

Qvintensen

NR 2 2020

TEMA CORONAPANDEMI/

Modellering av epidemier

Vad har forskare
och myndigheter
för kunskaper?

15 minuter
i rampljuset

SIDORNA 6-19

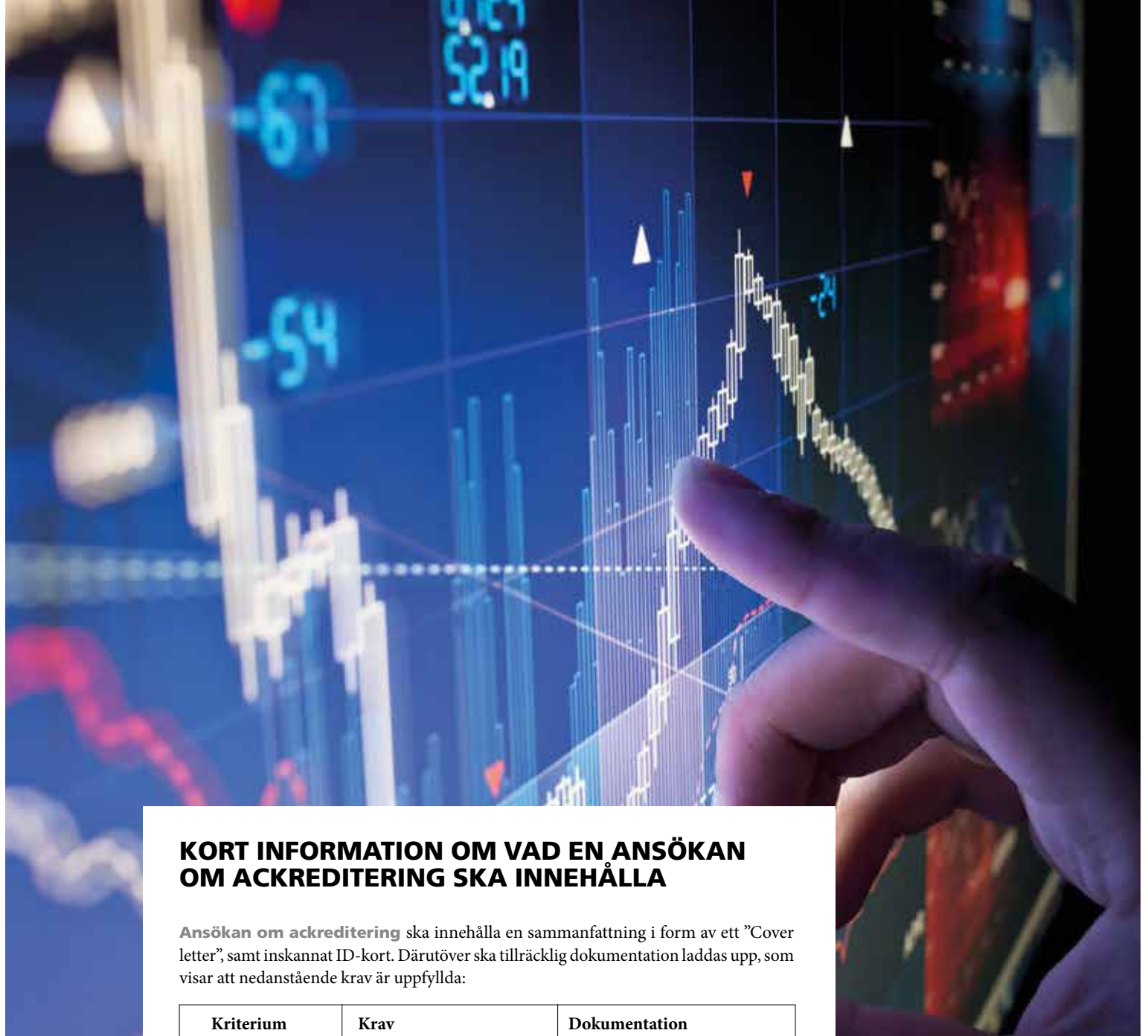
Låga sannolikheter
i forensiska
bedömningar

SIDORNA 20-23

Sannolikhetslärans
utmaningar inför
"Statistik 4.0"

SIDORNA 24-30





KORT INFORMATION OM VAD EN ANSÖKAN OM ACKREDITERING SKA INNEHÅLLA

Ansökan om ackreditering ska innehålla en sammanfattning i form av ett "Cover letter", samt inskannat ID-kort. Därutöver ska tillräcklig dokumentation laddas upp, som visar att nedanstående krav är uppfyllda:

Kriterium	Krav	Dokumentation
A Utbildning	Minst MSc i statistik/ matematisk statistik, eller motsvarande	Examensbevis Kursutdrag
B Yrkeserfarenhet	Minst 5 år (räknat i heltid) som statistiker	CV Rekommendationsbrev
C Professionell utveckling	Pågående utveckling i rollen som yrkesperson (OBS Ej nödvändigtvis inom statistik)	Lista CV Certifikat etc
D Kommunikation	Påvisad god förmåga att kommunicera statistiska begrepp.	Exempel på arbetsprodukter, rapporter, presentationer etc. Rekommendationsbrev (kan vara samma som under B)
E Etiska standarder	Ska följa alla relevanta lagar och etiska standarder	Sökanden visar kännedom genom att räkna upp de som gäller hen.
F Medlemskap	Medlem i en nationell förening	Kontrolleras vid granskning

Detaljerade instruktioner publiceras på FENStatS hemsida (www.fenstats.eu).

Ackreditering av statistiker införs nu av FENStatS

Efter ett omfattande arbete beslutade FENStatS årsmöte 21 september 2020 att införa ett system för ackreditering av statistiker, liknande de system redan används av AM-STAT och Royal Statistical Society.

Avsikten är att de olika systemen ska vara kompatibla. En portal för ansökningar om ackreditering har nu skapats. Den presenterades och öppnades den 20 oktober – på den internationella statistikdagen!

Ackrediteringens mål är att tydliggöra statistikers roll i samhälle och företag och på så sätt stötta, uppmuntra och gynna god statistik och goda möjligheter för statistiker. Framför allt riktar sig ackrediteringen till statistiker verksamma utanför högskolan, som på detta sätt – förhoppningsvis – ska känna stöd för att verka för högre kvalitet inom sitt yrke.

För att få ett certifikat som ackrediterad statistiker måste den sökande vara medlem i en nationell organisation, som är medlem i FENStatS, och som har valt att ansluta sig till FENStatS system. Svenska Statistikfrämjandet är en sådan organisation, som kommer att vara med från start.

En ansökningsavgift på 100€ kommer att tas ut, samt en årlig avgift på 20€. Av dessa avgifter kommer en viss del att stanna i den nationella organisationen samt en del att finansiera systemet för ackreditering. Ansökan måste göras via FENStatS speciella ansökningsportal.

Granskningen av ansökningarna kommer att genomföras lokalt, vilket gör det möjligt att även hänvisa till dokumentation på den sökandes lokala språk – detta har varit en förutsättning för att kunna genomföra detta i hela EU. Saknas dokumentation som behövs för att behandla en ansökan, kommer sådan att efterfrågas.

Granskningen ska inte innebära en ”peer review” utan blir en s.k. tredjepartscertifiering, där fokus ska

ligga på att dokument är korrekta och stöder ansökan, alltså inte en granskning av innehållet som sådant. Ej heller innebär godkänd granskning några garantier för den sökandes faktiska kompetens, utan det är enbart en sammanställning av dokumenterade meriter (jämför med t.ex. ISO-certifiering). Dock har de lokala bedömarena en viss möjlighet att t.ex. bedöma rimligheten i uppgivna yrkeserfarenheter, referenser eller rekommendationer.

Sedan ansökan granskats och godkänts utfärdas ett certifikat som gäller i fem år, varefter en begränsad omcertifiering måste göras. Detaljerna här är inte klara, men det handlar om att uppdatera meriter.

Ackrediteringen är frivillig. Men för dig som statistiker ger den en sammanfattning av dina meriter och ett slags sammanfattande yrkesbevis. Den visar att du har teoretisk kompetens, som du tillämpat i minst 5 år. Den visar också att du arbetat aktivt med din egen kompetensutveckling och att du har en roll som professionell yrkesperson. Vidare visar den att du har en medvetenhet om de etiska regler som gäller för din verksamhet, och att du förbinder dig att följa alla tillämpliga etiska regler. Ackrediteringen är en komplettering eller påbyggnad på examina från högskola eller universitet, som pekar på en systematisk utveckling och ansvarstagande hos den ackrediterade. Alltså en sammanställning av givna meriter under såväl utbildning som tiden efter att den avslutats. Mer information om anslutning etc. publiceras på FENStatS hemsida.

MAGNUS PETTERSSON
ordförande, FENStatS Accreditation
Committee

En missad chans?

När folk jag träffar får höra att jag är (eller har varit) statistiker, så märker jag att de ibland föreställer sig att en statistiker är detsamma som en tabellnörd. Jag brukar försöka invända att statistiker snarare är intresserade av hur uppgifter i en tabell har kommit till, hur mycket man kan lita på dem, och hur man kan dra slutsatser från dem. Att statistiker är intresserade av *metoder* och deras egenskaper snarare än av tabellstudier.

Sedan mars i år har vi dagligen fått statistiska uppgifter om utvecklingen av covid-19 i Sverige och i andra länder. Allmänheten har ett stort och berättigat intresse av sådan information. Men jag har märkt att de jag talat med inte förstår varifrån de numeriska värdena kommer. Ofta har man bara fått en vag uppfattning att någonting har gått upp (då är det oroväckande), eller att det har gått ner (då är det bra). Och jag har inte kunnat hjälpa dem, för jag har inte heller förstått särskilt mycket.

Jag anser att svenska statistiker med specialkunskaper borde ha gripit chansen att höja den statistiska allmänbildningen i landet genom att gå ut och berätta hur de tillkännagivna värdena har kommit till, och vilka begränsningar som finns när man försöker dra slutsatser från dem. Men jag har inte märkt några försök i den vägen. I varje fall inte försök riktade mot en bredare allmänhet.

En typ av data som regelbundet redovisats är "antal smittade i covid-19". Jag föreställer mig (kanske felaktigt) att värdena avser antalet nytillkomna personer som testats för förekomst av coronavirus, och där testet gett positivt utslag. Hur går det till i praktiken, när sådana värden samlas in från hela landet? Förekommer det det bortfall, alltså att vissa provresultat kommer bort i hanteringen? Vad vet man om testmetodens tillförlitlighet? (Man brukar ju diskutera sådant i termer av sensitivitet och specificitet.) Vad vet man om antalet smittade personer med så lindriga symtom att de aldrig kommer att utsättas för test?

En annan typ av data som också redovisats regelbundet är "antal döda i covid-19". Här föreställer jag mig att värdena avser



antalet nyrapporterade dödsfall, där dödsorsaken enligt någon form av läkarutlåtande uppges vara covid-19. Hur går det till i praktiken när sådana läkarutlåtanden samlas in från hela landet? Förekommer det bortfall, alltså att ett läkarutlåtande inte rapporteras? Hur stor är risken att ett utlåtande är felaktigt? Hur definieras dödsorsak i sådana fall där en avliden förutom covid-19 även hade andra livshotande sjukdomar?

I samband med sådant som smittorisk och immunitet så hänvisar media ibland (med tydlig respekt) till matematiker och deras modeller. Ibland får jag

intrycket att mediafolket har föreställningen att "matematik" är detsamma som "konsten att räkna rätt", och att det som sägs av en matematiker därför inte kan ifrågasättas. De modeller som det här är fråga om är matematiska uttryck, som ska försöka beskriva samband mellan olika storheter i den fysiska världen omkring oss. Vet man något om hur pass bra sådana här modeller fungerar, när man vill beskriva så komplicerade fenomen som smittspridning i ett modernt samhälle med komplicerade kontaktmönster? Använder man till exempel modeller som tar hänsyn till slumpmässig variation hos det fenomen som studeras, eller är det fråga om rent deterministiska modeller?

Till sist en fråga av mer internt intresse i statistikerkretsar: Varför talas det i media alltid om *matematik*, när corona och covid-19 är på tapeten? I själva verket handlar det ju om *statistik*. Data och modeller som diskuteras är ju av statistisk natur. Och varför presenteras till exempel en sådan som Tom Britton i media alltid som "professor i matematik"? Han kan naturligtvis en kolossal massa matematik, men han är ju egentligen professor i matematisk *statistik*. Det är det som är hans specialområde. Jag undrar om inte statistikerna har missat en chans att göra lite reklam för sig själva. De borde visa att de gör ett arbete av stor betydelse för samhället.

JAN WRETMAN

Qvintensen

Ansvarig utgivare

Joakim Malmadin

Redaktör

Jan Wretman, 070-781 78 35

Redaktion

Hans Alberg

Jens Malmros

Rolf Larsson

Marie Linder

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.

Annonssutskick 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.



SVENSKA STATISTIKFRÄMJANDETS STYRELSE

Ordförande

Joakim Malmadin 010-479 49 45,
ordforande.framjandet@gmail.com

Vice ordförande Mari Josefsson

Skattmästare Anders Haraldsson

Sekreterare Mattias Strandberg

Webbansvarig Jens Malmros

Representant Surveyföreningen

Marcus Berg

Representant FMS

Ida Hed Myrberg

Representant Industriell statistik

Hans Alberg

Representant Cramérskapskapet

Jesper Rydén

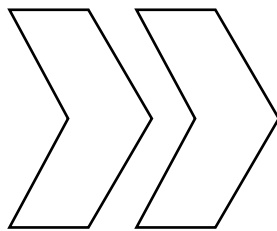
Adress

Svenska statistikfrämjandet

Sekreterare: Mattias Strandberg

E-post sekrfram@gmail.com

Webbplats www.statistikframjandet.se



Problemfrågeställningarna har ändrats i en dramatisk takt: Hur stort är R0? ... Vad är risken för en stor andra våg?»

TOM BRITTON, SIDORNA 17-19

Qvintensen

STATISTIKFRÄMJANDET

NR 2 2020

Innehåll

ACKREDITERING AV STATISTIKER 2-3

REDAKTÖRENS RUTA/

En missad chans?..... 4

TEMA: CORONAPANDEMIN

Modellering av epidemier 6-10

Vad har forskare och myndigheter för kunskap? 11-15

Ny europeisk arbetsgrupp för covid-19-statistik 16

"15 minutes of fame" orsakade av covid-19 17-19

LÅGA SANNOLIKHETER I

FORENSISKA BEDÖMNINGAR 20-23

SANNOLIKHETSLÄRANS UTMANINGAR

INFÖR "STATISTIK 4.0" 24-30

MINNESORD/

Bo Sundgren 31

SKRIV GÄRNA FÖR QVINTENSEN 35



Våra föreningar

SURVEYFÖRENINGEN

Ordföranden har ordet 32

Nya i styrelsen 32

Surveyföreningens kvalitetsseminarium 2020 33

FÖRENINGEN FÖR MEDICINSK STATISTIK

Ordföranden har ordet 34

INDUSTRIELL STATISTIK

Ordföranden har ordet 34

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 2020.

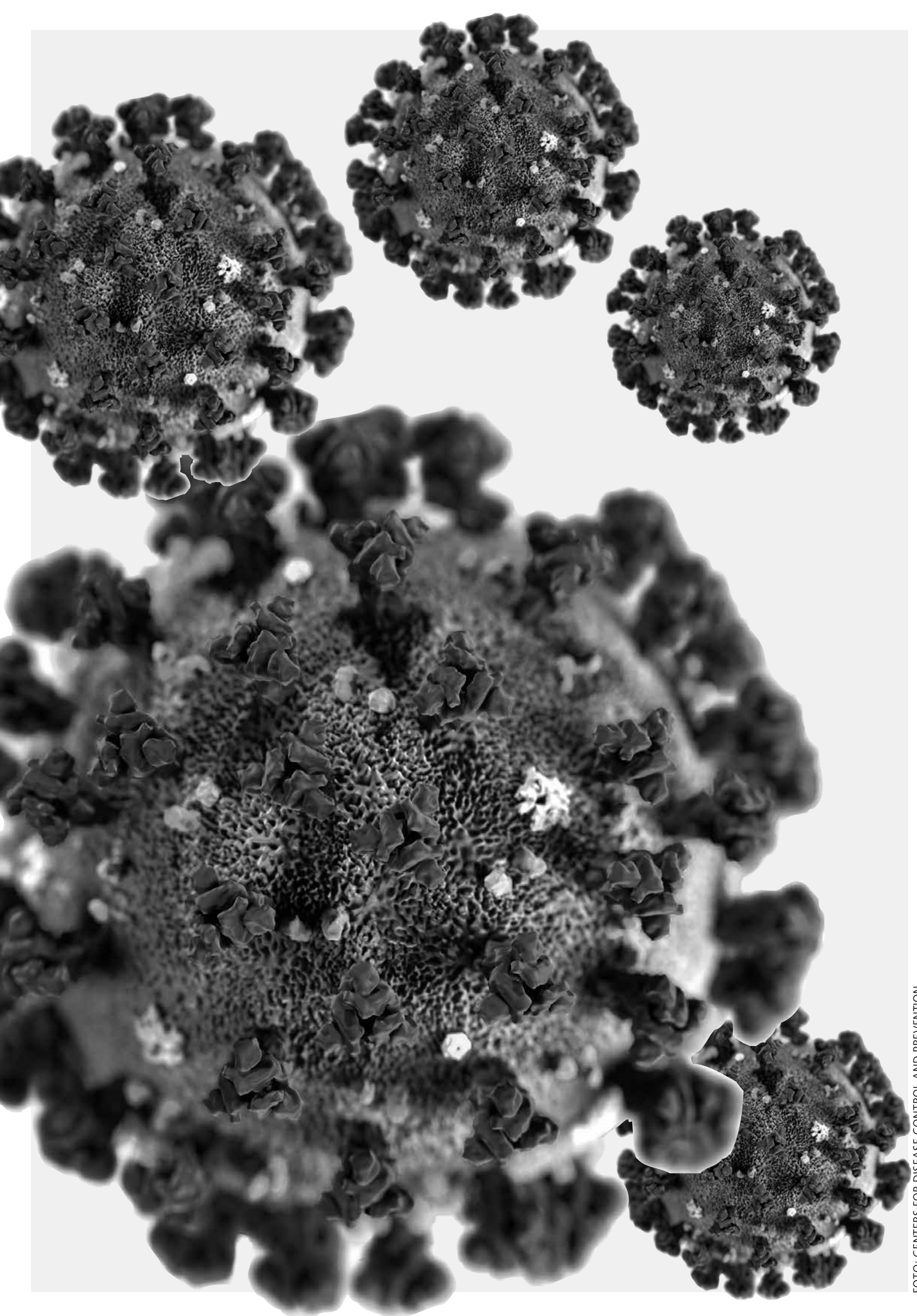


FOTO: CENTERS FOR DISEASE CONTROL AND PREVENTION

Modellering av epidemier

– en icke-matematikers tankar

Matematiska modeller som används för att förklara och förutsäga epidemiers förlopp och beräkna effekter av olika åtgärder har existerat sedan 1700-talet. Men användningen av denna typ av modeller, och det genomslag de fått, har ökat lavinartat och har nu en central roll i beslutsfattandet vid pandemier. Vår bild av covid-19-pandemin har påverkats starkt av matematiska epidemiologiska modeller, vid sidan av de rent statistiska beräkningarna av orsakssamband, trender och riskfaktorer.

Så påverkades till exempel beslutsfattandet av en modellstudie från Imperial College London, som i mars 2020 förutspådde utfallet av coronapandemin (Ref. 1). Deras modell visade att en *lockdown* var det enda tänkbara alternativet för att möta krisen, ett mildare angreppssätt skulle medföra ytterligare hundratusentals dödsfall. Detta utfall i en epidemimodell fick stora konsekvenser för Storbritanniens, och indirekt hela Europas fortsatta smittskyddsarbete (Ref. 2). Politikerna, med benäget bistånd av modellerare, drev fram en tidigare oprövad lockdown, åtgärder som blev stilbildande i stora delar av världen.

