisk statistik ett naturligt komplement till den rena matten. Det var sigma-algebror från allra första början i den teoretiska statistiken på den tiden, men även tillämpad statistik. Den skriftliga tentamensdelen hade två delar: en första dag med teoriproblem och en andra dag med ett mera praktiskt problem, som löstes genom beräkningar med Facit-snurror (utvecklingen hade då hunnit till elektriska sådana räknemaskiner). Den obligatoriska försäkringsmatematik som då ingick, fick mig att skippa den inriktningen för all framtid. Det som roade mest var sannolikhetsteorin; man läste inledningsvis "lilla Cramér" och "Feller vol 1", förträffliga böcker (Feller hade f.ö. varit en betydelsefull person vid Cramérs Institut, som gäst i flera år strax före världskriget). "Stora Cramér" (Mathematical methods…) dominerade tredje terminen i ämnet.

Men mat-stat var länge bara ett sidoämne, och jag blev forskarstuderande (lic-student) i matte en termin eller två (utan handledare, och utan forskningsämne – inget konstigt med det, då), innan jag insåg att jag varken hade tillräckligt intresse för ren matematik eller tillräcklig fallenhet för det. Jag ville ha något mer verklighetsnära, med matematiken som verktyg i stället för självändamål. Jag tog 1968 det kloka och viktiga steget att sadla om till att bli forskarstuderande i mat-stat (de kurser som fanns hade jag redan tenterat, och jag fick ett lic-stipendium). Efter en termin tog jag ett avgörande ytterligare steg i den riktningen – jag tog tjänst i Statistiska forskningsgruppen (SFG), dvs "Institutets konsulterande verksamhet", som det mer formellt benämndes. Sedan dess har jag varit lika engagerad i statistikvetenskapens tillämpningar som i dess teori. Vid SFG blev jag kvar under hela min forskarutbildningstid, även om jag periodvis hade reducerad tjänst för att hinna skriva på min avhandling (doktorandtjänster fanns inte på den tiden).

Institutets professor var framlidne Ulf Grenander, men han hade börjat tillbringa terminerna i USA och var väsentligen i Sverige bara under somrarna. Det blev några lågvattenår för Institutet, utan professor. Jag fick ett förslag till ett obetydligt forskningsämne av Grenander, i samband med Statistikersamfundets första sommarskola (den i Kall). Det föreföll inte särskilt tacksamt. Jag tog mod till mig och pratade med Per Martin-Löf, som instämde och föreslog att jag i stället skulle ägna mig åt ofullständiga data från exponentialfamiljer, där han tidigare hade tagit fram några första intressanta resultat. Det var ett mycket bättre ämne, men det gick trögt i början, på egen hand, eftersom även Per for till USA, dock bara för ett år. Det var före tillkomsten av webb, e-post och fax, så kommunikation var sparsam, och jag minns inte att jag ens växlade några brev med Per. Jag arbetade först med det som senare kom att kallas EM-algoritmen. Numeriska beräkningar var tidsödande – man skrev ett Algol- eller Fortran-program, gick till Stockholms datacentral, stansade hålkort, lämnade in ett jobb som kördes under natten, och hämtade resultatet dagen därpå. Tråkigt när detta var meddelandet "Syntax error".

Två helt olika händelser blev betydelsefulla: Dels införde regeringen forskarutbildning med doktorsexamen, i väsentligen nuvarande form, dels hade Per M-L disputerat (i matte, för den gamla doktorsgraden) och han fick en docenttjänst i mat-stat. Det blev bara för ett läsår, 1969–70, men då höll han sin berömda

föreläsningsserie "Statistiska modeller", som jag skrev föreläsningsnoter till (och han läste och rättade). Långsamt hade jag mognat till forskare, och nu lossnade det för mig i avhandlingsarbetet också. Våren 1972 disputerade jag, landets förste i mat-stat enligt den nya ordningen.

Jag sökte en utannonserad tjänst vid SCB:s metodenhet. Det var nära att jag fick den, men slutligen valde man Jan Hagberg, som disputerat vid KTH strax efter mig med ett mer samplinginriktat tema. Där tror jag det var nära att mitt liv tog en annan inriktning, men jag sörjde nog inte ens då över att jag blev kvar i universitetsvärlden.

Begreppet "post-doc" var ännu inte etablerat, men terminen efter disputationen tillbringade jag i Aarhus, som "adjunkt-vikar". Den teoretiska statistiken stod redan då på en fantastiskt hög nivå i Danmark, med många briljanta unga statistiker, så det var bra att få danska kontakter. Exempelvis ledde det till att jag blev "diskutant" vid Nordiska konferensen 1973 (som arrangerades i Danmark), vilket i sin tur resulterade i en artikel om dekomposabla modeller för kontingenstabeller, där jag bevisade en förmodan formulerad av konferensföreläsaren (Anders Holst Andersen).

