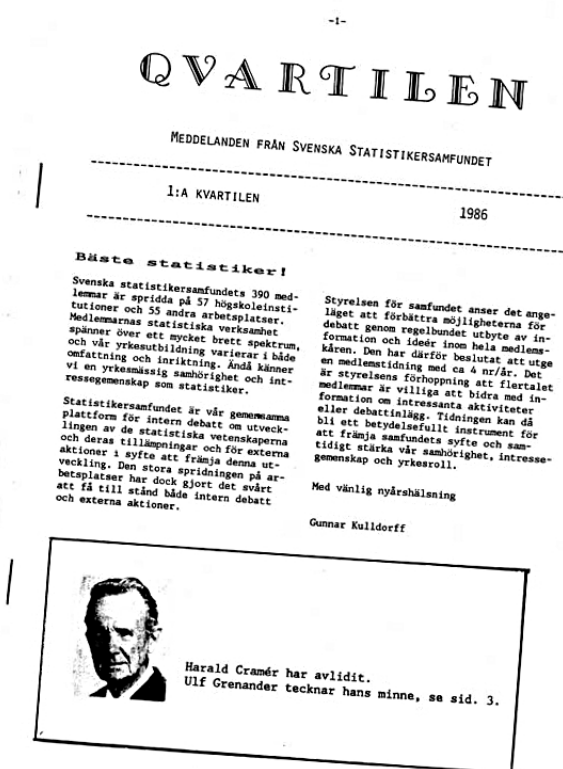


QVINTENSEN

Svenska statistikersamfundet och Statistiska föreningen

Produktionskostnad: 35 kronor

Årgång 1 • Nummer 1 • Februari 2008



Innehåll

Ordföranden har ordet	3	Mikrosimulering	16
Redaktören har ordet	3	Teoretiska artiklar	20
Styrelsen	4	Debatt	22
Surveyföreningen	4	Kåseri om statistik som tillämpning	24
Föreningen för medicinsk statistik	6	Hänt och hört	25
Industriell statistik	7	Livet som statistiker	28
Visualisering av kvantitativ information	9	Rapporter	33
Gästkolumnist	13	Bokanmälan	35
		Konferenskalendarium	36



Redaktör	Joakim Malmelin	Telefon	090 - 786 55 30	E-post	qvartilen@gmail.com
Redaktion	Magnus Arnér		046 - 36 28 70		magnus.arnér@tetrapak.com
	Göran Arnoldsson		090 - 786 52 12		goran.arnoldsson@stat.umu.se
	Anders Holmberg		019 - 17 69 05		anders.holmberg@scb.se
	Marie Linder		075 - 247 32 31		marie.linder@socialstyrelsen.se
	Johan Lyhagen		018 - 471 28 44		johan.lyhagen@dis.uu.se

Ansvarig utgivare: Eva Elvers och Mats Forsberg

Redaktör: Joakim Malmådin

Redaktion: Magnus Arnér (Industriell statistik), Göran Arnoldsson (Samfundet), Anders Holmberg (Surveysektionen), Marie Linder (FMS) och Johan Lyhagen (Samfundet)

Adress: Qvartilens redaktion
c/o Joakim Malmådin
Statistiska institutionen, Umeå universitet
901 87 Umeå
qvintensen@gmail.com

Riktlinjer för material

Figurer/bilder skickas i något av formaten eps eller tif. *Fotografier* önskas i formatet jpg med upplösningen 300 pixlar per tum och med fotografens namn angivet.

Enbart *text* skrivs i ASCII-format. Kursivering anges `\kursiv{kursiv text}`. Om bidraget innehåller annat än text och bild, önskas materialet i $\text{\LaTeX}$ -format. Svenskspråkiga personer skriver på *svenska* och övriga på engelska. Kontakta redaktören för eventuella undantag.

Antalet *referenser* bör begränsas för bästa läsbarhet. Tumregeln är maximalt fyra referenser per artikel.

Alla bidrag granskas av redaktionen och enklare skriv- eller syftningsfel ändras utan kommentarer. Författaren kan komma att kontaktas rörande förslag på ändringar, förtydliganden och omdispositioner. I mån av tid får författaren ett korrektur i form av en pdf-fil.

Annonspriser

	Hel-	Halv-	Kvartssida
MEDLEMMAR	4 000:-	2 000:-	1 000:-
ICKE-MEDLEMMAR	6 000:-	3 000:-	1 500:-

Annonser bokas med Qvartilens redaktör. *E-postutskick* och *adressetiketter* beställs däremot av Samfundets sekreterare. Priset är 4 000 respektive 6 000 kr för institutioner och företag som är medlemmar och 6 000 respektive 9 000 kr för övriga, 25% moms tillkommer. E-postutskick enbart för platsannonser i textformat utan bilagor. Utskicken når cirka 83% av medlemmarna.

Svenska statistikersamfundet

Samfundet har runt 1300 medlemmar verksamma inom industri, förvaltning, universitet och högskola.

Adress: Svenska statistikersamfundet
c/o Lars Rönnegård
Linneus Centrum för Bioinformatik,
Uppsala universitet
751 24 Uppsala
seksamf@gmail.com

Medlemsavgiften är 130 kr per år för enskilda personer och 1 300 kr för institutioner och företag med flera. Arbetsplatser med färre än fyra statistiker betalar halva avgiften. Vid betalning, använd postgironummer 65 41 95 - 7. Via Svenska statistikersamfundets hemsida, www.statistikersamfundet.se, är det möjligt att ansöka om medlemskap och att uppdatera personlig kontaktinformation.

Statistiska föreningen

Författare och andra som utan ersättning bidrar med material för publicering behåller upphovsrätten till sina verk. Svenska statistikersamfundet äger rätten att publicera dessa verk i valfri form och upplaga samt att behålla alla eventuella intäkter från sådan publicering, om inget annat avtalas. Författare och andra som utan ersättning bidrar med material för publicering garanteras rätten att själva få sprida sitt verk under förutsättning att Qvartilen anges som originalkälla med angivande av nummer och år.

Svenska statistikersamfundet äger upphovsrätten till publicerat material då det är framställt av redaktionsmedlem, Samfundets styrelse, någon till Samfundet knuten sektions styrelse eller person som erhållit arvode för sitt bidrag.

© Svenska statistikersamfundet 2007.

Styrelsen

		Telefon	E-post
Ordförande	Eva Elvers	08 - 506 947 14	eva.elvers@scb.se
Vice ordförande	Jesper Rydén	018 - 417 32 88	jesper.ryden@math.uu.se
Sekreterare	Lars Rönnegård	018 - 471 66 83	seksamf@gmail.com
Skattmästare	Annika Lindblom	019 - 17 60 86	annika.lindblom@scb.se
Övriga ledamöter	Michael Carlson	08 - 506 943 36	michael.carlson@scb.se
	Ingemar Sjöström	08 - 553 835 68	ingemar.g.sjostrom@telia.com
	Anders Holmberg	019 - 17 67 83	anders.holmberg@scb.se
	Mats Rudholm	031 - 703 73 72	mats.rudholm@statistikkonsulterna.se

Ordföranden har ordet

En nyutnämnd tidskriftsredaktör uttalade: "Jag vill behålla allt som är bra idag, men ändå ta ett steg framåt". När jag läste det reflekterade jag en stund över ordvalet *men ändå* – vad signalerar det i jämförelse med *och dessutom* ta ett steg framåt?

Information och kommunikation har jag på sistone tänkt mycket på. Vad är det som gör att information når fram och tas emot – alternativt inte gör det? Beror det på informationen, på sändaren och/eller på mottagaren? Att mottagaren är viktig är uppenbart, till exempel på en arbetsplats. Vissa håller sig ständigt uppdaterade med hjälp av intranät, samtal med mera. Andra förväntar sig information serverad på silverbricka. Jag minns det talande uttrycket sedan 1970-talet. Det var långt före dagens stora informationsflöde. Det gäller ännu att individen önskar information anpassad till sin egen situation. Hur skriver man så att det passar alla i en heterogen grupp? Hur vet sändaren vad som har nått fram? Hur ska tystnad tolkas – som samtycke eller som att informationen inte har mottagits?

Statistikersamfundets årsmöte i oktober beslutade om samgående mellan Statistiska föreningen och Svenska statistikersamfundet. Omröstningen på årsmötet och medlemsenkäten i somras visade båda en stark antalsmässig övervikt för samgående. I båda fallen var det en begränsad mängd medlemmar som uttalade sig. I båda fallen angavs det skäl för och skäl emot ett samgående.

Vad har ovanstående tre textstycken med varandra att göra?

Ordval och information: Enstaka ord kan ha stor betydelse för en text. Exemplet ovan är dock odramatiskt. Att jag stannade upp i just det läsandet berodde mest på att jag hade denna ordförandespalt och samgåendet i bakhuvudet.

Information och samgående: på årsmötet för två år sedan fanns en motion om en utredning av en ny sammanhållen förening för statistik och statistiker. Sedan har det fortsatt med utredningar, information under resans gång och förslag vid de två föreningarnas årsmöten. Har medlemmarna tagit till sig informationen? Kommentarer visar att några har gjort det, men långt ifrån alla och allt – varför?

Samgående och ordval: Det finns mycket i de två föreningarnas verksamheter som är uppskattat. Det är förstås angeläget att behålla uppskattad verksamhet och att utveckla den. Det kan och ska dessutom tillkomma nya aktiviteter. Jag säger här *och dessutom* om samgåendet.

Nu, kort efter årsmötet, har jag en stark vilja och övertygelse att fortsätta framåt på den väg som medlemmarna i de två föreningarna har pekat på, nämligen samgående. Vägen stakades ut grovt i mars, och nu driver vi – ett utskott om sex personer – vägarbetet vidare fram till mars 2008 då den nya föreningens första möte äger rum.

Eva Elvers

Redaktören har ordet

Det första numret av Qvartilen utkom för nästan 22 år sedan, närmare bestämt 1986. Då hade Samfundet 390 medlemmar, spridda på 57 högskoleinstitutioner och 55 andra arbetsplatser.

När du nu har det sista numret av Qvartilen i din hand har Samfundet runt 1300 medlemmar, fördelade på omkring 100 akademiska institutioner och 100 andra arbetsplatser. Tidskriften distribueras idag till läsare i *minst* 14 länder, på *minst* tre kontinenter.