Fanns det vetenskapligt stöd för detta? Hade modellerna rätt i sina förutsägelser av vad som skulle hända om inte drastiska åtgärder vidtogs? För många var det självklart att inte ifrågasätta dessa resultat, de ansågs representera objektiva fakta och upplevdes som sannare än traditionella expertanalyser. De baserades ju på matematiska beräkningar som borde vara fria från förutfat-

tade meningar. Men precis som med statistiska analyser kan val av metod, selektion av data och orealistiska antaganden påverka modellerna, och de kan användas både för att bekräfta en föreställning och för att motbevisa den. Det finns flera olika typer av epidemiologiska modeller, som är mer eller mindre verklighetsbaserade och mer eller mindre trovärdiga. Samma frågeställning kan ge radikalt olika utfall med olika val av modell.

Realismen i epidemimodellerna

Vad innebär det rent allmänt att formulera en modell, och hur förhåller den sig till den verklighet man vill undersöka? För det första måste modellaren definiera och avgränsa det system som ska modelleras. Det kan vara något så mekaniskt som en bil, ett flygplan eller en byggnad, men det kan också vara en människa, en population eller ett ekosystem. Då uppstår frågan om detta system i sin helhet kan representeras av en fysisk modell av verkligheten i miniatyr; en vindtunnel med bil- eller flygplansmodeller, en bassäng med små modellbåtar och vågrörelser, en liten stad med hus, parker och gator (SIM city) eller människoliknande avatarer i ett dataspel. Dessa utgör exempel på *fysiska* modeller.

Alternativt kan aspekter av systemet renodlas och abstraheras. Så kan till exempel trafikflöden eller vädersystem åskådliggöras och modelleras som flöden i tid och rum. Människors kontakter med varandra kan åskådliggöras med sociala nätverk av relationer mellan individer, och med olika typer av transportnätverk där platser knyts

samman. Dynamik i nätverk kan beskrivas med flöden över riktade länkar och i tidsnätverk. Därmed kan grunden läggas för modeller av hur smittsamma sjukdomar sprids. Sådana modeller är fortfarande representativa för ett verkligt system, även om de är det ur ett abstraherat och reduktivt perspektiv. Viktiga dimensioner, som tid eller rum kan ingå, eller vara reducerade i statistiska eller icke-spatiala modeller.

På det mest abstrakta planet kan man utgå från en teori, som anses styra det händelseförlopp man vill modellera. När det gäller epidemiska förlopp kan det vara idén att människor kan indelas i distinkta kategorier – mottagliga, infekterade och avförda från vidare smittspridning – och att övergången från den ena kategorin till den andra kan beskrivas med differentialekvationer i analogi med massverkanslagen för kemiska reaktioner. Denna typ av modeller kallas SIR-modeller, och många infektionsmodeller har sedan 1920-talet byggts enligt denna princip. De är ideala representationer – tankeexperiment – och inte reella. All kontext till smittspridningen saknas i SIR-modellernas rena form, bland annat en så grundläggande dimension som den geografiska.

SIR-modellen är *deterministisk*, den ger samma utslag varje gång den körs, om ingående data och parametervärden är desamma. Det mycket starka antagandet man har gjort i modeller av denna typ, är att individer i en population träffar på och reagerar med varandra på ett sätt som liknar molekyler i en lösning – individer som är homogent distribuerade, där alla kan träffa alla med samma sannolikhet. I verkligheten är kontaktmönster heterogena och det finns en mängd barriärer som begränsar och kanaliserar smittspridning.

För att föra in ett element av slumpmässighet

»»»



FOTO: IMPERIAL COLLEGE LONDON

Imperial Collage Londons modell visade att en *lockdown* var det enda tänkbara alternativet för att möta covid-19-krisen. Ett mildare angreppssätt skulle medföra ytterligare hundratusentals dödsfall.



»»» MODELLERING AV EPIDEMIER – EN ICKEMATEMATIKERS TANKAR

har *stokastiska* SIR-modeller konstruerats. I små populationer och tidigt i ett epidemiförlopp kan en stokastisk modell ge andra resultat än en deterministisk. Eftersom utfallet blir något olika för varje gång modellen körs, måste beräkningen upprepas många gånger för att få ett medelvärde och en varians. I stora populationer och genomsnittligt ger det dock mycket likartade resultat som de deterministiska modellerna.

Det nästan lika starka antagandet som med homogeniteten, är att bara tre klart åtskilda tillstånd existerar under en epidemi. I verkliga livet finns ofta flera grader inom de olika kategorierna – man kan till exempel vara mer eller mindre mottaglig för en infektion, vara mer eller mindre smittsam eller få mer eller mindre stark immunitet. Det går att förfina kategorierna i en SIR-modell genom att föra in fler diskreta kategorier och underkategorier, men det gör samtidigt modellen mer otymplig och inte nödvändigtvis mer användbar.

Vad är epidemimodeller bra för?

En viktig fråga för epidemiologer är vad som krävs för att en epidemi ska kunna komma igång. Logiskt sett måste en smittad individ överföra smittan till mer än en ny individ i nästa led för att det totala antalet smittade ska öka. Antalet nya smittade från en smittsam individ kallas smittans ursprungliga *reproduktionstal* eller R_0 , och det har fått en central betydelse för att förutsäga en epidemis initiering och förlopp. Är R_0 större än 1 ökar antalet smittade, och ju högre talet är desto mer omfattande blir epidemin. Är det

mindre än 1 avtar smittspridningen, och pendlar det kring 1 förblir antalet smittade konstant. Det finns dock en hake i detta resonemang – det förutsätter just en homogen spridning där en smittad individ har lika stor sannolikhet att träffa på alla mottagliga individer i en befolkning. Det speglar inte människors kontakter i verkliga livet. Vi vet alla att våra kontaktmönster huvudsakligen begränsas till familjemedlemmar, vänner, bekanta och arbetskamrater och bara enstaka tillfälliga kontakter. Detta gäller förstås i ännu högre grad vid social distansering.

»En viktig fråga för epidemiologer är vad som krävs för att en epidemi ska kunna komma igång.»

Det finns många anledningar att inte lägga alltför stor vikt vid reproduktionstalet. Det utgör ett genomsnitt av en mängd individuella reproduktioner. Ett outtalat antagande är att dessa individuella tal är snävt fördelade kring ett medelvärde. Flera smittor, och särskilt de som orsakas av coronavirus har i verkligheten visat sig variera mycket kraftigt vad gäller de enskilda reproduktionstalen. Man har observerat otaliga fall av så kallad *superspridning*, det vill säga att en smittad individ har kunnat smitta ett oproportionerligt stort antal andra, i storleksordningen ett 10-tal till ett 100-tal. Samtidigt är det många smittade som inte lyckats smitta ner en enda annan person, vilket gör att fördelningen av de enskilda reproduktionstalen får en lång svans. Denna fördelning kan bäst modelleras med en negativ binomialfördelning, där dispersionskoefficienten och inte bara reproduktionstalet blir avgörande för hur effektivt en epidemi sprids (Ref. 3). Få modeller som använts under coronapandemin har tagit hänsyn till den effekt

på epidemiförloppet som superspridning av olika grad medför, och vad som händer om superspridning kan förhindras.

När upphör en epidemi och hur omfattande kan den förväntas bli? Det är också en viktig fråga för epidemiologer. Tidigare trodde man att detta hade att göra med att smittämnet efter hand försvagades. En annan förklaring var att epidemin fortsatte tills alla mottagliga individer hade smittats, och det inte längre fanns några friska kvar. Kermack och McKendrick beskrev redan på 1920-talet i sin SIR-modell att inget av dessa alternativ stämde; de visade matematiskt att smittan stannade av innan alla mottagliga insjuknat, på grund av att det effektiva reproduktionstalet föll under 1 – de fortfarande mottagliga blev för få för att underhålla smittspridningen. Det är detta tröskelvärde som ligger bakom begreppet *flockimmunitet*, ett centralt begrepp, men inte helt oproblematiskt.

En population uppnår teoretiskt flockimmunitet när den effektiva reproduktionen sjunker under 1. Tröskelvärdet när antalet fall vänder från ökande till sjunkande kan formuleras som $p_c = 1 - 1/R_0$. Ju högre R_0 är, desto större blir omfattningen av epidemin innan flockimmuniteten uppnås. Tröskelvärdet har fått sin främsta användning som ett mått på hur stor vaccinationstäckningen bör vara för att skydda mot genombrott av den sjukdom man vaccinerat emot. För till exempel mässling behöver minst 95 % av befolkningen vara vaccinerad för att förhindra nya stora mässlingsutbrott, då mässling har ett mycket högt R_0 . Men eftersom verkliga befolkningar är heterogena i flera avseenden kan små utbrott alltid uppstå, och vaccinationsgränsen måste sättas med god marginal. Av samma anledning blir även det teoretiska tröskelvärdet



Picadilly circus våren 2020 under *lockdown*. En modellstudie från Imperial College London fick stora konsekvenser för Storbritanniens och indirekt hela Europas fortsatta smittskyddsarbete.

FOTO: WIKIMEDIA

för flockimmunitet flytande, och varierar brett för en och samma epidemi, inom olika områden och vid olika tidpunkter. Särskilt superspridning utgör en typ av heterogenitet som kan göra att en epidemi sätts igång även om $R_0 < 1$, och kan avstanna innan den teoretiska flockimmunitetströskeln uppnås – om superspridningen kan förhindras.

En central fråga för modellerare är förstas om modellerna kan förutsäga ett epidemiförlopp. Alla beslutsfattare vill ju veta vad de har att vänta sig, för att kunna förebygga och hålla en beredskap. Hur kan osäkerheten minskas, vilken prediktiv noggrannhet har de epidemiologiska modellerna? Det är flera faktorer som avgör detta – ofullkomligheter i modellen, ofullkomliga data och grundläggande gränser för förutsägbarhet. Frågan om var gränsen går för vad som över huvud taget kan förutsägas om infektionsförlopp på längre sikt har ofta undvikits, eftersom den inte är enkel att besvara.

Kortsiktiga prognoser är förstas alltid lättare att göra än långsiktiga, precis som med väderprognoser. Vissa fenomen, beteenden och sjukdomar är mer regelbundna och förutsägbara än andra, och det är viktigt att försöka uppskatta var gränsen för förutsägbarhet går. De kan skilja sig avsevärt mellan olika epidemiska sjukdomar (Ref. 4). Givet detta kan olika modeller ha olika prediktiv noggrannhet, och det är därför viktigt att alla val av parametrar och förutsättningar i modellen är redovisade och att osäkerheten i prediktionen anges, något som inte alltid är fal-

let. Modeller kan också vara onödigt komplexa, och därmed riskera så kallad *overfitting*, som också minskar prediktionsförmågan.

Slutligen används epidemimodeller för att uppskatta effekter av olika interventioner. Särskilt hur vaccinationer påverkar epidemiförlopp har studerats, och detta har kunnat

ge uppskattningar av hur många i befolkningen som måste vaccineras för att få en smittspridning att vända, givet ett vaccins skyddseffekt. Även andra läkemedelseffekter och icke medicinska åtgärder kan studeras. Under coronapandemin har man med hjälp av modeller undersökt effekter av social distansering, rese-restriktioner, nedstängningar etc., vilket inte är lika enkelt som när det gäller väldefinierade medicinska åtgärder. En mängd olika restriktioner har införts i olika ordningsföljd och under olika faser av epidemierna, vilket leder till stor osäkerhet om vad alternativa handlingslinjer får för konsekvenser. Trots denna osäkerhet har denna användning av modellerna fått stort genomslag.

»Frågan om var gränsen går för vad som över huvud taget kan förutsägas om infektionsförlopp på längre sikt har ofta undvikits, eftersom den inte är enkel att besvara.»

Olika typer av coronamodeller

Vilka typer av modeller har använts under coronapandemin och hur ändamålsenliga har de varit? Många modeller utvecklades och testades under H1N1-pandemin 2009, och fungerade något så när bra för influensa. Dessa modeller var ofta SIR-modeller (eller SEIR, där även ett exponerat men inte insjuknat tillstånd ingår). De har utvecklats på olika sätt under corona-

pandemin, med nya underkategorier, och i hybridmodeller med tillägg av resenätverk för att modellera förflyttningar (Ref. 5, 6). De är dock fortfarande behäftade med de grundläggande svagheter hos deterministiska modeller.

Tendensen har varit att gå från någon variant av SIR-modeller till mer realistiska men också mer komplexa modeller. Så var till exempel den tidigare nämnda brittiska modellen från mitten av mars, som fick sådant genomslag en så kallad *mikrosimulering*. Modeller av denna typ är individbaserade och tilldelar individerna ett hushåll, en arbetsplats eller en skola, och tillåter också kontakter på allmänna platser. Fördelning av de olika platserna, deras storlek och lokalisering, görs utifrån statistik från det land man vill studera. Dessa modeller blir i sin detaljrikedom och realism mer flexibla än de deterministiska, men också mycket beräkningstunga. Dessutom är individerna i modellen så kallade *representativa agenter* utan möjlighet till varierade attribut. I den brittiska studien införde man olika grader av social distansering, skolstängningar och karantänsregler. Resultatet från den simuleringen blev att sjukhuskapaciteten skulle överskridas mångfalt efter bara några veckor om inte en lockdown infördes. Att det inte blev så i till exempel Sverige, som inte införde dessa strikta åtgärder, vet vi nu. Men var låg felet i simuleringen – låg det i modellkonstruktionen, i parametervärderna eller i något annat implicit antagande?

En relativt ny typ av modell i epidemiologiska sammanhang, som prövats under coronapandemin är den så kallade *agentbaserade* modellen (ABM). Den har ytterligare en grad av flexibilitet jämfört med mikrosimuleringar, i och med att de individer eller *agenter*, som befolkar modellen har individuella attribut, och styrs av regler som

»»» MODELLERING AV EPIDEMIER – EN ICKEMATEMATIKERS TANKAR

kan förändras genom interaktion med andra agenter och även kontinuerligt anpassas till sin omgivning. På så sätt kan man till exempel införa superspridning i modellen, vilket också har gjorts och det har gett intressanta resultat (Ref. 7). ABM är en *bottom-up* ansats, som kan angripa problem med interagerande individer, en typ av problem som inte är omedelbart tillgängligt för statistisk analys. Den kritik som har framförts mot denna typ av modeller har bland annat gällt deras känslighet för ingångsvärdena, att de inte är särskilt transparenta och att det är svårt att reproducera resultaten (Ref. 8). En ny och spännande typ av epidemiologisk modell är den så kallade *dynamiska kausalmodellen*, utvecklad av Karl Fristons grupp vid University College London. De har tidigare arbetat med att modellera hjärnfunktioner, men har under pandemin tillämpat sina modeller på det epidemiska förloppet. Man kan tycka att det inte är närliggande områden, men egentligen handlar det om att beskriva komplexa tillstånd där många delar är beroende av varandra, och att uppdatera detta tillstånd med verkliga data. Ur detta kan en rad dolda tillståndsvariabler utläsas med bayesiansk inferens (Ref. 9). Att beskriva hur denna modell fungerar faller utanför mitt kompetensområde, men resultaten har varit av värde för förståelsen av coronapandemin. Genom att bland annat uppskatta den *effektiva befolkningen*, dvs. den del av den totala befolkningen som egentligen deltar i smittspridningen, har modellen visat att denna utgör en fraktion av totalbefolkningen, i en del fall inte mer än 20 procent. Om man för in denna uppskattade *effektiva befolkning* i den tidigare beskrivna ”domedagsmodellen” från Imperial College, får man mycket rimligare resultat än med antagandet att alla från början är mottagliga. Vad som ligger bakom detta fenomen med att en stor del av befolkningen inte verkar delta i epidemin (något som i media kallats ”coronapandemins mörka materia”) är ännu inte klarlagt, men det undersöks nu intensivt av immunologer och andra.

”Alla modeller har fel, men en del är användbara”

Ovanstående citat, som numera har blivit en kliché, har tillskrivits den brittiske statistikern George Box. Några olika felkällor som epidemiologiska modeller kan vara behäftade

med har berörts ovan, men hur är det med användbarheten?

Som vi sett kan epidemiologiska modeller svara på flera olika typer av frågor man kan ställa sig under en pandemi. Det kan gälla reproduktionstal och slutlig storlek, prediktioner av förloppet och effekter av olika interventioner. Givet alla abstraktioner, förenklingar och teoretiska antaganden som de olika modellerna innehåller, kan inte en enskild modell svara på alla tänkbara frågor. Det finns ingen universalmodell för att förklara en specifik pandemi, ännu mindre alla tänkbara pandemier. Vilken typ av modell man väljer, vilka grundantaganden man gör, får styras av den aktuella frågeställningen. Därför bör modelleraren ha tillgång till ett flertal olika modeller i sin arsenal att jämföra mellan, eller att bygga hybridmodeller med.

Det finns metoder för att jämföra olika modellers noggrannhet och komplexitet – deras epidemiologiska evidens – men det finns även inslag av subjektivitet i bedömningar av en modells lämplighet. Användbarhet är något som förhandlas fram. Det finns ingen gyllene standard, bedömningarna styrs av många saker – vilken forskargrupp som har mest trovärdighet, vilka resultat beslutsfattarna gillar och vad de tycker är genomförbart, eller eventuellt som bekräftelse på vad man redan bestämt sig för.

Därför måste man vara kritisk till beräkningarna och simuleringarna som kommer ur de epidemiologiska modellerna. För att kunna se genom modellerna med en kritisk blick och med sund skepsis, måste modellerna vara öppet redovisade och transparenta. Graden av osäkerhet måste vara redovisad, och olika gruppers resultat måste kunna jämföras konstruktivt och vetenskapligt, och inte på ett antagonistiskt och uteslutande sätt.

MARTIN HOLMBERG

Martin Holmberg är docent i infektionssjukdomar och har arbetat som infektionsläkare i 25 år. Han har även en bakgrund som molekylärbiologisk forskare och har under tio års tid arbetat på Socialstyrelsens tidigare smittskyddsenhet. Nyligen utkom hans bok *Pandemier och epidemier – från kolera till covid-19 i ett tvärvetenskapligt perspektiv* (Studentlitteratur, 2020).

Referenser

1. Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, m.fl. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College London; Mars 2020; Hämtad 2020-04-22 <http://hdl.handle.net/10044/1/77482>
2. Adam D. Special report: The simulations driving the world's response to COVID-19. *Nature*. 2020; 580(7803):316–318.
3. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005;438(7066):355–359.
4. Scarpino SV, Petri G. On the predictability of infectious disease outbreaks. *Nat. Commun.*; Dec 2019; Hämtad 2019-05-09. <http://www.nature.com/articles/s41467-019-08616-0>
5. Britton T, Trapman P, Ball FG. The disease-induced herd immunity level for Covid-19 is substantially lower than the classical herd immunity level. *Infectious Diseases (except HIV/AIDS)*; Maj 2020; Hämtad 2020-07-30 <http://medrxiv.org/lookup/doi/10.1101/2020.05.06.20093336>
6. Rocklöv J. COVID-19 health care demand and mortality in Sweden in response to non-pharmaceutical (NPIs) mitigation and suppression scenarios. *Epidemiology*; Mars 2020; Hämtad 2020-07-30 <http://medrxiv.org/lookup/doi/10.1101/2020.03.20.20039594>
7. Snejppen K, Taylor RJ, Simonsen L. Impact of Superspreaders on dissemination and mitigation of COVID-19 *Infectious Diseases (except HIV/AIDS)*; Maj 2020 Hämtad 2020-08-03. <http://medrxiv.org/lookup/doi/10.1101/2020.05.17.20104745>
8. Manzo G. Potentialities and Limitations of Agent-based Simulations. An introduction. *Rev. Fr. Sociol.* 2014;55(4):653–688.
9. Friston KJ, Parr T, Zeidman P, Razi A, Flandin G, Daunizeau J, et al. Dynamic Causal Modelling of Covid-19. 2020. arXiv Quantitative Methods.