Inför vt 1973 behövde Bengt Rosén en vikarierande lektor vid KTH. Det var brist på kvalificerat folk, och så var jag hastigt anställd där. Efter något år lystes lektoratet ut, och blev mitt. I tio år var jag vid KTH och lärde mig många nyttigheter, som hade saknats i min forskarutbildning, t.ex. försöksplanering, tillförlitlighetsteori, kvalitetskontroll och samplingteori. Framför allt engagerade jag mig i en kurs i linjära modeller och försöksplanering, vars kompendium i reviderad form fortfarande används (men numera bara vid SU), samt i en helt ny kurs för kemisektionen, där jag också skrev kompendier. De senare byggde på rätt grundliga undersökningar av vad kemitekniker kunde ha nytta av, och jag använde erfarenheterna också i externa kurser, t.ex. för Träforskningen, Cement- och Betonginstitutet, SUPRA, Pharmacia och Astra, som alla i gengäld fick bidra med exempel till undervisning och kompendier.

Min forskning fick ett uppsving efter att Phil Brown 1981 stannat till vid KTH på väg mellan Finland och England. Han höll ett seminarium om sin nyodlande forskning i multivariat kalibrering. Jag hade redan under min SFG-tid hjälpt Sveriges Geologiska Undersökning med ett kalibreringsproblem, och kemitillämpningar intresserade mig, så jag hakade på, och Phil och jag kom att arbeta rätt mycket ihop under resten av 80-talet.

»Min önskan att kombinera teori och tillämpningar i utbildningen har lett till att jag infört projektinslag i åtskilliga kurser, vanligen som gruppaktivitet.»

Ett par år senare blev det ett bra tillfälle att återvända till mitt "Alma Mater". Mat-stat SU lyste ut en docenttjänst, jag sökte och fick den (1983). Tid att koncentrera sig på forskning! Nej, det visade sig att en docent på plats gjorde att Bengt von Bahr, på professuren, tyckte att han kunde realisera sina planer att lämna tjänsten för att gå till försäkringsbranschen. För min del innebar det flera år som tf professor och avdelningsföreståndare (som vi skulle säga nu), innan Anders Martin-Löf kom från försäkringsbranschen. Jag hade senare en forskningsrådstjänst en tid, och när möjligheten uppstod att bli befordrad, så blev jag utnämnd till professor. Mest rörde min forskning multivariat kalibrering och regressionsmetoder, men periodvis dominerade annat, t.ex. stereologi en tid.

Min önskan att kombinera teori och tillämpningar i utbildningen har lett till att jag infört projektinslag i åtskilliga kurser, vanligen som gruppaktivitet. Jag införde det på KTH i regressions- och variansanalys samt i statistik för kemister redan på 70-talet, och jag har fortsatt på SU. Det tar tyvärr tid från den traditionella undervisningen och kan kräva mycket handledning och annat extraarbete, men jag anser att nyttan gör det motiverat i många statistikkurser. Grundutbildningens krav på graderad betygsättning kan dock kännas svårhanterligt i detta fall. Projektarbeten kan för övrigt gärna utformas så att de även ger träning i muntlig och skriftlig framställning.

Sedan länge hade jag haft vaga tankar på att vi borde gå ett steg längre och ha en regelrätt utbildning i statistisk konsultmetodik. Hösten 1998 gjorde jag ett första försök med ett par forskarstuderande. Kursen gick väsentligen ut på att ge erfarenhet från möten med klienter, att aktivt hjälpa en klient med något problem och skriva en rapport till klienten, och att diskutera litteratur om kommunikation och statistisk metodik med tanke på konsultrollen. Försöket utföll väl, och sedan dess har jag kört en sådan kurs mest varje år. Efter en första kursomgång vågade jag även släppa in masterstudenter, och därmed ökade underlaget för en årligen återkommande kurs. Ytterligare underlättades det när vi fick fakultetsanslag för statistikhjälp till naturvetare. Det gav ett tacksamt tillflöde av klienter till kursen. Det finns mycket mer att säga om kursen, men jag hänvisar till min artikel i Qvintensen 2013, "Kan man lära till statistisk konsult på en kurs?".

