

Qvintensen

NR 2 2021

Vad är
dataanalys?

SIDORNA 22-23

Våga medverka
i medierna!

SIDORNA 8-11

Kortfattat
om surveys

SIDORNA 14-17

Rankning
av universitet

SIDORNA 18-20

Willy Feller
i Stockholm

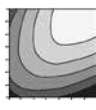
SIDORNA 4-7

SIDORNA 4-5

Innehåll

- 3 REDAKTÖRENS RUTA
Jan Wretman
- 4 WILLY FELLER – EN GIGANT INOM SANNOLIKHETSTEORIN
Jesper Rydén och Rolf Sundberg
- 8 VÅGA MEDVERKA I MEDIERNA!
Dan Hedlin
- 12 FRÅN UTBILDNING TILL ARBETSLIV
Jakob Börsum
- 14 KORTFATTAT OM SURVEYS
Jan Wretman
- 18 RANKNING AV UNIVERSITET
Hans Alberg
- 20 SIMULERINGSSTUDIE OM RANKNINGAR
Jan-Eric Englund
- 22 VAD ÄR DATAANALYS?
Fredrik Jonsson
- 24 FENSTATS AKTIVITETER INOM COVID-19
Hans Alberg

Våra föreningar



- 25 SURVEYFÖRENINGEN
Ordföranden har ordet
- 25 FÖRENINGEN FÖR MEDICINSK STATISTIK
Ordföranden har ordet
- 26 CRAMERSÄLLSKAPET
Ordföranden har ordet
Cramérpriset
- 27 INDUSTRIELL STATISTIK
Ordföranden har ordet

SVENSKA STATISTIKFRÄMJANDETS STYRELSE

Ordförande

Joakim Malmdin 010-479 49 45,
ordforande@statistikframjandet.se

Vice ordförande Maria Josefsson

Kassör Annika Tillander
kassor@statistikframjandet.se

Sekreterare Mattias Strandberg
sekreterare@statistikframjandet.se

Webbansvarig Jens Malmros
webmaster@statistikframjandet.se

Ledamot Magnus Pettersson

Representant Surveyföreningen

Åke Wissing

Representant FMS

Per Liv

Representant Industriell statistik
Hans Alberg

Representant Cramersällskapet
Anders Lundquist

Adress

Svenska statistikfrämjandet
Sekreterare: Mattias Strandberg

E-post sekreterare@statistikframjandet.se

Webbplats www.statistikframjandet.se

Ansvarig utgivare

Joakim Malmdin

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

Minnesvärda möten

Det var tänkt att det här numret av Qvintensen bland annat skulle innehålla diskussion om statistikundervisning. Nu har det inte blivit något av det. Trots önskan om bidrag har vi bara fått in ett enda, som presenteras i detta nummer. Det är från en yngre person, som nyligen avslutat sin statistikutbildning. Vi hoppas att det ska komma in fler bidrag om detta viktiga ämne till nästa nummer. Kanske kan vi hoppas på bidrag från Cramérsällskapets höstmöte om framtidens undervisningsmetoder, som (när detta skrivs) planeras äga rum i slutet av oktober 2021. Nog om detta.

En helt annan typ av artikel i detta nummer handlar om den unge (sedermera världsberömd) sannolikhetsteoretikern Willy Fellers vistelse i Stockholm under några år på 1930-talet. Jag vill inte föregripa artikelns innehåll genom att avslöja alltför mycket, men kan inte låta bli att nämna en sak som gripit mig speciellt. När Feller kom till Stockholm ville Harald Cramér ordna en fast tjänst åt honom på sin institution men lyckades inte med detta på grund av motstånd från antisemitiska krafter inom den dåvarande universitetsvärlden. (Feller hade judiska släktförbindelser bakåt i tiden.) Så Feller fann det så småningom säkrast att flytta vidare till USA, där han fick en universitetstjänst och med tiden kom att bli den världsberömda William Feller. (Om antisemitiska strömningar inom den svenska universitetsvärlden under den här tiden kan man få en uppfattning bland annat genom att studera skildringar av det beryktade "Bollhusmötet" i Uppsala 1939. Det kan man bland annat läsa om på nätet, där man också hänvisas till Forum för levande historia.)

Den som läser artikeln om Feller kanske först blir förvånad över att hans förnamn anges som Willy. Det kanske i förstone kan verka lite respektlöst. Numera är han ju känd som William. Men det finns en förklaring. Fellers förnamn genomgick flera förändringar, i korthet följande. Efter födelsen i Zagreb döptes han till Vilibald. Som studerande

i Zagreb ändrade han namnet till Vilim. Som ung forskare i Tyskland och Skandinavien ändrade han sitt namn på nytt, denna gång till Willy. I Stockholm var han alltså på papperet fortfarande Willy Feller. Det var först i USA som han blev William. Detta har jag läst på nätet och hoppas att det är åtminstone approximativt sant.

Nu kommer jag att tänka på Harald Cramér. Jag har bara hört honom föreläsa en enda gång. Det var på en så kallad sommarskola anordnad av Statistikersamfundet (den ena av Statistikfrämjandets två föregångare) i Kall i Jämtland. Kanske var det 1967, i alla fall mer än femtio år sedan. Jag och några andra yngre statistiklärare från Umeå befann oss bland deltagarna. Cramérs uppgift var att ge en inledande översikt över grundbegrepp inom sannolikhetslära och inferens. Vi unga från Umeå var bara nybörjare och hängde väl inte med på allting. Men jag fäste mig vid hans förmåga att föreläsa begripligt. Även när jag inte kunde hänga med i detaljerna, så fick jag ändå en klar bild av hur alltsammans hängde ihop i stort. En annan detalj som jag kommer ihåg speciellt är att Cramér vid ett tillfälle sa att han helst ville undvika att använda begreppet betingad sannolikhet. Jag kände en viss tillfredsställelse när jag hörde det, eftersom jag själv alltid hade funnit det svårt att komma på någon naturlig frekventistisk tolkning av detta begrepp. Då tyckte jag (kanske lite barnsligt) att jag hade stöd av själve Harald Cramér.

Jag minns att jag tillsammans med ett par andra deltagare en dag efter lunchen gjorde en kortare promenad i skogen bakom kursgården. Där mötte vi Harald Cramér på stigen. Vi stannade och pratade en stund utan högtidligheter. Inte om statistik, utan om sådant som man spontant pratar om när man möts på en skogsstig. Jag minns fortfarande hans vänlighet och naturliga anspråkslöshet.

JAN WRETMAN



»Den som läser artikeln om Feller kanske först blir förvånad över att hans förnamn anges som Willy. ... Men det finns en förklaring.»



Harald Cramér (1893–1985).

Harald Cramér utnämndes 1929 till professor i försäkringsmatematik och matematisk statistik vid Stockholms högskola. Efter några år hade han ett eget institut, Institutet för försäkringsmatematik och matematisk statistik, med egna lokaler vid Odengatan, och 1–2 anställda amanuenser. Där bedrevs en livlig verksamhet.

Cramér hade ett brett kontaktnät bland stora namn inom sannolikheteori. I hans utgivna memoarer, "Korta minnen från ett långt liv", kan man läsa om nära kontakter med t.ex. Lévy och Fréchet i Paris. I Institutets seminarier märks internationella namn som t.ex. Theodore (Ted) W. Anderson, Jerzy Neyman – och från 1934 en begåvad ung matematiker vid namn Willy Feller, som gästade Institutet under nära fem års tid.

Det är Feller som är ämnet för vår uppsats, en sedermera berömd probabilist, inte minst känd för sina två böcker i ämnet.

EN GIGANT INOM SANNOLIKHETSTEORIN PÅ SVENSK MARK:

Willy Feller vid Stockholms

Willy Feller – före Stockholms-tiden

Willy Feller föddes i juli 1906 i Zagreb, i dåvarande dubbelmonarkin Österrike-Ungern, nu i Kroatien. Han var bland de yngsta av tolv syskon, men familjen hade det gott ställt genom försäljning av en universalmedicin, som Willys far framställt – han var kemist. Farfadern var en jude som konverterat till katolicism när han gifte sig, och detta judiska arv hade betydelse för Willys framtid, som vi skall se. Efter snabba studier vid födelseortens universitet fortsatte han till Göttingen, där han snart (1926) hade sin doktorsavhandling i matematik klar, som 20-åring! I Göttingen var Courant och Hilbert stora inspirationskällor. Efter ett par år som assistent till Courant, flyttade Feller till universitetet i Kiel för att leda dess institut för tillämpad matematik. Januari 1933 kommer nazisterna till makten i Tyskland och börjar genast utsätta judar för diskriminering och förföljelser av olika slag.

Feller valde att lämna Tyskland detta

år, och tillbringade ett år i Köpenhamn, men fann inte möjlighet att bli kvar där. Valet av Köpenhamn hade nog flera skäl, det var inte bara relativt nära från Kiel, utan det spelade säkert roll att en av hans studenter i Kiel, Clara, var danska (och dessutom anti-nazist), ty hon och Willy blev så småningom ett äkta par (1938). I Köpenhamn lärde Feller känna bröderna Harald och Niels Bohr. Matematikern Harald Bohr var nära vän med Harald Cramér, och det blev betydelsefullt. Så här skriver Cramér själv i sina memoarer:

Längre fram på hösten 1934 inträffade en händelse, som fick stor betydelse både för mig själv och min grupp av studenter vid Högskolan, och inte bara för oss. Det var ju under de svåra åren efter Hitlers övertagande av makten i Tyskland, och vid de tyska universiteten fanns det många, som hade att välja mellan att lämna landet eller gå mot ett ännu ovisst öde, som med tiden skulle visa sig bli ännu värre än man då kunde förutse. Min vän Harald Bohr i Köpenhamn skrev och rekommenderade

en ung matematiker, Willy Feller, och jag kunde ordna ett blygsamt arvode åt honom för att hålla föreläsningar vid Högskolan.

Tiden i Stockholm

Cramér talar om ett blygsamt arvode, och Feller kom aldrig att få en fast tjänst. I de protokoll som fördes vid seminarier vid denna tid (mer om detta senare) omnämns han som "Institutets gäst, Privatdozent W. Feller". Utöver de rent matematiska, inomvetenskapliga, bidragen visade Feller intresse för tillämpningar, vilket också märks i en del av publikationerna från 1930-talet, t.ex. *Die Aufnahme von Glucose durch Bäckerhefe unter aeroben und anaeroben Bedingungen*, publicerad 1938 i *Naturwissenschaften*. Under Cramérs ledning arbetades i Stockholm för en fast tjänst, men antisemitiska uppfattningar förekom beklagligtvis vid denna tid. Cramér skriver:

Feller hade visat sig som matematiker kunna samarbeta med både biologer



Willy Feller tillsammans med Clara Nielsen, student hos Feller i Kiel. De gifte sig 1938.

högskola, 1934–1939

och nationalekonomer, och flera av oss vid Högskolan, bland andra John Runnström och Erik Lundberg, försökte skaffa honom en permanent ställning här. Men motståndet från den mer eller mindre antisemitiska gruppen, Carleman, Carlson och flera andra, var för starkt, och när han sommaren 1939 fick ett erbjudande från USA, vågade vi inte ta på vårt ansvar att avråda honom från att ta emot en anställning där.



William Feller, fotografi från 1936, när han var verksam i Stockholm.

Även Feller själv har kommenterat de stämningar som rådde. I ett brev 1934 till Richard Courant skrev han följande (i engelsk översättning):

I never catch a glimpse of pure mathematics, partly because it virtually does not exist as a university institution – the students are learning on their own, the professors are only doing the exams – partly because Carleman is of the touching [rührend] opinion that one should execute [an die Wand stellen] all Jews and immigrants (which, however, he only tells his assistant after consuming a nonnegative [nichtnegativ] amount of alcohol).

Fem år senare, i ett brev till Otto E. Neugebauer, skrev Feller:

The emotions here are aggravating from day to day. There is a flood of student revolutions, Nazism at the universities is mushrooming.

Dessa brevcitat återfinns i en bok av Siegmund-Schultze, se källförteckningen i artikelns slut.

Feller lämnade Sverige kort före krigsutbrottet 1939, resterande karriär skedde i USA, där han modifierade förnamnet till William. Han knöts först till Brown University, blev professor vid Cornell University 1945, och slutligen vid Princeton från 1950 fram till sin bortgång 1970. Mathematical Genealogy Project listar 21 disputerade studenter till Feller, mest känd kanske Billingsley. Feller skrev två böcker om sannolikhets teori och över hundra publicerade artiklar, varav majoriteten faller inom sannolikhets teori, övriga mest annan matematik, men ett mindre antal är mer tillämpningsinriktade, främst mot genetik (paradoxer i teorin för naturligt urval).

William Feller

– De två böckerna

Fellers mest kända verk är förmodligen de två volymerna av *An Introduction to Probability Theory and its Applications, I och II*, publicerade 1950 respektive 1966. Volym 1 kom i tre upplagor (den tredje 1968), och en andra utgåva av volym 2 utkom 1971. Översättningar av Volym I



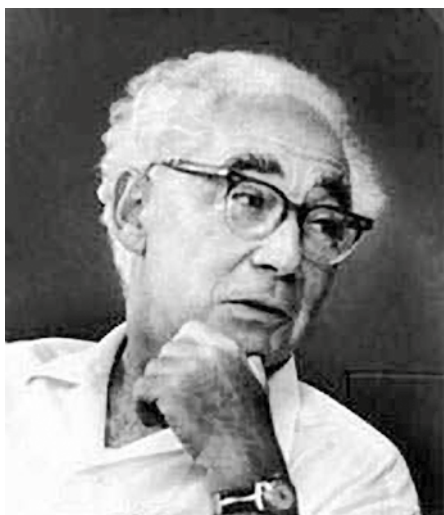
»»»» WILLY FELLER VID STOCKHOLMS HÖGSKOLA, 1934–1939

gjordes till ryska, japanska, kinesiska, polska, spanska och ungerska. Anmärkningsvärt är att den ryska översättningen av volym 1 var klar inom ett år efter att boken publicerats på engelska, vilket vittnar om ett mycket stort intresse. Böckerna blev också mycket väl mottagna och lovorden är många. Som ett exempel kan vi citera ur Ulf Grenanders recension i *Mathematical Reviews* 1957 av andra utgåvan av Volym I:

As in the first edition the exposition is mathematically rigorous and at the same time elegant and lucid. This fascinating book will remain a standard textbook of mathematical probability for many years to come.