STATISTIK I CORONA-TIDER:

Vad har forskare och myndigheter för kunskap?

Enskilda länder har valt olika strategier när de har drabbats av coronakrisen och hur de ska hantera pandemin i sitt land. Efter sitt vägval gör varje land allt för att försvara sina beslut och understryka sin egen förträfflighet. Oberoende av sina valda handlingsalternativ hittar man stöd med någon form av "vetenskaplig" förankring och/eller uttalanden från trovärdiga experter. När det gäller corona finns, mig veterligen, ännu ingen helt objektiv och vetenskapligt förankrad kunskap om pandemin som sådan eller hur smittspridning, immunitet, dödlighet, etc. utvecklas i samband med covid-19 (detta skrivs i slutet av oktober 2020). Det strategiska vägvalet baseras nu på uttalanden och råd från mer eller mindre sakkunniga (ibland självutnämnda) experter och/eller ännu inte verifierade forskningsresultat.

Det kommer att visa sig att flera av de "sakkunniga" har haft rätt på någon eller några punkter, men vi vet i förväg inte på vilka punkter de haft rätt. En viktig del i den erforderliga forskningen är statistiskt relaterat, både vad gäller långsiktiga (strategiska) och kortsiktiga (operativa) beslut.

Utvecklingen i Stockholmsområdet

Jag har personligen varit intresserad av att

följa utvecklingen av covid-19 i Sverige och har därvid noterat den dagliga statistik som har presenterats. Som undersökare vet jag att Stockholm vanligen ligger före övriga riket när det gäller olika typer av trender och utveckling. Det torde vara mer regel än undantag att det skulle gälla även när det gäller spridning av virus som drabbar samhället – speciellt om smitthärden kan antas komma genom anländande resenärer. Spridningen kan också vara beroende av stora folksamlingar. Det har därför fallit sig naturligt att fokusera på utvecklingen i Stockholmsområdet.

Jag har valt att i första hand titta på antal nyregistrerade smittade per dag samt antal med bekräftad covid-19 som har avlidit i länet.

Antal nyregistrerade anger hur många av de som har testats som har bekräftats positiva. Antalet påverkas av hur många som har testats. Till detta kommer ett svårbedömt mörkertal med dem som är smittade men inte har testats.

Det finns även andra statistiska mått som är av intresse, t.ex. antal behandlade inom intensivvården. Likaså kan det vara värdefullt att se på exempelvis relationstal och matematiska modeller vilket jag inte tagit hänsyn till i denna artikel.

Det förekommer viss eftersläpning vid registreringen av smittade när man ser på en enskild dag. Det har mindre betydelse om man vill stu-

dera utvecklingen över tiden. Något mer störande är svårigheter med enhetliga och adekvata definitioner samt eventuella felregistreringar som kan ha förekommit. Exempel på störande inslag har varit följande:

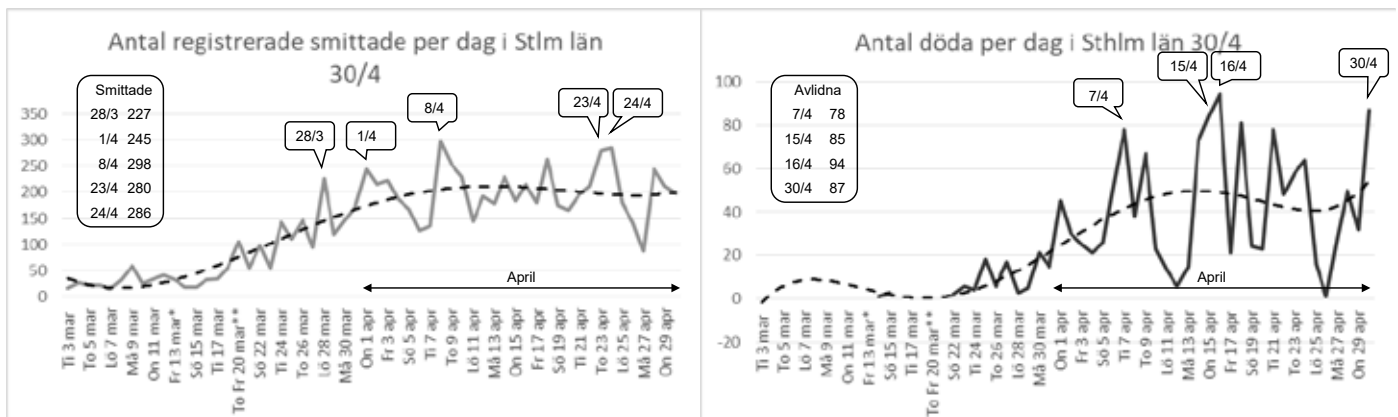
- Den 13/3 byter Folkhälsomyndigheten (FHM) strategi för provtagning.
- 20/2 byter Region Stockholm period för att koordinera med FHM.
- I april inleder SVA tester av covid-19 på människor. Det är segt att få igång testningen.
- 3/6 finner ett laboratorium 1 386 smittade som inte tidigare har inrapporterats.
- 18/6 övergår man från daglig redovisning till redovisning två gånger per vecka.
- 26/8 upptäcker man att 1 586 personer är felregistrerade.

Nedan följer diagram och några kommentarer månad för månad. Underlaget som jag återger här är hämtat från Region Stockholm (sll.se). Som trendlinje har jag valt att lägga in polynom vilket jag uppfattar som något mer långsiktigt och lättolkat. FHM/sll har valt att redovisa glidande medelvärde (7 dagar) vilket jag uppfattar vara mer populärt men något svårare och ibland svårare att tolka. Se diagram för april längst ner på denna sida.

Redan under andra halvan av mars kunde

»»»»

April – Dag för dag t.o.m. 30/4



»» STATISTIK I CORONA-TIDER: VAD HAR FORSKARE OCH MYNDIGHETER FÖR KUNSKAP?

vi konstatera att corona hade börjat ta fart i Stockholmsområdet med en stadig daglig tillväxt och en hög notering (den 28/3 med 227 nyregistrerade smittade). I början av april (8/4) fick vi sedan det vi under ett par veckor trodde, eller hoppades, skulle vara vårens toppnotering med 298 nyregistrerade smittade under ett dygn. Den glädjen, eller trygghetskänslan, varade inte länge eftersom vi redan den 15/4 och 16/4 fick toppnoteringar i antal avlidna i covid-19 och den 23/4 och 24/4 hade Stockholmsregionen nya höga noteringar i antal nyregistrerade smittade – i nivå med toppnoteringen den 8/4.

Både skärpta och lättade restriktioner

Den 27/3 skärpte regeringen reglerna för hur många personer som fick delta i en allmän sammankomst från 500 till 50. I övrigt inga nya restriktioner under april.

Den 14/4 konstaterade 22 forskare (på DN Debatt) att FHM redan vid fyra olika tillfällen hävdade att "smittspridningen har planat ut eller kommer att avta". Siffrorna för Stockholm tyder dock inte på att FHM hade fog för dessa påståenden – åtminstone inte i Stockholmsregionen, varken för antal nyregistrerade smittade eller för antalet registrerade avlidna i Stockholm. Min poäng här är inte sakfrågan utan hur man tolkar statistiken.

Vi kunde också i april konstatera att vissa företag och organisationer totalt ignorerade den allvarliga situation som Sverige nu befann sig i. Exempelvis ville friluftsanläggningar med stor publik ändå hålla öppet över påsk och resten

av vinter-/vårsäsongen. Stockholms fotbollförbund hade på fullt allvar för avsikt att dra igång seriespelet fredagen den 17/4. Se diagram för maj längst ner på denna sida.

Inte heller under maj syntes någon entydig och klar nedgång i vare sig antal registrerade smittade eller i antal avlidna.

I slutet av maj hade vi i Stockholmsområdet dessutom en trend som indikerade en uppgång i både antal registrerade smittade och antal avlidna. Den enda rekommendationen var att hålla sig hemma om man kände sig sjuk, hålla avstånd och tvätta händerna.

Stora delar av allmänheten visade att de tog den svåra situationen på fullt allvar. Samtidigt var många personer lika egocentriska som de motsträviga företagen/organisationerna som försökte ignorera det allvarliga läget.

Att den, som jag upplever, ganska uppenbara statistiken inte föranledde kraftigare åtgärder är för mig förvånande. Den positiva sidan av detta var att många företag (restauranger, barer, butiker, frisersalonger, etc.) inte behövde stänga. De kunde därmed överleva, samtidigt som allmänheten kunde röra sig ganska fritt utan några som helst förpliktelser. Detta var mer politiskt populärt än att stänga av verksamheter och införa förbud. I den svåra politiska avvägningen mellan att tillfredsställa allmänhet och företag jämfört med att få ner statistiken över antal smittade och

avlidna uppfattar jag, som statistiker, att man valt den mjukare linjen på bekostnad av antal nyregistrerade smittade.

Här kommer jag att tänka på en parallell när jag av Trafikverket blev ombedd att se hur de skulle kunna höja sitt NöjdKundIndex.

Jag tittade på vilka variabler som påverkade deras NKI och kom fram till att de borde vidta åtgärder som att öka/tillåta fri fart på motorvägar etc. Deras reaktion blev att då kommer ju trafikolyckor och antal döda i trafiken att öka. Ja, naturligtvis! Önskan var ju att öka kundnöjdheten – inte att minska olyckorna.

Har vi samma läge när det gäller covid-19? Då skulle regeringen hellre se en vinst i politiska poäng än att styra åtgärder som visar en minskning i statistiken över antal nyregist-

rerade smittade. Jag är inte rätt person att uttala mig i frågan om Sverige valt rätt eller fel strategi när det gäller hanteringen av covid-19. Om målsättningen har varit att så snabbt som möjligt minska antalet smittade, och därmed även antal avlidna,

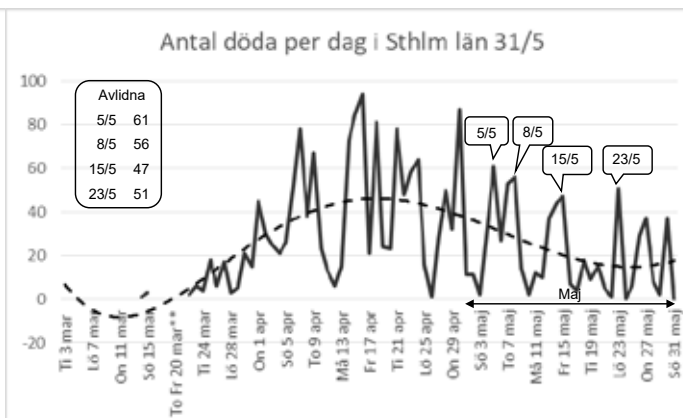
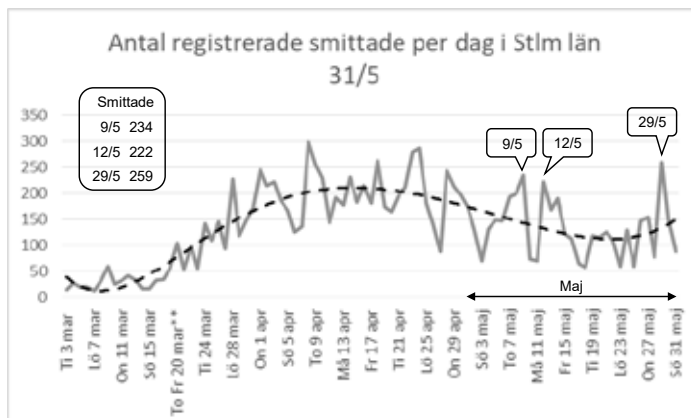
har implementeringen inte varit optimal. Med statistiken som beslutsunderlag borde man ha



Åke Wissing.

»Att den, som jag upplever, ganska uppenbara statistiken inte föranledde kraftigare åtgärder är för mig förvånande.»

Maj – Dag för dag t.o.m. 31/5



vidtagit åtgärder som hade varit mer obekväma för en stor del av allmänheten och företag än den mjukare linje man valde att gå. Se diagram för juni längst ner på denna sida.

Redan i slutet av maj och början av juni såg vi en ganska klar tendens till uppgång i antal smittade i Stockholmsområdet. Under resten av juni kom en riktigt kraftig ökning i antal nyregistrerade smittade.

I stället för stramare åtgärder förklarar regeringen den 5/6 att det nu är fritt fram att resa i Sverige. Detta togs sannolikt emot med stor glädje för alla som hade önskan och/eller planer på att resa under midsommarhelgen. Beskedet från regeringen kom exakt två veckor före midsommaraftonen (den 19/6). Den 18/6 registrerades en ny toppnotering i antal smittade i Stockholmsområdet med 386 personer, vilket är 30 % högre än toppnoteringen den 8/4.

Registreringen den 18/6 var den sista med dagliga uppgifter om antal smittade och avlidna! Efter den 18/6 redovisades resultatet endast två gånger per vecka med fyra respektive tre dagar som klumpsumma. I mina diagram har jag då tagit klumpsumman och dividerat med fyra respektive tre och använt det snittet för de aktuella dagarna.

Strax efter midsommar (23/6 – 25/6) hade vi den absoluta toppnoteringen av antal smittade i Stockholmsområdet. Totalt hade vi då 1 567 nyregistrerade smittade vilket gav ett snitt på 522 personer per dag. Detta visar inte vilken dag som hade den allra högsta registreringen och tyvärr inte heller hur hög den toppnoteringen

var. Myndigheten har inte varit villig att, efter två förfrågningar via mejl, svara på mina frågor om de dagliga siffrorna efter den 18/6.

Efter den 18/6 slutade jag ta fram diagram över antal avlidna. Det förekom perioder med mycket låga siffror (ingen, en eller några få avlidna) och jag kunde då inte veta hur många som avlidit en specifik dag. En topp vad gäller antal avlidna i covid-19 låg runt månadsskiftet juni/juli på närmare 40 personer. Några av dessa kan ha varit personer som funnits med i den toppnotering av antal smittade som registrerades efter midsommarhelgen.

Hur drogs slutsatser?

Den 24/6 presenterades en artikel i DN som kan få oss surveystatistiker att tappa hakan. Man diskuterade hur stor andel av befolkningen som man då trodde hade antikroppar mot covid-19. På Karolinska universitetslaboratoriet hade knappt 24 000 antikroppsprouver utförts, och av dessa var 17 % positiva. De som hade testats var personer från allmänheten, som själva valt att testa sig. Testningen var gratis för alla stockholmare.

Laboratoriets chef, Claes Ruth, förde då fram antagandet att: "Resultatet indikerar en utbredd smittspridning i samhället". Ruth angav inget om vad han menade med "en utbredd smittspridning". Han kritiserades i artikeln av Karin Tegmark Wisell, avdelningschef för mikrobiologi vid FHM. Hon "tror" inte att urvalet är ett

tvärsnitt av befolkningen – vi som är surveystatistiker vet att det inte kan vara ett tvärsnitt av befolkningen. Tegmark Wisell ville i stället hänvisa till FHM:s undersökning som visade att 10 procent av stockholmarna hade antikroppar.

De undersökta personer, som låg till grund för den siffran, var personer som själva valt att uppsöka öppenvården för provtagning. Enligt Tegmark Wisell var det "en grupp som normalt är representativ för befolkningen, men sannolikt

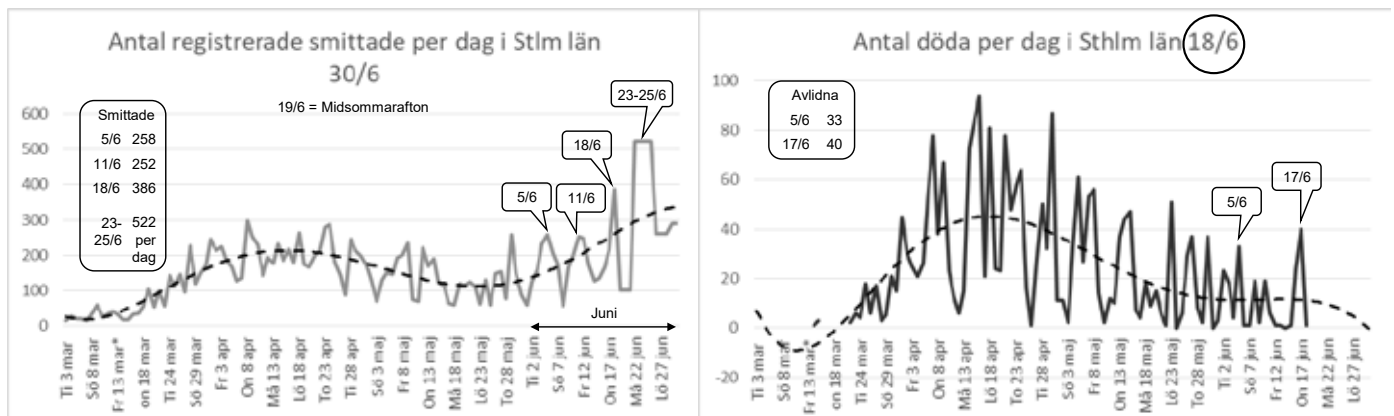
inte den här våren". Samtidigt nämndes i artikeln att företaget Werlabs utfört ca 50 000 tester och att det i Stockholm då var 14 %, Göteborg 11,5 % och i Malmö 5 % som utvecklat antikroppar.

Siffran 17 % från Karolinska är representativ endast för de ca 24 000 som testats där och siffran 10 % från FHM är endast representativ för dem som under vecka 17 till 21 valt att uppsöka öppenvården i Stockholmsområdet.

Om det är angeläget att veta hur stor andel av en given målgrupp som har antikroppar, skulle det krävas en regelrätt stickprovsundersökning i den givna målgruppen, med ett stickprov som är representativt för målgruppen. Det kan gälla den vuxna befolkningen i hela riket eller preciserade delar av befolkningen som definieras genom kön, ålder, geografisk begränsning etc. Skulle man nöja sig med en totalsiffra för exempelvis riket (eller t.ex. Stockholm) krävs inte särskilt sort urval. Vill man däremot kunna

»»»

Juni – Dag för dag t.o.m. 30/6 (antal döda t.o.m.18/6)



»»» STATISTIK I CORONA-TIDER: VAD HAR FORSKARE OCH MYNDIGHETER FÖR KUNSKAP?

uttala sig om delgrupper i målgruppen (t.ex. unga kvinnor i Mälardalen) behöver antalet undersökta personer ökas. Om det inte är angeläget att veta hur stor andel av befolkningen som har antikroppar finns det ingen anledning att munhuggas om vilken siffra man tror gäller för allmänheten i riket.

Den 26/6 markerade Hälso- och sjukvårdsregionrådet Anna Starbrink stor oro över den trängsel hon kunde konstatera vid Stockholms badplatser. Hon var frustrerad över att regionen inte hade några verktyg för att minska trängseln och önskade att regeringen skulle ta tag i detta. Den oron delades av professorn i infektions-epidemiologi på Karolinska institutet, Joakim Dillner. Han menade att risken för "superspridare" ökar på platser där många personer vistas samtidigt. En svårhanterlig grupp är, enligt Dillner, unga vuxna i 20 – 30 årsåldern som kan ha sämre förståelse för social distansering och som (enligt attitydstudier) visar att de för egen del inte är rädda att bli smittade (DN 16/7).

Utveckling och åtgärder under sommaren och hösten

Under juli och augusti såg vi en nedgång i antal nyregistrerade smittade i Stockholms län. Se översiktligt diagram längst ner på denna sida.

Efter den lugna semesterperioden började det, ur statistisk synvinkel, att framträda en bild som var både förvånande och oroande. Den 1/9 såg vi i Stockholmsregionen ett trendbrott vad gäller antalet nyregistrerade smittade. Och den trenden fortsatte att indikera en ganska alarmerande utveckling. Redan i mitten/slutet av september borde varningssignalerna ljuda.