Konsultmetodikkursen anknyter till min filosofi för statistikutbildning och statistisk forskning, så den försvinner inte ur bilden, när jag nu går över till det ämnet. Statistisk teori och tillämpningar är lika nödvändiga för ämnet, och för en sund utveckling av vårt ämne skall blivande statistiker tillgodogöra sig båda aspekterna i sin utbildning. Jag brukar citera Terry »»»

Speed, hur han säger sig bedriva statistisk forskning: "A very large amount of consulting, quite a bit of collaboration, and some research". Ett annat föredöme jag har är David Cox, för hans djupa insikter och ofantliga bredd. Första tillfället jag minns honom från är Nordiska konferensen i Mariehamn 1980, då jag avundsjukt tyckte han ställde en så bra fråga till Jöreskog efter dennes svåra föreläsningar. Jag har hållit doktorandkurser på hans böcker "Cox & Hinkley" och "Cox & Oakes", samt på 2000-talet "Principles of Statistical Inference". Djup och bredd i kombination har även jag aktivt eftersträvat, på min blygsammare nivå.

I konsultverksamhet krävs bredd. Bredd i forskningen är dock förenat med viss svårighet. Om man koncentrerar sig på en problematik och gräver där man står kan man komma djupt, och bli känd inom det området, och för det området. Med det senare syftar jag på vår benägenhet att placera kollegor (och andra) i fack, något som jag också tyvärr gärna gör. Bredd i forskningen kräver upprepat en initial ansträngning att komma in i ett område, så väl att man kan publicera sig inom det, och sedan går man till helt andra lockande problem. En del av lösningen på det dilemmat kan vara samarbete, med andra statistiker eller med forskare inom andra ämnen. En positiv aspekt på bredd i forskningen är förstås potentialen till korsbefruktning. Ett konkret sådant exempel för min del är en stereologi-sejour, som byggde på ett par artiklar jag strax innan skrivit om modellbaserad inferens i samplingundersökningar (i sin tur motiverade av problem från Byggforskningsinstitutet). Eva Vedel Jensen i Aarhus, med stereologi som specialområde, hade upplevt ett oklart förhållande mellan designbaserat och modellbaserat synsätt i stereologins inferensmetoder, och hon kontaktade mig för samarbete. Vårt viktigaste arbete hade arbetsnamnet "kaffe-plet"-manuskriptet, efter ett missöde med en kaffemugg.

En del statistiska käpphästar har jag skaffat mig genom åren, kanske återvänder jag till sådana i några senare nummer av Qvintensen. Den som eventuellt vill hålla koll på mina pågående vetenskapliga aktiviteter (ett bokmanus om exponentialfamiljsmodeller, samarbeten med klimatforskare och med arkeologer) finner mig på Researchgate. På min SU-profilsida www.su.se/profiles/rolfs nämns förstås det, att jag av Svenska Statistikfrämjandet fått utmärkelsen Årets Statistikfrämjare. För den hedersbetygelsen är jag mycket glad och stolt, och den har ju motiverat den här lilla essän.

ROLF SUNDBERG
PROFESSOR I MATEMATISK STATISTIK,
STOCKHOLMS UNIVERSITET

Jan Hoem.

MINNES-ORD ÖVER JAN HOEM

■ **Jan Michael Hoem** har avlidit i Stockholm efter en tids sviktande hälsa, i en ålder av 77 år. I Sverige utmärkte sig Jan Hoem som grundaren av Stockholms universitets demografiska avdelning, SUDA, och som den främste representanten för demografiämnet i Skandinavien under 1980- och 1990-talen.

Jan Hoem var född i Norge i Kristiansund där han disputerade 1969 vid Oslo universitet i aktuariestatistik. Efter en period vid universitet i Köpenhamn rekryterades han 1981 till Sverige som professor i demometri. I Stockholm var han drivande i utvecklingen av statistiska metoder för att analysera demografiska processer och demografiska data. I tillämpad forskning studerade han bland annat hur sociala och familjepolitiska förhållanden påverkade fruktsamhetsutvecklingen i Sverige.

Från 1999 och tio år framöver var Jan Hoem direktör vid Max-Planck-institutet för demografisk forskning i Rostock. Under den tiden rekryterade han och utbildade en helt ny generation av europeiska demografer. Max-Planck-institutet blev en knutpunkt för europeisk demografisk forskning.

Jan Hoem har satt ett stort avtryck på utvecklingen av demografiämnet både i Sverige och internationellt. Svensk och nordisk demografisk forskning gynnas av tillgången till exceptionellt goda demografiska data, den europeiska forskningen gynnas av möjligheten till jämförande perspektiv på hur demografiska processer påverkas av olika sociala och politiska förhållanden i Europa. Jan Hoem har varit drivande i att utveckla ny forskning som bygger på dessa komparativa fördelar.