Under Qvartilens 22-åriga historia har 88 nummer av tidskriften utkommit när vi nu som dess tionde redaktör sätter punkt. Antalet artiklar har varit många och visat Samfundets bredd inom ämnet.

Den nya föreningen, efter sammanslagningen med Statistiska föreningen, kommer också att ge ut en tidskrift. Vad som blir dess namn, hur ofta den ska komma ut och så vidare, funderas det på som allra bäst just nu.

I nummer 1-1 stod att läsa:

[...] Tidningen kan då bli ett betydelsefullt instrument för att främja samfundets syfte och samtidigt stärka vår samhörighet, intressegemenskap och yrkesroll.

När vi sänder nummer 22-4 till tryckeriet är det vår förhoppning att vi bidragit till att leva upp till dessa ord. Vi tror dock att en medlemstidskrift kan utvecklas på många sätt vad gäller att uppfylla det första numrets intention. Vi har gjort vad vi kunnat och vi önskar näste redaktör lycka till!

Vi vill tacka alla bidragsgivare till detta nummer och speciellt Göran Arnoldsson samt Britt och Anders Wallgren för deras intressanta artiklar om grafik, Eva Leander för hennes gästkolumn om undervisning i statistik, Tomas Pettersson för en omfångsrik artikel om mikrosimulering, Martin Karlberg och Hans Pettersson för två spännande rapporter om livet som statistiker utanför landets gränser.

Ett stort tack vill vi rikta till redaktionen för ett mycket gott samarbete under det gångna året. Avslutningsvis vill vi också passa på att tacka Samfundet för det förtroende som visats oss som redaktör för dess medlemstidskrift Qvartilen.

I mars finns inte längre Svenska statistikersamfundet, och följaktligen inte heller dess medlemstidskrift. Det som började i mars 1986 tar nu sin ände.

God Jul och Gott Nytt År!

Joakim Malmdin och Jens Olofsson

Styrelsen

Styrelsebeslut tagna på sistone

I det samgåendeuskott bestående av sex personer som Statistiska föreningen bildar tillsammans med Samfundet är Eva Elvers, Annika Lindblom och Jesper Rydén representanter för Samfundet.

En arbetsgrupp som leds av Michael Carlson kommer att jobba med den framtida hemsidan (tillfällig adress www.internetplatsen.se/statistiker/Home/tabid/579/Default.aspx) där bland annat information och debatt inför samgåendet kan läggas upp.

Vid årsmötet presenterades enkätundersökningen angående samgåendet med Statistiska föreningen. 151 medlemmar svarade på enkäten. De fem frågor som ställdes samt de vanligaste svaren var:

1. Vilka är de viktigaste skälen för dig att vara med i Svenska statistikersamfundet?
Svar: Viktigt nätverk, information, e-post utskick med lediga tjänster, Qvartilen.
2. Ett samgående mellan Statistiska föreningen och Svenska statistikersamfundet har utretts och föreslagits. Hur ser du på det?
Svar: 109 svarade positivt av de 117 som uttryckte en åsikt
3. Vad skulle du vilja se i framtiden i Statistikersamfundet eller den eventuella nya föreningen?
Svar: Nya sektioner, fler kurser, aktiva arbetsgrupper,

mötesforum, sommarskolor, nordiska möten, utveckla Qvartilen, m.m.

4. Skulle du själv vilja medverka aktivt i verksamheten i Statistikersamfundet eller den eventuella nya föreningen? På vilket sätt skulle du i så fall vilja göra det?
Svar: 38 av de 151 svarande ville medverka aktivt i den nya föreningen nu eller längre fram
5. Hur länge har du varit med i Statistikersamfundet?
Svar: 86 av de 151 svarande hade varit med i föreningen i mer än 10 år. 102 av de svarande var män och 49 var kvinnor.

En utförligare sammanfattning av enkäten läggs ut på www.internetplatsen.se/statistiker/Home/tabid/579/Default.aspx.

Styrelsen genom Lars Rönnegård

Nya medlemmar

Vi hälsar följande nya medlemmar välkomna:

Sofia Ekvall
Daniel Larsson
Mattias Östlund
Allan Henrysson
Kiat Wee Wong
Anna Törner
Xia Shen

Surveyföreningen

Surveysektionen har bytt namn till *Surveyföreningen* – för mer information om detta, se referatet av årsmötet längre fram i detta nummer av Qvartilen.

Styrelsen, i dess nya konstellation, höll sitt första möte per telefon den 31 oktober. Nästa möte hölls i Stockholm den 20 november. Ansvariga för föreningens webbplats är numera David Sundén och Sofia Ekvall.

Surveyföreningen ansökte tidigare i år om medlemskap i den internationella sammanslutningen IASS (International Association of Survey Statisticians). IASS är öppet för både enskilda medlemmar och institutionsmedlemmar. Det har emellertid visat sig i kontakterna med IASS att medlemsavgiften är så hög som 300 euro per år, vilket fick Surveyföreningens styrelse att besluta att inte ansluta föreningen till IASS. Näst intill alla institutionsmedlemmar i IASS är statistikproducerande myndigheter och företag, medan Surveyföreningen är en (del av en) ideell förening och lever på medlemsavgifter och intäkter från seminarier. Flera av medlemmarna i föreningen är dock redan enskilda medlemmar i IASS, vilket

kostar 20 euro per år. Gå alltså gärna med som enskild medlem i IASS!

Föreningens vårprogram är i skrivande stund ännu inte fastlagt, men vi hoppas kunna bjuda på ett intressant seminarium om så kallad medlyssning under början av 2008.

Styrelsen genom Jörgen Svensson

BLI MEDLEM I SURVEYFÖRENINGEN!

Både personer och organisationer är välkomna som medlemmar i Surveyföreningen. Läs om fördelarna med medlemskap på föreningens hemsida www.statistikersamfundet.se/survey/ under Medlemskap. Där går det också att anmäla intresse för medlemskap via ett webbformulär. Alternativt går det bra att kontakta föreningens sekreterare Sofia Ekvall, e-post: sekreteraren.survey@statistikersamfundet.se, telefon 013 - 48 94 023.

Surveysektionens årskonferens

av Tony Marañon, GfK Sverige AB

Surveysektionen höll en uppskattad årskonferens den 15 oktober på Stockholms universitet. En stor del av dagen ägnades åt intressanta föredrag inom surveyområdet. Ulf Isander och Åsa Jaktlund från SIFO Research International öppnade årsmötet med en diskussion om ett mycket relevant ämne: hur man når mobilabonnenter. Att dra ett fasttelefonurval är vanligt i marknadsundersökningar, men de senaste åren har vissa hushåll, framför allt hushåll där alla är under 25 år, valt att helt avstå från fast telefoni. Framtagning av urval utifrån fast telefon utesluter de hushåll som endast har mobiltelefon. Ulf och Åsa diskuterade olika surveyproblem kring mobilabonnenter, från svårigheter att få fram en acceptabel urvalsram till skillnader i sociodemografiska variabler och svarsfrekvens mellan mobiltelefonsegmentet och fasttelefonsegmentet. Det betonades att det är viktigt att fortsätta att noggrant följa utvecklingen av antalet mobilabonnenter utan fast telefon samt att utvärdera hur utvecklingen kommer att påverka telefonundersökningar. Två andra rekommendationer till framtidsarbete inom området telefonundersökningar var att utarbeta strategier för att kunna kontrollera bias och att utveckla en accesspanel med personer som endast har mobiltelefon.

Det andra föredraget hölls av Martin Ribe, erfaren statistiker på SCB, som talade om mätning av inflation. Han gav en omfattande beskrivning av olika typer av index och berättade om olika aspekter av prisindexkonstruktion och om hur inflationstakten beräknas med dessa index.

Marknads-, opinions- och samhällsundersökningar blir en alltmer globaliserad industri. I detta sammanhang var Stig Holmers föredrag om internationella standarder för surveyundersökningar riktigt informativt. Stig representerar SMIF, branschföreningen för de svenska undersökningsföretagen, som består av ett antal undersökningsföretag med en sammanlagd årsomsättning på cirka 1,3 miljarder kronor. Standarden ISO 20252 är en dokumenterad internationell referens för genomförande av marknads-, opinions- och samhällsundersökningar. Den innehåller en omfattande beskrivning av en effektiv undersökning "från ax till limpa".

Dalenius-föredraget

Det traditionsenliga Dalenius-föredraget hölls av Gunnar Kulldorff, Umeå universitet, som bjöd på en fascinerande kartläggning av den starka utvecklingen inom surveystatistiken som skedde för 50 år sedan. Gunnar berättade om utbildning, läroböcker samt pionjärforskningsarbeten inom surveystatistik från 1930- till 1950-talet.

Årets bästa uppsats inom surveyområdet

Utdelningen av priset "Årets bästa uppsats inom surveyområdet" var en annan höjdpunkt på sektionens årsmöte. Vinnare av uppsatspriset 2006/2007 blev Chandra

Adolfsson, SCB och Örebro universitet, med uppsatsen *Utvärdering av granskningssystem för SCB:s undersökningar Kortperiodisk sysselsättningsstatistik och Konjunkturstatistik över vakanser*. Motiveringen löd enligt följande:

Författaren lyfter fram och levandegör en viktig men undanskymd del av surveyprocessen. Ämnet framställs med klarhet och detaljrikedom. Förutom att delvis nya granskningsmetoder används, ges också intressanta diskussioner kring hur utvärdering av granskningssystem kan gå till och sedan hur de uppnådda resultaten kan tas till vara i det praktiska arbetet.

Efter prisutdelningen fick Chandra möjlighet att presentera den vinnande uppsatsen.



Vinnare Chandra Adolfsson, till höger, och prisutdelare Karin Dahmström.

Ett annat intressant föredrag handlade om analys av kvasi-experimentella undersökningsdata inom arbetstid och hälsa. Föredraget som hölls av Mats Nyfjäll, Statisticon, sammanfattade en stor studie som gjorts på uppdrag av Arbetslivsinstitutet, där det primära syftet var att klarlägga sambandet mellan reducerad daglig arbetstid och arbetstagares hälsa. Studien var kvasi-experimentell med två grupper: en interventionsgrupp där respondenterna gick ner i arbetstid och en referensgrupp. Mats berättade att reduktion av arbetstiden till sex timmar gav tydliga effekter på subjektiv hälsa, men i studien hittar man inte bevis för förbättrad fysiologisk hälsa.