Det är lätt att i efterhand kritisera och ond-göra sig över beslut som togs under resans gång, i en situation då man saknade säker kunskap om den framtida utvecklingen. Jag redogör här för några fakta som jag har kunnat konstatera i tiden runt september och under oktober. Jag vill härvid inte ta ställning i frågan huruvida beslut varit mer eller mindre lämpliga i det läge då de togs, och jag vill inte heller kommentera hur implementeringen gjorts.

29/8 Ordföranden för Sveriges kommuner och regioner, Anders Knappe, riktar skarp kritik mot att regeringen och FHM i våras var för långsamma med att be landets regioner att börja med

coronatestning. Hade testningen kommit igång tidigare, kanske också smittspårning och den allmänna smittspridningen hade kunnat begränsas, sa Knappe i Ekots lördagsintervju. Hans uttalande indikerade inte att man använt sig av den befintliga statistiken som beslutsunderlag.

15/9 Socialminister Lena Hallengren håller pressträff och informerar om att besöksförbudet på äldreboende kommer att hävas den 1 oktober. Sverige byter coronastrategi: Bara den smittade ska isoleras – inte de som absolut inte får bli smittade.

22/9 Larm från hälso- och sjukvårdsdirektören för region Stockholm, Björn Eriksson: "Vi ser mycket oroande tecken på att smittspridningen ökar".

23/9 Region Stockholm överväger att införa familjekarantän.

29/9 Besked om att höja det tillåtna publikantalet från 50 till 500 för idrotts- och kulturarrangemang skjuts fram från 1:a till 15:e oktober. Samtidigt planeras lättnader för restauranger och kaféer. Undantaget för dessa väntas träda i kraft 8:e oktober och en andra ändring den 15:e oktober.

Hälso- och sjukvårdsdirektören Björn Eriksson uttrycker åter att han ser allvarligt på läget. Antalet som testats positivt har fördubblats i Stockholm. De nya siffrorna tyder på att smittspridningen har ökat och det i sin tur på att allt för många har slutat följa FHM:s rekommendationer.

1/10 Beslut om familjekarantän fastställs av

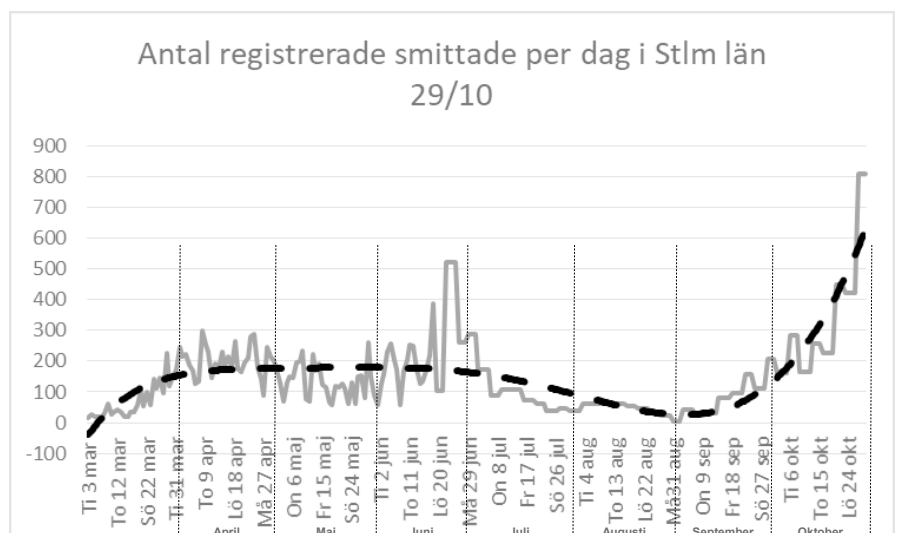
FHM. Besöksförbud på äldreboende hävs.

2/10 Nio forskare skriver på DN Debatt: "Svensk coronastatistik ger en skev bild av smittspridningen".

5/10 Forskare vid KTH konstaterar, genom studier av avloppsvatten i Stockholm, att virusmitten har fördubblats under september och nu ligger i nivå med de höga nivåerna i våras.

7/10 Anders Tegnell (FHM) uttalar att: "Stockholm har haft tydlig uppgång ett antal veckor och ligger nu på höga nivåer". Han understryker att viruset ökar mest bland unga vuxna och pekar ut studentfester, arbetsplatser, privata tillställningar och idrottssträningar, framförallt ishockey. Det kan bli tal om nya restriktioner, men han vill inte svara på frågan om vi nu har en ny våg. Hans prognos är en jul utan släktkalas, inga stora sammankomster under våren och fortsatt hemarbete under lång tid framåt. Julbord med arbetskamrater är inget undantag. Maria Rotzén Östlund, tf. smittskyddsläkare i Region Stockholm, pekar på att smittspridningen i Stockholmsområdet har ökat med hela 35 % under vecka 40 jämfört med den föregående veckan. Hon menar, precis som Anders Tegnell, att ökningen inte är början på en andra virusvåg.

8/10 Statistik för perioden 6:e – 8:e oktober (tisdag – torsdag v 41) presenteras med hela 855 nyregistrerade smittade, vilket ger ett snitt på 285 per dag. Det kan jämföras med 688 för tisdag – torsdag v 15 (7:e – 9:e april) med vårens topp-



**Statsepidemiolog
Anders Tegnell möter
media under Folk-
hälsomyndighetens
dagliga presskonfe-
rens våren 2020.**



FOTO: FRANKIE FOUGANTHIN/ WIKIPEDIA

notering den 8 april med 298 smittade. Det är inte orimligt, och t.o.m. troligt, att den dagliga noteringen någon av dagarna 6:e – 8:e oktober har överstigit den toppnotering som registrerades den 8 april.

Publikgränsen på 50 personer vid sport- och kulturevenemang ligger kvar tills vidare.

22/10 Särskilda råd till äldre (>70 år) och riskgrupper tas bort.

Danstillställningar med fler än 50 personer förbjuds fr.o.m. 1 november. Sammankomster får, fr.o.m. 1 november, ha högst 300 deltagare – inte 500 som tidigare.

29/10 Nya striktare råd till bl.a. Stockholm fr.o.m. 19/11. Exempelvis max 50 personer i stället för 300 som regeringen tidigare öppnat för.

Antal nyregistrerade smittade i Stockholmsregionen ökade med 79 % från v 42 till v 43.

Avslutande synpunkter

Jag har genom denna artikel försökt att belysa några aspekter kring hur man har använt sig av statistik i corona-tider. Mina begränsningar har varit att jag enbart tittat på statistik över antal nyregistrerade smittade (även antal avlidna fram t.o.m. 18 juni) och uteslutande använt sifferunderlag från Stockholmsområdet. Av tidsskäl har jag tyvärr behövt avbryta insamlingen av sifferunderlag den 29:e oktober – när det ur statistisk synvinkel tveklöst föreligger tydliga tecken på en andra våg.

Tre huvudpunkter som är värda att notera är:

- Hur man tolkat och använt sig av statistiken under pandemins första fas (t.o.m. juni).
- Hantering av läget runt midsommarhelgen.
- Hantering av läget efter trendbrottet 1:a september.

Varje uttalande, under april, maj och juni, om att smittspridningen planat ut eller börjat att avta står i strid mot den statistik man har haft tillgång till. Om regeringen och myndigheterna hade använt sig av den befintliga statistiken för att minska smittspridningen och de konsekvenser den medfört, så skulle utfallet inte ha setts som speciellt positivt – åtminstone inte regionalt i Stockholmsområdet. Den statistiken tycks ha uppfattats som ”nice to know”, snarare än att ligga till grund för beslut om vad, när och hur åtgärder skulle ha satts in.

Att två veckor före midsommarafton öppna för att man nu kunde få resa fritt i Sverige resulterade, inte oväntat, i en markant toppnotering av antal smittade några dagar efter den stora helgen.

Efter en nedåtgående trend i antal smittade under sommarmånaderna bröts mönstret vid månadsskiftet augusti/september och var i månadsskiftet september/oktober i nivå med det dystra läget under vårmånaderna april-juni. Inledningsvis, efter trendbrottet, avvaktade man och t.o.m. lättade något på restriktionerna i mitten av september. I slutet av september däremot stärkte man begränsningarna. Speciellt

»Varje uttalande, under april, maj och juni, om att smittspridningen planat ut eller börjat att avta står i strid mot den statistik man har haft tillgång till.»

på lokal nivå för Stockholm uttrycktes allvarlig oro.

Det har inte framkommit tydliga tecken på att den statistik som jag tittat på har legat till grund som beslutsunderlag för åtgärder till att få ner antal smittade i covid-19. Det gäller framförallt den sta-

tistik som presenterades under våren och runt månadsskiftet september/oktober. I slutet av oktober uppfattar jag åtgärder och allmänna råd som förvirrade. Man tycks vilja lätta på olika restriktioner samtidigt som man tycks inse allvaret med den markanta ökningen i antal registrerade smittade. Samtidigt föreligger en svår avvägning mellan samhällsnytta och hälsorisker.

Om hanteringen beror på hur man tolkat resultaten, hur åtgärder har implementerats eller om man inte tagit särskild stor hänsyn till den befintliga statistiken kan jag tyvärr inte uttala mig om. Däremot är det uppenbart att den statistik som presenterats efter sommarmånaderna har gett enskilda beslutsfattare underlag för att markera oro. Vi ser även tecken på att den redan har och troligen även kommer att ligga till grund för beslut om åtgärder och rekommendationer.

ÅKE WISSING,
WISSING & CO

NY EUROPEISK ARBETSGRUPP FÖR COVID-19-STATISTIK

Förkortningen FENStatS står för Federation of European National Statistical Societies. Det är en sammanslutning av nationella statistiska föreningar från 24 europeiska länder, för Sveriges del Svenska statistikfrämjandet. Ordförande är Walter Radermacher, tidigare chef för Eurostat. Han blev nyligen omvald för ytterligare en treårsperiod. (Det har nog inte undgått Statistikfrämjandets medlemmar att arbetet med ackreditering av statistiker har bedrivits av en arbetsgrupp inom FENStatS.)

Sommaren 2020 bildades en ny arbetsgrupp inom FENStatS, nämligen *arbetsgruppen för covid-19*. Svensk kontaktperson är Hans Alberg. Gruppens arbete leds av Katharina Schüller (som också leder ett konsultföretag inom statistik och "data science"). Hon är även känd som en av personerna bakom bloggen "Unstatistik des Monats", där man diskuterar och ifrågasätter statistik som presenteras i massmedia och på andra ställen.

Under sommaren har den nya arbetsgruppen hållit tre möten. På det första bestämdes att man skulle samla in adekvata länkar på "dashboards", tillgängliga data och studier. Detta har gjorts och resultatet finns i form av ett excelark. Hur man skall hantera materialet långsiktigt är ännu inte klart, men ett förslag är att lägga till material till Eursostats covid-19 sida. För närvarande pågår ett arbete att överföra informationen till FENStatS Wikisida.

Det andra mötet handlade om ett schweiziskt initiativ att främja "data literacy". Förslaget kom ursprungligen från en schweizisk läkarförening, där man insett behovet av kompetens för hantering av data och statistik. Diskussionerna har pågått i några år, men man såg i och med covid-

19 ett behov av att lyfta frågan till hög politisk nivå. Walter Radermacher har stött initiativet å FENStatS vägnar, och diskussioner pågår hur man skall lyfta initiativet till Europeisk nivå.

Det tredje mötet handlade om att ge kommentarer till ett UNWDF (United Nations World Data Forum) dokument "Global data community's response to COVID-19". Dokumentet var på rätt hög nivå och deadline för kommentarer var den 20 augusti.

Ett annat dokument som ska diskuteras inom gruppen är

National Academies of Sciences, Engineering, and Medicine (2020): *Evaluating Data Types: A Guide for Decision Makers using Data to Understand the Extent and Spread of COVID-19*. (Washington, DC: The National Academies Press). [<https://doi.org/10.17226/25826>].

Författarna av dokumentet har bett om feedback.

Den amerikanska artikeln är fokuserad på direkt mätbara storheter såsom antal avlidna och andel positiva testfall, och man går då igenom för- och nackdelar med olika mått. Intressant är att man även nämner mängden virus i avloppsvatten. Veterligen pågår ett antal projekt inom området, men endast Nederländerna redovisar resultat på sin websida. Dokumentet borde möjligen kompletteras med några härledda storheter såsom R-värdet och uppskattade antalet smittbärare. En fråga för FENStatS covid-19-grupp är också om man borde ta fram ett liknande dokument för Europa.

HANS ALBERG

(Notisen ovan beskriver läget i slutet av september 2020.)



Har du flyttat?

■ Du kan själv ändra dina uppgifter genom att gå till <https://www.membit.net/m4-member/login> där du loggar in med dina personliga inloggningsuppgifter (finns på din senaste inloggningsavi). Vid frågor kontakta Mattias Strandberg på sekrfram@gmail.com.

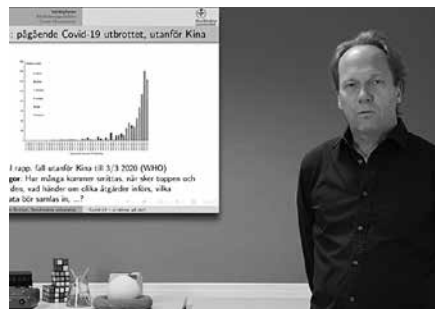
Jag, Tom Britton, har blivit ombedd att skriva ett bidrag till Qvintensen om COVID-19, och eftersom jag hållit oräkneliga föredrag på detta tema, och totala antalet artiklar på ämnet nu nog överstiger 10 000, så tänkte jag i stället vara lite mer personlig och berätta om mina erfarenheter av den mediala uppståndelse som jag fick vara med om under våren och sommaren 2020.

TOM BRITTON

”15 minutes of fame” orsakade av covid-19

För den som inte vet vem jag är, så kan jag nämna att jag är professor i matematisk statistik på Matematiska institutionen vid Stockholms universitet, och jag råkar dessutom vara ordförande i Statistikfrämjandets akademiska underavdelning Cramér-sällskapet. Som professor bedriver jag undervisning, handledning och forskning, samt utför en del administration (jag är bland annat sektionsdekanus för Matematisk-fysiska sektionen av Naturvetenskapliga fakulteten vid mitt universitet).

Jag disputerade 1996 vid Stockholms universitet och var därefter postdoktor i Australien, följt av att jag var lektor vid Uppsala universitet i 5,5 år innan jag år 2003 återvände till mitt alma mater. Min avhandling handlade om statistiska metoder för att upptäcka heterogen smittspridning, till exempel inom familjer. Sedan dess har jag till stor del forskat vidare om smittspridningsmodeller, både ur ett probabilistiskt och ett statistiskt perspektiv, men jag har även forskat om modeller och statistiska metoder för nätverk, fylogenetiska träd m.m. Även om min forskning får anses vara tillämpad, har merparten av mina bidrag inom smittspridning varit av metodologisk karaktär och mer sällan handlat om numeriska slutsatser för specifika sjukdomar (men jag har några sådana för bl.a. HIV, TB och nu COVID-19). Under dessa år har jag ett antal gånger träffat Anders Tegnell, Lisa Brouwers, Johan Giesecke och Jan Albert med



Efter matematiklektionen på Youtube i mars började olika medier kontakta Tom Britton.

flera på olika sorters möten. Johan Giesecke, och framför allt Jan Albert, har jag även haft vetenskapligt samarbete med.

När nyheter om smittspridning av COVID-19 i Kina nådde mig i januari var det inget som påverkade min forskningsinriktning – jag jobbade då med en teoretisk modell där individer ändrar sina sociala kontakter i takt med att närstående blir smittade (nu färdigt; se <https://arxiv.org/abs/2008.06375>). I mitten av februari organiserade jag och Etienne Pardoux (Aix Marseille Université) en workshop i vackra Luminy (utanför Marseille) om just smittspridningsmodeller och statistiska analyser av epidemidata. Antalet deltagare var drygt femtio, men när fem personer ställde in sitt deltagande med kort varsel p.g.a. att de inte ansåg sig kunna lämna sin pågående forskning, började jag inse skalan av utbrottet (som ju då i huvudsak hållit sig inom Kinas gränser,

enligt vad som då var känt). Första kvällen på workshopen ordnade vi en diskussionssession om COVID-19 och jag minns väldigt tydligt hur Julia Gog från Cambridge öppnade mina ögon helt, när hon inledde diskussionen med att be alla oss teoretiker att lägga åt sidan våra viktiga och intressanta forskningsproblem för att i stället ägna närmaste halvåret åt att hjälpa henne och andra mer tillämpade forskare att öka kunskapen om spridningen av denna farsot. Så från den 17 februari blev det klart för mig hur stort detta skulle bli, och jag beslöt mig för att inrikta min forskning mot problem av relevans för COVID under denna period.

Utöver workshopen i Frankrike hade jag ett annat evenemang om smittspridning inplanerat. På ”pi-dagen” den 14 mars skulle jag hålla ett populärvetenskapligt föredrag om smittspridning på Tekniska museet. När COVID blev aktuellt föll det sig naturligt att exemplifiera med just detta utbrott, men en vecka innan pi-dagen bestämdes det att hela evenemanget på Tekniska museet skulle ställas in på grund av just smittspridningen av COVID. Att helt ställa in ett föredrag om smittspridningsmodeller för bland annat COVID på grund av just smittspridningen av COVID kändes emellertid fel, varför jag och organisatören Elin Ottergren på Vetenskapens hus beslöt oss för att spela in föredraget och lägga ut det på YouTube. Eftersom vi ändå ordnat med professionell kamerautrustning och klippning m.m. beslöt vi att även spela in föredraget på engelska. Den 35 minuter långa »»»

»»» "15 MINUTES OF FAME" ORSAKADE AV COVID-19

matematiklektionen (på engelska och svenska) om smittspridningsmodeller har nu visats över 100 000 respektive över 10 000 gånger!

Ungefär samtidigt med YouTube-filmerna, och kanske delvis tack vare dem, så började olika medier kontakta mig för att be mig uttala mig om läget. Av naturliga skäl ville man nog ha in många olika röster som uttalade sig om vad som skulle kunna hända, och eftersom jag är en av få svenska smittspridningsmodellerare var det inte konstigt att även jag kontaktades. Den kanske vanligaste frågan gällde nog hur många som skulle komma att smittas, och när. Jag började därför studera data över det initiala förloppet (med fokus på utvecklingen av antalet döda – antalet inrapporterade fall bedömde jag och andra som alltför osäkert, med liten koppling till verkliga antalet smittade). Med hjälp av enkla statistiska metoder för en enkel epidemimodell gjorde jag även prediktioner över utvecklingen i Stor-Stockholm (eftersom smittspridningen där var störst), både för vad som skulle hända utan preventiva åtgärder, men också hur det skulle kunna gå med olika nivåer av prevention. Jag uttalade mig om dessa prediktioner i media och var samtidigt mån om att betona den stora osäkerheten i dem. I backspegeln kan sägas att mina (men även andras) prediktioner överskattade antalet som skulle smittas, såväl utan som med preventiva åtgärder. Huvudorsaken till överprediktionen ligger i att COVID spred sig mer heterogent än t.ex. influensa, som annars är den naturliga sjukdomen att jämföra med: vissa som smittas verkar inte ens smitta sina familjemedlemmar, medan andra kan smitta ett tiotal personer man bara kort vistats med i samma rum, s.k. klustersmitta. Jag sa bl.a. i medier att upp mot hälften skulle ha smittats innan sommaren, och ibland var jag inte tillräckligt tydlig med att jag menade i Stor-Stockholm (vilket dock tydligt framgår i rapporten), så några av dessa citat blir jag påmind om av personer som tycker att jag ska sluta uttala mig i medierna... Förutom att påpeka den stora osäkerheten i mina uttalanden har jag därefter även varit tydlig med att vissa av mina prediktioner varit felaktiga (en rimligare siffra är nog att ca 25 % av Stockholmarna hade smittats fram till midsommar, men det exakta svaret vet ingen).