Med värme och generositet har han inspirerat unga demografer att riva många av de murar som tidigare låg som hinder för forskningssamarbeten mellan företrädare för olika vetenskapliga discipliner och länder. Sedan 2009 var Jan Hoem återigen bosatt Stockholm och fortsatt aktiv som emeritus vid SUDA. Han publicerade sin sista vetenskapliga artikel ett halvår före sin bortgång.

GUNNAR ANDERSSON
OCH KOLLEGORNA VID
STOCKHOLMS UNIVERSITETS
DEMOGRAFISKA AVDELNING

Efterlysta svenskar på ESRA:s konferens 2017

European Survey Research Association (ESRA) bildades 2005 för att stärka metodutveckling och forskning inom surveyområdet i Europa. Organisationens två viktigaste aktiviteter är konferensen som hålls vartannat år sedan 2005 och tidskriften Survey Research Methods (SRM) som kom ut med sitt första nummer 2007. I somras var det dags för den sjunde konferensen, denna gång i Lissabon, Portugal.

Konferenserna brukar alltid inledas med kortkurser på måndagen och därefter ”vanlig” konferens resten av veckan. I år inleddes konferensen med ett föredrag av professor Edith de Leeuw. Ibland kan sådana föredrag vara slätstrukna och tråkiga, men Edith stack ut hakan och föreslog att alla som använder matrisfrågor (grid questions) i webbenkäter borde piskas offentligt. Jag tror att detta bidrog till att frågor och diskussioner i samband med olika föredrag blev lite mer energiska och vågade. I vanliga fall är de flesta så försiktiga och hövliga att det sällan blir några riktiga debatter under sessionerna.

På de tidigaste konferenserna handlade många av presentationerna om European Social Survey (ESS), men senare har det blivit en bättre blandning av olika sorters undersökningar och deltagare från fler länder. De flesta pass har nu ungefär 15 parallella sessioner, så det kan vara svårt att välja. Ämnena handlar till exempel om hur man behåller panelmedlemmar i webbpaneler, hur man kan ställa frågor till barn och ungdomar, hur man kan ställa frågor om känsliga ämnen, nya tekniska lösningar för undersökningar, testmetoder för enkätfrågor och webbenkäter, översättning av enkätfrågor, användning av öppna frågor i webbenkäter, effekten av att blanda datainsamlingsmetoder, och så förstås frågan om matrisfrågor eller inte i webbenkäter… Flera presentatörer instämde först med Edith de Leeuw, men sedan kom Mick Couper, som (med risk för att bli piskad offentligt) presenterade resultat som visade att bra matrisfrågor kan fungera bra medan dåliga matrisfrågor fungerar dåligt. Som vanligt i enkätbranschen finns det inget självklart rätt eller fel svar – det finns helt enkelt både fördelar och nackdelar, och det gäller att välja rätt för just den studie och de respondenter som är aktuella.

ESRA-konferensen är en trevlig tillställning. Tillräckligt stor för att locka till sig många världsledande forskare men tillräckligt liten för att man ska kunna hitta och hinna prata med dem man vill. Jag träffade denna gång bara tre andra svenskar, men desto fler nyfikna utländska deltagare som undrade var alla svenskar var. Och det undrar jag också – metodintresserade surveystatistiker var håller ni hus?

MARIKA WENEMARK

Survey-föreningens kommande kvalitets-seminarium 2018

Surveyföreningen anordnar varje år ett så kallat kvalitetsseminarium. Nästa års kvalitetsseminarium kommer att äga rum torsdagen den 15 februari 2018. Det kommer att belysa kvalitet ur två olika synvinklar, båda väsentliga för undersökningsbranschen: frågekonstruktion och opinionsundersökningar.

Frågeformuläret är en av de absolut viktigaste hörnpelarna i en undersökning. Det utgör grunden för att man ska kunna mäta det man avser att mäta (validitet) och att resultaten ska kunna analyseras och tolkas på rätt sätt. Baserat på två nyutkomna böcker kommer detta att belysas ur såväl teoretiskt som praktiskt perspektiv. Marika Wenemark delar med sig av sina erfarenheter som hon har summerat i sin bok Enkätmetodik och Lina Fjelkegård pekar på highlights ur boken Frågor och svar från SCB, som hon har varit med och skrivit. Dessutom kommer Dan Hedlin (professor i statistik vid Stockholms universitet) bland annat att peka på fällor och fel hämtade från verkligheten.