Dagens program, före årsmötesförhandlingarna, avslutades med en paneldiskussion om utbildningsläget in-

om surveyområdet. Panelen bestod av Eva Leander (Linköpings universitet), Annica Isaksson (Linköpings universitet), Thomas Laitila (Örebro universitet), Hans Nyquist (Stockholms universitet) samt Daniel Thorburn (Stockholms universitet).

Årsmötet antog nya stadgar varav de viktigaste förändringarna är att namnet *Sektionen för surveystatistik* (Surveysektionen) byts till *Föreningen för surveystatistik* (Surveyföreningen) och att föreningen blir formellt självständig från Svenska statistikersamfundet. Därmed

kan Surveyföreningen bli en egen juridisk person.

En ny styrelse valdes, med följande sammansättning:

Ordförande	Jörgen Svensson
Kassör	Anton Grafström
Sekreterare	Sofia Ekvall
Övrig ledamot	Karin Dahmström
Övrig ledamot	Anders Holmberg
Övrig ledamot	Niklas Lundvall
Övrig ledamot	David Sundén



Föreningen för medicinsk statistik

Stig Johan Wiklund – ordförande

För drygt elva år sedan anställdes jag på AstraZeneca, för att arbeta inom läkemedelsforskningen i Södertälje. Jag har sedan dess jobbat inom de flesta delarna av forskningen och innehaft ett antal olika roller, som statistiker, projektledare och gruppchef. För tillfället är jag Technical and Scientific Development Leader, en roll vars innehåll är nästan så oklart som det långa namnet antyder, men lyckligtvis kan den bland annat inrymma ett engagemang inom FMS.



Min utbildningsbakgrund är från Umeå universitet, där jag efter avslutad grundutbildning började på forskarutbildningen vid Statistiska institutionen. Tog doktorsexamen därifrån 1994 med en avhandling inom området statistisk kvalitetskontroll.

Jag bor sedan ett par år i Mariefred, tillsammans med fru och sexårig son. Försöker ha en ganska aktiv fritid, så gott det nu hinns med mellan jobb, pendling, villa och barn. Min chef har presenterat mig som en "avdankad gammal orienterare", vilket väl tyvärr är en alltför korrekt beskrivning av nuvarande status för mitt största fritidsintresse.

Eva Andersson – sekreterare

Jag heter Eva Andersson och är ny sekreterare i FMS. Jag bor i Göteborg sedan elva år tillbaka och arbetar på Arbets- och Miljömedicin på Sahlgrenska Universitetssjukhuset. En dag i veckan är jag på Statistiska forskningsenheten vid universitetet. För sex år sedan disputerade jag på icke-parametriska övervakningssystem för

cykliska processer och efter disputationen har jag fortsatt med forskning inom området statistisk övervakning. I våras blev jag antagen som docent i statistik.



Arbetet på AMM handlar om att studera samband mellan exponering och sjukdom/symtom. Jag får arbeta inom många områden: radon, UV-strålning, exponering inom petrokemisk industri, bullerexponering. De statistiska metoder som används är både grundläggande (till exempel t-test) och mer avancerade (exempelvis mixade modeller). Och så tillkommer ju den extra utmaningen att data inte så sällan är ofullständiga och obalanserade och kommer i mycket så mängder.



Mats Rudholm – ledamot

Mats Rudholm har arbetat som högskoleadjunkt i statistik vid Göteborgs universitet. För närvarande arbetar Mats som VD på Statistikkonsulterna. Mats har på Statistikkonsulterna profilerat sig som projektledare samt arbetat med principiell problemlösning, men även med statistisk bearbetning och analys.

Industriell statistik

Meddelanden från styrelsen

Omvandling av sektionen till en förening

Alltsedan Sektionen för Industriell Statistik grundades har den, som namnet antyder, utgjort en sektion inom Svenska statistikersamfundet. Eftersom Samfundet nu slås samman med Statistiska Föreningen ändras försättningsarna även för Industriell Statistik. Vid årsmötet beslutades därför att omvandla *Sektionen för Industriell Statistik* till *Föreningen industriell statistik*. För medlemmarna kommer de synbara förändringarna att vara ringa. Föreningen kommer att överta sektionens tradition att anordna två seminarier, ett på våren och ett på hösten. Från och med 2008 kommer årsmötet och seminariet i samband med det att hållas under våren. Mer information om vårens seminarium kommer efterhand på vår hemsida www.matstat.umu.se/indstat. En smärre medlemsavgift kan också komma att införas. Diskussion om detta blir det på årsmötet våren 2008.

S-ENBIS

Föreningen industriell statistik är medlem i ett nätverk som heter *European Network for Business and Industrial Statistics* (ENBIS). I moderföreningen ingår ett antal nationella föreningar, varav *Föreningen industriell statistik* är en. Nätverket anordnar regelbundet konferenser, se hemsidan www.enbis.org för vidare information. För våra medlemmar betyder kopplingen till ENBIS bland annat att information om dess verksamhet är att vänta via e-post.

Byte på sekreterarposten

Thomas Svensson från SP Bygg- och Mekanik i Borås, som redan tidigare varit ledamot av styrelsen, har tagit över sekreterarposten. Ingemar Sjöström från Sony Ericsson Mobile Communications, som tidigare varit sekreterare, övergår till att vara vanlig ledamot.

Styrelsen genom Magnus Arnér

Statistisk modellering av yttre belastningar på fordon

av Magnus Arnér

I samband med att Industriell statistiks årsmöte gick av stapeln så hölls också, precis som brukligt, ett seminarium. När vi i styrelsen diskuterade vad temat skulle vara blev vi relativt snabbt på det klara med att temat denna gång skulle vara laster. Alltför många, i synnerhet de som sett Gösta Ekman i *Mannen som slutade röka*, kommer dock att tänka på andra laster än de vi hade i åtanke, såvida inte temat skulle preciseras en aning. Seminariet kom därför benämnas *Statistisk modellering av yttre belastningar på fordon*.

Med tillförlitlighet menas ju, något förenklat, att en produkts styrka ska överstiga den last den utsätts för. Båda dessa, last och styrka, är stokastiska. Inom många företag ägnas tämligen mycket tid åt att studera den ena av dessa, nämligen styrkan. Anledningen är nog att den är förhållandevis enkel att studera. Man kan nämligen prova den i ett laboratorium. Den belastning produkten utsätts för studeras inte alls i samma omfattning trots att den vanligen har en variation som är av en magnitud större än spridningen i styrka mellan till synes likadana produkter. Ska man studera lasten så måste man ut på fältet, vilket inom fordonsindustrin innebär att man måste studera vägarnas beskaffenhet, förarnas beteenden och miljöbetingelserna under vilka fordonen används. Det kräver ett stort och omfattande arbete.

Till en väsentlig del är studiet av laster av statistisk natur, vilket märktes på föredragen under seminariet. Klas

Bogsjö från Scania talade om gupp och gropar i vägen. Gwenaëlle Genet, också hon från Scania, berättade om hur komplexa flerdimensionella laster kan omsättas till en skadeekvivalent last av en enkel natur. Magnus Karlsson från Volvo Lastvagnar höll ett föredrag om belastningars beroende av marknad, vägtyp och kund, vilket ni för övrigt kunde läsa om i förra numret av Kvartilen där han skriver om sin nyligen framlagda avhandling. De statistiska frågeställningar som uppkommer i dessa sammanhang är mångahanda, vilket vi också blev varse under seminariet. Det handlar om urvalsteori, hur man väljer ut vilka marknader, kunder eller vägar man ska studera. Ska de väljas slumpmässigt eller ska man välja på annat sätt? Vilka osäkerheter tillför man genom att extrapolera från korta mätningar till längre tider och körsträckor? Kan variansanalys användas när man studerar bidrag till variation i last från marknad, vägtyp och kund? Är det exponentialfördelat avstånd mellan gropar i vägen? Vilken fördelning har antalet lastcykler under en given tidsperiod? Vi som statistiker måste kunna beskriva detta på ett så enkelt sätt att man kan omsätta det i kravspecifikationer, laster vid prover och input i beräkningar. Samtidigt så får beskrivningarna inte bli så förenklade att de blir meningslösa.

Seminariet avslutades med en diskussion ledd av Karl Erik Olsson som bland annat anknöt till det arbete kring laster som utfördes på VCE på 1980-talet.

Ett par reflektioner är värda att nämna. Den ena är att flertalet av föredragen var statistiska tillämpningar inom utmattning. Den andra är frågan hur andra branscher än just fordonsindustrin studerar sina laster. Kan det måne vara så att fordonsindustrin ligger steget före? Och är det bara en tillfällighet att utmattning är mer välrepresenterat i sådana här sammanhang än andra felmoder?

Jag har under min tidigare verksamhet, då jag varit verksam vid Saab Automobile, noterat ett par saker som jag tror kan intressera läsaren. Det ena är att det fokus som industrin numera har på kunden har lett till ett ökat intresse för att förstå lasterna. Dessutom har man inom in-

dustrin förstått att statistisk teori och praktik i det sammanhanget är en jättegrej. Det återstår dock att se om det innebär omfattande rekryteringar av statistiker och andra statistiskt kunniga till fordonsindustrin.

Bland seminariedeltagarna fanns folk från såväl industri som högskola samt en och annan konsult. Seminariets rubrik ledde naturligtvis till att fordonsbranschen var den mest välrepresenterade industrigrenen, men det saknades inte folk från andra branscher. Av den diskussion som följde så tycks intresset för laststudier vara omfattande. ■

Annons

SCB och Örebro universitets sommarskola 2008

19 – 21 AUGUSTI VID ÖREBRO UNIVERSITET

Sommarskolan 2008 kommer att hållas mellan 19-21 augusti vid Örebro universitet. Rubriken för skolan är *Utvärderingsstudier och analys av registerdata*. Skolan innehåller två föreläsningsserier. Den ena serien ges av Professor Xavier de Luna, Umeå universitet, och den andra ges av Anders och Britt Wallgren, SCB.