Hur som helst blev det under mars-maj väldigt många medieframträdanden med bl.a.



Tom Britton på SVT den 31 mars ...



... och på Nyhetsmorgon i TV 4 den 1 april.

sex besök i Aktuelltstudion, men också väldigt många intervjuer i andra medier, främst nationellt men också internationellt – eftersom Sverige stack ut i sin hantering av COVID uppstod ett stort intresse även utomlands ifrån. Jag gjorde under samma period fem publika Facebook-inlägg varav de senare fick väldigt stor spridning (tusentals delningar och likes). Tillsammans skapade alla dessa faktorer en blandning av rus och stress hos mig. På Facebook fick jag en uppsjö av (mest positiva) kommentarer och jag fick väl ca 100 mejl om dagen, blandat beröm, enkla frågor och mer omfattande förfrågningar där jag ombads studera avsändarens text och simuleringar. Jag minns att min fru ringde mig en dag kl 17:23 och det var fjortonde samtalet den dagen, men det första från ett nummer som jag kände igen – de tidigare telefonsamtalen hade varit från journalister samt några vanliga medborgare. Tiden var således spännande men skapade också en hel del stress.

Under maj och därefter kom jag igång med lite mer forskning. Bland annat en artikel som visar att flockimmunitetsnivån snarare ligger runt 40–45% än tidigare uppgifter om 60%. Orsaken till denna skillnad är att när immuniteten uppstår via smittspridning, så blir de socialt aktiva i högre grad smittade (och därefter immuna), medan siffran 60% baseras på vaccinering då det inte sker någon dylik selektion av immuniteten. Artikeln publicerades i tidskriften *Science*, och med ett sådant tema under COVID-pandemin

och i den tidskriften, blev uppståndelsen minst sagt stor. Jag blev under sommaren intervjuad 50–100 ggr av olika internationella medier, bl.a. CNN, NY Times, Daily Mail, Brasiliens 60 Minutes m.m. Eftersom mitt universitet (liksom andra universitet) uppmanats att öka antalet utbildningsplatser under sommar och höst höll jag och en kollega en web-baserad sommarkurs om matematisk-statistiska smittspridningsmodeller där vi fick begränsa antalet deltagare till 150 personer... Jag medverkade även i två sommarskolor, vid University of Washington respektive Peking University, så sommaren blev liksom våren väldigt hektisk men också rolig och intressant. Nu under hösten har läget för mig lugnat sig lite och jag har på nytt fått fart på min forskning, denna gång om risken för en andra våg och hur den beror på R_0 och immuniteten i befolkningen (den nyfikne får hålla utkik på MedRxiv). Jag har den senaste veckan (i slutet av september) noterat att modelleringsintresset i medierna verkar vara på väg tillbaka, så kanske blir det även en andra medievåg för mig under hösten ...

Vilka är då mina erfarenheter av media och av att medverka i media? Mitt huvudintryck är mycket positivt: jag tycker merparten medier är proffsiga, många är väl pålästa, och de har ingen dold agenda som de försöker lura in mig i. Rent personligen tycker jag att jag har klarat mig bra, dvs mina huvudbudskap om osäkerhet och att med väldigt enkla medel försöka förklara slutsatserna tycker jag överlag har nått fram väl. En sak jag noterade var att SVTs Aktuellt har väldigt stort genomslag. Dagen efter kan man förvänta sig en hel del respons i mejl och sociala medier, och dessutom att andra medier hör av sig om intervjuer. Det kommer också förfrågningar om att hålla föredrag: jag har hållit ca 50 föredrag sedan i mars (två i veckan i snitt under hela perioden), vetenskapliga eller mer populärvetenskapliga, i en mängd olika svenska och internationella sammanhang. Jag brukar försöka svara på alla mejl, men har tyvärr inte haft tid att läsa alla utkast till artiklar och simuleringstudier som jag fått mig tillskickat, trots att många av dem verkat intressanta – min tid har helt enkelt inte räckt till.

En annan insikt är att min korrespondens som statstjänsteman är offentlig. Detta har jag såklart känt till sedan länge men aldrig reflekterat över, och inte heller har min korrespondens

tidigare begärts ut. Men nu... Under våren hade jag ganska tät kontakt med Tegnell, Giesecke, Albert med flera i olika trådar där allehanda saker diskuterades. Inga hemligheter, men ibland uttryckte sig folk inte så diplomatiskt, och nu har dessa olika korrespondenser begärts ut av ett flertal individer (journalister men även tillsynes "vanliga medborgare") och figurerat i såväl kvällstidningar som sociala medier. Bara till min institution har vi fått in ca 10 begäranden om mejlkorrespondens jag haft med olika andra adressater. Jag kan inte säga att jag hade betett mig väsentligt annorlunda om jag tänkt på att mejlen skulle komma att begäras ut, men jag hade nog inte vävt in mer privata kommentarer i samma mejl som COVID avhandlas i, och så hade jag sorterat min korrespondens lite bättre ... Visserligen har jag möjlighet att ta bort privata kommentarer i jobbmejl innan de lämnas ut, men när det handlar om >1 000 mejl har i alla fall inte jag orken att gå igenom dessa mejl ...

Även forskningsmässigt har det senaste halvåret varit spännande. Problemfrågeställningarna har ändrats i en dramatisk takt: Hur stort är R0? Hur många kommer att smittas före sommaren, med och utan restriktioner? Vilken är effekten av olika preventiva insatser? Vad är risken för en stor andra våg? Nya frågeställningar har blivit aktuella: Vad är effekten av en total lock-down? Har det någon effekt att införa reseförbud? Viktiga lärdomar har också dragits om spridningen av COVID, även om många osäkerhetsfaktorer fortfarande finns. COVID har en mer heterogen smittspridning än t ex influensa: många smittas inte någon men en del tycks vara väldigt smittsamma, R0 varierar ganska mycket mellan olika länder men också inom länder, effekten av olika preventiva åtgärder är väldigt svår att skatta, immunitet erhållen via smittspridning minskar smittspridningen i högre grad än immunitet från vaccinering, m.m. Mest häpnadsväckande är för mig ändå den väldiga omställning som i princip hela världen snabbt genomfört och även till stor del stått ut med. Det inger förhoppning inför andra stundande utmaningar, såsom klimathotet!

TOM BRITTON



Nya förhållningsregler i detaljhandeln. Den väldiga omställning som i princip hela världen till stor del stått ut med är det mest häpnadsväckande, menar Tom Britton.

Anders Nordgaard, som utsetts till 2020 års statistikfrämjare, berättar här om hur man kan resonera med hjälp av subjektiva sannolikheter, när man som statistiker i forensiska eller rättsliga sammanhang ska bedöma styrkan hos bevis.

Låga sannolikheter i forensiska bedömningar

Bakgrund

Vad kan det vara som gör att man en dag kommer till insikt om att man hellre använder sig av den subjektiva tolkningen av sannolikhetsbegreppet än den frekventistiska? Professor Mattias Villani brukar ibland lägga till epitetet "Natural Born Bayesian" i sin kommunikation, och kanske är det så det är för oss som ansluter oss till det Bayesianska paradigmet. Men i mitt eget fall upptäckte jag inte detta förhållande senare i livet. Som ung amanuens och sedermera forskarstuderande undervisade jag i sannolikhetslära utifrån den klassiska inledningen om slumpmässiga försök som kunde upprepas oberoende av varandra och under samma förutsättningar, och vars utfall inte var kända på förhand. Med enfasett upplyste jag mina studenter om att det inte går att uttala sig om sannolikheten för att man cyklar omkull på vägen hem från universitetet eller att man klarar nästa tentamen, eftersom det inte rörde sig om försök som kan upprepas under samma förutsättningar.

På ett sätt var jag nöjd med att göra detta, eftersom det fanns en definition att hela tiden relatera till. På ett annat sätt var jag inte alls nöjd. Varför kan vi tala om risken att man cyklar omkull och chansen att man klarar en tentamen, men inte om sannolikheten för någondera? På den tiden var det heller inte gnijselfritt att lyfta tankar om subjektiva sannolikheter eller om det Bayesianska paradigmet.

Mötet med SKL

Så kom det sig en vacker dag att jag

blev involverad i forensiska frågeställningar och skickades på konferenser, för att med min kompetens i statistik sätta mig in i om den pågående utvecklingen inom forensisk bevisvärdering var något som dåvarande Statens kriminaltekniska laboratorium (SKL) skulle närma sig (idag heter det Nationellt forensiskt centrum, NFC¹). Det tog ett tag att få rätsida på detta. Jag kom till dessa konferenser fortfarande aktiv inom miljöstatistisk forskning och till viss del med en föresats om att datorintensiva statistiska metoder (som på den tiden var bootstrap och korsvalidering) var lösningen på många problem. Och även om sådan metodik i modern form nu har börjat användas även inom det forensiska området, så var det inte alls att tänka på vid det laget (och då pratar vi om tiden runt sekelskiftet – det senaste alltså).

Jag kom till SKL med uppdrag att utbilda och vara statistisk konsult – det var ett avtal mellan SKL och Linköpings universitet – och jag kom raskt fram till insikten att det i detta sammanhang inte var fråga om att arbeta med ett stort antal mätningar. Skulle man jämföra ett glasfragment säkrat från en misstänks kläder med glaset i en krossad ruta, så gjorde man två eller tre mätningar av brytningsindex på respektive glas, och resultatet av dessa mätningar var det underlag på vilket man sedan skulle bedöma om glasfragmentet härrörde från den krossade rutan eller inte. Det stod klart att traditionella signifikanstest inte var att tänka på, och det var inte särskilt meningsfullt att föreläsa

med påhittade räkneexempel av detta slag, trots en aldrig så ambitiöst vald kontext hämtad från laboratoriets verksamhet för att underlätta förståelsen. Ansatsen föll pladask när jag i exemplen gjorde antaganden om runt 10 mätningar av en storhet – några frihetsgrader måste det ju finnas för att man ska kunna demonstrera hur ett testresultat blir signifikant.



Anders Nordgaard.

Om forensiska bedömningar

Forensisk verksamhet, men också juridisk, är behäftad med osäkerhet, som inte kan ignoreras. Samtidigt är osäkerheten svår att mäta, då den egentligen finns i en persons *föreställning* om något och inte som en objektiv storhet, som betyder samma sak för envar. Jag tror att det var här någonstans, som jag själv kom till uppfattningen att sannolikheter i detta sammanhang måste tolkas som subjektiva, vilket förstås inte utgör något hinder för att flera personer kan ha lika stor sannolikhet för något, och att en sådan uppfattning kommer från ett insamlat datamaterial, som alla har tillgång till. Om vi betraktar den ultimata hypotesen i ett rättsfall, dvs. att den tilltalade är skyldig, så lär de flesta hålla med om att personen endera är skyldig eller oskyldig till det brott hen anklagas för, men att ingen förutom möj-

ligen den tilltalade och dennes brottsoffer kan med säkerhet veta. Sannolikheten att personen är skyldig är ju inte ett okänt tal, för vilket ett närmevärde med hög noggrannhet skulle kunna bestämmas efter att situationen som frammanat misstankarna mot den tilltalade har upprepats under samma betingelser ett stort antal gånger.

Ändå handlar det om att åklagaren måste bevisa att det är ställt utom rimligt tvivel att den tilltalade är skyldig, och att måttet på tvivel har sådana egenskaper att det kan ses som ett sannolikhetsmått. Språkbruket – oavsett inställning till sannlikhetsbegreppet – är sådant att vi ibland kan säga oss ”vara 100 % säkra på något”. Oavsett om den är 100 % eller något annat så har måttet därför en övre gräns, och funderar vi över uppräknelig additivitet så tycker vi nog att även den egenskapen gäller för mått på tvivel. Det finns förstås andra teorier om bevis, t.ex. Dempster-Shafer’s evidensteori², men att sannolikheter inte skulle kunna tillämpas på en situation som inte kan upprepas under i princip identiska betingelser har jag inte sett framföras utanför det frekventistiska paradigmet.

Sannolikheten att en tilltalad är skyldig är enligt min uppfattning lika med tilltron till att den tilltalade är skyldig hos den som har att förhålla sig till denna frågeställning, till exempel hos en domare i tingsrätten. Med detta som grund faller det sig naturligt att dela upp bevisningen i komponenter, som relaterar till varandra sannolikhetsmässigt. En sådan komponent är utfallet av en forensisk undersök-

ning, till exempel jämförelsen mellan ett glasfragment och glaset i en krossad ruta. Det är svårt (för att inte säga ogörligt) att bedöma hur sannolikt utfallet av en sådan jämförelse är om vi utgår från att den tilltalade är skyldig jämfört med hur sannolikt det är om vi utgår från att den tilltalade är oskyldig. Men om vi utgår från att den tilltalade är skyldig (kanske i detta fall till grov stöld, vilken är brottsrubriceringen vid inbrott), så har det inverkan på bedömningen av sannolikheten att den tilltalade krossade fönsterrutan. Om vi sedan utgår från att den tilltalade krossade rutan, så har det förstås inverkan på bedömningen av sannolikheten att ett säkrat glasfragment från den misstänktes kläder kommer från rutan. Om vi slutligen utgår från att glasfragmentet kommer från rutan, så har det inverkan på bedömningen av sannolikheten att få det utfall man fick i glasjämförelsen.

Sannolikhetsresonemang fungerar när det gäller att utreda hur osäkerhet i olika steg inverkar på den bedömning som slutligen ska göras om huruvida åklagaren har bevisat att det är ställt utom rimligt tvivel att den tilltalade är skyldig till den gärning hen åtalas för. Det är förmodligen att slå in öppna dörrar när jag påstår att domstolen inte med nödvändighet kommer fram till vad som är sant eller ej. När domstolen avger en

fällande dom så ska det uppfattas som ett beslut att beviskravet är uppfyllt, men att det är fortfarande ett beslut taget under osäkerhet.

Resonemang kring ett räkneexempel

I en känd rättegång för några år sedan stod en person åtalad för terroristbrott. Del av bevisningen utgjordes av en mobilfilm säkrad på ett USB-minne hemma hos personen ifråga. Filmen visade en avrättning, och flera personer delaktiga i handlingen var synliga, dock med maskerade ansikten. I några delar av filmen syn-tes tydligt handen (ovansida och insida) på en person. Vid en jämförelse mellan

handen på filmen och motsvarande hand på den åtalade observerades flera överensstämmelser, bland annat beträffande linjemönster i huden och storlek och placering hos ett ärr. Sådana överensstämmelser är förstås inte särskilt förväntade, och man fann nu fyra

olika överensstämmelser av detta slag. Några avvikelser kunde inte observeras. Det krävs egentligen inga ingående sannolikheteoretiska analyser, för att komma fram till att sannolikheten att träffa på en person vars hand har fyra specifika kännetecken, som direkt motsvarar de man ser på filmen, är mycket liten. Handläggande forensiker vid NFC bedömde att

»När domstolen avger en fällande dom så ska det uppfattas som ett beslut att beviskravet är uppfyllt, men att det är fortfarande ett beslut taget under osäkerhet.«

»»» LÅGA SANNOLIKHETER I FORENSISKA BEDÖMNINGAR

denna sannolikhet var mindre än en på 250 miljoner.

NFC rapporterar sina slutsatser på en niogradig ”utlåtandeskala”³ som går från grad -4 till grad +4, där varje grad motsvarar ett intervall av så kallade Bayesfaktorer (som regel likelihoodkvoter; i NFC:s utlåtanden benämns de *resultatvärden*). I en artikel i Qvintensen 2013:1 redogjorde jag för hur värdering av forensiska resultat görs vid NFC (som då fortfarande hette SKL). För den som inte läst den artikeln kan man någorlunda kortfattat säga att bevisvärdering utifrån ett Bayesianskt synsätt omfattar beslutsorienterad statistisk inferens. I ett brottmål förväntas rättsens ledamöter fatta ett beslut som minimerar förväntad social kostnad. Beviskravet⁴ är sedan 40 år uttalat som att det ska vara ställt utom rimligt tvivel att den tilltalade är skyldig och att en tillräcklig utredning har gjorts. Detta tolkas av de flesta som att sannolikheten för att den tilltalade är skyldig givet vad som framkommit vid rättegången måste vara mycket hög, även om få vill specificera en numerisk gräns. Ur inferenssynvinkel handlar det hur som helst om att bedöma sannolikheter för hypoteser som åtminstone i teorin kan användas för att beräkna förväntad social kostnad. Det ovan nämnda beviskravet signalerar därför att den sociala kostnaden för att en oskyldig blir fälld är avsevärt mycket större än den sociala kostnaden för att en skyldig blir friad. Bedömning av sannolikheter för hypoteser hör typiskt till det Bayesianska paradigmet för statistisk inferens och bedömningarna görs genom att utgå från initialt satta sannolikheter för hypoteserna och successivt uppdatera dem med hjälp av Bayes sats⁵.

NFC:s uppgift i detta är att värdera sina undersökningsresultat genom att uppskatta likelihooderna för två motstående hypoteser. Dessa kan vara formulerade som förklaringar endera till ursprung hos de bevisfakta (spår och annat omstritt material) som undersökts, eller till hur bevisfakta har uppstått. Exempel kan vara ”Glasfragmenten som säkrats på den misstänkte kommer från det krossade fönstret”, ”De två beslagen av amfetamin kommer från en och samma tillverkningsplats” och ”Den misstänktes tröja har varit i kontakt med förarsätet i bilen (vilket förklarar varför bomulls-fibrer överensstämmande med de i tröjan har påträffats på bilsätet)”. Kvoten av dessa båda likelihooder benämns som

sagt resultatvärde och skall användas i utredningen för att uppdatera de sannolikheter för hypoteserna som där bedömts utifrån annan information än de forensiska resultaten. Bedömningen av resultatvärdets storlek är med några undantag svår att göra med hög precision, och därför rapporteras endast det intervall av resultatvärden – ett av de nio förutbestämda intervall som framgår av utlåtandeskalan – i vilket det aktuella resultatvärdet bedöms ligga.

Intervall för grad +4 går från en miljon till oändligheten. En sannolikhet om en på 250 miljoner för ett oskyldigt sammanträffande innebär att gränsen för grad +4 passerats med mycket god marginal. Polisiära utredare, åklagare, försvarsadvokater och domare är väl medvetna om att dna-bevisning rörande ursprung hos ett biologiskt spår (blod, sekret mm.) nästan alltid redovisas med den högsta slutsatsgraden, dvs. +4. I princip ingen ifrågasätter idag dna-bevisning rörande ursprung, men inte särskilt många är medvetna om att underliggande data kommer från drygt 400 personer och att skälet till de höga slutsatsgraderna beror på de (verifierade) populationsgenetiska modeller som används.

I utlåtandet om de överensstämmande observationerna på handen framgick inte överskattningen en på 250 miljoner, utan endast slutsatsgrad +4 redovisades. Överskattningen stod dock att läsa i det arbetsblad som kan begäras ut. När detta blev känt hos försvaret blev NFC öppet kritiserat i TV och andra media⁶ för att ha värderat sina resultat *högre* än vad som görs vid dna-bevisning. Försvaret refererade till att dna-bevisning brukar nå grad +4 och gränsen går ju vid en miljon, i princip tolkningsbart som att sannolikheten för ett oskyldigt sammanträffande är högst en på en miljon. Här var det alltså hiskeliga en på 250 miljoner. ”Detta kan inte vara korrekt, det kan ju inte finnas så mycket data över utseendet hos händer.” ”Fingeravtrycks-expert har ju sagt att linjemönstret inte kan värderas med ackrediterade metoder, så hur kan ni då ändå värdera detta?”. Citaten är inte äkta, men återspeglar andemeningen i den kritik som framfördes.