Opinionsundersökningar – och då framförallt partisympatiundersökningar – är de undersökningar som främst påverkar hela undersökningsbranschens image, på både kort och lång sikt. Hur dessa mätningar hanteras av undersökningsföretagen och behandlas när de publiceras kan i stor utsträckning påverka förtroendet för branschen som sådan eller skapa misstroende mot såväl opinionsundersökningarna själva som mot helt andra typer av undersökningar. Hur de stora aktörerna bygger upp och hanterar sina opinionsundersökningar kommer att beskrivas av SCB, Sifo och Ipsos. Dessutom kommer vi att få en översikt över vad som görs i Sverige, samt insyn i det svenska valforskningsprogrammet från Göteborgs universitet. Vi avslutar med en paneldebatt där några av de främsta journalisterna som bevakar och kommenterar opinionsundersökningar deltar. De kommer bland annat att ge sin syn på de mätningar som kommer att bli aktuella inför valet 2018.

Praktiska upplysningar om det kommande kvalitetsseminariet (anmälning, kostnad, lokal med mera) kommer att läggas ut på Surveyföreningens hemsida: http://statistikframjandet.se/survey. Boka redan nu in den 15 februari i almanackan. Om du vill försäkra dig om en plats kan du redan nu förhandsanmäla dig till **ÅKE WISSING** via e-post: ake@wissing.se.

FMS

FMS firar 30 år
Föreningen för medicinsk statistik fyller i år 30 år, vilket medlemmarna kommer att påminnas om vid ett möte i Uppsala 23 november.

Riksstroke – ett svenskt kvalitetsregister för strokevård.

Marie Eriksson, ny ledamot i FMS styrelse

Jag har jag nyligen tillträtt som biträdande prefekt för enheten för Statistik, Handelshögskolan, Umeå Universitet. Jag är också föreståndare för det svenska kvalitetsregistret för strokevård, Riksstroke, som har sitt kansli på Norrlands Universitetssjukhus.

Efter en magisterexamen i matematisk statistik från Umeå universitet, gick flyttlasset till Södertälje och ett statistikerjobb på AstraZeneca (dåvarande Astra Pain Control). Där trivdes jag bra och jag lärde jag mig mycket av att arbeta med många statistikerkollegor. Jag började också med volleyboll och rekryterade en ”riktig” volleybollspelare till statistikernas lag Red Forearms. Stor framgång! Vi vann Astras internserie och jag gifte mig med volleybollspelaren. Det som från början var ett kortare vikariat blev till slut 5 år innan flytten gick norrut igen.

Väl tillbaka i Umeå arbetade jag som biostatistiker på det nystartade bioteknikföretaget UmanGenomics. Efter att ha arbetat på ett stort företag var det spännande att arbeta på ett mindre bioteknikföretag och ha det statistiska metodansvaret. Parallellt arbetade jag också deltid som statistikkonsult vid enheten för Medicin, Umeå Universitet. Efter några år ville jag ha nya utmaningar och disputerade i medicin 2008. Nu arbetar jag främst med registerbaserad forskning och leder ett forskningsprojekt där vi undersöker sociala ojämlikheter i strokevården, och ett projekt som handlar om förbättrade statistiska metoder för att mäta hälsa och vårdkvalitet.

Marie Eriksson.

Jag undervisar olika studenter, och en favoritkurs är statistik för läkarstudenter som ska göra sitt vetenskapliga arbete. Där får jag många frågor från studenterna. Jag försöker också hinna med några konsultuppdrag eftersom det är roligt och utvecklande. Sammantaget trivs jag väldigt bra med mina kanske lite spretiga arbetsuppgifter.

På fritiden är jag gärna med mina barn som nu hunnit bli 12 och 14 år. Jag har nyss förnyat mitt sportkort på IKSU och tänker att jag ska komma igång att träna, vi får se hur det går.

MARIE ERIKSSON

MATEMATISK STATISTIK SOM UNIVERSITETSÄMNE

– Vart är vi på väg?

Lite personlig ”nutidshistoria”

I början av 80-talet, när jag skulle söka in på universitetet, visste jag att jag ville läsa matematik. Efter att ha tagit del av en trevlig broschyr som låg i min gymnasieskolas bibliotek fastnade jag för matematikerlinjen på Stockholms universitet.

Under det första året läste vi kurser i matematik (analys och linjär algebra), datalogi och numerisk analys. År två fördjupades studierna i dessa ämnen och kompletterades med kurser i sannolikhetslära och statistik samt Markovprocesser. Sedan var det dags att välja inriktning för år tre. Jag fastnade för matematisk statistik, eftersom jag gillade matte mer än data och ville läsa något som var användbart och kunde leda till ett jobb. Ja, sedan utökades inriktningen med ytterligare ett år (det fjärde), och efter det fortsatte jag som doktorand i matematisk statistik. Men det är en annan historia.