Data från register används i allt större utsträckning inom forskning och vid produktion av officiell statistik. Istället för urvalsundersökningar kan totalräkning av register över populationen användas för produktion av statistik. Ansatsen bidrar till minskad uppgiftslämnarbörda samtidigt som statistik kan produceras med högre kvalitet till en lägre kostnad. I utvärderingssyften används register för att belysa och mäta effekter av behandlingar, exempelvis effekten av arbetsmarknadsutbildning på arbetslöshetstidens längd eller effekten av rehabilitering på sjukskrivningstidens längd. Sommarskolan behandlar teoretiska och praktiska aspekter på användning av register vid utvärderingsstudier och för produktion av officiell statistik. Sommarskolan ger ett utmärkt tillfälle för statistiker, forskare, utredare, universitetslärare och andra att träffas och lära sig mer om möjligheter och problem med användningen av registerdata.

Mer information om sommarskolan samt inbjudan kommer under inledningen av 2008. Frågor om sommarskolan kan ställas till Thomas Laitila (e-mail: thomas.laitila@scb.se, tel: 019 - 17 62 18, 019 - 30 10 63).

Organisationskommitté:
Martin Axelson
Thomas Laitila

Ann-Marie Flygare
Elisabeth Svensson

Visualisering av kvantitativ information

Grafik och regression

av Göran Arnoldsson, Umeå universitet

Grafiska metoder är ett viktigt inslag i regressionsanalys. Men, kan man göra en regressionanalys enbart med hjälp av grafikbaserade metoder? Frågan var utgångspunkt för den utveckling av metoder för *grafisk regression* som förknippas med arbeten av R. Dennis Cook och Sanford Weisberg [1], [2], [3] och [5]. Det allmänna svaret på frågan är nog ett nej. Men grafiskt regression erbjuder en metodik där grafik används aktivt under hela modellbyggnadsprocessen. Detta i kontrast med mer traditionella ansatser baserade på successiva testbaserade modellanpassningar (till exempel stegvis regression) där grafiska metoder används huvudsakligen i början och slutet (till exempel matrisplottar, analys av residualer).

Låt y vara responsvariabel och x en p -dimensionell vektor av förklaringsvariabler. Sambandet mellan y och x sägs ha *strukturell dimension* q ($\leq p$), eller q -d-struktur, om $y \perp x \mid \eta^t x$, för någon $p \times q$ -matris η . Ett alternativt uttryck är att säga att grafen $(y, \eta^t x)$ är *tillräcklig* för sambandet mellan y och x . Vanlig multipel regression $y = \beta^t x + \varepsilon$ är ett exempel på 1-d-struktur. Det funktionella sambandet mellan y och $\eta^t x$ behöver nödvändigtvis inte vara linjärt.

Grafisk regression blir datorintensiv och underlättas av tillgång till programvara för dynamisk grafik med till exempel möjligheter att länka grafer till varandra. Lisp-Stat [6] är en lisp-dialekt för statistiska tillämpningar med sådana möjligheter. Den mjukvara, Arc, som utvecklats för grafisk regression baseras på Lisp-Stat. Arc är fritt tillgängligt via internet [4]. Boken *Applied Regression Including Computing and Graphics* [3] baseras på användning av Arc. En tidigare variant av programmet, R-code, utgör grunden för alla analyser i An

Introduction to Regression Graphics [2].

Ett exempel

För 206 träd av arten (*Cupressus Lusitanica*) mättes total höjd HT (m), diameter i brösthöjd D (cm) och trädets volym VT (merchandiseable volume, m^3) [7]. Responsvariabel är VT . En matrisplott visar att de direkta sambandet inte är linjärt. De inversa graferna (x_j, y) indikerar att sambandet bör ha dimension 1. Om sambandet har 1-d-struktur måste varje invers regression, x_j mot y vara på formen

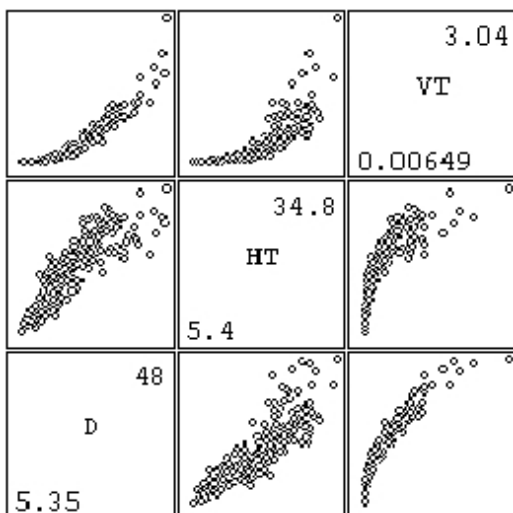
$$E(x_j|y) = E(x_j) + \alpha_j m(y)$$

$$V(x_j|y) \approx V(x_j) - \alpha_j^2 s(y)$$

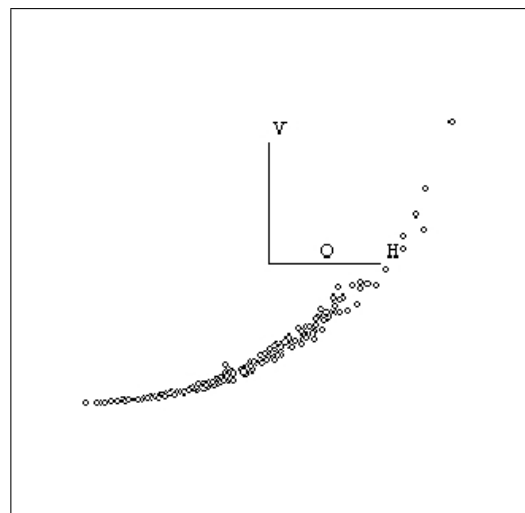
för två funktioner $m(y)$ och $s(y)$ [3, s.435–6]. Matrisgrafens (Figur 1) tredje kolumn indikerar detta ganska tydligt.

Efter att ha konstaterat att 1-d-struktur är rimlig gäller det att försöka hitta den. Detta görs genom att rotera den 3-dimensionella grafen (VT, D, HT) runt den vertikala axeln tills en tillfredsställande graf hittats. När data roteras kommer den horisontella komponenten av dess projektion på grafytan (skärmen) att representera en linjärkombination $h = b_0 + b_1 D + b_2 HT$. Här är transformationen $h = -1.753 + .04116D^* + .03264HT^*$ (se Figur 2) där D och HT har translaterats och skalats om för att data ska fylla ut den 3d-kub som används i den grafiska programvaran.

Om modellen är korrekt ska nu, enligt tidigare, y vara oberoende av x för varje givet värde på $h = b^t x$. I så fall gäller också [1, s.64] att $y \perp b_o^t x \mid b^t x$ där $b_o^t x$ är

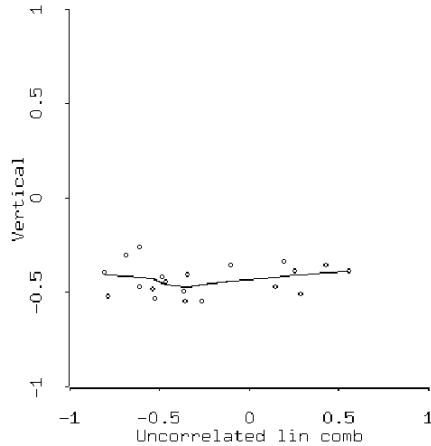


Figur 1: Matrisgrafen (VT, HT, D)



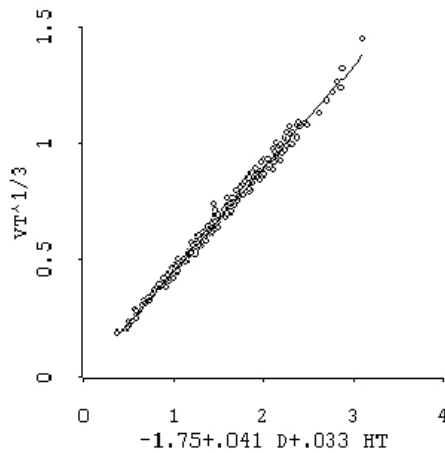
Figur 2: Grafen (VT, h)

okorrelerad med $b_o^t x$ i den meningen att $b^t \text{Var}(x) b_o = 0$. Ett sätt att undersöka detta är att undersöka grafen $(y, b_o^t x)$ inom lagom stora fönster $b^t x \in (h_-, h_+)$ (*slicing*). För varje fönster ska något samband mellan y och $b_o^t x$ inte synas. En *lowess*-utjämning eller liknande inom varje fönster kan underlätta. Figur 3 visar en sådan graf.



Figur 3: Grafen $(VT, b_o x)$ (VT centrerad)

Genom att transformera VT till $VT^{1/3}$ erhålls ett linjärt samband (se Figur 4). Den anpassade räta linjen är $VT^{1/3} = 0.00807330 + 0.442579h$.



Figur 4: Grafen $(VT^{1/3}, h)$

Dimensionreducerande underrum

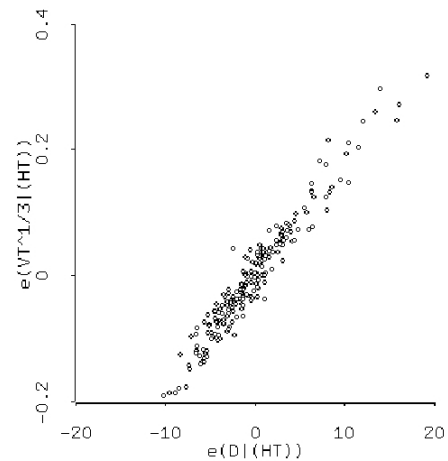
Låt $S(\eta)$ vara kolumnrummet till matrisen η . I relationen $y \perp x \mid \eta^t x$ är η av sekundär betydelse i den meningen att den kan ersättas med alla matriser med samma kolumnrum som $S(\eta)$. Om $y \perp x \mid \eta^t x$ sägs $S(\eta)$ vara ett *dimensionsreducerande underrum* (dimension reducing subspace, DRS) för regressionen av y på x . Ett DRS är ett *minimalt* DRS (MDRS) om dess dimension är det minsta bland alla DRS. Dimensionen av ett MDRS sägs vara den *strukturella dimensionen* för regressionen av y på x . Slutligen, ett MDRS är ett *centralt* DRS (CDRS) om $S(\eta) \subset S_{drs}$ för alla DRS S_{drs} . Det kan visas att ett CDRS, om det existerar, är entydigt.