Av någon anledning accepterar olika aktörer runt ett rättsfall att NFC redovisar starka slutsatser när det gäller dna-bevisning och fingeravtryck (enligt ackrediterad metodik), men så fort det

görs gällande att en stark slutsats bygger på en uppskattad låg sannolikhet, så är det många som blir ”självutnämnda experter”, och resonemang om sannolikheters tillkomst går alltid tillbaka till hur relativa frekvenser av sådan storleksordning kan ha beräknats – det frekventistiska paradigmet lyser på ett sätt igenom i det att man anser att det måste finnas tillförlitliga databaser från vilka sannolikheterna uppskattas.

När det gäller överskattningen en på 250 miljoner, så kommer man ganska snabbt ned till sådana nivåer, om man har belägg för att flera observationer inte är relevanta för varandra i ett allmänt perspektiv. I det aktuella ärendet är det till exempel problemfritt att hävda att linjemönstret i en persons handflata inte kan vara relevant för huruvida personen har ett ärr på handen eller inte, och motsvarande gällde även för de övriga två egenskaperna. Bedömningen av sannolikheten för att göra de observationer som gjorts på en slumpmässigt vald hand skulle kunna göras genom en relativt enkel studie bland exempelvis personalen på NFC. När det gäller det observerade ärrer kan hänsyn tas till att den aktuella filmen återgav en avrättning i terroristssammanhang, och att ärr på personer involverade i sådana aktiviteter kan antas vara något mer sannolika än i andra referensgrupper. Övriga observerade egenskaper kan knappast påstås vara stratifierade i någon referenspopulation.

Säg att man för var och en av de fyra egenskaperna ansätter en icke-informativ apriori-fördelning för proportionen individer som uppvisar egenskapen ifråga i den population av individer som kan tänkas förekomma på en film av sådant slag som den som undersökts. En lämplig sådan apriorifördelning är en likformig fördelning mellan 0 och 1. En sådan fördelning är egentligen orimlig, eftersom envar lär anse att det inte kan vara fler (utan snarare mycket färre) än i genomsnitt varannan person, som har egenskapen ifråga. Valet av denna fördelning kompenserar å andra sidan med råge att ett ärr skulle vara vanligare hos personer med anknytning till terroristaktiviteter och att valet av referenspopulation (NFC:s personal) kan leda till underskattningar. NFC har drygt 450 medarbetare stationerade i Linköping (där bildundersökningarna görs), och det mest förväntade utfallet, om alla dessa medarbetare skulle under-

sökas, är att ingen skulle uppvisa ett ärr av samma storlek och placering som det ärr som syns på handen på filmen. Utan att gå in på några detaljer är det heller inte förväntat att någon bland NFC:s medarbetare skulle uppvisa övriga tre egenskaper som överensstämde mellan handen på bilden och den åtalades hand, och som bedömts inte vara förväntade i allmänhet. Det är inte möjligt att empiriskt bevisa att en egenskap hos en person är unik, men den samling av egenskaper vi har att göra med här hör till sådant som många skulle anse vara unikt.

Bayes sats på normalform för uppdatering av sannlikhetsfördelningar⁷ kan nu användas för att uppdatera den ansatta fördelningen. Den likformiga fördelningen mellan 0 och 1 är detsamma som en betafördelning⁸ med parametrar $a = 1$ och $b = 1$ och betafördelningen är i sin tur konjugerad till binomialfördelningen. Det innebär att om vi som apriorifördelning ansätter en betafördelning med parametrar a och b för en proportion θ av en viss egenskap, och observerar x individer med denna egenskap i ett slumpmässigt urval om n individer från en mycket stor population, i vilken proportionen positiva är θ , så blir aposteriorifördelningen också en betafördelning med parametrar $a + x$ respektive $b + n - x$. I denna fördelning är väntevärdet $(a + x)/(a + b + n)$. Detta är också den punktskattning av θ som ger minsta förlust vid kvadratisk förlustfunktion. På engelska kallas den *posterior mean* och på svenska skulle vi kunna säga "förväntat värde *ex post*" som är den språkkonstruktion jag kommer att använda nedan. Man

skulle alternativt kunna säga "förväntat värde *a posteriori*" eller ännu krångligare "aposterioriväntevärde", men min åsikt är att vi egentligen borde försöka hitta ett försvenskat uttryck både för *a priori* och *a posteriori* för att slippa ta till latinet i olika former.

Om vi nu utgår från att ingen av NFC:s medarbetare i Linköping uppvisar de aktuella egenskaperna, så blir för varje egenskap den förväntade proportionen *ex post* $1/452 \approx 0,0022$ i populationen av individer som kan tänkas förekomma på en film av sådant slag som den som undersökts (i det fortsatta avses just denna population för beräknade proportioner). Utifrån bedömningen att de fyra egenskaperna inte är relevanta för varandra blir den förväntade proportionen individer *ex post* som uppvisar alla fyra egenskaper ungefär $2,3 \cdot 10^{-11}$, en proportion som (än så länge) inte skulle kunna observeras som en relativ frekvens med ens hela jordens befolkning som rampopulation. Om varje egenskap hittas hos exakt en medarbetare så blir den förväntade proportionen individer *ex post* med alla fyra egenskaper cirka $3,8 \cdot 10^{-10}$. Den skulle bli runt en på 250 miljoner – det vill säga lika med den sannolikhet som bedömdes i ärendet – om 3-4 medarbetare skulle uppvisa respektive egenskap.

Nu kan man ju invända mot att dessa proportioner har räknats fram genom att multiplicera det förväntade värdena i aposteriorifördelningarna för proportionerna för de enskilda egenskaperna. Om vi istället beräknat uppåt begränsade 99%-iga kredibilitetsintervall för var och en av proportionerna så hade

alla dessa blivit $(0 - 0,0102)$ om ingen av 450 medarbetare uppvisar någon av egenskaperna. Med användande av den övre gränsen i detta intervall – istället för den förväntade proportionen – för var och en av egenskaperna blir bedömd proportion individer som uppvisar alla fyra egenskaperna $1,08 \cdot 10^{-8}$, eller cirka en på 92,5 miljoner. Om 3 medarbetare skulle uppvisa respektive egenskap så blir motsvarande uppåt begränsade kredibilitetsintervall $(0 - 0,022)$ och bedömd proportion individer med alla fyra egenskaperna blir $2,34 \cdot 10^{-7}$ eller cirka en på 4,2 miljoner. Om nu sannolikheten för att träffa på en person vars hand har fyra specifika kännetecken, som direkt motsvarar de man ser på filmen, sätts lika med uppskattad proportion, blir den i bägge fallen högre än en på 250 miljoner, men långt ifrån så hög att gränsen för grad +4 i NFC:s utlätandeskala inte skulle nås.

Det ska slutligen sägas att bedömningen en på 250 miljoner i ärendet byggde på mer än bara de rent statistiska uppskattningar man kan få fram genom kalkyler motsvarande de som redovisats ovan. Poängen är dock att det via tankeexperiment och rimlighetsbedömningar går att resonera med sannolikheter utan att det behövs finnas adekvata databaser tillgängliga.

Avslutningsvis vill jag passa på att tacka för ha blivit utsedd till årets statistikfrämjare, en hedervärd utnämning som jag är mycket stolt över.

ANDERS NORDGAARD

Forensisk specialist
vid Nationellt forensiskt centrum
(NFC)

Noter

¹ nfc.polisen.se

² Dempster A.P. (1968). A Generalization of Bayesian Inference. *J. R. Statist. Soc. B*, 30(2): 205-247. ; Shafer G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

³ nfc.polisen.se/tjanster/utlätandeskala/

⁴ Beviskrav och utredningskrav. Se till exempel "Lambertz G. (2010). Bevisprövning och utredningskrav i brottmål. *Festschrift till Torkel Gregow*. Norstedts juridik, Stockholm."

⁵ Om H står för en hypotes som åklagaren vill bevisa och $\neg H$ för dess motsats, samt E står för den information (t.ex. forensiska undersökningsresultat) som ska användas för uppdatera en sannolikhet för H , kan Bayes sats på

s.k. oddsform användas för uppdateringen enligt $\frac{P(H|E)}{P(\neg H|E)} = \frac{P(H)}{P(\neg H)} \times \frac{P(E|H)}{P(E|\neg H)}$. Kvoten $\frac{P(E|H)}{P(E|\neg H)}$ kallas Bayesfaktor och är ofta en likelihoodkvot (de ingående sannolikheter tolkas då som likelihooder för de två hypoteserna).

⁶ Se t.ex. "https://www.expressen.se/tv/nyheter/brottscentralen/hela-program/brottscentralen-21-mars---se-hela-avsnittet-har/" (besökt 9 september 2020). För ytterligare bemötande av kritiken, se "Nordgaard, Rasmusson (2017). Professionell värdering av forensiska fynd borgar för rättssäkerhet. *Juridisk Tidskrift* 29(1): 228-232."

⁷ $q(\theta, x) \propto f(x, \theta) \cdot p(\theta)$ där $p(\theta)$ är apriorifördelningen hos parametern θ , $q(\theta, x)$ är aposteriorifördelningen hos θ givet data (x) och $f(x, \theta)$ är likelihooden för θ given av dessa data.

⁸ $f(y, a, b) = \frac{y^{a-1}(1-y)^{b-1}}{B(a, b)}$, $0 < y < 1$

Det talas en del om "den fjärde industriella revolutionen" eller "Industrie 4.0", där tillgången till data är en viktig aspekt. Detta kan onekligen innebära nya utmaningar för statistikerna och även påverka sannolikhetsläran. Jag har försökt att utröna hur olika statistiska institutioner tänkt hantera dessa utmaningar, men har haft svårt att få någon klar bild av detta. Men genom att studera historien kan man kanske få en känsla för hur sådan påverkan kan se ut.

Sannolikhetslärans utmaningar inför "Statistik 4.0"

Introduktion

Trots att jag egentligen inte haft statistik och sannolikhetslära som huvudinriktning under min studietid eller yrkesverksamma tid så har jag hyst ett visst intresse för ämnet och även kommit i kontakt med det i olika sammanhang. Detta har lett till en del funderingar som jag nu försöker få ned på papper, med förhoppningen att de ska inspirera till fortsatta undersökningar.

Även om det finns en förhistoria med Girolamo Cardano (1501–1576) och andra, så brukar man ange korrespondensen om diverse hasardproblem, mellan Pierre de Fermat (1607–1665) och Blaise Pascal (1623–1662) som startpunkten för sannolikhetslära som en vetenskaplig disciplin. Några generationer senare så kunde Pierre Simon de Laplace (1749–1827) konstatera att denna lärda debatt lett till insikter som påverkat ett flertal vetenskaper. Denna utveckling fortsatte efter Laplace; man behöver bara nämna namn som Gregor Mendel (1822–1884), William Sealy Gosset (mest känd som "Student" 1876–1937) och Agner Krarup Erlang (1878–1929), vilka lade grunden till nya viktiga tillämpningsområden.

Många kan i likhet med Tage Danielsson (1928–1985) uppfatta sannolikhetslära som något flummigt. Om meteorologen säger "imorgon kommer det att regna mellan 5 och 10 mm i Stockholm", så är det lätt att efteråt konstatera om hen hade rätt eller fel. Om prognosen däremot säger att "det är 60 % risk för regn" är det svårt att säga om utsagan var rätt eller fel. Risken kanske var 40 % eller 80 %. Möjliga

skulle man kunna samla alla "sextio procent-utsagor" och i efterhand kontrollera hur stor andel av dessa som slagit in. Ännu knepigare blir det naturligtvis med utsagor som att ett visst klimatscenario för år 2100 kommer att inträffa med 1 procents sannolikhet.

Likväl måste mänskligheten, och även djuren, haft någon form av intuitiv känsla för sannolikheter och risker, även om den inte alltid var exakt. Hade våra förfäder varit totalt orädda så hade de blivit ett lätt byte för sabeltandade tigrar. Hade de varit helt obenägna att ta risker hade de inte fått tag i mat, eller vågat att föröka sig. Å tredje sidan så drar man gärna alltför stora växlar på anekdotiska exempel, som när en inbiten rökare nämner madame Jeanne Calment (1875–1997). (Hon anses vara den person som blivit äldst, och vars ålder var säkerställd.) Hon rökte fram till några år före sin bortgång. Det finns således ett behov av att hantera sannolikheter på ett mer systematiskt sätt.

Det ligger naturligtvis långt över denna artikels ambitionsnivå att beskriva sannolikhetslärans historia, liksom alla dess nuvarande och möjliga framtida tillämpningsområden. Man kan exempelvis hänvisa till en bok med den talande titeln. *The empire of chance: How probability changed science and everyday life* (Gigerenzer m.fl., 1990). Man skulle även kunna hänvisa till Harald Cramérs (1893–1985) personliga betraktelser i Cramér (1976).

Som Guttorp och Lindgren (2019) konstaterade finns det i Sverige professurer i statistik vid ett antal olika fakulteter såsom teknisk, handels,

naturlivsvetenskaplig och medicinsk. Därutöver finns en professur i kemometri vid Umeå universitet.

Exempel på sannolikhetslärans industriella tillämpningar

Det finns ett antal intressanta tillämpningar av sannolikhetslära inom för Sverige viktiga industrigrenar. Jag väljer att nämna några som jag själv kommit i kontakt med; därutöver finns bl.a. läkemedelsindustrin.

Rent allmänt kan sägas att det inom flera industrigrenar finns en potential att öka användningen av metoder baserade på sannolikhetslära, exempelvis statistisk försöksplanering. I många fall har verksamheten bedrivits av eldsjälar. Några hinder för ökad användning har troligen varit bristande kompetens hos ledningsgrupper samt brist på data.

Det senare torde avhjälpas i och med "Industrie 4.0" (Ett tyskt uttryck som myntades i samband med Hannovermässan 2011 som ett samlingsnamn för en rad teknologier och koncept inom automation, processindustriell it och tillverkningsteknologier. Konceptet har koppling till begreppet "internet of things", som innebär att varje produkt i produktionskedjan bär med sig information om vart den ska gå vidare i kedjan.

Under sextio- och sjuttioalet var det långt ifrån självklart att använda datorer för utvecklingsarbete inom traditionell verkstadsindustri. Utvecklingen drevs då oftast framåt av engagerade medarbetare på "mellannivå". Motsvarande

ningar

skulle kunna tänkas ske i samband med den "fjärde industriella revolutionen".

Kemi samt gruv- och mineralnäring

Gruv- och mineralnäringen har länge varit en viktig industrigren för Sverige, vilket troligen förklarar landets tidiga framgångar inom kemin, med namn som Carl Wilhelm Scheele (1742–1786). Jöns Jakob Berzelius (1779–1848) och Svante Arrhenius (1859–1927).

Kemin skulle definitionsmässigt kunna betraktas som en gren av fysiken, men har av hävd ansetts som en egen disciplin, något mindre beroende av matematiken. Exempelvis så var matematikkurserna för kemistudenter på KTH något mindre krävande än de för teknologer på flertalet andra sektioner. Det finns kanske skäl att tänka om?

I början av sjuttioalet myntade den då nyblivne doktorn Svante Wold (1941–) termen kemometri, som i första hand handlar om att analysera datamängder från kemiska experiment medelst multivariat analys. Det har dock redan tidigare funnits nära band mellan kemi och matematisk statistik. Bl.a. kan nämnas att Harald Cramér studerade kemi i början av sin bana, liksom att den numera tydligen helt bortglömde tyske professorn Martin Hengst (1907–?) kombinerade kemi och statistik på ett fruktbart sätt, bl.a. vad gäller utvärdering av livsmedelskvalitet.

Kännedom om fördelningen av partikelstorlek kan vara viktig för att förstå ett materials fysikaliska och kemiska egenskaper. Relativt

många artiklar har skrivits om ämnet. Bl.a. så lyfter Hengst (1967) fram metodiken att anpassa en uppmätt fördelning till någon teoretisk (exempelvis log-normal) fördelning. Ofta har man försökt anpassa till den s.k. Rosin-Rammler-fördelningen, som egentligen är ett annat namn på Weibull-fördelningen, uppkallad efter den svenske ingenjören och vetenskapsmannen Waloddi Weibull (1887–1979). Denna fördelning definierades i anslutning till grundläggande forskning om metallers hållfasthet och livslängd. Svensken Jonas Svensson (1918–?) studerade undersökningar om partikelstorlekar, utförda vid Jernkontoret, varvid han definierade en mer generell fördelning, (Svensson 1955).

En annan viktig tillämpning av sannolikhetslära inom gruv- och mineralnäring är s.k. kriging, uppkallad efter den sydafrikanske gruvingenjören och professorn Daniel Krige (1919–2013). Kriging innebär att med statistiska metoder utifrån provbörningar försöka bedöma sannolikheten för större mineralförekomster.

IT och telekom

Den danske ingenjören Agner Krarup Erlang (1878–1929) lade grunden till teletrafikteorin, vilken senare tillämpades av bl.a. Ericsson, och som bidrog till koncernens framgångar. Conny Palms (1907–1951) klassiska studier brukar ofta framhållas som en pionjärinsats inom köteori, vilket i sin tur numera utgör en viktig hörnsten inom datorvetenskapen. Även Norge hade en framstående pionjär inom teletrafikområdet: Tore Olaus Engset (1865–1943), som tog fram

en formel för beräkning av sannolikheten för upptagetton under vissa förutsättningar.

Som bl.a. Bonald & Roberts (2012) påpekar är dagens internettrafik betydligt mer komplex än den teletrafik som Erlang studerade, men hans klassiska metoder och resultat kan fortfarande klargöra inverkan av nätutveckling. De inleder sin artikel med följande uttalande: "We demonstrate that the Internet has a formula linking demand, capacity and performance that in many ways is the analogue of the Erlang loss formula of telephony. Surprisingly, this formula is none other than the Erlang delay formula."

Inom telekommunikation har det varit viktigt att synkronisera tidsmätning mellan olika stationer. För detta syfte uppfann fysikern Allan (1936–) ett speciellt variansmått.

Maskin- och fordonsindustri

Vibrationer orsakade av exempelvis jordbävningar, vind och vågor kan betraktas som stokastiska. Stephen Crandall (1920–2013) anses vara en av föregångarna inom området. Crandall et. al. (1962) refererade till Arvid Palmgrens (1890–1971) delskadeteori. Palmgrens doktorsavhandling från 1930 behandlade den statistiska bärformågan hos kullager. Åren 1937–1955 var Palmgren chef för SKF:s tekniska avdelning.

Exempel på andra tillämpningar

Förutom de industriella tillämpningarna finns en mängd andra användningsområden av sannolikhetslära. På följande sidor ges några exempel för att påvisa bredden av dessa.

»»»



Försäkringsmatematik har varit en klassisk tillämpning av sannolikhetslära. För att kunna sätta rimliga försäkringspremier behövs kvantitativa uppskattningar av risker. Där torde ständigt uppkopplade bilar kunna erbjuda nya möjligheter till förfinade uppskattningar, och man kan naturligtvis också ha synpunkter vad gäller personlig integritet.

Astronomi

Astronomi kan synas vara en rätt deterministisk vetenskap: stjärnorna är där dom är och förflyttar sig enligt bestämda naturlagar. Astronomer började samla in stora datamängder på ett tidigt stadium. Kunskaper i sannolikhetslära och statistik behövdes för att behandla osäkerheterna i dessa datamängder. Dessa kunskaper kunde, som exempelvis Stamhuis (2012) visat med fallet Jacobus Kapteyn (1851–1922), tillämpas på andra områden. Andra exempel på personer som kombinerat statistik och astronomi är Pehr Wargentin (1717–1783), Adolphe Quetelet (1786–1875) och Carl Charlier (1862–1934).

Bildanalys

Analys av digitala bilder med algoritmer får en allt större betydelse, inte minst för medicinska tillämpningar.