Grenander hade rätt även i den sista frasen. Vid SU under 60-talet var Fellers bok kurslitteratur, både för 1 och 2 betyg i matematisk statistik, och den äldre av oss (RS), som läste den då, fann den utomordentligt stimulerande. Läs den! När den avsevärt mer avancerade Volym II kom, körde vi studiecirkel på den i forskarutbildningen. Den innehöll en hel del fel att upptäcka och försöka rätta, men det hade vi overseende med, för att den var så intressant att läsa. I andra upplagan var felet i stort sett borta.

Under rubriken "Books worth reading" inledde Terry Speed i *IMS Bulletin* Vol 34, issue 2 (2005) några kolumner om läsvärda böcker



William Feller, fotografi från 1960-talet, troligen.

med Fellers böcker, "My all time favourites", och fyllde en sida med superlativer om vad som gjorde dem så bra. Han avslutar med en uppmaning att pröva att läsa dem, "you won't regret it".

En minnesartikel som inleder *Annals of Mathematical Statistics* 41:6, beskriver utförligt Fellers gärningar och betydelse som probabilist. Och hela 1970 års volym av *Annals* dedicerades till Fellers minne! Minnesartikeln innehåller även en fullständig publikationslista, och en annan länk till den finns i litteraturlistan nedan. Här begränsar vi oss till att diskutera hans aktiviteter under Stockholmsåren.

Arbeten under Stockholmstiden

Perioden i Stockholm var mycket produktiv. Feller fullbordade ett 20-tal publikationer, de flesta som ensamförfattare. Flitigaste medförfattaren var Herbert Busemann, som även han disputerat i Göttingen (1931). Liksom Feller flydde han 1933 till Köpenhamn; från 1936 var han verksam i USA. Dessa arbeten är i differentialgeometri och publicerades i *Matem. Tidskr. B* och *Acta Math.*

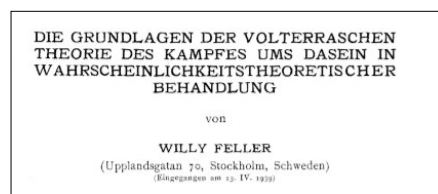
Merparten av artiklarna faller dock inom matematisk statistik -- fundamentala begrepp som stora talens lag, centrala gränsvärdesatsen och stokastiska processer förekommer exempelvis bland titlarna. Cramér skriver i en nekrolog (1970):

Among his publications during the Stockholm years are two outstanding papers, one containing his celebrated necessary and sufficient conditions for the validity of the Central Limit Theorem, and the other a development and generalization of the Kolmogorov theory of the class of stochastic processes which are now known as Markov processes.

Det ovan nämnda intresset för tillämpningar återspeglas i *On the logistic law of growth and its empirical verifications in biology*, publicerad 1940 i *Acta Biotheoretica*. Artikelns innehåll sammanfattar Feller själv 26 år senare i Volym II, sid 52.

Liksom i våra dagar förekom en viss eftersläpning från färdigställandet av manus till färdig publikation. I en artikel publicerad 1936, mottagen av tidskriften i maj 1935, anges i den färdiga artikeln kort och gott "Von Willy Feller in Stockholm". I en artikel publicerad i *Acta*

Biotheoretica 1 maj 1939 ger han sin hemadress (se bild).



Artikel publicerad 1939 i *Acta Biotheoretica*.

För ovannämnda artikel från 1940 uppges affilieringen Brown University, där Feller då verkade, men i slutet av artikeln står tryckt "Stockholm, Institute of math. Statistics"; manus mottaget av redaktionen i maj 1939.

Förlaget Springer hade redan i början av 1930-talet föreslagit att Cramér skulle skriva en bok i deras s.k. gula serie om sannolikhetskalkyl och matematisk statistik. Cramérs hustru, Marta, berättar i "Historien om en bok" att Feller var en "utomordenligt lämplig person att resonera med". Hon skriver vidare (om åren kring 1937-1939):

Samtidigt fortsatte Harald sina resonemang med Willy Feller och de hade en tid planer på att med Springers goda minne samarbeta vid författandet. Därav blev emellertid inget. Hitlerregimen och sedan andra världskrigets utbrott klippte av både samarbetet med Feller och förhandlingarna med Springers.

Sedermera tog bokprojektet form i *Mathematical Methods of Statistics* (hos Almqvist och Wiksells förlag 1945, Princeton University Press 1946).

Ett fruktbart möte ägde rum 1937 mellan Feller och Conny Palm, verksam vid Ericsson från 1936. Palm besökte seminarieserien, och höll själv ett föredrag den 12 maj 1938: Några matematiska problem rörande telefontekniken. I början av protokollet anges att "föredraget berörde pågående undersökningar angående samtalsfrekvens, vilka utföras för Telegrafverkets räkning av Ing. Palm". I föredraget figurerar, enligt Ove Lundbergs protokoll, bland annat Erlangs formel. Feller skrev följande 1950:

Waiting time and trunking problems for telephone exchanges were studied long before the theory

Protokoll

hålllet vid Försäkringsmatematiska seminariets sammanträde den 29 september 1938.

Dr W. Feller: Något om "statistical estimation" enligt Neyman och Pearson.

Föredragshållaren angav först problemställningarna.

En fördelningsfunktion $F(X; \vartheta_1, \vartheta_2, \dots, \vartheta_g)$ är given, där X är en k -dimensionell stokastisk variabel och ϑ_i äro parametrar. Funktionssammanhanget antas känt men ej värdena på parametrarna. Man förfogar öfver n st. värden på variabeln X , vilka kunna uppfattas som stickprov ur en oändligt stor population med fördelningen F .

Den ena frågeställningen är då: Hur skall man kunna ^{/finna} gränser för värdena på ϑ_i ?

Tal. avvisade metoden att införa hypotesen, att parametrarna ϑ_i äro stokastiska variabler med vissa (apriori)-fördelningar, och att studera fördelningsfunktionerna för ϑ_i under betingelsen att X antagit de n värdena X_1, X_2, \dots, X_n . Apriorifördelningarna måste alltid vara okända och de kunna ej för ändliga n , vilket i praktiken är det enda relevanta fallet, bortskaffas ur formlerna för de betingade fördelningarna.

Seminarieprotokoll från seminariet den 29 september 1938 (första sidan).

Sammanfattning

Med denna artikel har vi främst uppmärksammat att en briljant probabilist, William Feller, i början av sin karriär vistades en längre period i Sverige; hur han här arbetade såväl med teoretiska problemställningar som interdisciplinära samarbeten inom såväl bio- som ingenjörsvetenskaper och nationalekonomi. Dessutom måste framhållas att Harald Cramér, utöver sina vetenskapliga bidrag, var en stor humanist som tog strid för att behålla Feller i Stockholm.

JESPER RYDÉN OCH ROLF SUNDBERG
(Varmt tack till Georg Lindgren, som sökt efter information kring ett förmodat besök av Feller i Lund. Tack även till Allan Gut och Rolf Larsson för värdefulla upplysningar.)

of stochastic processes was available and had a stimulating influence on the development of the theory. In particular Palm's impressive work over many years has proven useful to several authors.

Mer om Conny Palm finns att läsa i en artikel av Haugen (1995).

Seminarier vid institutet

Som nämnts hölls en seminariereserie vid Cramérs institut. Här presenterades egna arbeten, men även sammanställningar av andra forskares. Rapporter från besök vid konferenser eller kongresser förekom också. För denna epok finns för varje föredrag ett referat, eller seminarieprotokoll, där Institutets amanuens i ord och formler sammanfattat föredraget.

På bilden nedan framgår de tillfällen då Feller talade:

1934	24 oktober, 31 oktober, 7 november: <i>Über die Grundlagen der Wahrscheinlichkeitsrechnung</i>
1935	13 februari: <i>Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung</i> 10 oktober, 24 oktober, 7 november: <i>Differentialekvationer inom riskteorin</i>
1937	23 september: <i>Om sannoliketskalkylens grundvalar</i> 4 november: <i>Berättelse från sannoliketskonferensen i Genève 1937</i>
1938	29 september: <i>Något om "statistical estimation" enligt Neyman och Pearson</i>
1939	20 april: <i>Om full-monotona funktioner och deras användning i sannoliketskalkylen</i>

Seminarietillfällen i Stockholm med William Feller som talare.

Bilden till höger ses som illustration första sidan av protokollet från den 29 september 1938.

Protokollen är typiskt sakligt skrivna, och även eventuella diskussioner mot slutet ingår. På sina ställen har protokollen humoristiska drag; här är ett exempel från den 4 november 1937 med temat "Berättelse från sannoliketskonferensen i Genève 1937". Bland annat behandlade Feller föredrag av Hopf, Neyman, Pólya och Steinhaus. Ove Lundberg, vid protokollet, avslutar detsamma med:

Seminarieret hade emellertid fått höra åtskilligt, både sant och intressant, kryddat med lustiga poänger, så att de, för de flesta åhörarna, abstrakta "storheterna" fingo drag av mänsklighet.

Här kan det passa att citera ur en minnesartikel av Joe Doob (1972), där han avslutningsvis skriver om Feller som föredragshållare:

But apart from his mathematics those who knew him personally will remember Feller most for his gusto, the pleasure with which he met life, the excitement with which he drew on his endless fund of anecdotes about life and its absurdities, particularly the absurdities involving mathematics and mathematicians. To listen to him deliver a mathematics lecture was a unique experience. No one else could generate in himself as well as in his auditors so much intense excitement. In losing him, the world of mathematics has lost one of its strongest personalities as well as one of its strongest researchers.

Källor till artikeln om Willy Feller

Cramér, H. (1970). William Feller, 1906-1970. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute* Vol. 38, No. 3 (1970): 435-436.

Doob, J.L. (1972). William Feller and twentieth century probability. *Berkeley Symposium on Mathem. Statistics and Probability*, Vol 6.2. Finns online på Project Euclid.

Haugen, R.B. (1995). The life and work of Conny Palm – some personal comments and experiences. *Elektronikk 2* (3): 50-55.

Sigmund--Schultze, R. (2009). *Mathematicians Fleeing from Nazi Germany*. Princeton University Press, Princeton and Oxford. Original på tyska: *Mathematiker auf der Flucht vor Hitler*, publicerades 1998 Braunschweig, Wiesbaden Vieweg.

Speed, T. (2005) Terence's stuff: Books worth reading. *IMS Bulletin* 34 (2)}.
Finns på https://imstat.org/wp-content/uploads/Bulletin34_2.pdf

Länk till Fellers publikationer:
<http://www.motapa.de/feller/feller-bibliography.shtml>

Länk till kroatisk hemsida om Fellers liv och gärningar:
<http://www.croatianhistory.net/etf/feller.html>

Länk till översikt om seminarier vid Cramérs avdelning 1934-1950:
<https://www.su.se/matematiska-institutionen/om-institutionen/om-oss/seminarier-1934-1950-institutet-för-försäkringsmatematik-och-matematisk-statistik-1.564950>

DAN HEDLIN:

Våga medverka i medierna!

Sveriges radios enorma entréhall väntar jag på att bli insläppt. Radio P5 Stockholm har en sändning bakom en glasruta. Där satt jag en gång. Det var dagen efter att Donald Trump vunnit valet. En vakt leder mig genom korridorerna till en studio. Bakom en glasruta sitter programledarna i pågående sändning. Ljudet som går ut i radio hörs i högtalare i rummet där jag väntar. Bakom det breda mixerbordet sitter en ljudtekniker med producenten bredvid. I det inre rummet sitter de medverkande vid ett runt bord. En stor mikrofon framför munnen. Hörlurarna gör att de hör hur det låter när det går ut i sändning.

Bör man som lärare i statistik på universitetet medverka i medierna? Det kallas ”samverkan” på högskolorna, eller mer informellt ”tredje uppgiften”. Samverkan är ett brett begrepp. På vår institution brukar vi ta emot gymnasieelever och prata om vad statistik är och hur det är att studera statistik. Det är också samverkan. Det finns inget krav på att man ska medverka i medierna.

Många statistiker säger till mig att det är bra att jag tar på mig det här. Annars hörs inte statistikernas röst, säger en del, vilket förstås är överdrivet. Under pandemin har Tom Britton medverkat otaliga gånger. Många andra statistiker hörs också i medierna.

En del kanske är rädda för att någon med en

journalistisk stil som Janne Josefsson ska dyka upp med mikrofonen. Att man med mikrofonen under näsan och strålkastarna i ansiktet märker att nu sitter man i skiten. Eller att medierna bandar 20 minuter och väljer ut de fem sekunder man sa något dumt. Men så går det inte till. Jag uppfattas som expert. Jag kan faktiskt tycka att jag blir behandlad lite väl okritiskt. Vetenskapsjournalisten Anna Bäsén på Expressen har sagt att det finns en rädsla för journalister hos forskare, i synnerhet bland yngre forskare. När jag en eller två gånger har blivit fulciterad, har det varit av andra forskare, inte av journalister.

Journalister gillar telefonsamtal. När det ringer i min arbetstelefon brukar det vara en journalist. Jag pratar gärna med journalister och har nästan bara goda erfarenheter. De som ringer mig är säkert inte något representativt utsnitt av journalister. De kan inte så mycket om statistiska metoder och statistisk inferens, men är bra på att ställa rätt frågor och går in för att verkligen förstå själva. Ofta blir jag imponerad av hur bra de förklarar saken i artikeln och vilka belysande exempel de har utformat.

»När jag en eller två gånger har blivit fulciterad, har det varit av andra forskare, inte av journalister.»

Det finns ett sätt att framställa en fråga som inte är helt ovanligt och som alltid gör mig skeptisk. Det är när man beskriver ett påhittat scenario och formulerar frågan inom det scenariot. Det är meningen att jag ska förstå att det är påhittat. Syftet är antagligen att jag inte ska styras av förutfat-

Ett axplock av Dan Hedlins medverkan i medier.



tade meningar. Kanske också att hemlighålla sitt ”gräv”. Frågan kanske handlar om huruvida personer med icke-svenskt namn döms oftare i rättegångar än personer med svenskt namn, och hur man skulle visa det statistiskt. Det påhitade scenariot handlar kanske om äpplen och päron eller nåt. Som statistiker vet man att det är jätte viktigt att känna till i vilket sammanhang en fråga ställs. Naturligtvis är det bra att i både forskning och undervisning renodla problem, att så att säga skala av allt utom det nakna, teoretiska problemet, så att det syns vilket problemets kärna är, och därmed öppna möjligheten till generalisering. Men i tillämpning måste man förstå sammanhanget.