Kanske känner du som läsare igen dig i min berättelse, även om du inte tillhör min generation? Mitt intryck är att man fortfarande kan göra ungefär som jag gjorde, även om det nu inte heter matematikerlinjen utan kandidatprogrammet i matematik, följt av något mastersprogram med inriktning mot matematisk statistik. Och de flesta av oss som har gjort så här tycker kanske att det är en naturlig ingång till att bli statistiker i någon form. Det finns förstås också andra vägar, till exempel genom att plocka ihop lämpliga kurser på ett civilingenjörsprogram. Så finns ju också ämnet statistik, med kurser där man inte har samma förkunskapskrav i matematik. Vissa lärosäten erbjuder statistikerlinjen, där statistikämnet kompletteras med kurser i matematik, programmering med mera. I denna text har jag valt att främst fokusera på matematisk statistik.

Om jag jämför de grundläggande kurser i matematisk statistik som jag läste i början av 80-talet med dem som jag undervisar på nu i Uppsala, slås jag av att innehållet egentligen inte skiljer sig särskilt mycket. Först ges grunderna i sannolikhetslära, som Kolmogorovs axiom, fördelningar med mera. Därefter går man igenom skattning, konfidensintervall och test. Man stöter fortfarande oftast på datamaterial med en storlek på högst ett par hundra observationer och är ganska fixerad vid signifikansnivån fem procent. I läroböckerna bryr man sig sällan om att diskutera masssignifikansproblemet, trots att detta fanns i statistikernas medvetande redan långt innan jag själv var student. För att få fram fördelningskvantiler använder man fortfarande ofta tabeller, trots att man sedan länge enkelt har kunnat använda datorer för detta ändamål.

Rolf Larsson.

Samtidigt ser vi att tillämpningarna av statistik utvecklas vindsnabbt. Den kraftigt ökande datorkraften har gjort det möjligt att lösa helt andra problem än man förr kunde drömma om. Numera kan mycket stora datamaterial analyseras med komplicerade ansatser och modeller. Detta kräver ett helt nytt tänkande hos en statistiker. Att anpassa en enkel modell och titta efter p-värden mindre än fem procent leder mycket sällan ända fram till målet. Ändå är det detta som vi fortfarande lär ut på våra grundläggande kurser.

Statistik och IT

Nuförtiden arbetar jag på matematiska institutionen i Uppsala. Nära intill oss håller den stora IT-institutionen hus. Vi samarbetar med dem inom det så kallade Centrum för Interdisciplinär Matematik (CIM). I praktiken innebär detta bland annat en och annan konsultation med forskare från andra delar av den teknisk-naturvetenskapliga fakulteten samt gemensamma lunchseminarier. Ämnena för seminarierna är ofta mer eller mindre av statistisk natur. När jag går dit slås jag ofta av hur många som är där och lyssnar. Ofta är det 20-30 stycken. Givet det faktum att vi på fakulteten bara är tre seniora personer som är statistiker till namnet finns det tydligen ett mycket större intresse för statistik än vad man skulle kunna tro. Hur kan det komma sig?

Naturligtvis är det den snabba utvecklingen på datorområdet som ligger bakom. Vid sidan av oss vanliga statistiker har det öppnats en forskningskultur på IT-sidan (även kallad maskininlärning) där man hanterar stora datamaterial. Synsättet förefaller vara mer algoritmiskt än hos oss, där vi gärna sätter den statistiska modellen i det första rummet. Språkbruket är också ett annat, till exempel ”tränar” man data när vi vanliga statistiker säger att vi skattar en modell.

Hur ska vi förhålla oss till detta? Det verkar dumt att sticka huvudet i sanden och fortsätta som vanligt. Istället borde vi uppmärksamma att det här finns en stor kompetens inom IT (t.ex. inom reglerteknik och beräkningsvetenskap) som vi statistiker kan ha nytta av att ta del av. Och omvänt vore det nog bra för IT-folket att tillägna sig lite av vårt tänkande när det gäller modeller och metoder inom statistisk inferens. För all del ser man redan nu mycket av det i litteraturen, alltså teoretiska analyser av komplexa metoder som MCMC, Lasso med mera. Så kanske är vi redan på rätt väg. Men det gäller att fortsätta att ”haka på tåget”.