Brownsk rörelse

En rätt välkänd tillämpning av stokastik är "Brownsk rörelse": en partikel nedsänkt i en vätska utför en slumpmässig rörelse beroende på att den kolliderar med vätskemolekyler. Fenomenet upptäcktes av den skotske botanisten Robert Brown (1773–1858) och analyserades senare av fysikern Albert Einstein (1879–1955).

Epidemiologi

Epidemier såsom covid-19 våren 2020 innebär ofta stor osäkerhet, och som tidigare nämnts handlar statistik och sannolikhetslära i stor utsträckning om att hantera osäkerheter. Det är därför inte speciellt märkligt att covid-19 medförde ökad efterfrågan på tabeller, diagram etc. Till en början handlade det rätt mycket om att tolka den officiella statistiken, men efterhand så ökade insikten om stora osäkerheter, varför sannolikhetsrelaterade frågeställningar som "R-värde", "falsa negativa testresultat" och "randomiserade tester" dök upp. I en del studier presenterades (ofta ganska vida) konfidensintervall.

Forensisk forskning

Aitkin et.al. (2010) påpekade i en skrift utgiven av Royal Statistical Society att "statistical evidence and probabilistic reasoning today play an important and expanding role in criminal investigations".

Försäkringsmatematik

Försäkringsmatematik har varit en klassisk tillämpning av sannolikhetslära. För att kunna sätta rimliga försäkringspremier behövs kvantitativa uppskattningar av risker. Där torde ständigt uppkopplade bilar kunna erbjuda nya möjligheter till förfinade uppskattningar, och man kan naturligtvis också ha synpunkter vad gäller personlig integritet.

Genetik

En annan tidig tillämpning av sannolikhetslära var genetiken. Dansken Wilhelm Johans-

sen (1857–1927), mest känd för att ha myntat ordet ”gen”, var en av pionjärerna vad gällde att behandla genetiska frågor med matematisk-statistiska metoder. En annan pionjär var nederländskan Jantine (”Tine”) Tammes (1871–1947). Hon tillämpade statistiska metoder (som hon lärt av astronomen Kapteyn) på genetiska frågor. De viktigaste insatserna under 1900-talet inom området gjordes förmodligen av Ronald Fisher (1890–1962), då han med matematiska metoder knöt ihop Mendels läror med teorin om det naturliga urvalet. Med dagens DNA-teknik ökar naturligtvis mängden genetiska data explosionsartat.

Meteorologi

Numera brukar SMHI och andra vädertjänster ange osäkerheter i prognoserna på något sätt. American Meteorological Society gav exempelvis ut riktlinjer AMS (2008). Exakt hur osäkerheterna bör redovisas är inte självklart, och man kan konstatera att olika vädertjänster (såsom SMHI och norska yr.no valt något olika metodik; personligen föredrar jag den senare).

Operationsanalys

Under min studietid på KTH (1971–76) så upplevde jag operationsanalys som en ”rätt stor grej”.

Sommaren 1974 hade jag också möjligheten att arbeta i en forskningsgrupp på dåvarande Svenska Träforskningsinstitutet, som arbetade med bl.a. simulering av produktionen i sågverk. På mitt första jobb efter examen träffade jag då och då på kollegor som simulerade produktionsflöden i verkstäder. Därefter har jag inte hört talas om några som hävdat att de arbetat med operationsanalys eller optimering. Jag har inte fått riktigt klart för mig om det beror på att approachen inte motsvarat förväntningarna eller på att det blivit del av den dagliga rutinen. Artikel av Ackoff (1979) kan vara av visst intresse i sammanhanget.

Det har påståtts att ämnet har starkare ställning i andra europeiska länder. KTH har dock en hyfsat stark ställning (37) i QS ranking av ”statistics and operational research” (2020). Då operationsanalys bl.a. handlar om att hantera osäkerheter med matematiska metoder, så finns naturligtvis en stark koppling till sannolikhetsläran. Med tanke på den ökade tillgången till data är det möjligt att ämnet får en renässans. Villani (2020) nämner inte operationsanalys explicit men det torde rymmas under punkt 1 i hans lista på områden där statistikämnet borde förändras: ”Prediktion och beslut”.

Toxikologi

Hengst (1967) pekar på en klurig metod att ur djurförsök försöka bedöma risken att ett djur från en given population avlider av en viss dos av ett visst gift. Metoden togs fram av medicina-

ren Behrend Walter Behrens (1895–1969) och beskrivs i Behrens (1929)).

Metodaspekter

Sannolikhetslära och statistik

I första hand måste kopplingen till statistiken framhållas. Statistik innebär bl.a. hantering av stora datamängder behäftade med större eller mindre osäkerhet. För att kunna handskas med osäkerheter fordras kunskaper i sannolikhetslära.

Statistik och ”big data”

Tillgången till data ökar dramatiskt i våra dagar. Detta innebär, som bl.a. Hartford (2014) påpekar, såväl stora möjligheter som risker. Ökad tillgång till data möjliggör kvantitativ bedömning av risker, s.k. probabilistisk riskanalys. Det har också framhållits att ”big data” eventuellt skulle minska behovet av teori. Harford (2013) påpekar dock vikten av att inte glömma de läxor man lärt sig från den klassiska statistiken, vilket även understryks i Villanis (2020) nyligen publicerade artikel i Qvintensen 2020:1.

p-värdet

Det s.k. p-värdet används inom många discipliner där man har behov av att pröva hypoteser med hjälp av insamlade data. Under senare år har det pågått en debatt om detta; bl.a. ägnades en stor del av Qvintensen 2019:2 åt detta ämne. Wasserstein och Lazar (2016) gick ut med ett uttalande å ASA:s (American Statistical Society) vägnar. Artikelns har fått mycket stort genomslag, men en brist som framförts är att artikeln i huvudsak påtalar viss problematik utan att ge så mycket riktlinjer för hur man borde göra istället. Svaret är förmodligen att det inte finns någon enkel metod att omvandla data till vetenskapliga insikter, vilket i sin tur innebär att det krävs närmare samarbete mellan statistiker och företrädare för andra discipliner.

Försök till kvantitativ studie av tillämpningar.

Med syftet att försöka få en kvantitativ känsla för hur ”tunga” olika tillämpningar av sannolikhetslära är, gjorde jag ett antal sökningar med ”Google scholar” den 21 maj 2020.

För varje sökning noterade jag dels antal träffar, dels det mest citerade verket bland de tio första träffarna.

Resultatet redovisas i uppställningen på följande två sidor.

Lite överraskande var att så pass många lärda verk handlar om ”particle size distribution”. Vi ser att rätt många träffar erhålls för ”probability + matematiknära discipliner som fysik, computer science och genetik”. Förhållandevis lite träffar erhålls för ”random vibrations”, då maskin- och byggnadsindustrin inte är så forskningstunga.

Ännu i början av 2000-talet upplevde jag det som svårt att få data om vind-inducerade vibrationer.

Slutsatser

Redan Laplace påpekade betydelsen av sannolikhetslära för andra vetenskaper. Denna betydelse har ökat i och med utveckling av discipliner såsom genetik, teletrafikteori, utmattningsskallfasthet m.m. Bristen på adekvata data från fältet har ofta varit hämmande för utvecklingen. Introduktion av sensornätverk torde kunna råda bot på detta. Förr Eurostatschefen Walter Radermacher använder i sin nyligen publicerade avhandling (Radermacher, 2020) begreppet ”Official statistics 4.0”, uppenbarligen inspirerat av begreppet ”Industrie 4.0”.

Många viktiga insatser inom sannolikhetslära och statistik har gjorts av statistik-kunniga personer som egentligen haft annan huvudinriktning: astronomer (t.ex. Quetelet, Kapteyn), ingenjörer (t.ex. Crandall, Palmkvist, Krige), genetiker (t.ex. Fisher, Tammes, Johanssen) och kemister (t.ex. Wold). Villani (2020) betonar betydelsen av forskning i gränslandet mellan statistik, machine learning och artificial intelligence.

Radermacher påtalar att ”More influence means more responsibilities: It is the duty of statisticians to explore the link between statistics, science, and society, and to lead intellectual reflections on the possible risk of over-reliance on data-centrism”. Nya möjligheter och risker uppstår för statistiker i och med den dramatiska ökning av data som ”Statistik 4.0” innebär. Som Radermacher påpekar, behövs troligen närmare samarbete mellan statistiker och statistikintresserade personer med annan huvudinriktning.

Villani (2020) försöker förutsäga statistik-ämnets framtid. Som han påpekar, så påverkas hans slutsatser av hans närhet till en teknisk högskola. Troligen har han rätt i sin slutkläm ”Min prediktion är att statistikämnet kommer att genomgå stora positiva förändringar under detta nya decennium, delvis som ett resultat av det upplevda hotet från angränsande ämnen som maskininläring”. Förutom maskininläring torde en del andra, såväl nya som gamla, tillämpningar som nämnts i detta inlägg ha påverkan.

Då statistikforskningen skett inom många olika fakulteter finns det risk för att utbytet av kunskap hämnas. Där skulle ”Svenska Statistik-främjandet” och liknande organisationer kunna spela en viktig roll, exempelvis genom att initiera en mer fullödig kartläggning av tillämpningar av sannolikhetslära och statistik.

HANS ALBERG

Sökord	Träffar Google Scholar	Mest citerade verk	Antal citeringar
Probability	5 000 000	Billingsley, P. (2008). <i>Probability and measure</i> . John Wiley & Sons.	13 789
P-value	4 600 000	Westfall, P. H., & Young, S. S. (1993). <i>Resampling-based multiple testing: Examples and methods for p-value adjustment</i> . (Vol. 279). John Wiley & Sons	3 458
Probability + physics	3 710 000	Rosenkrantz, R. D. (Ed.). (2012). <i>ET Jaynes: Papers on probability, statistics and statistical physics</i> (Vol. 158). Springer Science & Business Media.	566
Probability + industry	2 960 000	Mitchell, W. (1989). Whether and when? Probability and timing of incumbents' entry into emerging industrial subfields. <i>Administrative science quarterly</i> , 208-230.	901
Probability + genetics	2 570 000	Levene, H. (1949). On a matching problem arising in genetics. <i>The annals of mathematical statistics</i> , 20(1), 91-94.	1 226
Probability + "computer science"	2 400 000	Trivedi, K. S. (1982). <i>Probability and statistics with reliability, queuing, and computer science applications</i> (Vol. 13). Englewood Cliffs: Prentice-hall.	4 475
Probability + insurance	2 370 000	Johnson, E. J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. <i>Journal of risk and uncertainty</i> , 7(1), 35-51.	1 094
Probability + epidemiology	2 690 000	Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., & Rothman, N. (2004). Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. <i>Journal of the National Cancer Institute</i> , 96(6), 434-442.	1 584
"Particle size distribution"	1 830 000	Arya, L. M., & Paris, J. F. (1981). A physicoempirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data 1. <i>Soil Science Society of America Journal</i> , 45(6), 1023-1030.	1 376
"Randomized clinical trial"	1 750 000	Jensen, P. S. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. <i>Archives of general psychiatry</i> , 56(12), 1073-1086.	3 679
Probability + chemistry	1 830 000	Lindstedt, R. P., Louloudi, S. A., & Váos, E. 2. (2000). Joint scalar probability density function modeling of pollutant formation in piloted turbulent jet diffusion flames with comprehensive chemistry. <i>Proceedings of the Combustion Institute</i> , 28(1), 149-156.	181
"Game theory"	1 260 000	Osborne, M. J., & Rubinstein, A. (1994). <i>A course in game theory</i> . MIT press.	8 979
Probability + astronomy	1 500 000	Groth, E. J. (1975). Probability distributions related to power spectra. <i>The Astrophysical Journal Supplement Series</i> , 29, 285-302.	229
"Mathematical statistics"	1 150 000	Serfling, R. J. (2009). <i>Approximation theorems of mathematical statistics</i> (Vol. 162). John Wiley & Sons.	7 518
Probability + "machine learning"	1 020 000	Cestnik, B. (1990, August). Estimating probabilities: a crucial task in machine learning. In <i>ECAI</i> (Vol. 90, pp. 147-149).	666
Econometrics	903 000	Granger, C. W., Newbold, P., & Econom, J. (1974). Spurious regressions in econometrics. <i>Baltagi, Badi H. A Companion of Theoretical Econometrics</i> , 557-61.	8 774
Probability + toxicology	1 090 000	Elandt-Johnson, R. C. (1971). <i>Probability models and statistical methods in genetics</i> . New York, London, Sydney, Toronto: John Wiley & Sons, Inc..	376
Probability + "artificial intelligence"	908 000	Hutter, M. (2004). <i>Universal artificial intelligence: Sequential decisions based on algorithmic probability</i> . Springer Science & Business Media.	726

Sökord	Träffar Google Scholar	Mest citerade verk	Antal citeringar
Probability + linguistics	833 000	Hatch, E., & Farhady, H. (1982). Research design and statistics for applied linguistics.	1 386
Probability + "image processing"	588 000	Duin, R. P. W. (1976). On the choice of smoothing parameters for Parzen estimators of probability density functions. <i>IEEE Transactions on Computers</i> , (11), 1175-1179.	432
"Brownian motion"	650 000	Karatzas, I., & Shreve, S. E. (1998). Brownian motion. In <i>Brownian Motion I and Stochastic Calculus</i> (pp. 47-127). Springer, New York, NY.	3 798
Probability + forensic	570 000	Evvett, I. W., & Weir, B. S. (1998). <i>Interpreting DNA evidence: statistical genetics for forensic scientists</i> (Vol. 244). Sunderland, MA: Sinauer Associates.	851
Biostatistics	562 000	Rosner, B. (2015). Fundamentals of biostatistics. Nelson Education.	8 968
Probability + meteorology	461 000	Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. <i>Journal of Applied Meteorology</i> , 8(6), 985-987.	581
Probability + "image analysis"	422 000	Srivastava, A., Joshi, S. H., Mio, W., & Liu, X. (2005). Statistical shape analysis: Clustering, learning, and testing. <i>IEEE Transactions on pattern analysis and machine intelligence</i> , 27(4), 590-602.	303
"Design of experiments"	322 000	Fisher, R. A. (1936). Design of experiments. <i>Br Med J</i> , 1(3923), 554-554.	7 445
Kriging	219 000	Stein, M. L. (2012). <i>Interpolation of spatial data: some theory for kriging</i> . Springer Science & Business Media.	3 763
Chemometrics	192 000	Miller, J., & Miller, J. C. (2018). <i>Statistics and chemometrics for analytical chemistry</i> . Pearson education.	10 983
Probability + "search engine"	177 000	Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. <i>ELECTROPHORESIS: An International Journal</i> , 20(18), 3551-3567.	8 567
"Software reliability"	128 000	Musa, J. D., Iannino, A., & Okumoto, K. (1990). Software reliability. <i>Advances in computers</i> , 30, 85-170.	3 053
Probability + "big data"	114 000	Mitzenmacher, M., & Upfal, E. (2017). <i>Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis</i> . Cambridge university press.	3 011
Probability + cryptology	90 100	Van Tilborg, H. C. (2012). <i>An introduction to cryptology</i> (Vol. 52). Business Media.	145
"Queueing theory"	74 400	Gross, D. (2008). <i>Fundamentals of queueing theory</i> . John Wiley & Sons.	6 881
"Statistical sampling"	68 200	Wunderlich, R. E., Wensch, T. F., Falsafi, B., & Hoe, J. C. (2003, May). SMARTS: Accelerating microarchitecture simulation via rigorous statistical sampling. In <i>Proceedings of the 30th annual international symposium on Computer architecture</i> (pp. 84-97).	663
"Probabilistic risk assessment"	34 100	Henley, E. J., & Kumamoto, H. (1996). Probabilistic risk assessment and management for engineers and scientists. <i>IEEE Press</i> (2nd Edition).	907
"Random vibrations"	19 400	Newland, D. E. (2012). <i>An introduction to random vibrations, spectral & wavelet analysis</i> . Courier Corporation.	1 903
"Allan variance"	13 900	El-Sheimy, N., Hou, H., & Niu, X. (2007). Analysis and modeling of inertial sensors using Allan variance. <i>IEEE Transactions on instrumentation and measurement</i> , 57(1), 140-149.	675

»»»»

Referenser

- Ackoff, R. L. (1979). The future of operational research is past. *Journal of the operational research society*, 30(2), 93-104
- Aitken, C., Roberts, P., & Jackson, G. (2010). Fundamentals of probability and statistical evidence in criminal proceedings: guidance for judges, lawyers, forensic scientists and expert witnesses. RSS.
- AMS, 2008: Enhancing weather information with probability forecasts. Bull Amer. Meteor. Soc., 89, 1049-1053
- Behrens, B. (1929). Evaluation of digitalis leaves in frog experiments. *Archiv for Experimentelle Pathologie und Pharmakologie*, 140, 237-256.
- Bonald, T., & Roberts, J. W. (2012). Internet and the Erlang formula. *ACM SIGCOMM Computer Communication Review*, 42(1), 23-30.
- Cramér, H. (1976). A century with probability theory: Some personal recollections. *The annals of probability*, 4(4), 509-546.
- Crandall, S. H., Mark, W. D., & Khabbaz, G. R. (1962). *THE VARIANCE ON PALMGREN-MINER DAMAGE DUE TO RANDOM VIBRATION*. MASSACHUSETTS INST OF TECH CAMBRIDGE.
- Gigerenzer, G., Swijtink, Z., Porter, T., & Daston, L. (1990). *The empire of chance: How probability changed science and everyday life* (Vol. 12). Cambridge University Press.
- Guttorp, P., & Lindgren, G. (2019). Why Distinguish Between Statistics and Mathematical Statistics—The Case of Swedish Academia. *International Statistical Review*, 87(1), 110-126.
- Harford, T. (2014). Big data: A big mistake?. *Significance*, 11(5), 14-19.
- Hengst, M. (1967). *Einführung in die mathematische Statistik und ihre Anwendung*. Bibliograph. Inst.
- Palmgren, A. (1930). *Untersuchungen über die statische Tragfähigkeit von Kugellagern*. Kungliga Tekniska Högskolan
- Radermacher, W. J. (2020). Science and Society: A Reflexive Approach to Official Statistics. In *Official Statistics 4.0* (pp. 53-117). Springer, Cham.
- Stamhuis, I. H. (2012)., Statistiek en waarschijnlijkheidsrekening in het werk van Tine Tammes. *GEWINA/TGGNWT*, 15(3), 195-208.
- Svensson, J. (1955). *A new formula for particle size distribution of products produced by comminution*. Elanders boktr. [H. Lindstahls bokhandel i distribution, Stockholm].
- Villani, M., (2020) Statistikämnet är hotat – tack och lov. *Qvintensen* (2020:1)
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129-133.
- ### Litteratur som ej nämns i texten
- Behrens, B. (1929). Zur Auswertung der Digitalisblätter im Froschversuch. *Naunyn-Schmiedebergs Archiv für Experimentelle Pathologie und Pharmakologie*, 140(3-4), 237-256.
- Bojkovic, Z., Bakmaz, M., & Bakmaz, B. (2009). Electrical Engineering Hall of Fame: Originator of Teletraffic Theory [Scanning Our Past]. *Proceedings of the IEEE*, 98(1), 123-127.
- Boroviczeny, K. G. V. Prof. Martin Hengst 80 Jahre alt.
- Crandall, S. H., & Mark, W. D. (2014). *Random vibration in mechanical systems*. Academic Press.
- Crandall, S. H., & Zhu, W. Q. (1983). *Random vibration: a survey of recent developments*.
- Elishakoff, I. (2017). *Probabilistic Methods in the Theory of Structures: Strength of Materials, Random Vibrations, and Random Buckling*.
- Engset T., O. (1918) *Die Wahrscheinlichkeitsrechnung zur Bestimmung der Wählerzahl in automatischen Fernsprechämtern*, In *Elektrotechnische Zeitschrift*, Heft 31., Original manuscript (1915) also translated in Arne Myskja, *On the Calculation of Switches in an Automatic Telephone System*. In *Teletronikk*, 94(2):99-142, 1998.
- Fisher, R. A. (1936). Design of experiments. *Br Med J*, 1(3923), 554-554.
- Guttorp, P. (2019). How We Know that the Earth is Warming. *CHANCE*, 32(1), 36-41.
- Johanssen, W. L. (1926). *Elemente der exakten Erblchkeitslehre: mit Grundzügen der biologischen Variationsstatistik*. G. Fischer.
- Miller, J., & Miller, J. C. (2018). *Statistics and chemometrics for analytical chemistry*. Pearson education.
- Paez, T. L. (2006). The history of random vibrations through 1958. *Mechanical Systems and Signal Processing*, 20(8), 1783-1818..
- Palm, C. (1943). Intensitätsschwankungen im Fernsprechverkehr. *Ericsson Technics*.
- Palmgren, A. (1924). Die Lebensdauer von Kugellagern. *Zeitschrift des Vereines Deutscher Ingenieure*, 68(4), 339.
- Palmgren, A.G. (1924) Die Lebensdauer von Kugellagern (Life Length of Roller Bearings or Durability of Ball Bearings). *Zeitschrift des Vereines Deutscher Ingenieure* (ZVDI), 14, 339-341.
- Stewart, R., Stobart, S., Carter, S., & Danishyar, S. (2019). Probabilistic risk assessment transition to industry. In *AIAC18: 18th Australian International Aerospace Congress (2019): HUMS-11th Defence Science and Technology (DST) International Conference on Health and Usage Monitoring (HUMS 2019): ISSFD-27th International Symposium on Space Flight Dynamics (ISSFD)* (p. 332). Engineers Australia, Royal Aeronautical Society..
- Svensson, J., & Murkes, J. (1957). An empirical relationship between work input and particle size distribution before and after grinding. In *International Mineral Dressing Congress*.
- Tammes, T. (1907). "Der Flachsstengel. Eine Statistisch-Anatomische Monographie." *Natuurkundige Verhandelingen Hollandsche Maatschappij van Wetenschappen* 3, VI. 4. Haarlem, 285 pp.
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2), 109-130.
- Wong, E. W., Zalesky, A., & Zukerman, M. (2007). On generalizations of the Engset model. *IEEE communications letters*, 11(4), 360-362.
- Zaretsky, E. V. (1997). A. Palmgren revisited: A basis for bearing life prediction.
- Zhou, T. (2010). Numerical comparisons of bioassay methods in estimating LC50.