Varför medverkar jag i medier? Det är främst nyfikenhet som driver mig. Av samma skäl har jag alltid gärna tagit på mig konsultuppdrag. Jag tycker att det är så roligt att se nya saker och lära mig nytt. Kanske jag har lite för lätt att säga ja också. Men om jag ser på det utifrån skulle jag hävda att det är viktigt att statistiker och statistikämnet syns i medierna och i debatten. Statistik är så viktigt för vår verklighetsuppfattning. Man skulle tro att de som är statistiker till yrket är särskilt medvetna om det. Men som Mats Wadman, då överdirektör för SCB, sa när jag intervjuade honom i Qvintensen nr 3/2009: ”Statistiker underskattar ofta betydelsen av statistik”.

Ett område där forskare gjort ovärderliga insatser i massmedier och sociala medier är klimatet. Hade inte forskare lagt så mycket tid och engagemang på det, hade många kanske även idag trott att den pågående klimatförändringen har naturliga orsaker.

På universiteten verkar många som jobbar på presstjänst utgå från att vi lärare och forskare ska prata om vår forskning. Egentligen gör jag det sällan. Jag forskar inte på opinionsundersökningar, som är det enskilt vanligaste ämnet jag blir intervjuad om. Ibland handlar frågorna om bortfall i undersökningar och då kommer det närmare det jag forskar på.

En vanlig typ av fråga som jag får handlar om hur någon undersökning tolkats och rapporterats. Journalisten brukar redan ha förstått vad problemet är eller vad felet består i och vill bara kolla med mig om det stämmer och hitta någon som vill säga det.

När Negra Efendić på Svenska Dagbladet skrev kritiska artiklar om medicinska åldersbedömningar med knäckkamera kollade hon ofta med mig att hon förstått rätt. Då är jag glad om jag kan hjälpa till. Det var bara enstaka gånger det ledde till en intervju i SvD och i de fallen sa hon till i förväg att hon kommer att skriva om det.

Det finns intressanta etiska aspekter i att uttala sig i medier. Är det helt okomplicerat för en forskare att uttrycka sina åsikter? I nr 4/2013 av Qvintensen intervjuade jag två forskare, Stefan Jansson och Olle Häggström, om etiska aspekter av att uttala sig i medier om forskning.

Man ska vara medveten om att journalisten kan börja skriva anteckningar direkt. Jag kan ibland höra att journalisten knattrar på tangentbordet medan jag pratar. I råd om kontakt med medier brukar det stå att ”off the record” inte finns. Min erfarenhet är att jag som svarar ändå förstår om det är off the record eller inte.



»»»»

»»»» VÅGA MEDVERKA I MEDIERNA!

Tar det inte en massa tid då? Jo, det är klart att det tar tid. Det kan ta ett par timmar att granska en rapport. Att åka in till radiohuset eller någon annan plats tar förstås tid, även om man brukar bli bjuden på taxi. För mig överväger fördelarna. Jag lär mig något på det. Jag använder mycket i undervisningen. Ett av många exempel som jag ofta berättar om är när Therese Rosenvinge frågade mig om en återkommande undersökning en kvällstidning utförde. De ställde frågor till tre slumpmässigt utvalda ishockeyspelare i varje lag i elitserien. Jag tror att hon funderade på om urvalet var tillräckligt stort. Jag sa att det är helt ok. Det blev inget inslag i SR:s program Medierna. Journalister är ofta fokuserade på problem. Är det inget problem tiger hälsan still, som det heter.

Jag hänvisar gärna till andra statistiker när medier kontaktar mig. Men jag vet inte allting vem som skulle kunna medverka.

I något enstaka fall har jag varit med om att en journalist bestämt sig för vilka slutsatser hen vill dra av sina data och inte vill ändra sig om jag går emot. En som jobbade på Tv4:s Kalla fakta hade samlat in data, kommit fram till en slutsats och ville kolla med mig om det var rätt. Efter att ha gått igenom hans siffror svarade jag i mejl att han hade räknat fel. Han ringde senare till en av mina kollegor på institutionen, och drog en lögn när han sa att jag inte hade tid. Som eloge till hans journalistiska integritet vill jag dock nämna att han i det färdiga inslaget i Kalla fakta inte hade med något alls om sina mödosamt insamlade data. Inslaget verkade balanserat och bra, såvitt jag kunde se. Men som sagt, mina erfarenheter av journalister är nästan bara goda.

Som alltid när man har med människor att

göra måste man respektera att andra kan se saker på annat sätt. I kontakt med medier måste man respektera att de väljer vinkeln. Aktuellt filmade mig i mitt rum på universitetet. Det tog minst en halvtimme. Fotografen monterade en liten strålkastare och tog ljud- och bildprov. Reportern satt med mikrofon framför mig. Jag satt med bokhyllan som bakgrund. Det blev ett inslag på högst åtta sekunder. Det redaktören valde att fokusera på var att jag sa att ”medierna har också ett ansvar” (nämligen att rapportera korrekt om statistik). Resten av inslaget i Aktuellt var en debatt mellan två nyhetschefer om mediernas ansvar.

Vid ett tillfälle filmade Rapport i parken utanför där jag bor. Det handlade om Moderaternas enligt min mening missriktade förslag om att göra en gammaldags folkräkning för att hitta folk som har fel uppgifter i folkbokföringen. Men någon större nyhet fick företrädere, så det blev bara några sekunders inslag där den enda röst som hördes var reporterns. Har man svårt för att medier kan prioritera bort en själv ska man inte medverka.

»Som alltid när man har med människor att göra måste man respektera att andra kan se saker på annat sätt. I kontakt med medier måste man respektera att de väljer vinkeln.»

TV-medier vill ha variation i sina bilder. När intervjun är slut går fotografen bort tio meter och filmar när man pratar om vad som helst med reportern. Eller så tar hen närbilder på reportern. En gång skulle jag sitta vid datorn och skriva. I ett anfall av fåfänga tog jag inte på glasögonen.

Jag såg ingenting av vad jag skrev. Jag frågade varför de vill ha bilder där man låtsas skriva eller läsa i en bok. ”Därför att om inslaget börjar med att du pratar, kommer åskådaren bara att titta på hur du ser ut och inte lyssna på vad du säger”, blev svaret.

En gång frågade en reporter på Sveriges

radio mig kanske tio gånger om samma sak. Hon varierade frågan lite grann varje gång. Vi satt i mitt rum på universitetet med en liten mikrofon på bordet. Jag förstod medan det pågick varför hon envisades. Hon ville att jag skulle arbeta mig fram till ett ordentligt och tydligt svar.

Det har inte hänt mig att en journalist bara ställer konstiga frågor. Men vore det så, skulle jag säga: ”jag upplever att du ställer fel sorts frågor, det som är intressant i det här sammanhanget är...”

I en direktsändning i Sveriges radio kände jag att jag inte hade kunskap nog att svara på en viss fråga. ”Du kan spekulera”, sa journalisten. Jag gjorde inte det. Efteråt sa producenten, som följde med mig ut till säkerhetsörrarna, att han var glad att jag inte gjorde det. Man kan säga nej. Man bör tänka efter hur säker man vill vara för att uttala sig. Det gäller inte bara vid kontakter med media. Man kan likna det med en skala som går från 0 (man kan prata om precis vad som helst, kallas ibland för att ”killgissa”) till 1 (man ska vara absolut säker, vilket i princip betyder att man aldrig kan säga något). Senast fick jag en fråga från en frilansande vetenskapsjournalist om vad ”tau-b” är. Ett alternativ till p-värde, undrade han. Jag sa som det var: att jag aldrig hört talas om det. Jag skäms inte för att jag kanske framstår som okunnig.

En gång började jag svara på en fråga och glömde bort vad jag skulle säga. Jag sa något i stil med ”nu kom jag av mig”. Det gick ut i direktsändning. Men vad spelar det för roll? Journalisten är beredd med en följdfråga.

Man ska respektera andras bristande kunskap. Det finns de som verkar tycka att alla borde kunna grundläggande högstadiematte. Ja, vore det så, skulle världen vara bättre. Men kan du själv allt man fick lära sig i grundskolan? Tror du det, tror jag att du har en uppblåst självbild.

En del anser att det inte går att säga något som är både meningsfullt och kortfattat om komplicerade frågor. Och därför kan man inte vara med i medier alls, för då blir det bara just kortfattat. Tom Britton sa i ett anförande i Statistikfrämjandet att man måste ”ljuga” lite grann.



»Säg inte nej till sminket. Sminkösen är ett proffs som lägger på något som neutraliserar studios snedvridande effekt. Man får ett rättvisande utseende, som man inte annars hade haft».

FOTO: SHUTTERSTOCK

Jag skulle vilja lägga en mer filosofisk syn på det: man är alltid kortfattad. Den till hundra procent fullständiga beskrivningen av något existerar inte, kan inte existera.

Debatter i radio ställer jag inte upp på, eftersom de debatter jag hör i radio nästan alltid är torftiga. Sveriges radio brukar tala om i förväg vilka fler som kommer att vara med i samma inslag. En gång har jag sagt att "då kommer jag inte", för det var helt klart att det skulle bli debatt. Kompromissen blev att jag bandades strax före direktsändningen.

Om du vill höra ett exempel på kritisk granskning av en statistisk undersökning kan du till exempel lyssna på ett inslag från Medierna 2018-09-29. "Barn pratar hellre med ingen alls när de har problem än med sin pappa". Inslaget slutar på ett överraskande sätt. Jag tror att du kan hitta det genom att googla på medierna P1 KP och mitt namn, eller bläddra igenom alla program som Medierna har sänt på SR:s webbplats. Min egen insats vill jag verkligen inte uttala mig om. Jag gör så gott jag kan.

Jag har mycket svårt för att lyssna på mig själv i radio. Och att se mig själv i tv är en svår plåga. Jag talar om för vår duktiga studievägledare, som är bra på att skriva, att det finns en länk. Så får hon avgöra om det är något att skriva om på institutionens webbplats. Jag vill inte fatta sådana beslut.

Jag avslutar med några råd. Säg inte nej till sminket. Sminkösen är ett proffs som lägger på något som neutraliserar studios snedvridande effekt. Man får ett rättvisande utseende, som

man inte annars hade haft. Det syns inte efteråt att man är sminkad.

- Journalister är ofta tacksamma om man själv föreslår en fråga. Kom ihåg att när det gäller statistik rör de sig på marker som de känner sig osäkra på. Man kan säga "jag skulle gärna vilja prata om...". Blir man under intervjun osäker på något, kan man be att få kolla upp det och återkomma. Eller helt enkelt bara säga att man inte vet. I en bandad intervju kan man säga "äh det där blev inte bra, kan vi ta om det".
- Journalister brukar fråga "vad är viktigast". Det kan vara den första frågan man får. Till exempel "Du har tidigare kritiserat den här metoden. Vilket är det viktigaste skälet till att den inte fungerar?" Kan man inte avgöra vad som är viktigast kan man säga: "Det finns många skäl, men jag skulle kunna nämna...". Alldeles nyss hörde jag en cancerläkare svara på frågan om vad som är viktigast att tänka på när man är ute i solen. Hon räknade raskt upp inte en utan tre saker.
- Var beredd på att den du pratar med lägger om rösten när mikrofonen slås på. TV-personer och radioreportrar har ofta en professionell röst.
- Journalister som ringer frågar alltid om man har tid för ett samtal. Det går utmärkt att säga att de kan ringa tillbaka om tjugo minuter. Då ringer de om exakt 20 minuter.

- Ge gärna positiv återkoppling. Om du tycker att artikeln blev bra, skriv och säg det. I sociala medier, sms eller mejl.
- Om det känns som om man gjorde en slät figur i tv, så ska man inte hänga upp sig på det. De flesta av oss är benägna att vara alltför självkritiska. Om en artikel inte blev så bra är det inte ditt fel.
- Media arbetar i högt tempo. Svarar man "kanske" på frågan om man kan vara med i Aktuellt på kvällen, kommer de att fortsätta ringa till andra och behålla ens namn som reserv. Har man då ägnat eftermiddagen åt förberedelser kommer man bli besviken. Bättre att säga ja eller nej på en gång. Eller be att få att få återkomma om en kvart.
- I kurser i medieträning kan man få höra att man inte ska svara på frågor utan istället säga det man vill föra fram. Det rådet följer jag inte. Jag kan inte förstå varför man inte ska svara på frågor. Det skulle möjligen vara om man förespeglats inför intervjun att man ska få berätta om sin forskning och journalisten envisas med att fråga om något annat.

Som sagt, jag hänvisar gärna till andra statistiker när medier kontaktar mig. Men jag skulle behöva veta vilka som jag kan tipsa journalisten om.

DAN HEDLIN,
STOCKHOLMS UNIVERSITET

Från utbildning till arbetsliv

Första gången jag satte mig i hörsal 3 i Ekonomikum vid Uppsala universitet visste jag inte vad statistik var för något. Inte vad det *egentligen* var. Innan kursen började var statistik för mig, liksom för många andra, något nödvändigt ont som man behövde slutföra för att få fortsätta med ekonomiutbildningen. Det skulle vara grått, tråkigt och tungrodd matematik. Kursarna var dock överens om att en halv termins elände fick det vara värt för att senare få återgå till nationalekonomin. En tydlig bild hade formats innan kursen ens hade startat. En diskurs som spred sig som en löpeld genom korridorerna men som endast var grundad i hörsägen. Detta var hösten 2016 och det var under den perioden jag upptäckte statistiken.

De introducerande föreläsningarna på grundnivån gjorde ett bra jobb med att fånga upp oss studenter. I sannolikhetsläran studerade vi sannolikheten att få kåk i en pokerhand eller att få yatzy på första kastet. Vad behöver vinsten och dess sannolikhet vara för att man som riskaversiv ska kunna rationalisera ett deltagande i lotteriet? Jag upptäckte snabbt att det fanns många områden där man kunde använda statistiken för att svara på problem man tidigare lek-

mannamässigt grubblat på, utan att kunna lista ut hur man bör gå till väga för att lösa dem. Triviala men intressanta problem. Ytterligare fascination uppstod när föreläsaren visade upp stillbilder från gårdagens TV-nyheter. Nyhetsankaret beskrev ett linjediagram som särskilt hiskeligt. Linjen såg ut som höga berg och djupa dalar och var i illröd färg. Sedan fick vi studenter se ytterligare en stillbild från en nyhetssändning, denna gång från en annan produktion. Samma undersökning framställdes där i form av ett linjediagram men kommunicerades nu på ett helt annat sätt. Grafens skala på y-axeln hade fördubblats och linjen, som tidigare gallskrek i dess röda lynne, var nu av mer timid och grå karaktär. Linjens berg och dalar hade förvandlats till platt mark och den panikkänsla som uppstod från den tidigare nyhetssändningen var som bortblåst. Det var tydligt att ett ansvar låg i kommunikation av data och resultat. Ett ansvar som grundade sig i att noggrant analysera data, anpassa relevanta modeller, kommunicera resultat så objektivt som möjligt samt tillämpa känslighetsanalyser. Det låg en utmaning däri som jag ivrigt accepterade. Jag lämnade ekonomiutbildningen bakom mig och fortsatte med statistiken istället.