Grafisk regression

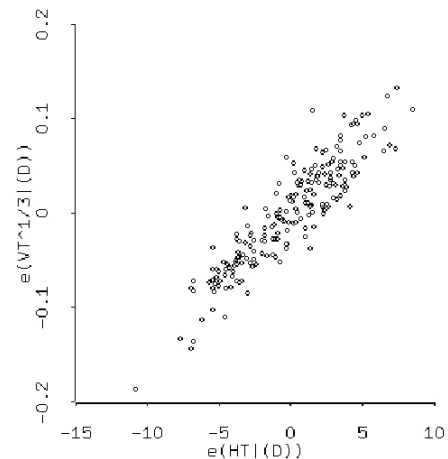
Syftet med en regressionsanalys formuleras som att utifrån data (y, x) söka eller uppskatta ett CDRS $S_{y|x}$, av lägre dimension än p , om något sådant existerar. I grafisk regression genomförs detta med hjälp av grafer $(y, A^t x)$ av dimension 2 eller 3.

Det kanske viktigaste hjälpmedlet i processen är *added variable plots* (AVP) i två eller tre dimensioner, en parallell till modellbyggnad baserad på partiell korrelation. Låt x_1 vara de prediktorer som redan ingår i modellen och låt x_2 vara en kandidatvariabel. En 2-d AVP är grafen $(e_{y|1}, e_{2|1})$ där $e_{y|1}$ är residualerna från regressionen av y på x_1 och där $e_{2|1}$ är residualerna från regressionen av x_2 på x_1 . En 3-d AVP är en graf $(e_{y|1}, e_{2|1}, e_{3|1})$.

För exemplet visas båda 2-d AVP med $V^{1/3}$ som respons i Figur 5 och 6.



Figur 5: AVP för $VT^{1/3}, H|D$



Figur 6: AVP för $VT^{1/3}, D|HT$

Naturligtvis visar båda att både D och HT bör inkluderas, givet att den andra prediktorn redan ingår i modellen och leder i slutänden till samma modell som ovan.

Den som vill sätta sig in mer i teorin och ytterligare metoder för grafisk regression hänvisas till främst referen-

serna [1] och [3]. ■

Referenser

- [1] Cook, R. D. (1998). *Regression Graphics. Ideas for Studying Regressions through Graphics*. Wiley.
- [2] Cook, R. D. and Weisberg, S. (1994). *An Introduction to Regression Graphics*. Wiley.
- [3] Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. Wiley.
- [4] Cook, R. D. and Weisberg, S. (2004). *Arc Software*. <http://www.stat.umn.edu/arc>.
- [5] Cook, R. D. and Wetzel, N. (1993). Exploring regression structure with graphics (with discussion). *TEST*, 2:33–100.
- [6] Tierney, L. (1990). *LISP-STAT An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*. Wiley.
- [7] Berhe, L. (2007). *Data från skogsinventering*. In-samlade på en skogsplantering i Etiopien.

Konsten att förmedla ett budskap – diagram är viktigare än du tror!

av Britt och Anders Wallgren, SCB

Det var en gång för länge, länge sedan två statistiker som fick ett intressant uppdrag. En grupp kirurger som under flera år samlat in data kring sina patienter med en viss form av tumörer behövde hjälp med statistisk analys. Man hade data avseende tvåårsöverlevnad och ett antal förklaringsvariabler och ville ha sambanden utredda. Statistikerna fann flera intressanta samband och presenterade stolt dessa resultat vid en träff med kirurgerna genom att visa ett antal flervägsindelade tabeller. Kommentaren blev "Detta var ju intressant, men det behövs väl fler analyser?" sagt med mycket neutral röst.

Två slokörade statistiker gick därifrån och undrade hur skulle de ha gjort för att få läkarna att förstå vilka intressanta slutsatser som fanns? Till nästa träff gjordes inga fler analyser, men däremot ett antal handritade diagram. Då blev det klang och jubel! "Detta vara verkligen sensationellt, detta skall vi publicera och berätta om på konferenser!" Sedan den dagen har de två statistikerna intresserat sig för statistisk grafik.

Skol- och universitetskurser

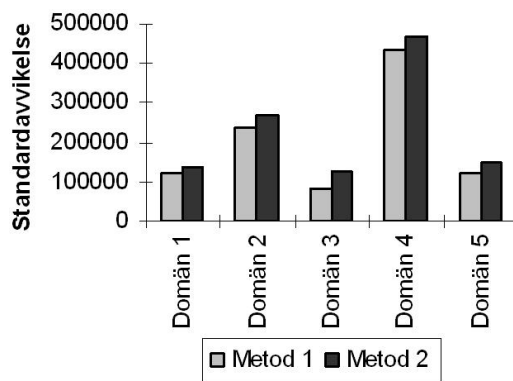
Under våra egna statistikstudier ingick diagram under den första delkursen på ettbetygsnivån, och därefter aldrig mer. Vi fick läsa Huffs bok [1] och lära oss om några enkla diagramtyper i kursboken. Alla diagram måste göras för hand och därför övades det i ringa omfattning.

När vi sedan själva arbetade som universitetslärare i statistik så var det sällan diagram gjordes i samband med statistikstudier och uppsatser i statistik. Det var först 1977 när vi började arbeta med den nya projektinriktade ekonomlinjen som våra studenter fick arbeta praktiskt med att göra diagram, fortfarande för hand. Vi hade till exempel *Långtidsutredningen 1980* (LU80) som kurslitteratur och studenterna arbetade med olika delproblem från LU80 (illustrerade med diagram) och gick sedan till källorna och gjorde egna analyser med den officiella statistik de fann. I sina uppsatser jämförde de LU80 med sina resultat som illustrerades med egna handgjorda diagram. Vår kursbok [2] skrevs för denna typ av statistikundervisning och kom att innehålla ett kapitel om

diagram. Såväl diagrammen i LU80 och studenternas, belyste viktiga samhällsproblem och hade tydliga budskap. Vi kan rekommendera denna typ av statistikundervisning där statistiska data och statistisk metodik och diagram får en naturlig och viktig roll. Detta borde gälla både på gymnasier och universitet.

Vetenskapliga tidskrifter

Hur är det med förmågan att skapa diagram inom statistikerkåren? Om man bläddrar igenom våra internationella vetenskapliga tidskrifter och böcker har man skäl att bli bedrövad. Välgjorda diagram är sällsynta, många diagram är arbetsdiagram som tagits direkt från en körning med ett statistikpaket. Dessa diagram är inte gjorda för presentation utan skulle behöva redigeras före publicering. Vanliga fel är hopplösa tidsaxlar med tidpunktsnummer i stället för årtal, eller månadsnamn på snedden under tidsaxeln. Rubriken på y-axeln är ofta vertikal, det vill säga diagrammen är så kallade "ugglediagram" som förutsätter att läsaren har samma rörlighet som ugglor när det gäller huvud- och ögonrörelser. Diagrammet nedan har vi gjort själva, men det har inspirerats av flera faktiska diagram.



Figur 1. Medelfel för skattningar för olika domäner

Vi skulle föredra ett diagram med liggande staplar där ritytan har ram och rubriken står *ovanför* diagrammet.

Bild 1. Medelfel för skattningar för olika domäner

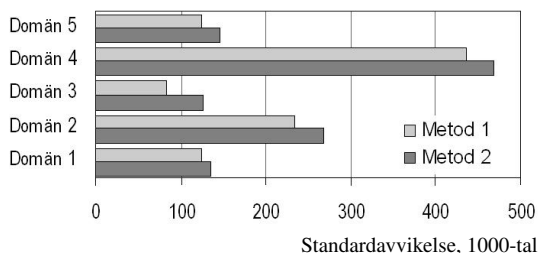


Diagram i officiell statistik

Vår kollega Ulf Jorner startade arbetet med [3] på 90-talet. Orsaken var att SCB:s diagram inte var enhetligt utformade och att många diagram i SCB:s publikationer var mindre lyckade. Det fanns behov av en diagrambok som kunde fungera som inspiration och rättesnöre för dem som arbetade med att skapa diagram. När sedan boken kom ut så förväntade vi oss att kvaliteten på SCB:s diagram skulle öka. Till vår stora förvåning såg vi ingen skillnad trots att många inom SCB läste boken och visade sin uppskattning.

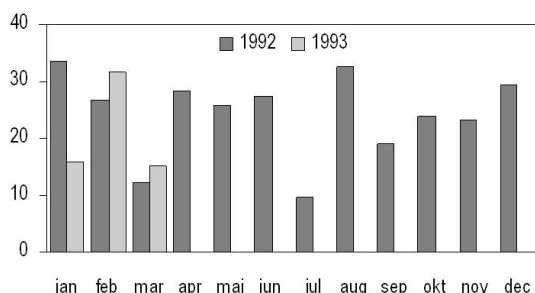
Detta mysterium fick sin förklaring när vi började ha kurser med boken som bas. Det visade sig då att det var Bill Gates som var ansvarig för många diagram på SCB. Excels default-format dominerade av två skäl, dels krävs det uppövad blick för att se bristerna i ett diagram, dels hade de flesta svårt att veta hur man skall ändra formateringarna. Vi utvecklade därför tillsammans med en IT-kollega ett Excel-tillägg, *SCB:s Diagramverktyg*, som hjälper användaren att skapa diagram som i enlighet med principerna i [3]. Detta hjälpmedel, som är gratis, har vi spritt till många som har deltagit i våra diagramkurser under årens lopp.

Diagram i företagsvärlden

Vi har under många år också arbetat som konsulter åt några industriföretag. Den dramatiska konjunkturedgången under 1990-talet medförde att behovet av tidsserieanalyser och rättvisande diagram blev stort. En vanlig diagramtyp inom företag är diagrammet nedan. Med denna typ av diagram sitter man och försöker bedöma den aktuella utvecklingen – tyvärr så har vi lärt oss att ekonomistyrning ofta baseras på tidsseriebrus där man ofta jämför med motsvarande månad föregående år.