Organisatoriskt vet jag att man på den statistiska institutionen i Linköping har gjort detta mer konkret. Inom institutionen för datavetenskap har man där en avdelning som heter ”statistik och maskininlärning”, där statistikerna håller hus. En liknande struktur fanns för några år sedan på Uppsala universitet. Institutionen hette informationsvetenskap, och bland de fyra avdelningarna fanns bland annat dator- och »»»

systemvetenskap (DSV) och statistik. Men i detta fall kom det aldrig till något forskningssamarbete mellan statistik och någon av de andra avdelningarna, och institutionen splittrades.

Tillämpad matematik och statistik

När Allan Gut pensionerades från sin professorsstol i matematisk statistik på Uppsala Universitet efterträddes han av Takis Konstantopoulos. Takis är en utpräglad probabilist, och det visade sig snart att han hellre ville rikta sig mot matematisk analys än mot statistisk inferens i sin forskning. Detta bidrog till att matematisk statistik bröts upp som forskningsområde på institutionen. Vi har numera tre så kallade forskningsprogram: ett i algebra, ett i matematisk analys och sannolikhetsteori och ett i tillämpad matematik och statistik. Det är i det senare som vi tre seniora statistiker (jag själv, Silvelyn Zwanzig och Jesper Rydén) ingår. Förutom oss finns där bland annat en relativt stor grupp av forskare under ledning av en framgångsrik professor i tillämpad matematik, David Sumpter. Denna grupp sysslar mycket med biologiska tillämpningar, där system av differentialekvationer dyker upp. Parametrarna i dessa system behöver förstås skattas på något sätt, och här kommer de statistiska metoderna in. Så det är kanske inte helt onaturligt att vi sitter i samma forskningsgrupp som dem.

I kölvattnet av denna uppdelning har vi även startat en inriktning av institutionens mastersprogram mot tillämpad matematik och statistik. Det finns också en allmän inriktning samt en inriktning mot finansiell matematik. Man kan även disputera i tillämpad matematik och statistik på institutionen. Däremot kan man numera inte ta vare sig master- eller doktorsexamen med inriktning mot matematisk statistik i Uppsala.

Man kan tycka att detta är sorgligt. Men kanske är det mer en fråga om namn än innehåll? Våra kurser i matematisk statistik finns ju kvar. Fortfarande finns förkunskapskrav om sannolikhetsteori för att läsa kurser i statistisk inferens.

Man skulle kunna tro att det som har hänt i Uppsala är en engångsföreteelse, men så är inte fallet. Samma sak har också inträffat på Chalmers i Göteborg.

Matematisk statistik och statistik

Under många år var jag anställd som professor i statistik på Uppsala Universitet. Vid ett tillfälle bjöds jag och prefekten Bo Wallentin in till samtal med prefekterna på matematik och IT. På grund av flytten av vissa institutioner från Polacksbacken till det nybyggda Ångströmslaboratoriet hade lokaler frigjorts, och det fanns möjlighet för statistiska institutionen att flytta in. Prefekterna på matematik och IT ville höra om vår inställning. Detta ledde till en diskussion på dåvarande statistikavdelningen. Motståndet mot idén var ganska stort.

Senare flyttades ju även matematik och stora delar av IT till Ångström, så tanken var kanske aldrig egentligen realistisk. Men huvudskälet till att de flesta på statistik var negativa var att en flytt från Ekonomikum, där man satt, skulle kunna bli väldigt opraktisk för alla lärare. Detta eftersom flertalet av statistiks studenter på grundnivå läser på ekonomprogrammet, och dessa har alla sina kurser på Ekonomikum. Så sannolikt skulle en flytt innebära mycket flängande ett par kilometer fram och tillbaka mellan tjänsterummet på Polacksbacken och undervisningslokalerna på Ekonomikum för lärarna. Alternativt kunde man förstås tänka sig att statistikkurserna flyttades till Polacksbacken så att istället studenterna fick ta sig fram och tillbaka. Men då var farhågan att andra ämnen på samhällsvetenskaplig fakultet, t.ex. företagsekonomi och nationalekonomi, skulle dra öronen åt sig och istället ordna sina egna statistikkurser.

»Visst vore det ibland bra för dig som statistiker att vara korridorsgranne med en filosof?»

Situationen med två snarlika ämnen på olika fakulteter, alltså statistik på samhällsvetenskaplig fakultet och matematisk statistik på (teknisk-) naturvetenskaplig, är ju lite speciell ur ett internationellt perspektiv. Mig veterligt är det bara i Danmark som vi kan se något liknande. Historien bakom hur det har blivit så i Sverige är mycket intressant, men jag tänker inte närmare ge mig in på den här. Intresserade läsare hänvisas till en artikel av Lindgren och Guttorp (se t.ex. länk från Georg Lindgrens hemsida).