Masterutbildningen i statistik på Uppsala universitet började 2018 med en rivstart. Linjär algebra och alla dess faktoriseringar mötte oss efter sommarlovet och för många blev det en rejäl uppförbacke. Detta problem grundade sig nog i de lägre antagningskraven för kurserna på grundnivån. För att fånga upp fler studenter från gymnasiet, samt plocka in eventuella intressenter från ekonomi- och statsvetenskapsutbildningen på universitetet, krävs numera endast matematik 3 från gymnasiet. Jag

vill tro att detta sänkta krav hämmar utbildningen i någon mån. Mer än en gång fick vi höra att vissa delar i kurslitteraturen var utanför kursen på grund av deras intrikata matematiska formulering. Ibland uppstod känslan att vi saknade en djupare förståelse för statistiska fenomen på grund av den avancerade matematikens frånvaro. Samtidigt har jag de sänkta antagningskraven att tacka. Utan dem hade jag inte kunnat inleda statistikutbildningen och antagligen inte landat på den arbetsplats jag stormtrivs på idag.

Kurser i sannolikhetslära, inferens och strukturella ekvationsmodeller följde på varandra under höst och vår. Den kurs som fick mig att särskilt stanna upp och fundera var den i kausal inferens. Redan från högstadiet får man höra att korrelation inte innebär kausalitet och det var en sanning som upprepades på alla våra kurser där olika relationsmått avhandlades. För hitintills var det endast olika relationsmått som beskrivits, fram tills nu. Vad är kausalitet? Hur bevisar man ett kausalt samband? Hur skattar man en kausal parameter? Dessa frågor hänförde mig och var något jag ville grotta ned mig i ytterligare. Tillämpningar inom biostatistik var extra intressant. Det mynnade ut i en masteruppsats i form av en tillämpad registerstudie på lymfompatienter där jag undersökte den kausala effekten av en särskild cancerbehandling på risk för hjärtsjukdom. Frågeställningarna var givetvis intressanta ur ett metodologiskt perspektiv men var även del av ett större, väldigt viktigt, arbete. Det var en ynnest att som statistiker få hjälpa till med det.

Det fantastiska med statistik är att området inte är branschspecifikt. De flesta institutioner och företag har nu-



Ekonomikum i Uppsala.

»Självfallet krävs en fallenhet för siffror i någon mån, men för mig lockar de varierande tillämpningarna och vad man faktiskt kan utröna ur, till synes, stora och invecklade datamängder. «

mera stora mängder data att analysera. Universitetet gör ett bra arbete med att följa hur arbetsmarknaden för statistiker utvecklar sig och har möjlighet att justera sitt utbud av utbildningar och kurser sett till efterfrågan. Extra känsligt må det vara i statistikutbildningen då efterfrågan av spetskompetenser på arbetsmarknaden ser olika ut beroende på bransch. Nya tillämpningar av artificiell intelligens och maskininläring fordrar en relevant kurs på masterprogrammet i statistik och det är något jag gärna ser mer av. Ett ypperligt exempel är biostatistik och överlevnadsanalys. Just nu finns det inget masterprogram i biostatistik i Sverige. Detta medför att information kring potentiella arbetsroller och statistisk modellering inom olika områden behöver komma bort ifrån de mer "generella" utbildningarna som har ett helhetsperspektiv på statistik. Utan detta famlar studenterna lätt i mörkret när det kommer till vilka arbetsplatser man kan söka sig till. Samtidigt uppstår problem när efterfrågan från studenterna från början är liten. Det är inte hållbart för universitetet att utforma och hålla kurser när inga önskar gå dem. Det finns en balansgång här som är svår att få till och jag är osäker på vem ansvaret ligger på. Jag vill inte tro att intresset för biostatistik är lågt, snarare har studenterna inte fått upp ögonen för det än.

Utbildningen i statistik har möjlighet att marknadsföra sig bättre och tydligare. Denna möjlighet är inte begränsad till Uppsala utan gäller för många andra statistikutbildningar. Under min utbildning frågade vänner och familj vad jag faktiskt hade för mig på universitetet. Jag svarade alltid sanningsenligt och stolt att jag studerade på masterprogrammet i statistik. Reaktionen som följde var dock

alltid densamma: "Oj, då tycker du om siffror!" Självfallet krävs en fallenhet för siffror i någon mån, men för mig lockar de varierande tillämpningarna och vad man faktiskt kan utröna ur, till synes, stora och invecklade datamängder. När jag lyft fram exempel på sådana tillämpningar väcks intresset bland de tidigare avvaktande frågeställarna. Många har en bild av statistik som ett nödvändigt ont och glömmer hur brett och rikt området faktiskt är. Om det budskapet skulle förmedlas och om tydligare vägar till olika arbetsroller inom statistik utstakas hade vi möjligen sett en tillökning av intresserade studenter. Ytterligare hade tydligare inriktningar inom statistiken förberett oss än mer inför framtida arbetsuppgifter.

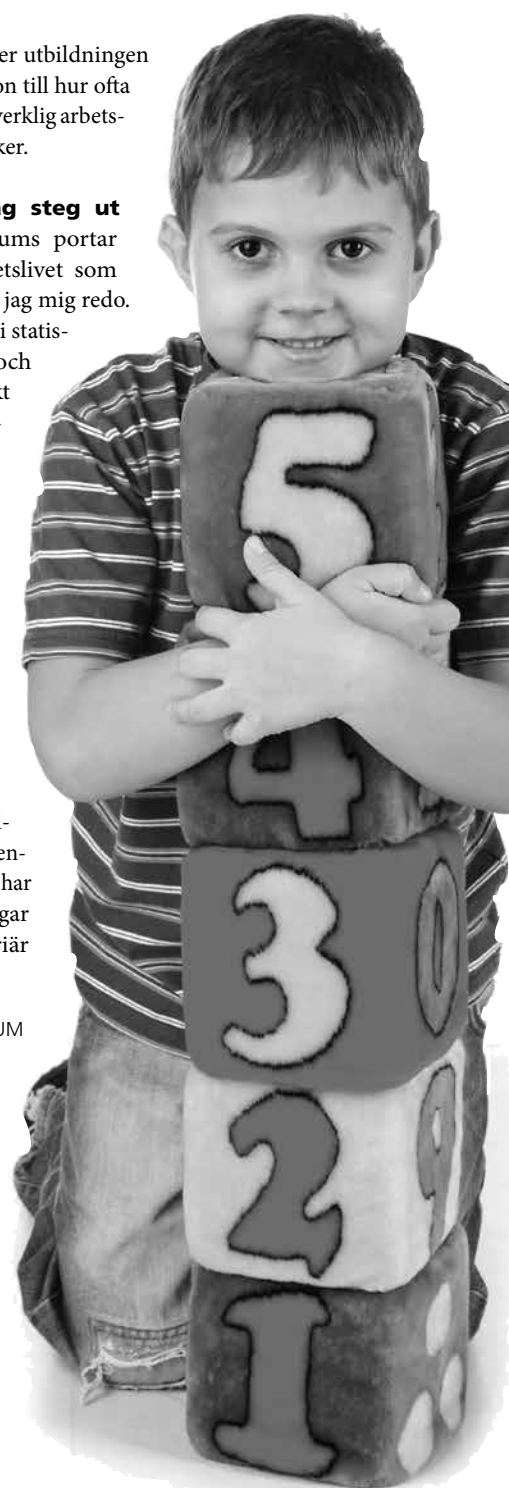
Efter att jag examinerades Uppsala våren 2021 arbetar jag nu som statistikkonsult på SDS Life Science. I mitt arbete hjälper jag främst till med observationsstudier och registerstudier. Det kan både vara mindre forskningsgrupper och större läkemedelsföretag som behöver hjälp med statistiken. Arbetet kan variera väldigt och är för mig mycket givande. Vad som däremot är återkommande i alla typer av kundprojekt är varierande datakvalitet. Missing values, inkonsekventa variabelformat och andra allmänna datahanteringsfrågor är något vi konsulter inom life science arbetar med dagligen. Det är en del förberedelser som krävs innan man kan anpassa modeller och tolka resultat. Dessa typer av förberedelser var något jag saknade i utbildningen. Givetvis är varje litet beslut angående data omöjligt att undersöka i utbildningen, som till exempel att korrekt imputationsteknik är avgörande för ett giltigt resultat. Sådana frågeställningar dök upp som en paren-

tes här och var under utbildningen men inte i proportion till hur ofta de förekommer i en verklig arbets-situation för statistiker.

Sista gången jag steg ut genom Ekonomikums portar för att inleda arbetslivet som biostatistiker kände jag mig redo. Med en bred grund i statistisk modellering och analys, en utmärkt programmeringsförmåga, som var icke-existerande innan utbildningen, och ett otroligt givande examensarbete i bagaget har alltsammans bidragit till en professionell självsäkerhet. Med ett nätverk av kursare, där samtliga nu gått sin egen väg i statistiken, universitetslektorer och engagerade professorer har jag alla förutsättningar för en lyckad karriär som statistiker.

JAKOB BÖRSUM

Siffror – en medfödd kärlek ?



Kortfattat om surveys

Det som kommer att sägas här är inga märkvärdiga saker. Det handlar om sådant som nog är välkänt för flertalet statistiker med samhällsvetenskaplig inriktning. Men för statistiker (inklusive matematiska statistiker) med annan bakgrund och annorlunda arbetsområden är det här kanske främmande mark. Mitt intryck är att teori och praktik avseende stickprov från ändliga populationer har hamnat lite i skymundan i universitetens statistikundervisning, trots att det är ett stort tillämpningsområde. Därför denna mycket korta, förenklade, icke-tekniska (och något subjektiva) översikt, som utgår från termen "survey".

Vad är en survey?

Det engelska ordet *survey* används mer och mer i statistiska sammanhang även i svenska språket. Med en survey menas då en speciell typ av undersökning. Men vad är det som kännetecknar en survey? Ordet är inlånat från engelskan och kommer (via franskan) från latinets *super* (= över) och *videre* (= se). Svenska Akademiens ordbok (SAOB) ger följande något föråldrade definition:

(i fackspr., i sht *statist.* o. *sociol.*)

1) om metod bestående i en icke experimentellt anlagd, stickprovsmässig frågeundersökning i stor skala med beskrivande l. förklarande syfte; äv. om olika modifikationer av sådan metod. [...] Vissa metoder att insamla data förknippas vanligen med survey: observation och intervju. [...].

En nyare och mer statistiskt färgad beskrivning ges av Särndal och Lundström (2005):

The objective of a survey is to provide information about unknown characteristics, called parameters, of a finite collection of elements, called a population, of individuals, or households, or enterprises. A typical survey involves many study variables and produces estimates of different types of parameters.

Termen *survey* används mest i samhällsvetenskapliga sammanhang. Men det verkar inte finnas någon mer detaljerad definition av vilka krav

en undersökning ska uppfylla för att få kallas *survey*. Typiska kännetecken är dock följande:

1. Det finns en väldefinierad *ändlig population* av *objekt*, som man är intresserad av. "Objekten" kan t.ex. vara individer, hushåll, skolor, företag eller andra objekt i samhället.
2. Man vill ha information om vissa sammanfattande, kollektiva, numeriska storheter hos populationen (ibland, möjligen något oegentligt, kallade "parametrar").
3. Man väljer ut ett antal objekt (ett *stickprov*) från populationen. Urvalet kan göras på olika sätt. Det mest primitiva är att bara ta ett antal tillgängliga objekt, utan att bry sig om varför just dessa objekt blir utvalda. Det teoretiskt mest välbehandlade fallet är däremot s.k. *sannolikhetsurval*, som innebär att varje objekt i populationen har en i förväg känd sannolikhet att bli utvalt.
4. Genom någon form av *datainsamling* försöker man erhålla önskade uppgifter från de utvalda objekten; numera ofta genom postenkät eller webbenkät. I regel är det omöjligt att få alla tänkta uppgiftslämnare att medverka och svara på de frågor som ställs. Det förekommer alltså *bortfall*. Observera att man måste skilja mellan (1) det ursprungliga stickprovet av objekt, som dragits från populationen, och

(2) den delmängd av detta stickprov som man lyckas få svar från. Även om det ursprungliga stickprovet har erhållits genom sannolikhetsurval, så gäller detta inte för den slutliga svarsdelen av detta stickprov. Vi känner ju inte till sannolikheten för att en utvald person ska svara.

5. Med användande av erhållna data beräknas *skattningar* av de populationsstorheter man är intresserad av. Ett problem i samband med bortfall är att det finns risk för *bias* hos skattningarna. (Se vidare stycket om bortfall i nästa avsnitt.)

En del saker är oklara för mig, t.ex.: Kan en undersökning utan sannolikhetsurval i boten kallas för en survey? Vilka villkor måste i så fall vara uppfyllda? Måste det vara en viss minimiandel (eller ett visst minimiantal) svarande objekt för att man ska kunna tala om en survey? Kan en totalundersökning ses som ett specialfall av en survey? Var går gränsen mellan bra och dåliga surveys? I resten av denna artikel får läsaren stå ut med att termen *survey* används på ett lite trevande sätt.

Komplikationer i praktiken

I survey-verkligheten tillstöter alltid komplikationer, som man får försöka lösa från fall till fall:

Ramproblem. För att kunna göra ett sannolikhetsurval av element krävs i princip en förteckning över samtliga element i populationen, en s.k. *urvalsram*. Men den urvalsram man har tillgång till stämmer ofta inte helt överens med den population man egentligen är intresserad av. Man brukar tala om *målpopulation* och *rampopulation*, och man måste vara medveten om att dessa i praktiken kan skilja sig åt, mer eller mindre.

Bortfallsproblem. Bortfallet har ökat markant under de senaste decennierna och upplevs som ett växande hot. Man måste vara medveten om bortfallets eventuellt snedvridande effekt. Ef-

tersom förekomst av bortfall innebär att vi inte har ett sannolikhetsurval, så kan vi egentligen inte använda sådana begrepp som väntevärdesriktighet eller konfidensintervall i samband med surveyskattningar. Man är hänvisad till approximativa metoder när det gäller att redovisa skattningars säkerhet. Bortfallsproblemet kan i någon mån angripas genom att dels på olika sätt försöka minska bortfallets storlek, dels på olika sätt försöka beakta bortfallet vid beräkning av skattningar och deras osäkerhet (t.ex. genom utnyttjande av tillgänglig hjälpinformation). Mindre nogräknade undersökare kanske inte ens låtsas om bortfallets existens.