Bild 2. Ordergång i AB Industri 1992 – mars 1993

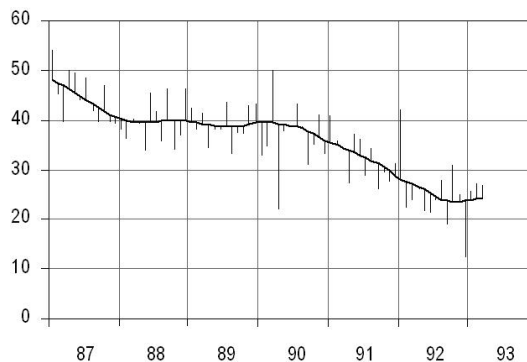
Miljoner kronor per månad



Genom vettig tidsserieanalys och bra diagram kan statistiker göra nytta genom att rättvisande bilder av både den långsiktiga utvecklingen och det aktuella förändringsmönstret. Vi har lärt oss att diagram av det slag som visas nedan är ett effektivt sätt att förmedla budskap till företagsledningar.

Bild 3. Ordergång i AB Industri 1992 – mars 1993

Miljoner kr per månad, säsongrensade och utjämnade värden



Slutord

Grunden för diagram är statistisk analys. Det är ju analysresultaten som skall vara budskapen i diagrammen. Vill man lära sig göra bra diagram måste man arbeta med analys och helst öva sig genom att presentera resultaten inför publik. Det är då man lär sig vad som gör att en presentation går hem. ■

Referenser

- [1] Huff, D. (1964). *Hur man ljugar med statistik*. Natur och Kultur, 6:e upplagan.
- [2] Wallgren, A. och Wallgren, B. (1983). *Undersökningar och statistik*. Natur och Kultur.
- [3] Wallgren, A., Wallgren, B., Persson, R., Jorner, U. och Haaland, J. A. (1996). *Statistikens bilder*. Publica, Nordstedts Juridik.

Tack

Quartilen är mycket uppskattad av sina läsare. Redaktörerna och redaktionen lägger ned ett omfattande arbete på såväl idéer och kontakter som återkontakter och redigering i flera led fram till en tidskrift som är välfylld till innehåll och yta. Vi framför här ett stort tack för detta arbete under 2007.

Styrelsen genom Eva Elvers

Ny hemsida

En hemsida för den nya föreningen – efter samgåendet mellan Statistiska föreningen och Svenska statistikersamfundet – är under uppbyggnad. Den nås via Statistikersamfundets nuvarande hemsida, alternativt kan man gå direkt till www.internetplatsen.se/statistiker/Home/tabid/579/Default.aspx.

Gästkolumnist

I första numret för i år (22-1) av Qvartilen skrev Harry Khamis, Wright State University, USA, om skillnader mellan Sverige och USA med avseende på undervisning och till andra numret (22-2) ställde vi frågan: Hur skulle du vilja bedriva undervisning i ämnet (matematisk) statistik på grundnivå? till Robert Lundqvist, Luleå tekniska universitet. Till detta nummer har frågan ställts till Eva Leander, Linköpings universitet. I förra numret besvarade Kenneth Carling, Högskolan Dalarna frågan: Hur skulle du vilja bedriva undervisning i ämnet (matematisk) statistik på forskarnivå?

Hur jag skulle vilja bedriva undervisning i ämnet (matematisk) statistik på grundnivå

av Eva Leander, Linköpings universitet

Hur ska jag egentligen svara på frågan "Hur skulle du vilja bedriva undervisning i ämnet (matematisk) statistik på grundnivå". Är det verkligen rätt fråga till en erfaren lärare, tro? Efter att jag faktiskt har undervisat i statistik inom högskolan – inte minst på grundnivå – under ett helt arbetsliv, så borde kanske snarare frågan som jag besvarar vara "Hur undervisar du på grundläggande nivå?" Åtminstone i den mån min föreställning om hur det bör vara kunnat påverka hur det faktiskt blivit. Alltnog, jag kommer att tala om vad jag anser vara särskilt viktigt i grundläggande statistikundervisning och lite om vad som påverkat mig i detta avseende. Som redan framgått, så ligger min undervisningserfarenhet inom statistik och inte inom matematisk statistik men mina tankar är åtminstone delvis säkert också relevanta där.

Innan jag fortsätter vill jag gärna uppehålla mig vid frågan "Vilka problem står statistikutbildningen inom högskolan för i dag?" ty svaret på den frågan borde vara viktigt för hur undervisningen ska utformas. Denna fråga stod bitvis i fokus i den utvärdering av statistikämnet som Högskoleverket (HSV) genomförde för ett par år sedan. I den självvärdering som vi gjorde här vid Linköpings universitet som en del av HSV:s utvärdering framhöll vi de två huvudproblemen, dels alltför knappa resurser, dels alltför dålig studerandetillströmning. Problemen på övriga universitet och högskolor som utvärderades var likartade. Resurssidan har blivit bättre åtminstone på vissa håll, och de nya planer som tycks finnas för tilldelning av resurser till högskolan kanske till dels löser kvarvarande resursproblem. Kvarstår alltså frågan om hur man ska påverka tillströmningen till ämnet.

Statistik är tråkigt

Varför är det då svårt att rekrytera studenter till statistikutbildning? Det finns säkert många olika svar på den frågan men för egen del är jag övertygad om att det faktum att många människor, inte minst gymnasister, uppfattar statistik som tråkigt är en av de viktigaste bovarna i dramat. Vad kan man då göra som ambitiös lärare, som vill göra bästa möjliga insats för sitt ämne, men som inte har tid, möjlighet eller lust att påverka utbildningspolitiken eller innehållet i skolans utbildning etcetera genom politiskt engagemang eller debatter. Det uppenbara svaret är förstås att man måste försöka påverka sina studenter. Om många av dem som faktiskt läser statistik inom

högskolan finner ämnet intressant och stimulerande, så kommer detta på sikt leda till en större förståelse för ämnet och bidra till att tvätta bort tråkighetsstämpeln.

Alltså kommer jag att fortsättningsvis försöka besvara frågan "Hur påverkar du (eller vill du påverka) dina nybörjarstudenter, så att de finner att statistik är intressant och spännande?". Jag tänker då i första hand på starten av statistikstudierna. Det kan handla om den första 5- eller 10-poängskursen (gamla poäng, varav 20 motsvarar en termins studier) för dem som väljer att läsa en 20-poängskurs eller ett statistikprogram. För andra studenter som läser statistik som en liten del i en annan utbildning handlar det kanske snarare om de första undervisningstillfällena.

Min egen bakgrund

Kommen så här långt så vill jag ett ögonblick se tillbaka på mina egna grundläggande statistikstudier vid Stockholms universitet i början av 1960-talet. Då hade jag först praktiserat inom socialvården i två år efter studenten. Egentligen hade jag tänkt läsa bara ett betyg i statistik, som det hette på den tiden, för att därefter påbörja sociala studier på föregångaren till Socialhögskolan i Stockholm. Men roliga kurskamrater och stimulerande lärare fick mig att fortsätta med två betyg och sedan – efter studier i andra ämnen – fler betyg och forskarstudier.

Den förste läraren som jag mötte var Avo Raud, född och uppvuxen i Estland och en person med stark utstrålning. Avo var en mångsidigt begåvad person. Han spelade piano i nästan proffssklass och tillhörde Sverigemästarklassen i schack. Han slutade sin karriär som läkare, och avled i lungcancer före sextio års ålder. Avo påverkade mig starkt i min inställning till statistikämnet. Han fick mig bland annat att se hur man med statistikens hjälp faktiskt kunde bli klokare och förstå samhällsutvecklingen bättre. Det jag särskilt minns är hans uppmaning att alltid börja med att noga studera datamaterialet. Detta var tiden före datorerna, så att studera datamaterialet innebar ofta att rita bilder för hand.

Statistisk analys

Ettbetygsundervisningen i Stockholm var vid denna tid upplagd så att man inledde med *Statistisk analys*, en

fyra-veckors-kurs baserad på delar av [1]. Kursen var rent deskriptiv, olika datamaterial utgjorde hela det material som man var intresserad av och slumpen stötte man inte på vad jag minns. Den följdes sedan av kurser i sannolikhetslära, och senare under terminen stickprovsteori och statistisk inferens.

I statistiska analyskursen behandlade vi vid varje undervisningstillfälle diskussionsuppgifter ur [1] eller liknande uppgifter som Avo Raud gjort själv eller hittat på andra håll. Säkert är det många av mina kurskamrater som minns exempelvis uppgiften om Laymans skumsläckningsmetod, en metod att släcka bränder som ansågs synnerligen framgångsrik därför att värdet av det brunnna var betydligt lägre än motsvarande värden under tidigare bränder och man hade i sin tolkning helt missat det faktum att alla dyrbara byggnader redan brunnit när metoden började användas. Beskrivande mått, vägda medelvärden, variabeltransformationer och mycket annat ingick också. Mitt intresse för standardvägning härrör från den tiden. Vem minns inte jämförelsen mellan medelantalet barn hos godsägare och jordbruksarbetare (godsägarna hade till allmän förvåning i medeltal fler barn, något som de facto förklarades av en snedfördelning i det undersökta materialet när det gällde äktenskaps varaktighet) eller dödligheten hos präster och järnvägsmän (där prästerna dog i högre grad, trots att järnvägsmännens dödlighet var högre i varje åldersklass).

Den beskrivna inledande kursen hade i sig stor betydelse för mitt val att fortsätta med statistik. Den väckte mitt intresse och min nyfikenhet och jag kände hur jag för var dag blev en liten aning klokare. Jag har nu efter ett helt yrkesliv huvudsakligen i statistikutbildningens tjänst nått slutet av min lärargärning. Och till skillnad från många av kollegerna så har jag med tiden blivit alltmer intresserad av att undervisa på teoretiskt allt mindre avancerade kurser. Att påverka människor till att försöka bli klokare genom att utnyttja den information som finns i alla de data som sköljer över oss har jag sett som en stor och viktig utmaning.