Många med mig har länge försökt argumentera för att statistik och matematisk statistik borde vistas under samma tak. Men, bland annat av skäl som ovan, har jag förstått att detta är mycket svårt att få till. Åtminstone verkar det svårt att helt slå ihop ämnena. Ekonomstudenterna utgör ju en viktig ”kassako” som de statistiska institutionerna inte kan vara utan. Eftersom man kan läsa på ekonomprogram med lägre förkunskaper från gymnasiet i matematik än vad som krävs för tekniska och naturvetenskapliga utbildningar så måste det finnas kurser i statistik som är anpassade för sådana studenter. Och det vore naturligtvis olyckligt om det inte var statistiker som undervisade på dessa kurser. Och det kan ju också vara en fördel om dessa undervisande statistiker har en liknande bakgrund som studenterna.

Samtidigt är det ju, på de matematiska institutionerna, naturligt att dra nytta av studenternas matematikkunskaper vid kurser i matematisk statistik. Visst kan samläsning mellan studenter i statistik och matematisk statistik ändå fungera, men i så fall sker detta med fördel på lite högre kurser av tillämpat slag.

Drömmen om metodfakulteten

Tillåt mig ändå till sist att frigöra mig från bindande ekonomiska aspekter och samtidigt höja blicken en smula. Hur hänger egentligen alla vetenskaper som representeras på våra universitet ihop? Låt mig här begränsa mig till de vetenskaper som har någon kvantitativ, eller empirisk, komponent. Grunden för alla dessa går att hitta inom teoretisk filosofi. Teoretisk filosofi ”tillämpas” sedan i sin tur på metoder och verktyg inom matematik, (matematisk) statistik och/eller IT. Så varför inte samla de nu uppräknade ämnena i en egen fakultet, metodfakulteten?

Fördelarna är uppenbara. Förutom att man på detta sätt äntligen skulle föra ämnena statistik och matematisk statistik närmare varandra så skulle man få ”kortare väg” mellan ämnen som metodmässigt är mest närbesläktade. Visst vore det ibland bra för dig som statistiker att vara korridorsgranne med en filosof? Eller, om du har ett knepigt programmeringsproblem, med en programmeringsexpert från IT?

Den stora faran skulle förstås vara att man fjärmade oss statistiker från tillämpningsämnena. Om det blev för långt till statistikerna för till exempel en nationalekonom skulle hen kanske lita till sina egna kunskaper i större utsträckning och konsultera statistikern mindre? Men konstruktioner som CIM på ”teknat” i Uppsala har visat att det inte måste vara så. Och visst kan studenter vara flyttbara! Ett exempel är ett civilingenjörsprogram som jag undervisar på: System i Teknik och Samhälle (STS). Dessa studenter läser kurser både hos oss på Ångströmlaboratoriet/Polacksbacken och på det två kilometer norrut belägna Ekonomikum. Jag har inte hört någon klaga på det. Så visst går det att flytta på stora studentmassor om man bara vill.

Ja, det här var lite ganska spridda och personliga tankar om läget nu och i framtiden för universitetsämnet (matematisk) statistik. Vare sig du som läsare håller med eller inte vore det roligt att få ta del av dina åsikter. Så skriv direkt till mig, eller skriv ett svar i något kommande nummer av Qvintensen!

ROLF LARSSON
PROFESSOR I TILLÄMPAD MATEMATIK OCH STATISTIK VID UPPSALA UNIVERSITET

Forskare i tillämpad matematik och statistik använder matematikens kraft för att förstå hur världen fungerar. De löser problem inom en mängd tillämpningar – allt från biologi och djurs beteende till ekonomi och sociologi.

Qvintensen

POSTTIDNING B

SVERIGE
PORTO BETALT
PORT PAYÉ

Avsändare/Sender: Svenska statistikfrämjandet
c/o John Öhrvik, Västmanlands sjukhus Västerås, Centrum
för klinisk forskning, 721 89 Västerås, Sweden

Bli medlem i Svenska statistikfrämjandet

Svenska statistikfrämjandets syfte är bland annat att främja sund användning av statistik som beslutsunderlag och att väcka och sprida intresse för statistik i samhället.

Om du önskar bli medlem i Svenska statistikfrämjandet skicka ett e-brev till Linnea Johansson Kreuger på sekrfram@gmail.com. Du får Qvintensen i brevlådan och platsannonser via e-post.

Det ställs inga krav för att bli medlem; alla som är intresserade av statistik och vill stödja statistikens roll i samhället är välkomna.

Statistik-
främjandet