Mätproblem. Att konstruera en frågeblankett är en konst för sig. Man vill ju att uppgiftslämnaren (1) inte ska avskräckas och kanske låta bli att svara, och (2) inte ska missuppfatta frågorna och lämna felaktiga svar, och (3) inte ska lämna avsiktligt lögnaktiga svar. Det här är viktiga saker, men i traditionell statistikundervisning sägs inte mycket om sådana saker. I de kommersiella undersökningar där jag själv deltar som intervjuperson finner jag ofta dåligt formulerade frågor.



Tore Dalenius (1917-2002), svensk pionjär.

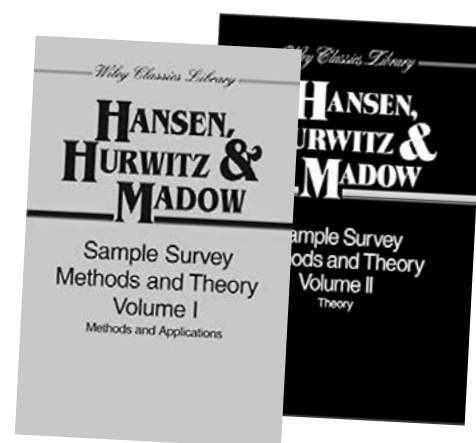
Kort om metodutveckling

Den moderna (randomiseringsbaserade) surveyteorin började få slutlig form på 1930-talet; Jerzy Neyman är ett känt namn. Det började bli möjligt att uttala sig om undersökningsresultatets säkerhet på ett matematiskt och objektivt sätt, med användande av begrepp från statistisk inferensteori. Det började också bli möjligt att dimensionera undersökningar så att man uppnådde önskad (men inte onödigt stor) noggrannhet. Det var ett stort vetenskapligt framsteg. Vid mitten av 1900-talet började denna surveyteori få genomslag på olika tillämpningsområden, bl.a. inom officiell statistikproduktion. Vid censusbyrån i USA fanns pionjärerna ”Hansen-Hurwitz-Madow”, vilkas lärobok (i två delar) blev ett standardverk. En svensk pionjär var Tore Dalenius.

Teorin utvecklades vidare. Men världen förändrades, och det visade sig så småningom att den ursprungliga teorin inte riktigt kunde hantera problemet med ökande bortfall. I vissa tillämpningar var bortfallet en betydande, men okontrollerbar, felkälla. Här hade man ju ingen känd randomiseringsmekanism att resonera kring. Ansvarsmedvetna statistikproducenter har med tiden utvecklat egna metoder att i någon mån komma tillrätta med estimationsproblemet i samband med bortfall. Men någon standardmetodik finns inte. Ofta är det fråga om att genom utnyttjande av tillgänglig hjälpinformation (t.ex. registerdata) försöka minska bortfallets eventuellt snedvridande effekt, så gott det går.

Surveyverksamhet vid SCB

Låt oss nu hastigt se på några exempel på tillämpningsområden, där termen ”survey” brukar användas. Vi börjar med Statistiska centralbyrån (SCB), som är den centrala statliga statistikmyndigheten. På statligt uppdrag producerar man s.k. officiell statistik. (En del officiell statistik produceras numera också av andra statliga myndigheter.) Vid sidan av den officiella sta-



Hansen-Hurwitz-Madows lärobok (i två delar) blev ett standardverk.

tistikproduktionen gör SCB dessutom många undersökningar på uppdrag av olika kunder.

En del av SCB:s undersökningar kan karakteriseras som surveys. I SCB:s surveyverksamhet har den statistiska teorin för urval från ändliga populationer haft stor genomslagskraft. Den ger stöd vid resonemang om hur undersökningar ska utformas och dimensioneras. Den ger också stöd vid resonemang om undersökningsresultatets säkerhet.

Liksom vid alla surveytillämpningar uppstår praktiska problem, t.ex.:

- Det ökande bortfallet är ett problem som är svårt att hantera.
- Ökande krav på att kostnaderna måste hållas nere kan innebära att intern metodutveckling kommer i andra hand.
- Ibland anlitas kanske mindre seriösa ”underleverantörer”.
- Krav på ansvarsfull hantering av personuppgifter måste noggrant beaktas.

Man kan ibland märka en nyfikenhet på att i en mer eller mindre avlägsen framtid gå utöver det traditionella handfasta arbetssättet med san-



»»»» KORTFATTAT OM SURVEYS



Vid vallokalsundersökningar får ett antal personer, som just lämnar vallokalen anonymt svara på ett antal frågor, bl.a. om hur de nyss har röstat. Bilden från en vallokalsundersökning i Kiev, Ukraina.

nolikhetsurval, datainsamling, estimation och kvalitetsdeklaration. Här är det nu datorernas framtida möjligheter som hägrar. Termer som "bayesiansk inferens", "big data", "data science", "artificiell intelligens", maskininlärning" och liknande dyker upp. En (såvitt jag förstår) ännu obesvarad fråga är hur dessa saker ska förenas med den officiella statistikens krav på objektivitet, genomskinlighet och största möjliga exakthet.

Till sist ska påpekas att en stor del av den officiella statistiken bygger på annan metodik än surveymetodik. Några exempel är nationalräkenskaper, befolkningsstatistik och konsumentprisindex, samt allehanda statistik baserad på administrativa register, som från början tagits fram för andra ändamål än statistikproduktion.

Surveys rörande partisympatier

Så kallade *partisympatiundersökningar* utförs regelbundet av flera opinionsinstitut. De tilldrar sig stort intresse, varför de kan vara värda ett eget avsnitt. Varje institut gör här sin egen undersökning. Typiska kunder är olika media (såsom dagstidningar och tv), som vill publicera nyheter om medborgarnas politiska åsikter. Undersökningsresultaten brukar presenteras i media med angivande av vilket institut som utfört undersökningen. Detta ses nog som en sorts reklam för institutet ifråga. (Om jag förstått rätt, så lär det också förekomma att enskilda politiska grupperingar ibland beställer egna specialundersökningar, där resultaten inte presenteras för allmänheten.)

I partisympatiundersökningar brukar uppgiftslämnarna vanligen uppmanas att ange vilket politiskt parti de skulle rösta på "om det vore val idag". Det kan diskuteras om en sådan undersökning ska betraktas som en opinionsundersökning eller ej. Man frågar ju om hur respondenten skulle "handla" i en tänkt situation, inte om vad han eller hon egentligen "tycker". Det kan ju t.ex. vara så att man av strategiska skäl säger att man skulle rösta ("stödrösta") på ett annat parti än det man egentligen sympatiserar med, eller att

»Användande av s.k. webbpaneler gör att det ofta är svårt att få grepp om hur urvalet av medverkande egentligen har gått till. Bortfallshantering kan vara svår att få grepp om.»

man med avsikt svarar ”fel” i undersökningen för att inte avslöja vilket parti man egentligen sympatiserar med.

I samband med riksdagsvalet vart fjärde år brukar den närmast föregående partisympatiundersökningen ibland betraktas som en sorts inofficiell valprognos. Detta är lite diskutabelt, eftersom det kan vara osäkert om den tillfrågade i praktiken kommer att rösta enligt sin tidigare angivna ”partisympati”. Det kan ju vara så att man ändrar sympati under tiden fram till valdagen. Ibland verkar det vara en sorts inofficiell tävling mellan olika institut om vem som ”kommer närmast valresultatet”.

Partisympatiundersökningar utförda av privata ”institut” kritiserar ibland av representanter för den mer vetenskapliga teorin för urval från ändliga populationer. Det kan gälla bristfällig redovisning av hur undersökningarna gått till, samt att urvalsmetoderna inte uppfyller villkoren för sannolikhetsurval. Användande av s.k. webbpaneler gör att det ofta är svårt att få grepp om hur urvalet av medverkande egentligen har gått till. Bortfallshantering kan vara svår att få grepp om. I sin presentation av resultaten använder undersökningsföretagen ofta statistiska termer och begrepp, som såvitt jag förstår saknar mening i sammanhanget (t.ex. ”felmarginal” och ”statistiskt säkerställda skillnader”).

Förutom opinionsinstitutens undersökningar så finns också partisympatiundersökningar utförda av SCB. De görs bara två gånger per år, och de gör inte anspråk på att ses som valprognoser. Där frågas också hur utvalda personer skulle rösta ”om det vore val idag”. Här lägger man ner åtskillig möda på att utföra undersökningarna enligt den randomiseringsbaserade surveyteorins principer. Dessa undersökningar kostar betydligt mer än de enskilda institutens undersökningar, och mitt intryck är att instituten betraktar SCB:s partisympatiundersökningar med viss respekt.

En helt annan typ av undersökning avseende valbeteende är SVT:s s.k. *vallokalsundersök-*

ningar. En sådan görs vart fjärde år i samband med riksdagsvalet. Enligt någon för mig okänd urvalsprincip väljs ett antal personer, som just lämnar vallokalen, och dessa får (anonymt) svara på ett antal frågor, bl.a. om hur de nyss har röstat. Målet är att snabbt (redan under valvakan) kunna göra en prognos av valresultatet. Jag är osäker på om denna undersökning kan kallas för en *survey*.

Annan surveyverksamhet med samhällsanknytning

Statistiska undersökningar av surveytyp görs ibland av myndigheter, företag, organisationer och andra. Dessutom finns det privata undersökningsföretag, vilkas hela affärsidé är att utföra undersökningar (av vilka vissa kallas *surveys*) på uppdrag av betalande kunder. En del av de här företagen kallas ibland *opinionsinstitut*. De undersökningar som utförs kan ofta karakteriseras som opinionsundersökningar och marknadsundersökningar.

Bland praktiska problem i samband med opinionsinstitutens verksamhet kan nämnas:

- Undersökningarna måste av konkurrensskäl vara snabba och billiga.
- Kunderna saknar ofta kunskap om statistiskt tänkande och nöjer sig med vad de får.
- Användande av s.k. webbpaneler gör att det ofta är svårt (eller omöjligt) att få grepp om hur urval och bortfallshantering egentligen har gått till.
- De urval som görs kan oftast inte ses som erhållna genom sannolikhetsurval, vilket gör det svårt att bedöma resultatens kvalitet.

I samband med opinionsundersökningar kan man också nämna SOM-institutet, där ”SOM” står för Samhälle, Opinion och Medier. Det är en oberoende och icke-kommersiell forsknings- och undersökningsorganisation med anknytning till Statsvetenskapliga institutionen och Institutionen för journalistik, medier och kommunikation vid Göteborgs universitet. Med

vetenskapliga ambitioner utför man varje år olika *surveys* för att studera befolkningens åsikter och värderingar i frågor rörande politik och samhälle.

Avslutande kommentarer

Det här var bara några få exempel på tillämpningsområden. *Surveys* med krav på hög kvalitet kan förekomma på många andra håll, t.ex. inom ramen för samhällsvetenskaplig forskning vid universitet och högskolor. SCB anlitas här ofta för planering och genomförande.

Till sist vill jag nämna att det naturligtvis finns studier inom samhälls- och beteendevetenskap av ett helt annat slag, som inte alls har med *surveys* eller statistik att göra. Jag tänker på sådan metodik som ibland brukar sammanfattas under termen ”kvalitativa metoder”. Jag vill poängtera att jag inte alls kritiserar denna metodik. Den har sitt berättigande vid frågeställningar, där det statistiska synsättet inte är aktuellt.

Men vad är då ett statistiskt synsätt? Jag ser statistikvetenskapen som en vetenskap huvudsakligen om *metoder och deras egenskaper*. Det handlar om metoder att (1) *samla in data*, (2) *dra slutsatser eller fatta beslut från insamlade data* och (3) *uttala sig om resultatens osäkerhet*, i regel med hjälp av sannolikheteoretiskt resonemang. Det uppfattar jag som statistikvetenskapens kärna. Andra kanske tycker annorlunda.

JAN WRETMAN

(Jag vill tacka Sixten Lundström och Carl-Erik Särndal för värdefulla synpunkter på tidigare utkast.)

Referenser

- Svenska Akademiens ordbok, SAOB (1998). Aktuellt ord: *survey*. (Sökord på webben: svenska.se)
- Särndal, C. E. och Lundström, S. (2005). *Estimation in Surveys with Non-response*. New York: Wiley.

Rankning av universitet

För några år sedan roade jag mig med att skriva om ranking av universitet i Qvintensen. Det som nu fick mig att nu följa upp artikeln var dels Hanne Kjöllers bok *Kris i forskningsfrågan eller vad FAN får vi för pengarna*, dels en artikel av Jan Wretman om ranking i senaste Qvintensen.

Kris i forskningsfrågan

Hanne Kjöllers huvudbudskap är att Sverige spenderar alltmer pengar på forskning men ändå tappar terräng, bl.a. gentemot några andra mindre europeiska nationer vilka hon nämner flera gånger: Danmark, Nederländerna och Schweiz. Som en viktig förklaring

uppges hon att Sverige pytsar ut mycket resurser på en mängd mindre högskolor. Liknande synpunkter framförs även av företrädare för Kungliga Vetenskapsakademien i ett debattinlägg i SvD 2021-05-25. Kjöllers nämner dock på ett ställe att enligt Riksrevisionen så var Högskolan i Kristianstad ett av sex lärosäten som uppnådde ”full resurseffektivitet under samtliga sex mätår”. Detta torde indikera att det borde finnas förbättringsmöjligheter även för mindre lärosäten.

Vad jag uppfattar att Kjöllers förbigår är studentaspekten: de mindre lärosätena sänker sannolikt tröskeln för personer från mindre studievana miljöer att inleda akademiska studier. För att bedriva akademisk undervisning behövs någon form av forskning vid mindre lärosäten, även om den har svårt att nå världsklass. Hur mycket den ena eller andra aspekten bör vägas in är inte självklart, och beror bl.a. på politiska preferenser, men bör ändå beaktas i diskussionen.

En aspekt som bör intressera medlemmarna i Statistikfrämjandet är att Kjöllers vid ett flertal tillfällen lyfter fram vikten av goda statistikkunskaper. Hon nämner John Ioannidis ett antal gånger. I ett av sina slutsatskapitel skriver hon också: ”Däremot verkar många alltför dåligt utbildade i statistik för att ens själva förstå att resultaten inte håller”. Kjöllers förlitar sig också rätt mycket på statistik över citeringar o.dyl. samtidigt som hon framför viss kritik mot syste-

met. Man bör ha Godhaarts lag i åtanke ” Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes.”