Statistik är inte obegripligt

En av de kurser från senare år som jag särskilt gärna minns gavs inom ramen för det treåriga utbildningsprogrammet *Social omsorg* vid Campus Norrköping, en del av Linköpings universitet. Studenterna på programmet utgjordes till stor del av personer som jobbat i årtal inom exempelvis hemtjänsten eller på behandlingshem. De flesta var kvinnor och de flesta var äldre än de studenter, som jag annars undervisade. Vanliga arbeten efter genomgången program var som biståndshandläggare i en kommun (de personer som beslutar om våra äldre ska få hemtjänst, färdtjänst etcetera) eller som mellanchefer inom äldreomsorgen. Studenterna var ofta nära nog skräckslagna inför kravet att läsa statistik och utgick ifrån att kursen skulle bli såväl obegriplig som oanvändbar.

Lärarens tillfredsställelse var stor, när studenternas räds- la avtog och intresset tilltog, i takt med att de fann att

man faktiskt kunde bli klokare om man försökte analysera sina data med relativt enkla medel. Ofta inledde jag kursen med något aktuellt exempel. I några år var min favoritfråga "hur stor tror ni att löneskillnaden är mellan männen och kvinnorna som arbetar på Linköpings universitet?". När de studerande fick veta att männens genomsnittliga månadslön var cirka 8000 kronor högre än kvinnornas blev upprördheten ofta ganska stor och alla blev väldigt intresserade. En bättre inledning till att tala om vägda medeltal, störande faktorer och standardvägning kan man knappast tänka sig. Eftersom jag i ett antal år under början av 2000-talet fungerade som konsult åt personalchefen när det gällde att beskriva universitetets löner för JämO, så visste jag mycket om löneskillnaderna mellan könen. Ett bättre sätt att introducera analys av kvantitativa data i grupper av människor som fått lära sig i sin tidigare utbildning att kvalitativ analys är överlägsen kvantitativ får man leta efter.

Liknande erfarenheter har jag från den kvantitativa strimman (5 gamla poäng) inom det i övrigt starkt kvalitativa programmet *Samhälls- och kulturanalys* och från uppdragskurser för mellanchefer i en kommun liksom för analytiker på dåvarande Integrationsverket. I samtliga fall handlade det om att förmedla förståelsen att man kan bli klokare och fatta bättre beslut om man kan plocka ut – åtminstone en del av – den information som finns fördold i de data som finns tillgängliga.

De ovan beskrivna exemplen avser kurser för personer som läser bara en liten statistikkurs inom en utbildning som med något undantag är yrkesinriktad. Nog kommer slumpen och testen in där också men tyngdpunkten ligger ändå på det deskriptiva. När det gäller kurser för studenter som valt att läsa mera statistik går det enligt min mening mycket bra att tillämpa samma modell. Som exempel kan nämnas utbildningsprogrammet *Statistik och dataanalys* där de studerande läser 60 eller 80 poäng. Den första halvan av den första 10-poängskursen är mycket inriktad på att visa statistiks mångfald genom exempel som väcker intresse och vi skjuter slumpen framför oss den första månaden.

Om jag skulle göra en punktlista med några av mina viktigaste råd till framtidens undervisare när det gäller hur man ska inleda den grundläggande statistikutbildningen med målet att få de studerande intresserade och nyfikna, så skulle den kunna se ut så här:

- *Statistik kan vara roligt även som mer deskriptiv disciplin.* Visa redan från början att statistik är roligt och spännande och inte bara svårt, gärna med hjälp av bra exempel.
- *Verka för goda kontakter med kolleger i "tillämpningsämnen".* Tala inte illa om kvalitativ analys, framhåll i stället vikten av att använda både kvantitativa och kvalitativa metoder i analys av samhällsproblem.
- *Hämta exempel ut den pågående samhällsdebatten i tidningar och andra media.* Är det verkligen så att man bör undvika säkerhetsbälte om man vill klara sig

bra i trafiken, som DN tycktes mena när man skrev i rubriken på debattsidan *Ordentliga förare i flest dödsolyckor* baserat på att "hela 62% av de omkomna använde bilbälte"? Utifrån den artikeln gjorde jag en mycket nyttig övningsuppgift.

- *Börja med populationer, inte med urval.* Det är minst lika lätt att tänka kring populationer och på köpet slipper man börja öka på obegripligheterna med $n - 1$ redan första dagen. Och introducera spridning med hjälp av genomsnittlig (absolut)avvikelse, som är mycket lättare att förstå sig på.

- *Se till att räknandet är så enkelt som möjligt.* Konstruera uppgifter med små datamaterial med heltalsmedeltal etcetera. Då förenklar man räknandet för hand utan att behöva introducera kalkylformler. Som läsaren förstår tror jag på att studenterna ska räkna själva först, innan man börjar med datorer och statistikprogram.

Självklart anser även jag att datorer är viktiga i dataanalys och i statistikundervisning även om jag sett många exempel på hur enkelheten att göra komplicerade analyser med datorns hjälp av allt att döma ofta leder till olämpliga slutsatser.

I nummer 22-2 av *Qvartilen* skrev Robert Lundquist om hur han ser på grundläggande statistikutbildning. I mycket av det han skriver instämmer jag. Här har jag fokuserat på den inledande delen eller möjligen den inledande kursen i grundläggande statistik, men självklart måste det finnas en fortsättning där man hanterar sannolikheter, urval, skattningar och test. Liksom Robert anser jag "mer deskription, mindre inferens" (rubrik i Roberts artikel) och liksom han hänvisar jag gärna till böcker av D S Moore liksom av Anders och Britt Wallgren, tidigare uppskattade kolleger.

Nu det känns lite grann som om cirkeln blivit sluten. Min statistiska karriär började genom att jag blev fascinerad av statistik tack vare en entusiasmerande lärare som visade hur man kan lära sig genom att studera sina data och den slutar – åtminstone själva undervisningen – på samma sätt. Hur kan man få studenter – även sådana utan alltför goda förutsättningar och utan eget val att läsa statistik – att bli intresserade av och utveckla förståelse för statistik, särskilt analys av kvantitativa data? ■

Referenser

- [1] Wallis, W. A. and Roberts, H. V. (1956). *Statistics: A New Approach*. Free Press.



Statistik konsulterna
JOSTAT & MR SAMPLE AB

Statistik konsulterna är en av Västsveriges största och snabbast växande arbetsplats för statistiker! Vår vision är att bli ledande i Skandinavien. Vi har kunder inom näringsliv och offentlig förvaltning med tyngdpunkt på Västsverige

Vi arbetar inom områdena:

- **MARKNAD OCH SAMHÄLLE**
- **INDUSTRIELL STATISTIK**
- **BIOSTATISTIK**

Vi söker ständigt kontakt med statistiker inom olika områden. Likaväl som vi vill attrahera intressanta kunder och projekt vill vi skapa en bra miljö för intressanta personer. Våra konsulter måste ha pedagogisk skicklighet, förmåga att generalisera och en liten entreprenör inombords

Statistik konsulterna Jostat & Mr Sample AB • Gårdavägen 1 • 412 50 Göteborg • 031-703 73 70 • info@statistikkonsulterna.se • www.statistikkonsulterna.se

Mikrosimulering

SESIM, en dynamisk mikrosimuleringsmodell

av Tomas Pettersson, Finansdepartementet

Mikrosimulering kan idag sägas vara en vedertagen metod för analyser av skatte- och transfereringssystemen, det vill säga såväl socialförsäkringar som olika bidrag, såväl i Sverige som i ett flertal andra länder. Det är framför allt statiska mikrosimuleringsmodeller som idag är allmänt förekommande men även dynamiska modeller blir alltmer vanliga. Vid Finansdepartementets ekonomiska avdelning används båda modelltyperna i det dagliga arbetet. I det följande ges en kort introduktion till mikrosimulering med fokus på Finansdepartementets dynamiska mikrosimuleringsmodell SESIM.

Statisk och dynamisk mikrosimulering

Som prefixet mikro indikerar utgår denna typ av modeller från den enskilda individen och hushållet. Detta gör det möjligt att inte bara studera medelvärden och summor för olika målvariabler utan även deras fördelning. Den statiska mikrosimuleringsmodellen utgår typiskt från en datakälla innehållande detaljerade uppgifter om individer och hushåll, till exempel olika typer av inkomstuppgifter och uppgifter om hushållets sammansättning. I modellen finns detaljerade beskrivningar av de aktuella skatte- och transfereringssystemen vilket möjliggör att individernas skatter, transfereringar och disponibla inkomster kan beräknas.

Det primära syftet med den statiska mikrosimuleringsmodellen är att studera egenskaper hos de offentliga skatte- och transfereringssystemen, till exempel att utvärdera effekter av olika reformförslag. Med effekter avses såväl totala budgeteffekter för den offentliga sektorn som effekter för den enskilde individen. Vem vinner respektive förlorar på reformen och hur bidrar reformen till en omfördelning av ekonomiska resurser mellan hushåll i olika inkomstlägen? En annan frågeställning är hur starka de ekonomiska incitamenten till arbete är för olika typer av hushåll och hur reformen påverkar detta.

På grund av eftersläpningen i de administrativa registren är de mest aktuella databaserna ungefär två år gamla. För att kunna göra beräkningar avseende aktuella förhållanden finns därför ett behov av att kunna göra prognoser till dagens datum. Vid dessa prognoser tillämpas ett "statiskt åldrande" där modellpopulationens sammansättning förändras utan att egenskaperna hos de ingående individerna påverkas. Detta åstadkoms genom en justering av individernas urvalsvikter varpå förändringar i demografisk sammansättning, sysselsättning och andra nyckeltal kan kontrolleras.

Ett exempel på en statisk mikrosimuleringsmodell är Statistiska centralbyråns (SCB) FASIT som i flera decennier har använts inom regeringskansliet och som under de senaste åren även fått en spridning utanför statsförvaltningen.

Den primära skillnaden mellan statiska och dynamiska mikrosimuleringsmodeller är det sätt på vilken modellpopulationen framskrivs. I den statiska modellen konstanthålls alltså individernas egenskaper under simuleringen. I en dynamisk modell uppdateras däremot egenskaperna hos den ingående modellpopulationen: individer blir äldre och utsätts för olika händelser, till exempel att flytta hemifrån, att börja studera, att få en inkomstförändring och så vidare. De processer som bedöms viktiga för att beskriva populationen och dess förändringar, och som är nödvändiga för de analyser som ska företas, estimeras med hjälp av statistiska modeller. De statistiska modellerna används sedan för att återge de i datakällorna observerade händelseförloppen genom Monte Carlo-simulering.