Professor Bo Sundgren.

Till minne av Bo Sundgren

Professor Bo Sundgren, Solna, har avlidit i en ålder av 73 år. Närmaste anhöriga är hustrun Irina Sundgren och hennes dotter Irina Franzén med familj.

Bo anställdes 1967 vid Statistiska centralbyrån (SCB). Han satte myror i huvudet på de äldre statistikerna med nya idéer. Tankarna på betydelsen av grundläggande begrepp i den statistiska informationsbehandlingen rimmade väl med dåvarande generaldirektören Ingvar Ohlssons visioner om ett statistiskt system baserat på fasta urvalsramar, "databaser". Bo kom att bli drivande i utvecklingen mot dagens SCB.

Bo var genom sin utåtriktade och positiva personlighet samt det skarpa intellektet, en aktiv och uppskattad deltagare i det omfattande internationella samarbetet kring utveckling av metodik för central statistikproduktion.

Hans doktorsavhandling "An infological approach to data bases" rönt stor internationell uppmärksamhet och kom att bidra till SCB:s anseende som en ledande statistikbyrå. Den lade grunden för vidare forskning, som Bo kom att driva som professor inom flera akademiska

institutioner, såsom universiteten i Uppsala och Linköping, Handelshögskolan och Högskolan Dalarna.

En för SCB:s framtid mycket viktig insats var utvecklingen av statistiska databaser. Bo bidrog starkt till att SCB blev en av de första statistikbyråerna som gjorde statistik tillgänglig för användarna genom internet. Det inleddes i liten skala men kom under 80-talet att utvecklas till att täcka så gott som all statistik som SCB producerade.

Bo hade alltid mod att säga sin mening. Han arbetade hårt och metodiskt även under tider av sjukdom och personliga tragedier.

Bos resor inom östblocket mitt under det kalla kriget hjälpte hans anställda till ökad förståelse och eftertanke kring hur den kommande internationella utvecklingen skulle te sig. Bo sökte medarbetare med begåvning och talanger och fann arbetsuppgifter även för dem som hade allvarliga svårigheter av olika slag. Bo var en pålitlig vän.

CHRISTER ARVAS,
HANS BLOCK,
LARS OLSSON
(från DN 2020-06-16)



Styrelseledamöter efter årsmötet 2020

Främjandet Joakim Malmudin, ordf.
Maria Josefsson, vice ordf.
Mattias Strandberg, sekr.
Anders Haraldsson, skattm.
Jens Malmros, webbansv.
Marcus Berg, repr. Surveyföreningen.
Ida Hed Myrberg, repr FMS
Hans Alberg, repr. Ind. stat.
Jesper Rydén, repr. Cramérsamf.

Ind. stat. Hans Alberg, ordf.
Ingemar Sjöström, sekr.
Mats Franzén, kassör
Göran Lande
Leif Nilsson
Peter Örtlund

Surveyför. Marcus Berg, ordf.
Rikard Gard, vice ordf.
Natalia Ogorodnikova, kassör
Yingfu Xie, webbansv.
Paul Fuehrer
Johan Martinsson
Mats Nyfjäll
Linda Theining
Åke Wissing

FMS Ingeborg Waernbaum, ordf.
Per Liv, sekr.
Fredrik Norström, kassör
Andreas Gustavsson, webbansv.
Marie Linder, Qvintensen
Therese Andersson
Ida Hed Myrberg
Henrik Renlund
José Sanchez, repr. EFSPI
Anna Torrång, repr. EFSPI

Cramérsamf. Tom Britton, ordf.
Jesper Rydén, sekr.
Anders Lundquist, kassör
Annika Lang
Yuli Liang
Fredrik Olsson



ORDFÖRANDEN HAR ORDET

Nya medlemmar i Surveyföreningens styrelse

Paul Fuehrer

– Jag arbetar som lektor i sociologi och är avdelningsföreståndare för sociologiämnet på Södertörns högskola. Under doktorandtiden på Stockholms universitet hade jag kursansvar för en del kurser i utrednings-sociologi. Jag undervisar idag i metod på grundläggande och avancerad nivå på Södertörns högskola. Mina specialintressen är utbildning i enkätkonstruktion och kvantitativ dataanalys, samt teoretiska och metodologiska aspekter på webbundersökningar.

Natalia Ogorodnikova

– Jag är företagsekonom som för närvarande utökar min kompetens genom att läsa mastersprogrammet i statistik på Örebro Universitet. Vid sidan av studierna arbetar jag med företagsanalys på avdelningen för ekonomisk statistik på SCB. Jag är intresserad av surveymetodik, och Surveyföreningen ger mig möjlighet att träffa människor som har liknande intressen. Jag ser fram mot att delta och utveckla föreningen!

Linda Thelning

– Jag jobbar som senior analytiker på Demoskop sedan 10 år tillbaka. Arbetar till största delen med varumärkesanalyser, men även inom Demoskops andra områden. Har tidigare jobbat på Rikspolisstyrelsen samt Centrum för folkhälsa, och Centrum för arbets- och miljömedicin inom Stockholms läns landsting. Har examen i psykologi och statistik från Linköpings universitet. Ser fram mot ett spännande år!

Svåra val kan underlätta längre fram

Det är lätt att vara efterklok, men att vara klok i stunden är alltid svårt. Jag minns så väl styrelsemötet där vi beslutade att anordna föreningens första digitala årsmöte. Det var ett svårt beslut att fatta, men också rätt beslut, med facit i hand. Jag har ofta kommit på mig själv med att tänka tillbaka till den dagen, den stunden, under de senaste månaderna när jag suttit i frivillig isolering tillsammans med min familj.

En visdom jag fick med mig hem till Sverige från min tid när jag bodde i Japan är orden "Easy choices, hard life. Hard choices, easy life". Gör de lätta valen och få ett svårt liv, gör de svåra valen och få ett enkelt liv – de svåra valen idag banar väg för en morgondag med ett lättare sinne. Och det var det svåra valet som jag gjorde tillsammans med resten av styrelsen där i tidiga mars, när det enkla valet hade varit att bara köra på ett fysiskt möte och hoppas att ingen sjuk-utan-symptom hade närvarat.

Men vad var det som gjorde det beslutet så svårt? Avsaknad av data och osäkerhet, skulle jag säga. Två saker som vi statistiker alla är skolade i att hantera och väga in i våra analyser, förvisso, men det gör det inte lättare i stunden när man sitter med för lite data eller för mycket osäkerhet och ändå måste fatta ett beslut. Jag är glad att vi fattade rätt beslut då, och jag hoppas att det kan bana väg för en mer digital Surveyförening under åren som kommer.

Vi har tre nya styrelsemedlemmar i år: Paul Fuehrer, Natalia Ogorodnikova och Linda Thelning. Här är hela styrelsen som valdes på årsmötet 2020:

- Marcus Berg, Swedbank (ordförande, omval)
- Paul Fuehrer, Södertörns Högskola (nyval)
- Rikard Gard, SCB (omval)
- Johan Martinsson, SOM-institutet Göteborgs Universitet (omval)
- Mats Nyfjäll, Statisticon (omval)
- Natalia Ogorodnikova, EMOS-student, Örebro Universitet (nyval)
- Linda Thelning, Demoskop (nyval)
- Åke Wissing, Åke Wissing & Co AB (omval)
- Yingfu Xie, SCB (omval)

Ann-Marie Flygare, Gösta Forsman och Karin Schneller avgick, och vi tackar dom varmt för deras tid och engagemang!



MARCUS BERG

Surveyföreningens kvalitetsseminarium 2020

Surveyföreningens traditionella kvalitetsseminarium ägde i år rum den 20 februari, som vanligt på Westmanska palatset i Stockolm. (Det här var på den gamla goda tiden före corona-restriktionerna, då man fortfarande kunde träffas i större grupper.) Årets tema var *Hur gör vi bra surveyundersökningar 2020?*

Konferencier var **ANN-MARIE FLYGARE**. Fem föredragshållare medverkade, och det hela avslutades med en paneldiskussion.

DAN HEDLIN (Stockholms universitet) inledde med att diskutera frågan "Är kvoturval en vetenskaplig metod?" Hans avsikt var inte att ge ett enkelt svar (ja eller nej) på frågan, utan snarare att diskutera olika tolkningar av frågan. Vad menas idag med termen "kvoturval"? När termen infördes, för kanske 70-80 år sedan, avsågs en typ av icke-sannolikhetsurval, där varje intervjuare på egen hand skulle välja ut och intervjua ett visst antal personer, med restriktionen att bestämda procenttal av intervjupersonerna skulle skulle tillhöra vissa i förväg bestämda kategorier. Är det så att termen "kvoturval" idag används för fler typer av förfaranden än tidigare? T.ex. i samband med webbpaneler, där en svarande ibland kan få besked av typen "Tack för visat intresse, men vi har redan fått tillräckligt många svar från din åldersklass"?

En annan fråga är "Vad menas med vetenskaplig metod?" Kanske en metod som används inom allmänt respekterad vetenskaplig forskning. Och vad menas då till sist med "vetenskaplig forskning"? Kanske ett samlingsnamn för olika sätt att systematiskt studera och skaffa sig kunskap om världen? Dan hänvisade till ett välkänt kunskapsbegrepp, som brukar förknippas med Platons dialog *Theaitetos*, nämligen att kunskap är välgrundad sann övertygelse, "justified true belief".

Kvoturval ska inte förväxlas med andra urvals- och estimationsmetoder, där man också på något sätt utnyttjar förhandskunskap om populationens sammansättning eller annan förhandskunskap om populationen. T.ex. stratifierat urval, poststratifiering, regressionsestimation, kalibrering. I dessa fall utgår man från att man har ett sannolikhetsurval. Kvoturval är däremot inte något sannolikhetsurval. Men i och med att bortfallen i många fall nu är så stora att det som från början var tänkt att vara sannolikhetsurval ofta inte kan betraktas som ett sådant, så kanske någon moderniserad form av kvoturval skulle kunna vara ett tänkbart alternativ, undrade Dan. Jag avstår från att försöka referera de mer tekniska delarna av Dans resonemang, men jag ser fram mot en skriftlig framställning där man i lugn och ro kan fundera över hans tankar.

De övriga föredragen handlade om

speciella undersökningar, där föredragshållarna varit verksamma. **SANNA WALLIN** (BRÅ) berättade om hur man utfört förändringsarbete avseende Nationella Trygghetsundersökningen (NTU). Förändringarna hade att göra med datainsamlingsmetoder och frågekonstruktion. **TORBJÖRN SJÖSTRÖM** (Novus) resonerade mer allmänt om undersökningar (speciellt politiska opinionsundersökningar) och om deras tillförlitlighet. **FOLKE CARLSSON** (SCB) berättade om det arbete som utförs internt på SCB med syfte att anpassa verksamheten efter principiella önskemål och praktiska möjligheter. **PETER SANTESSON** (Demoskop) berättade om de studier som under senare år bedrivits vid Demoskop avseende de politiska väljarbarometrarna. Mest anmärkningsvärt är väl att arbetet lett till att man nu slutar att använda telefonintervjuer och går över till att enbart förlita sig på uppgifter från en "onlinerekryterad" webbpanel. Mer om detta kan man läsa i en artikel på DN Debatt, som publicerades den 15 november 2019, och som återgavs i Qvintensen 2020:1 (sid. 28-29).

JAN WRETMAN



ORDFÖRANDEN HAR ORDET

Digitalt möte valde ny styrelse

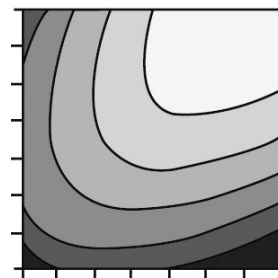
I nom FMS samlas personer med anknytning till medicinsk statistik inom universitet, statliga myndigheter och det privata näringslivet. I EU-sammanhang representerar vi svensk medicinsk statistik i European Federation of Statisticians in the Pharmaceutical Industry (EFSPI).

Våren 2020 har präglats mycket av omständigheter som beror på covid-19. Därför omvandlades, i all hast, årsmötet den 19:e mars till ett elektroniskt årsmöte. Det vetenskapliga programmet ställdes in och själva årsmötet genomfördes via zoom. Den föregående styrelsen beviljades ansvarsfrihet och en ny styrelse valdes. Ingeborg Waernbaum, Henrik Renlund, Fredrik Norström och Andreas Gustavsson (tidigare EFSPI) valdes om. Även Anna Torrång blev omvald som representant för EFSPI. Vi avtackade styrelsemedlemmarna Catrin Wessman, Jonas Häggström, Sara Ekberg och Anna Nilsson som alla gjort viktiga insatser i FMS. Nya representanter i styrelsen är Per Liv (ny sekreterare), Ida Hed Myrberg (representant i Statistikfrämjandet), Therese Andersson och José Sanchez som är ny EFSPI representant.

Den 29:e maj genomfördes ett elektroniskt vårmöte via zoom med ett bredare program. Paul Dickman presenterade planer för doktorandutbildning och nyheter gällande masterutbildningen i Biostatistik på MEB, KI. Anna Torrång presenterade covid-19-relaterade nyheter från EFSPI. Vi fick lyssna på en presentation och workshop om datavisualisering av Mathias Lundin från Statistik, Handelshögskolan vid Umeå universitet. Två medarbetare från KI, Alessandro Gasparini och Daniel Olsson presenterade delar av sin forskning. Mötet avslutades med en workshop (den andra i vår serie) om nya visioner för FMS ledd av visionsgruppen med Fredrik Norström, Anna Torrång och Ingeborg Waernbaum. Av naturliga skäl, p g a läget med pandemin, präglades mötet av ett starkt engagemang för biostatistikens roll i samhället.

Det ordinarie höstmötet som skulle hållits i Malmö tillsammans med vår danska systerorganisation DSBS har flyttats fram till 2021 och vi har istället beslutat att ha ett höstmöte online den 13:e november med tema covid-19. Håll utkik efter mer nyheter och program för höstmötet på vår hemsida (www.fmsbiostat.se).

INGEBORG WAERNBAUM



ORDFÖRANDEN HAR ORDET

Indstat – en plattform för nya idéer

Föreningen för Industriell statistik (Indstat) har nära samarbete med såväl Svenska Förbundet för Kvalitet (SFK) som med Svenska statistikfrämjandet. SFK delar, sedan flera årtionden ut ett pris för bästa examensarbete inom relevanta ämnesområden till minne av Olle Jonson. Det finns kanske inte så många kvar som minns honom men vi vet att han alltid underströk användning av fakta och statistisk analys som beslutsunderlag. Indstat har en viktig roll att se till att dessa tankegångar lever vidare.

Detta år, som råkat sammanfalla med 200-års jubileet av Florence Nightingales födelse, har i stor utsträckning präglats av covid-19, vilket bl.a. medfört en plötsligen ökad efterfrågan på statistik i realtid. Oaktat detta har det också kommit andra tecken på statistikens ökade betydelse. Som exempel kan nämnas att Hanne Kjölter, känd samhällsdebattör och tidigare ledarskribent i DN (och även utbildad sjuksköterska), nyligen har kommit ut med boken *Kris i forskningsfrågan – eller vad FAN får vi för pengarna*. I ett av slutsatskapitlen skriver hon om forskare: "Däremot verkar många alltför dåligt utbildade i statistik för att ens själva förstå att deras resultat inte håller".

I början av förra decenniet så introducerades konceptet "Industrie 4.0", vilket bl.a. innebär ökade datamängder. Om dessa hanteras på ett förnuftigt sätt, enligt Olle Jonsons tankegångar, så finns det alla möjligheter att vinna konkurrensfördelar. Indstat kan bli en plattform för utbyte av idéer om hur man kan uppnå detta.

HANS ALBERG



Skriv gärna för Qvintensen

Det är olika slags bidrag från Statistikfrämjandets medlemmar som gör att vi får en levande medlemstidning med intressant innehåll för just statistiker i Sverige.

Qvintensen är en medlemstidning, inte någon "vetenskaplig" tidskrift. Ingen referee-granskning av manuskript förekommer. Den redaktionella granskningen och bearbetningen är huvudsakligen av språkligt och presentationsmässigt slag.

Varje skribent ansvarar själv för vad som sägs i bidraget. Skribenter garanteras rätten att själva sprida sitt bidrag vidare, under förutsättning att Qvintensen anges som originalkälla med angivande av nummer och år.

Vi har inga stränga regler för hur manuskript ska vara utformade. Men för att vi lätt ska kunna hantera manuskript bör de följa en enkel standardformatmall, utan konstigheter. (Alla manuskript kommer ändå i sinom tid att omformateras i samband med layout-arbetet.)

Tänk på följande:

- Inte avstavning.
- Inte rak högermarginal.
- Inte flera spalter.
- Inte fotnoter.

Eventuella tabeller och diagram

som inte ryms inom Qvintensens spaltbredd kommer att lyftas ut och placeras utanför själva texten. Tabeller och diagram ska vara numrerade och försedda med rubrik.

Foton, tabeller och diagram ska bifogas som separata filer. Bildformat bör vara .jpg, .eps eller .tif. Foton ska innehålla så många kb som möjligt.

Färgtryck är möjligt bara på omslagets främre och bakre insida.

Är det något du undrar över, ta kontakt med redaktören eller redaktionskommittén.

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.

För att bli medlem, gå till <http://www.statistikframjandet.se> och läs mer i högerspalten under "Vill du bli medlem?". Har du frågor kontakta Mattias Strandberg 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.