För att sammanfatta så tror jag att Kjöllers lyfter fram många viktiga po-
änger, men det behövs att någon med bättre insikter om det akademiska systemet recenserar hennes bok.

Tankar kring rankinglistor

Wretmans artikel tar egentligen upp en specifik rankinglista, vilket också titeln antyder, nämligen om hur näringslivet i 290 svenska kommuner upplever företagsklimatet. Men de frågeställningar som tas upp torde dock vara av generell intresse, exempelvis för ranking av universitet. En sådan fråga är om det går att beräkna någon form av konfidensintervall för placeringarna. När man studerar ranking av universitet kan man få viss känsla för detta genom att jämföra olika listor med varandra (se till exempel Tabell 1 och 2).

Då jag råkade mailväxla med en professor emeritus i matematik på Paris Saclay (f.d. Paris Sud) lade jag märke till att nämnda universitet tagit första platsen detta ämne i Shanghai-rankingen (ARWU). I en annan ranking, QS, så erhöll Paris Saclay en mer modest placering, nämligen plats 86, vilket innebär att man hamnade efter exempelvis KTH och fem andra franska universitet,

När man tittar på Shanghai-ranking-

Hanne Kjöllers kritiserar utfallet av svenska forskningspengar.



FOTO: BENGT OBERGER/WIKIPEDIA

Tabell 1: Total ranking av svenska lärosäten enligt fyra rankinglistor:

- CWUR: Center for World University Rankings
- THE: World University Ranking från Times Higher Education
- ARWU: Academic Ranking of World Universities ("Shanghai ranking")
- QS: Quacquarell Symonds

	CWUR 2021	THE 2022	ARWU 2021	QS 2022
KI (Karolinska Institutet)	35	39	42	
UU (Uppsala universitet)	88	131	78	124
KTH (Kungliga Tekniska Högskolan)	140	201-250	201-300	89
LU (Lunds universitet)	141	116	151-200	87
SU (Stockholms universitet)	164	176	74	148
GU (Göteborgs universitet)	173	185	101-150	180
UmU (Umeå universitet)	324	351-400	401-500	365
LiU (Linköpings universitet)	362	251-300	301-400	329
CTH (Chalmers Tekniska Högskola)	400	250-300	401-500	121
SLU (Sveriges lantbruks- universitet)	404	301-350	201-300	
HHS (Handelshögskolan i Stockholm)	432		401-500	
ÖrU (Örebro universitet)	764	401-500	601-700	
Luleå Tekniska Universitet	1134			
Linné- universitetet	1278		801-900	
Mittuniversitetet	1781			
Karlstads universitet	1811	801-1000		
Malmö universitet	1840			
Jönköpings universitet	1886			

en för statistik så kan man notera att fyra svenska lärosäten nu hamnat på listan, att jämföra med två stycken för några år sedan. Framförallt så har KI nu hamnat bland de 100 främsta i världen, vilket är intressant om man betänker att de ännu inte fått till något masterprogram i biostatistik.

Det kan vara intressant att notera att Sverige enligt QS-rankningen intagit första plats inom ett tämligen viktigt ämnesområde nämligen odontologi (åtminstone uppfattar jag det som viktigt, då min mor arbetade närmare femtio år som tandsköterska). Göteborg intog första platsen, KI plats tio och Malmö plats 24. Orsaken till dessa förnämliga positioner kan vara nog så intressant för Statistikfrämjandets medlemmar. Dekanus Peter Lingström säger nämligen:

Svensk odontologisk forskning har en stor fördel mot andra länder. Här kommer barn och unga regelbundet till tandvården, och de flesta fortsätter även med regelbundna tandvårdsbesök som vuxna. Data samlas in och registreras, vilket ger oss unika förutsättningar att genomföra studier och få svar på frågor ur ett befolkningsperspektiv.

Naturligtvis räcker det inte med bra data för att åstadkomma forskning av världsklass, men det verkar som om svenska odontologer ändå har en fördel framför duktiga utländska kollegor. Det bör tilläggas att enligt ARWU-rankningen så intog de svenska lärosätena något blygsammare, men ändå

**Tabell 2:** Ranking av svenska lärosäten med avseende på matematik och statistik.

	QS Stat&OA 2021	ARWU Stat 2021	QS Mat 2021	ARWU Mat 2021	THE Mat&Stat 2022
KTH	51-100		49	32	151-175
CTH	101-150	151-200	151-200	101-150	251-300
LU	101-150		201-250	201-300	78
SU	101-150	151-200	251-300	401-500	83
UU	151-200	151-200	151-200	101-150	126-150
GU				101-150	301-400
KI		76-100			
LiU			301-350	301-400	201-250
UmU				501-600	301-400
KaU					601-800
ÖrU					601-800

Referenser

Alberg, H., (2019). Ranking av universitet. *Qvintensen* (2019:1)

Göteborgs Universitet. Topprankning för odontologin i Göteborg
<https://www.gu.se/nyheter/topprankning-for-odontologin-i-goteborg>

Hammarfelt, B. (2020). Välkommen granskning av forskningens villkor. (Recension av Hanne Kjöllers bok "Kris i forskningsfrågan, eller: Vad FAN får vi för pengarna?") *Respons*, (6), 67-69.

Kjöllers, H., (2020) Kris i forskningsfrågan,

eller vad FAN får vi för pengarna?, *Fri tanke*.

Kungliga Vetenskapsakademien: När forskningsstatsningen blev regionalpolitik. (Debattartikel publicerad i Svenska Dagbladet 2021-05)

https://s3.eu-de.cloud-object-storage.appdomain.cloud/kva-image-pdf/2021/06/210525_debatt_SvD_forskningsstatsning.pdf

Wretman, J. (2021). Tankar kring en rankinglista. *Qvintensen* (2021:1)

hedrande positioner: GU plats 26, och KI och Malmö inom spannet 51–75.

Professor Martin Neovius är inne på liknande tankegångar vad gäller ”public health”. I detta kompetensområde ingår bl.a. epidemiologi, ett ämne som livligt diskuterats det senaste året. Där ligger KI, enligt ARWU, på plats 7 (efter tre amerikanska och tre brittiska universitet). Fyra svenska lärosäten finns bland de 100 främsta (KI, LU, UU och UmU). Detta kan jämföras med två tyska och ett franskt men hela fem nederländska. Neovius nämner tillgången till registerdata som en viktig framgångsfaktor för Sverige inom området. En synpunkt som Statistikfrämjandet kanske borde trycka hårdare på.

Kjöller nämner i sin bok åtskilliga gånger Nederländerna, Danmark och Schweiz som exempel på relativt små länder som visat sig vara framgångsrika som forskarnationer. Om man betänker dels hennes påpekande

om vikten av statistikkunskaper, dels Karl Pearsons uttalande att ”statistics is the grammar of science”, så kan det vara intressant att studera hur dessa länder står sig i rankningen i ämnet statistik. Man kan från Tabell 3 konstatera att förutom det faktum att USA och andra engelskspråkiga länder dominerar, så har Nederländerna och Schweiz flera universitet som blivit högt rankade enligt ARWU. ETH i Zürich är rankat som nummer fyra i världen enligt QS, efter de amerikanska prestigeuniversiteten MIT, Harvard och Stanford inom statistik och operationsanalys. Enligt ARWU så hamnar ETH på sjuttonde plats inom ämnet statistik. Köpenhamns universitet får också en hygglig placering. Man kan nog konstatera att de länder som Kjöller nämner som framstående forskarnationer verkar väl rustade vad gäller forskning och utbildning i statistik.

HANS ALBERG

Tabell 3: Topprankade lärosäten inom statistik, statistik&OA och matematik&statistik, fördelade på land.

	Statistik, ARWU 2021	Statistik&OA, QS 2021	Matematik&Statistik, THE 2022
USA	47	33	35
UK	11	12	10
Kina	8	11	7
Australien	5	7	5
Canada	5	6	3
Frankrike	5	2	6
Nederländerna	4	7	5
Schweiz	4	1	4
Tyskland	2	5	10
Belgien	1	3	1
Danmark	1	1	1
Israel	1	0	0
Saudi Arabien	1	1	1
Singapore	1	2	2
Sydkorea	1	2	0
Spanien	1	0	0
Sverige	1	1	2
Finland	1	0	0
Italien	0	3	2
Japan	0	2	2
Nya Zeeland	0	1	0
Hong Kong	0	0	2
Ryssland	0	0	2

Simuleringsstudie om rankningar

Qvintensen 2021:1 innehåller en artikel av Carl-Erik Särndal med titeln ”Om rankningar”. Han uppehåller sig speciellt vid Svenskt Näringslivs (SvN:s) årliga (sedan 20 år) rankning av Sveriges 290 kommuner med avseende på företagsklimat. Rangordningen är där baserad på 17 olika numeriska ”faktorer” för varje kommun. Utifrån en utförlig rapport av Sixten Lundström om SvN:s rankning tolkar Särndal (på sid. 10) förfarandet så här:

För var och en av de 17 faktorerna rankas sedan kommunerna från 1 till 290. Den definitiva rankningen utförs på en poäng beräknad, tycks det, som en lineärt viktad kombination av de 17 rankingstalen för kommunen, med ett ”obegripligt” viktsystem. Det dubbelbottnade slutresultatet är en ”rankning baserad på rankningar”.

En fråga man kan ställa sig är om slutresultatet skulle bli annorlunda ifall man för varje kommun först beräknar ett genomsnittligt faktorvärde, och därefter rangordnar kommunerna efter dessa värden.

Jag redovisar här en simuleringsstudie under mycket förenklade förutsättningar för att undersöka hur mycket metoderna skiljer sig åt. Med inspiration från undersökningen tänker vi oss 290 observationer (motsvarar kommunerna) och 17 variabler (motsvarar faktorerna), men nu är det inte längre fråga om Sveriges kommuner och SvN:s faktorer.

I min studie är det 290 oberoende och likafördelade observationer som vardera består av 17 oberoende och likafördelade normalfördelade stokastiska variabler. Detta är långt ifrån förutsättningarna för de verkliga kommunerna, men det kan belysa om det är någon skillnad mellan de två metoderna att rangordna observationer.

Metod 1: Bilda 17 separata rangordningar av de 290 observationerna, en för varje variabel. Beräkna sedan för varje observation ett medelvärde av de 17 variablernas rangtal. Rangordna slutligen de 290 observationerna efter storleken på dessa rangmedelvärden. För att få en unik rangordning i den här studien, så randomiseras ordningen i de fall där två (eller kanske flera) observationer råkar få samma rangtal.

Metod 2: Bilda för varje observation medelvärdet av de 17 variablerna. Rangordna sedan de 290 observationerna efter storleken på dessa medelvärden.

Ingen av de två metoderna ovan överensstämmer med det som SvN gör. Även om Metod 1 är den som mest liknar SvN:s metod, så är skillnaderna mycket stora. SvN använder en (icke förklarad) viktning av de 17 rankingstal som är erhållna för en kommun. Dessa rankingstal är långt ifrån normalfördelade, utan de är osäkra skattningar med stort urvals- och bortfallsfel inom varje kommun. Det finns givetvis också en faktisk skillnad mellan kommunerna som inte finns mellan observationerna i den här simuleringen.

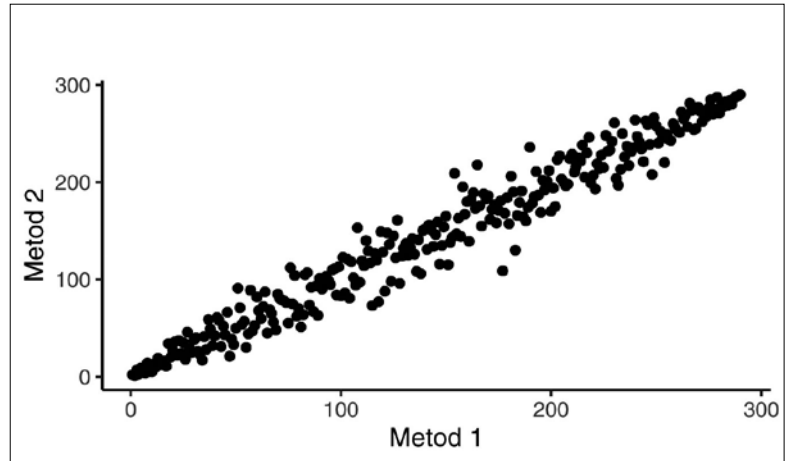
Efter att ha genererat slumpstal från 290×17 likafördelade normalfördelade stokastiska variabler utförs en rangordning enligt Metod 1 och Metod 2. Detta ger 290 stycken (x_1, x_2) -värden, där x_1 = rangtalet enligt Metod 1, och x_2 = rangtalet enligt Metod 2. Korrelationen mellan x_1 - och x_2 -värdena visar sig, inte oväntat, ligga nära 1, nämligen 0,979. Figur 1 visar resultatet i en (x_1, x_2) -plot.

Trots att korrelationen är nära 1 kan skillnaden i rangplats för en enskild observation, alltså $|x_2 - x_1|$, bli ganska stor. I denna första simulering är det största observerade värdet av $|x_2 - x_1|$ lika med 68.

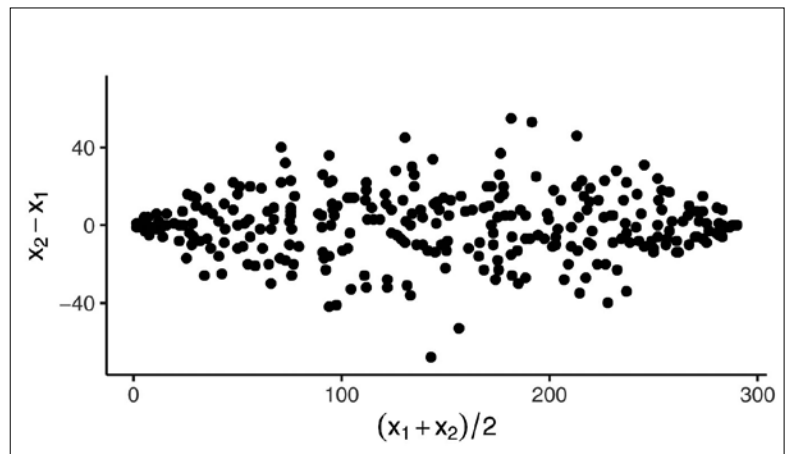
Värdet på $x_2 - x_1$ för samtliga 290 observationer visas i Figur 2, där varje observerat värde på $x_2 - x_1$ har plottats mot $(x_1 + x_2)/2$.

För att närmare studera hur stor skillnaden i rangplats med de två rangordningsmetoderna kan bli för en och samma observation, så gör jag sedan 1000 upprepningar av den ovan beskrivna simuleringemetoden, och $\max(|x_2 - x_1|)$ i varje simulering illustreras i Figur 3. Det största observerade värdet på $|x_2 - x_1|$ är 128, medelvärdet av $\max(|x_2 - x_1|)$ för alla simuleringarna blir 76 och det minsta observerade värdet på $\max(|x_2 - x_1|)$ i en simulering är 45, vilket är förvånansvärt högt.

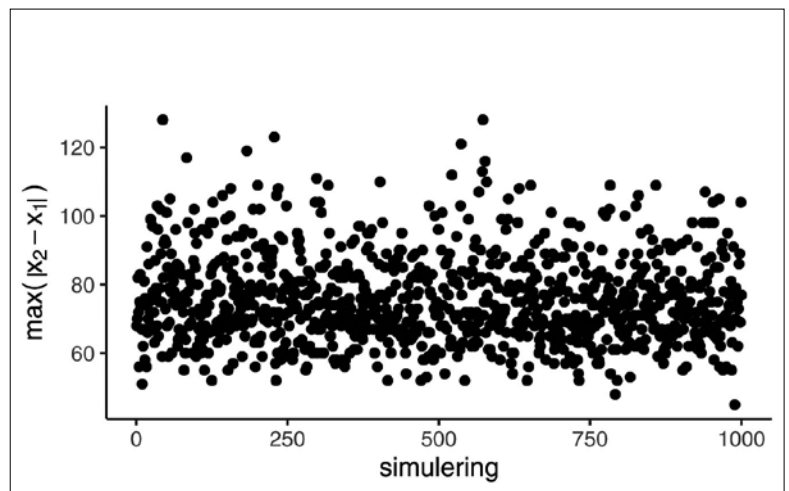
Kan man dra några praktiska slutsatser från en sådan här simuleringstudie? Vad gäller undersökningen och jämförelsen med SvN:s årliga ranking blir svaret "nej", eftersom studien är gjord under helt andra förutsättningar. Men simuleringemetodiken kanske kan vara av intresse för någon matematisk statistiker och studien visar också ett exempel på hur man kan påvisa att två metoder inte är så likvärdiga som man först tror.



Figur 1. Jämförelse mellan metodernas ranger för den första simuleringen.



Figur 2. Skillnaden mellan metodernas ranger för den första simuleringen.



Figur 3. Maximala skillnaden i ranger för respektive simulering.

Vad är dataanalys?

Namnet *John W. Tukey* är förknippat bland annat med så kallade boxplots (eller lådagram). För egen del väcker det minnen från tiden strax innan jag disputerade i matematisk statistik vid Uppsala universitet 2012. Jag kom då i kontakt med hans artikel *The Philosophy of Multiple Comparisons* (Tukey 1991). Även på senare tid har jag blivit upplyft av att läsa andras tankar och formuleringar i skrift, ibland från helt andra epoker, men som tangerar frågor som jag själv har funnit anledning att fundera över. I skrivande stund, då jag som statistiker på Skolverkets analysavdelning arbetar med en registerstudie om elevers övergång från grundskolan till gymnasiet, får jag för mig att jag kanske borde bilda mig en uppfattning även om Tukeys artikel *The Future of Data Analysis* (Tukey 1962).

Statistisk slutledning (inferens)

De flesta som läser en grundkurs i statistik stöter på begreppet statistisk inferens. Alm & Britton, i sin lärobok *Stokastik* (2008), introducerar begreppet med följande ord: ”vi har observerat utfall i ett slumpförsök och vill försöka uttala oss

om den underliggande sannolikhetsmodellen”. Det är en bred, tämligen teknisk, matematisk beskrivning, som inte direkt uttalar sig om användning och tillämpning, utan snarare om en uppsättning verktyg, och möjligen angreppssätt för framtida verktyg. Intressant i sammanhanget är att statistisk inferens ofta brukar avgränsas gentemot dataanalys. Dataanalys avser i sådana fall metoder och begrepp för att sammanfatta och beskriva tillgängliga statistiska material, med såväl numeriska mått som grafiska framställningar. I inledningen till Lehmann & Casellas bok *Theory of Point Estimation* (1998) görs tillägget att dataanalys genomförs ”essentially without extraneous assumptions”. Varför denna begränsning, och i vilken kontext ska den förstås?

Modellbaserad inferens

Först mot slutet av sin bok *Principles of Statistical Inference* (2006) nämner Cox randomiserade experiment och teorin för sannolikhetsurval från ändliga populationer. Utgångspunkten i de tillämpningar Cox har i åtanke är inte i första hand strikt kontrollerade studier, utan istället situationer där det antas förekomma, kanske i något idealiserad form, en bakomliggande, ej direkt observerbar, datagenererande mekanism. Angreppssättet, och det detektivarbete som det kan innebära för den statistiker som känner sig hågad, brukar kallas för modellbaserad inferens. Cox själv är inte minst känd för artikeln *Regression Models and Life Tables* (1972), i vilken den numera berömda Cox's Proportional Hazard Model introducerades.

Modellera eller inte?

Titeln på Särndal, Swensson & Wretmans bok *Model Assisted Survey Sampling* (1992) antyder att frågan om användningen och nyttan av statistiska modeller inte alltid har varit så enkel att

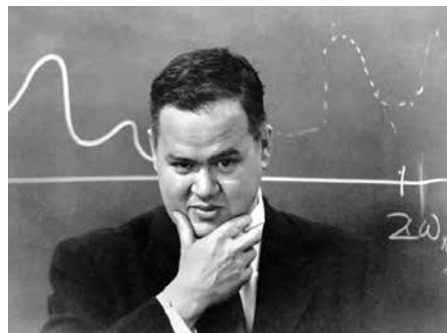
besvara. En bit in i boken förs ett resonemang om väntevärdesriktighet och variansreduktion med den föreslagna regressionsestimern, som författarna vill kalla modellassisterad. Exempelvis kan approximativ väntevärdesriktighet antas gälla, även om valda modellansatser inte har så mycket att tillföra. Från ett helt annat perspektiv skriver Tukey (1962): ”*Many seem to find it essential to begin with a probability model containing a parameter, and to ask for a good estimate for this parameter*”. Jag tror att man i just det sammanhanget kan ana liknelsen med hemmafisaren som använder hammaren i allt han tar sig för, eller villaägaren som söker ett avancerat multiverktyg som kan användas till alla trädgårdsarbeten.

Men vad är då dataanalys?

På sätt och vis har väl ett svar redan givits här tidigare; dataanalys avser metoder och begrepp för att sammanfatta och beskriva tillgängliga statistiska material, med såväl numeriska mått som grafiska framställningar. Om vi nu hoppar rakt på min egen situation så är jag inte helt främmande för att tillämpa logistisk regression för att beskriva elevers förutsättningar att klara av sina gymnasiestudier, utifrån de nationella register över skolungdomars studieval och betyg som jag har att tillgå. Men samtidigt tvekar jag över om det verkligen behövs; om det passar det aktuella syftet; medveten om att det också kräver att jag funderar över hur väl dessa modellstrukturer fångar det aktuella datamaterialet. En del av dessa överväganden leder säkert till ytterligare dataanalys och frågor till datamaterialet.

Dataanalys, inferens, eller både och?

Låt oss anta att jag tillämpar logistisk regression för att beskriva elevers förutsättningar att klara av sina gymnasiestudier. Har jag därmed tagit steget från dataanalys till inferens? Frågan före-



John Wilder Tukey (1915–2000). Med glimten i ögat.

faller mig inte helt självklar att besvara. Kanske menar vissa att svaret är ja, utifrån valet av analysinstrument. Andra kanske istället förknippar inferens med specifika forskningsprocesser, samt hur man förhåller sig till de resultat som framkommer. Ytterligare andra kanske menar att jag ägnar mig åt inferens oavsett vilket val jag gör, så länge jag har för avsikt att göra uttalanden om den aktuella populationen. Åter till artikeln i fokus, (Tukey 1962), så kan jag inte låta bli att fundera över de inledande orden: "For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched how mathematical statistics is growing, I have cause to wonder and to doubt."

Ett annat exempel

Walter Radermacher utgår i sin nyligen utkomna bok *Official Statistics 4.0* (2020) från digitalisering och globalisering som paraplybegrepp för pågående förändringar i vår samtid. Ett konkret exempel, som ofta brukar nämnas gällande den officiella statistiken, är den successivt ökande användningen av försäljningsuppgifter (transaktionsdata) vid beräkning av konsumentprisindex. Dessa förändringar till trots så behöver statistikbyråerna dock själva fortsatt stå för den enhetliga klassificeringen av varor och tjänster. Frågan är dels hur kodningen ska gå till, men jag tänker framför allt på utmaningen i att utvärdera och övervaka kodningsprocessernas tillförlitlighet. Skillnaderna är många jämfört med det förra exemplet, om gymnasieungdomars studieförlopp, men båda fallen väcker funderingar kring hur man ställer relevanta frågor till stora datamaterial.

Avslutningsvis

Denna text bygger på ett subjektivt och spretigt urval av referenser. Flera av de texter jag valt att

lyfta fram har åtskilliga år på nacken, eller så har de starka rötter bakåt i historien, vilket jag som sagt tycker har varit en del del av det roliga. Det finns också intressanta och delvis annorlunda böcker som jag hade velat lyfta fram men inte riktigt fått plats med, såsom Göran Anderssons *Lögn och sanning med statistik* (1982).

FREDRIK JONSSON,
SKOLVERKETS ANALYSAVDELNING

Referenser

- Alm, S. och Britton, T. (2008). *Stokastik*. Liber.
- Andersson, G. (1982). *Lögn och sanning med statistik*. Liber.
- Cox, D. (2006). *Principles of Statistical Inference*. Cambridge University Press.
- Lehmann, E. och Casella, G. (1998) *Theory of Point Estimation, second edition*. Springer.
- Radermacher, W. (2020). *Official Statistics 4.0*. Springer.
- Särndal, C., Swensson, B. och Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer.
- Tukey, J. (1962). *The future of data analysis*. Ann. Math. Statist., 33: 1-67.
- Tukey, J. (1991). *The philosophy of multiple comparisons*. Statist. Sci., 6: 100-116.



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 sekreterare@statistikframjandet.se.

FENStatS aktiviteter inom covid-19

Som framgått av tidigare Qvintensen så startade FENStatS 2020 en arbetsgrupp för covid-19. Det bestämdes på ett tidigt stadium att man skall ta fram en kurs betitlad "Data-informed decision making in a pandemic". Därigenom hoppas man få ett större genomslag än genom att bara leverera ett dokument.

Gruppens ledare Katharina Schueller har fått till stånd ett samarbete med KI-Campus (ett tyskt kompetenscenter för artificiell intelligens). KI-Campus finansierar redan nu några studenter som arbetar extra på hennes företag STAT-UP, och man kommer dessutom att överta marknadsföring och förvaltning av kursen. Studenterna får nu stöd av en grupp seniora personer, förutom Katharina Schueller:

- Walter Radermacher, ordförande för FENStatS, tidigare chef för Eurostat. Har under senare tid skrivit en del inlägg om vikten av officiell statistik.
- Marleen De Smedt, har varit ansvarig för hälsorelaterad statistik på Eurostat, samt rådgivare till Walter Radermacher.
- Fulvia Mecatti, professor i statistik i Milano. Har bl.a. varit editor för en bok om sampling, vilket torde borga för att statistikspekterna inte tappas bort i materialet.
- Lura Lasa, associate professor i biostatistik i Ljubljana.

Enligt tidsplanen skall kursen vara klar för att testas på användare vid årsskiftet 2021/22 och färdigttestad i slutet av januari 2022. Därefter överlämnas materialet till KI-Campus. Kursen skall vara på engelska och omfatta 16 timmar. Kursen vänder sig till beslutsfattare, journalister etc. Dessa förutsätts ha behörighet för högskolestudier, men inte nödvändigtvis postgymnasiala kunskaper i matematik och statistik.

Kursen baseras på erfarenheter från covid-19-pandemin, men mycket av materialet skall kunna tillämpas på andra situationer där statistiskt underlag erfordras för att fatta beslut. Exempel på aspekter som skall tas upp enligt projektspecifikationen är:

- Allmänna statistiska indikatorer
- Indikatorer för att förstå infektionsförlopp och vaccinationsprogress
- Orientering om vad officiell statistik innebär
- Optionellt: R-värde, sensitivitet, specificitet
- Statistisk osäkerhet, konfidens- och prognosintervall
- Styrkan hos covid-indikatorerna
- Effektivitet hos vaccinationer
- Effekten av "lock-down strategies"
- Medvetenhet om "publication bias"

Behovet av en sådan kurs framgår av följande exempel. På Nederländernas regerings hemsida finns en intressant rapport som visar jämförelse mellan tre olika metoder att uppskatta antal avlidna i covid-19:

- Direkt rapportering till GGD (Gemeentelijke gezondheidsdienst, lokala hälsovårdsmyndigheter)
- Överdödlighet enligt CBS (motsvarande SCB)
- CBS dödsorsaksregister

Fram till slutet av februari 2021 har följande antal rapporterats med dessa tre metoder: 15 818, 18 634 respektive 27 056. Det förstnämnda värdet är det man hittar om man går in på exempelvis worldometer, medan det sistnämnda värdet förmodligen ligger närmast sanningen.

I Sverige har Socialstyrelsen 2021 gjort en liknande analys, där man jämför sin egen rapportering med Folkhälsomyndighetens. Resulta-

tet sammanfattas så här: "Socialstyrelsens analys visar att antal rapporterade dödsfall för covid-19 enligt SmiNet och dödsorsaksintygen stämmer väl överens överlag. Antalet rapporterade fall under april var något fler enligt dödsorsaksintygen, medan antal rapporterade fall under december var fler enligt SmiNet".

Det skulle vara intressant att se om liknande analyser gjorts i andra länder, och vilka slutsatser man dragit.

Förutom kursmaterialet har arbetet med FENStatS hemsida fortskridit. Tanken är att hemsidan skall ses över med jämna mellanrum för att exempelvis undersöka om länkarna är aktuella.

HANS ALBERG

Referenser

FENStatS Covid-19 sida. <https://www.fenstats.eu/guides-resources>

Mecatti, F., Conti, P. L., & Ranalli, M. G. (Eds.). (2014). *Contributions to sampling statistics*. Springer.

Nederländernas regering (2021) "What is the COVID-19 mortality rate based on cause of death?" <https://corona-dashboard.government.nl/artikelen/wat-is-het-covid-19-sterftecijfer-op-basis-van-doodsoorzaak>

Socialstyrelsen (2021). Jämförelse av Socialstyrelsens och Folkhälsomyndighetens statistik över avlidna i covid-19 under april och december. (Dnr: 6.7-7387/2021) <https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/dokument-webb/statistik/datakallor-avlidna-covid19-210225.pdf>

ORDFÖRANDE HAR ORDET

2022 kommer inte bli lugnt...

När det gäller föreningsaktiviteter har vi haft ett förhållandevis mycket lugnt år, och vi har (som många andra) varit märkbart drabbade av pandemin. Men styrelsearbetet har fortsatt i normalt tempo med digitala möten, översyn av hemsidan, planering av kommande aktiviteter o.s.v. En av de mest centrala delarna av föreningens verksamhet har i år varit det fortsatta arbetet med vår kortfattade statistiska ordlista – se gärna ordlistan på vår hemsida.

I september bjöd vi, i samarbete med Statistikfrämjandet, på ett mycket välbesökt webinarium med titeln "På spaning om research – Framtiden för undersökningsbranschen". Talare

var tre av Sveriges mest framstående visionärer som såg om framtiden för oss som producerar och använder oss av statistik som beslutsunderlag. Cirka 150 medlemmar i Statistikfrämjandet och/eller Surveyföreningen lyssnade till:

Lilli Japac (PhD, Senior Scientific Advisor, SCB) som talade om det nya surveylandskapet – utmaningar och möjligheter.

Karin Nelsson (Vd Demoskop) som presenterade nya förutsättningar för undersökningsbranschen.

Annie Lidesjö (Operativ chef på NTM) som berättade om värdet av data och analys i det operativa arbetet.



Nu ligger vi i startgroparna för att få till ett riktigt bra program till "Årets kvalitetsseminarium", som preliminärt är planerat till februari 2022. Samtidigt är det tid för att börja tänka på aktiviteter med anknytning till att vi under nästa år har ett valår med minst sagt högt tryck när det gäller opinionsundersökningar. Om det ur surveystatistisk synvinkel har varit relativt lugnt under förra och innevarande år så kan vi redan nu garantera att det inte kommer att vara det under 2022.

ÅKE WISSING

FÖRENINGEN FÖR MEDICINSK STATISTIK

ORDFÖRANDE HAR ORDET

Stort behov av fördjupad utbildning av biostatistik

Inom 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).

Under 2021 har covid-19 fortsatt att prägla arbetslivet och vardagslivet för de flesta. För andra gången genomfördes årsmötet den 25:e mars som ett elektroniskt årsmöte. På årsmötets vetenskapliga program presenterade Linda Vidman, Umeå Universitet, Peter Ström och Sara Ekberg (bägge KI) delar av sina avhandlingsarbeten. Måns Magnusson höll ett föredrag om maskininlärning för tillämpade statistiker. Den föregående styrelsen beviljades ansvarsfrihet och en ny styrelse valdes. Ingeborg Waernbaum,

Per Liv, Therese Andersson, Andreas Gustavsson och Ida Hed Myrberg valdes om. Även Anna Torrång och José Sanchez blev omvalda som representanter för EFSPI. Vi avtackade styrelsemedlemmarna Fredrik Norström och Henrik Renlund för deras insatser i FMS. Nya representanter i styrelsen är Stefan Franzén som sekreterare och Lovisa Björnberg som ledamot. Vi delade ut det nyinstituerade priset för bästa uppsats: Föreningen för Medicinsk Statistiks Uppsatspris, 5000 kr. Pristagaren 2021, Enoch Chen, Karolinska institutet, fick priset för sin utmärkta uppsats "Extrapolating cancer patient survival: a comparison of the flexible parametric model and the rolling-over algorithm".

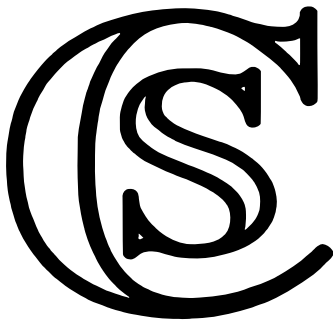
För andra året i rad pausade vi utlysningen av vårt stipendium som brukar kunna sökas för att t.ex. finansiera konferensresor. Som det

ser ut nu tror vi att nästa års utlysning kommer att bli av. Vanligtvis lyser vi ut stipendiet under vårterminen med deadline i juni. Håll uppsikt under våren!

I styrelsen och visionsgruppen fortsätter vi att arbeta med att främja utbildning inom biostatistik. Vi vet att det finns en stor efterfrågan på biostatistik och ett behov av fördjupad utbildning av biostatistik. Därför vill vi uppmuntra både till ett ökat utbud av statistikkurser med biostatistisk profil på universitet och högskolor, men även en masterutbildning i biostatistik tror vi skulle vara betydelsefull.

INGEBORG WAERNBAUM





ORDFÖRANDEN HAR ORDET

En serie virtuella kollokvier

Cramérsällskapet har inlett en serie virtuella kollokvier med tanken att låta svensk-baserade forskare inom statistik och sannolikhetsteori berätta om sin forskning eller forskningsområde för en bredare publik.

Först ut var professor Yudi Pawitan (KI) som talade under titeln "Epistemic condence, the Dutch Book and relevant subsets" den 4 februari 2021. Näst i tur var Henrike Häbel (också KI) som talade den 26/5 på temat: "Image-assisted survival analysis of stage III melanoma". Under höstterminen inledde vi med Måns Magnusson (Uppsala) den 29/9 under titeln "Statistical analysis of large textual data". När detta skrivs planeras också ett föredrag av Jeff Steif (Chalmers & GU). Se även Cramérsällskapets hemsida.

När detta skrivs planeras dessutom årets höstmöte med temat "Framtidens undervisningsmetoder i statistiska ämnen med ett års digitalisering i bakfickan". Mötet avses bli fysiskt och äga rum 26-27 oktober i Linköping (med möjlighet att lyssna från distans). Se även Cramérsällskapets hemsida.

På årsmötet valdes följande styrelse för det kommande året:

Tom Britton (SU), ordförande
Annika Lang (Chalmers/GU), sekreterare
Anders Lundquist (UmU), kassör
Behnaz Pirzamanbein (LU)
Yuli Liang (ÖrU)
Martin Singull, (LiU)

TOM BRITTON

Sebastian Rosengren fick 2020 års Cramérpris

Cramérprisför år 2020 gick till Sebastian Rosengren som disputerade 2020 med avhandlingen "Random graph and growth models" i ämnet matematisk statistik vid Stockholms universitet under handledning av Mia Dejifjen. Motiveringen till priset lyder:

"I avhandlingen ger Sebastian Rosengren prov på eleganta lösningar för olika problem inom diskret sannolikhetsteori. Arbetena visar på stor självständighet och i den sista artikeln byter han inriktning mot maskininlärning och kopplar detta på ett elegant sätt till perkolationsproblem."

Priset utgörs av ett diplom och ett pris på 10 000 kr.

Med anledning av Cramérpriset arrangerades en minikonferens samma dag som årsmötet, den 25/3 2021. Konferensens tema var densamma som avhandlingen, nämligen Random graph and growth models. Konferensen samlade dryga 40 åhörare och innehöll föredrag av följande talare:

Sebastian Rosengren (SU), Lasse Leskelä (Aalto), Tatyana Turova (LU), Victor Falgas-Ravry (UmU) och Daniel Ahlberg (SU).

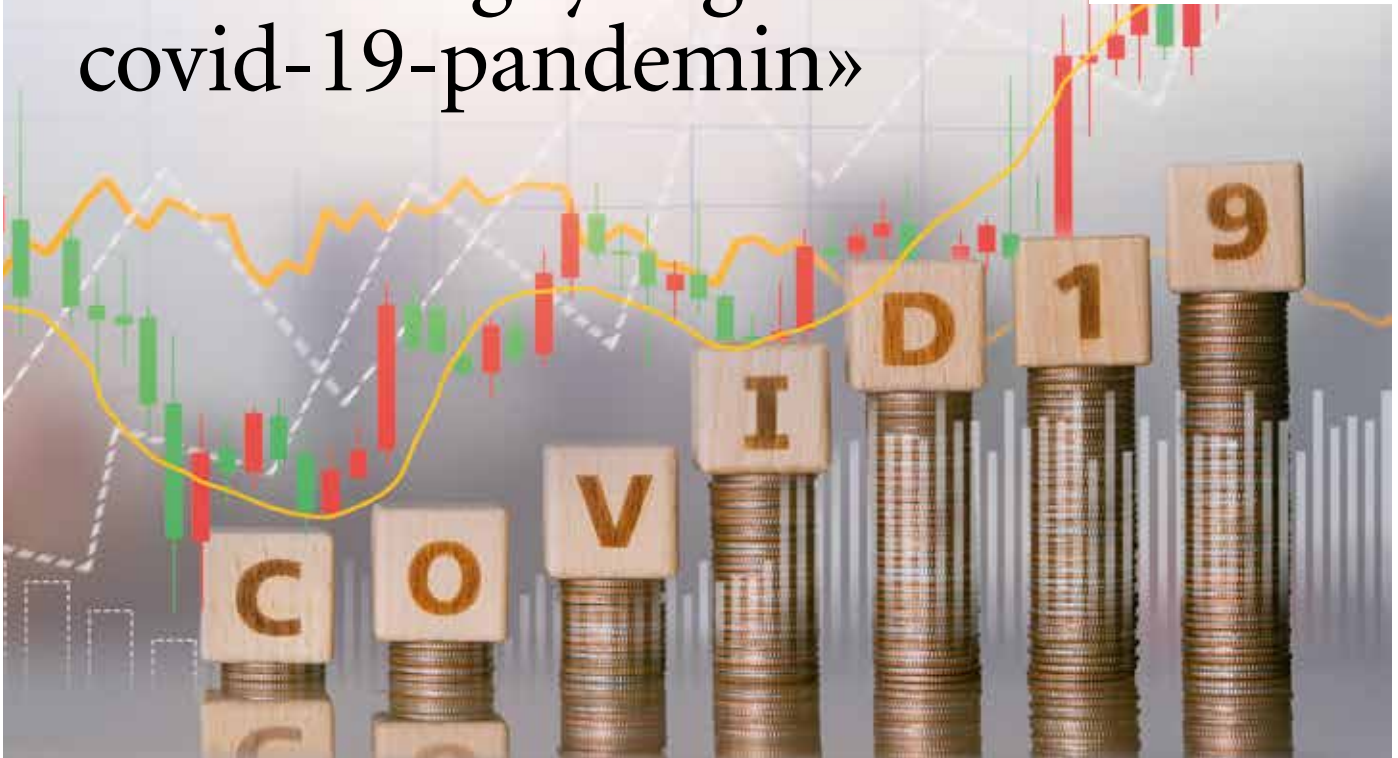
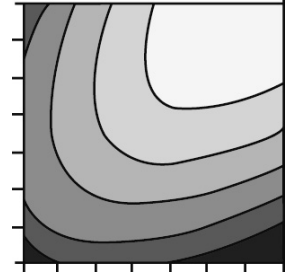


Sebastian Rosengren som disputerade 2020 med avhandlingen "Random graph and growth models" fick det årets Cramérpris.

TOM BRITTON

ORDFÖRANDEN HAR ORDET

»Behovet av god statistik har visat sig tydligt under covid-19-pandemin»



Det har nu gått sex år sedan jag åtog mig att representera Föreningen för Industriell Statistik i styrelsen för Svenska Statistikfrämjandet. Vad har jag lärt mig under dessa år? Förhoppningsvis en hel del. Under mina yrkesverksamma år arbetade jag egentligen aldrig med statistik och sannolikhetslära, men såg många möjliga tillämpningar.

Behovet av statistik och, framför allt, statistik av god kvalitet har visat sig tydligt under covid-19-pandemin. Ytterligare en tankeställare om nyttan av statistik fick jag när jag som återmatning på en lista över internationella statistiska organisationer fick tips om ytterligare en organisation:

International Biometric Society (IBS).

Detta påminde mig om att det finns ett antal discipliner i gränstrakterna mellan statistik och andra vetenskaper såsom:

- Biometrics (931 000)
- Econometrics (999 000)
- Pharmacometrics (26 300)
- Chemometrics (245 000)
- Psychometrics (181 000)

- Bibliometrics (132 000)
- Jurimetrics (16 600)
- Geostatistics (158 000)
- Operations research (2 400 000)
- Astrostatistics (1 690)

Siffrorna i parentes anger antal träffar med Google.scholar. Det vore kanske en idé att få in artiklar om dessa discipliner i kommande nummer av Qvintensen. Det kan kanske vara intressant att notera att astrostatistik framställs som ett nytt forskningsområde (se exempelvis Hilbe, J. M. (2014). *Astrostatistics*. Wiley *Stats-Ref: Statistics Reference Online*, 1–5.), samtidigt som vi vet att astronomin varit drivande vad gäller utveckling av statistik och sannolikhetslära.

Tack vare mitt medlemskap i Statistikfrämjandet fick jag också en del intressanta kontakter inom FENStatS. Dess ordförande Walter Radermacher har formulerat en del frågor, vilka enligt min uppfattning förtjänar en djupare diskussion, varav en del också är relevanta för industrisidan:

- Hur bör man förstärka den officiella statistikens infrastruktur för att befrämja en framåtblickande statistikutveckling?

- Hur åstadkommer man ett intensivt samarbete mellan akademiska metodutvecklare och praktiska användare?
- Hur åstadkommer vi aktivt synergier mellan statistik, datavetenskap och AI?
- Vilken digitalpolitik behövs för att stärka den officiella statistiken?
- Varför skall statistiker involveras tidigare och mer regelbundet i politiska expertpaneler?
- Hur kan man göra mer för statistisk bildning och kultur?

Jag tror att en del av de ”gränsdiscipliner” som nämnts tidigare kommit till för att användarna saknat metodstöd från ”grunddisciplinen”.

Samarbetet på Europa-nivå har varit mycket givande, men mitt engagemang har hittills främst gällt FENStatS. Industriell Statistik ingår i ENBIS och vi borde engagera oss mer även i detta samarbete, speciellt som Sören Knuts nyligen valts till vice ordförande. För att orka driva detta samarbete behöver dock fler engagera sig. Hör gärna av er om ni har idéer.

HANS ALBERG

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 sekreterare@statistikframjandet.se.

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.