Liksom hos de statiska modellerna finns de offentliga systemen avbildade, dock ofta något mindre detaljerat. Det primära syftet med de dynamiska modellerna är att göra simuleringar över längre sikt för att analysera system där tidsaspekten är avgörande, till exempel ålderspensionssystem eller studiefinansieringssystem. För analyser av sådana longitudinella system är dynamiska modeller att föredra eftersom kumulativa processer är svårare att modellera i en statisk modell. En allmän diskussion av dynamisk mikrosimulering ges i [1].

Modellstruktur

SESIM består av ett stort antal stokastiska modeller och algoritmer som används för att uppdatera egenskaperna hos modellpopulationen. Uppdateringen tar hänsyn till de processer från den verkliga världen som bedöms påverka intressanta målvariabler. Modellpopulationen består av ett urval om cirka 110 000 individer ur 1999 års panel av LINDA-databasen (se [2]). Vid behov kan urvalet göras större. I vissa studier har cirka 300 000 individer använts. Modellpopulationen inkluderar även ett antal utlandsboende individer med svensk pensionsrätt, vilket har betydelse vid beräkningar av framtida pensionsutbetalningar.

Modellen simulerar de olika processerna en efter en i en rekursiv simulering. Simuleringen löper i diskreta steg avseende ett kalenderår, ett val som i huvudsak styrts av att också de tillgängliga datakällorna uppdateras årligen. Inom ett simuleringsår simuleras varje process en gång för varje individ. Somliga händelser leder till absorberande tillstånd, till exempel dödsfall och utflyttning från föräldrahemmet, och kan således inträffa maximalt en gång per individ.

Den rekursiva simuleringen innebär ett antagande om kausalitet mellan de olika processerna. De händelser som antas vara de mest basala, till exempel demografiska processer som mortalitet och fertilitet, placeras därför tidigt i sekvensen. Det är viktigt att man vid estimering

av de ingående modellerna beaktar den ordning i vilken de simuleras då riskpopulationen för en specifik process kan påverkas av utfallet av tidigare processer. Om till exempel händelserna A och B simuleras i en följd och riskpopulationen för B påverkas av utfallet av A bör riskpopulationen även vid estimering av B betingas på A. Tabell 1 beskriver översiktligt SESIM:s innehåll av modeller och deras inbördes ordning.

Eftersom SESIM huvudsakligen används för longitudinella analyser är det också viktigt att modellen på ett korrekt sätt återskapar den tidsdynamik som de olika variablerna uppvisar empiriskt. Därför är alla ingående modeller, med några få undantag, av longitudinell karaktär, allt från enklare transitionsmatriser (fast parameteriserade som logit- eller probitmodeller) till *random effects*-modeller för kontinuerliga eller diskreta variabler. Att använda modeller som estimerats på tvärsnittsdata vid longitudinell simulering tenderar att överdriva den intertemporala variationen. De flesta modeller har estimerats från LINDA-databasen, några från HEK-databasen och ytterligare några modeller från andra datakällor. HEK-databasen baseras på *Hushållens ekonomi*, en urvalsundersökning som genomförs årligen av SCB.

Simuleringarna resulterar i att varje individ i modellpopulationen klassificeras efter individens huvudsakliga

sysselsättning, eller status (barn, ålderspensionär, student, förtidspensionär, föräldraledig, arbetslös, sysselsatt, emigrerad och övrig) under året. Statusindelningen ger en förenklad bild av verkligheten eftersom många individer har flera sysselsättningar under ett år. Till exempel är det vanligt att studenter arbetar under delar av året, i synnerhet de år då studierna påbörjas och avslutas. Dessutom kan olika sysselsättningar pågå simultant, till exempel deltidstudier och förvärvsarbete. Förenklingen begränsar dock modellens komplexitet betydligt och gör den mer överblickbar.

Eftersom modellen simulerar enskilda individer utgör det också den minsta analysenheten. En av fördelarna med mikrosimuleringsmodeller är just flexibiliteten att kunna analysera godtyckliga grupper, allt ifrån den enskilda individen till hela populationen. En viktig faktor i de flesta analyser är hushållen och dess sammansättning. I baspopulationen utgörs modellhushållen av faktiskt observerade hushåll, varv vissa hushåll justeras dock initialt så att deras sammansättning bättre stämmer överens med det så kallade kosthushållet. De observerade hushållen ersätts successivt med syntetiska hushåll allteftersom simuleringen pågår. Nya hushåll bildas då individer flyttar ihop eller lämnar föräldrahemmet och splittras då individer avlider eller separerar. Information om hushållstillhörighet och hushållens

Tabell 1: Moduler/modeller i SESIM

demografi	mortalitet, adoption, emigration, immigration, fertilitet, flytt från föräldrahemmet, parbildning, separation, förtidspension (samt rehabilitering)
regional migration	flytt från region, flytt inom region, allokering till ny region
utbildning	avhopp från gymnasium, gymnasium till universitet, avhopp från universitet, arbetsmarknad till universitet, arbetsmarknad till KOMVUX, KOMVUX till universitet
arbetsmarknad	ålderspensionsbeslut, arbetslöshet, sysselsättning, sektorstillhörighet, antal dagar med sjukpenning, förvärvsinkomster
förmögenheter och bostäder	äga/hyra bostad, finansiell förmögenhet, real förmögenhet, privat pensionssparande, skuld, avkastning/utdelning, kapitalvinster, skuldränta
skatter och transfereringar	inkomstskatt, förmögenhetsskatt, fastighetsskatt, skatt på kapitalinkomster, socialavgifter, ålderspension, sjuk/aktivitetsersättning, efterlevandepension, arbetsmarknadsstöd, sjukpenning, föräldrapenning, barnbidrag, bostadsbidrag, bostadstillägg till pensionärer, underhållsstöd, studiestöd, socialbidrag, äldreförsörjningsstöd, återbetalt studielån, avtalspensioner
offentlig konsumtion	grundskola, gymnasium, universitet, vuxenutbildning, barnomsorg, äldreomsorg, arbetsmarknadsåtgärder, sjukvård, läkemedel
hälsa och vård	hälsoindex, slutenvårdsdagar, funktionshinder, äldreomsorg (inklusive överdödlighet), närhet till anhörig

sammansättning är viktig då många av SESIM:s modeller använder både individ- och hushållsvariabler. I inkomstfördelningsanalyser används också vanligen hushållets inkomster eftersom det ger en bättre bild av individernas ekonomiska standard. I hushållets inkomster ingår bl.a. hushållstransfereringar som bostadsbidrag eller socialbidrag, vilka är knutna till ett hushåll snarare än till en individ. Vid beräkningar av hushållens skatter och transfereringar används i SESIM kod som har hämtats från FASIT-modellen och anpassats till SESIM:s programspråk och variabelnamn. I sammanhanget kan det nämnas att SESIM är programmerad i Visual Basic 6.0 och innehåller cirka 40000 rader kod och i den nuvarande versionen finns cirka 220 variabler knutna till individen, cirka 85 till hushållet och cirka 300 makrovariabler.

SESIM utgör i sig inte en komplett beskrivning av samhällsekonomin. Till exempel simuleras populationens arbetsutbud utan att efterfrågan på arbetskraft beaktas. Ett annat exempel är att efterfrågan på vårdplatser modelleras men inte tillgången på sådana. Implicit antas att den modellerade komponenten är styrande, det efterfrågas så mycket arbetskraft som finns till hands och det finns så många vårdplatser som efterfrågas. Det är därför ofta lämpligt att använda en annan makroekonomisk bild än den som endogen genereras av SESIM om den makroekonomiska aspekten är viktig för analysen. Den makroekonomiska modellen FIMO, för en närmare beskrivning, se [3], används ofta för att generera makroekonomiska scenarier som SESIM sedan kalibreras till. Ett liknande problem gäller den antagna demografiska utvecklingen. SESIM kan i sig ses som en modell för befolkningsframskrivningar, betydligt mer disaggregerad än de kohort komponent-modeller som traditionellt används i detta syfte. Ofta är det dock att föredra att kalibrera SESIM till den av SCB producerade officiella befolkningsprognosen för riket. Med exogen bestämd makroekonomisk och demografisk utveckling används de aktuella modellerna i SESIM i första hand för att fördela olika ekonomiska och demografiska händelser till 'rätt' individer, till exempel att det inom ramen för en given total dödlighet i en viss åldersgrupp avlider fler individer i en undergrupp som har högre risk än åldersgruppens genomsnitt. För vidare diskussion om kalibrering av mikrosimuleringsmodeller, se till exempel [4].

För en mer detaljerad presentation av SESIM se [5] eller appendix A till [6].

En dynamisk mikrosimuleringsmodell tenderar att bli mycket komplex. De många ingående stokastiska modellerna och deras relationer kan vara svåra att överblicka. Detta ställer också höga krav på att modellerna, de ingående delmodellerna såväl som hela mikrosimuleringsmodellen, löpande valideras och rimlighetsbedöms. En detaljerad validering av SESIM:s inkomstvariabler presenteras i [6]. Eftersom SESIM simuleras från år 1999 finns möjligheter att jämföra modellresultaten mot faktiska utfall under några år.

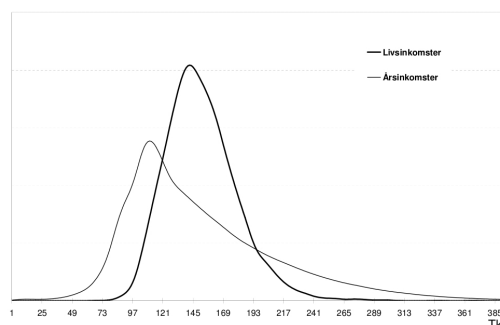
Något som är relevant vid validering, såväl som vid skarpa analyser, är inferensaspekten. Detta bortses vanligen

från, som bland annat påpekas i [4], på grund av de dynamiska modellernas komplexitet. Det gör tyvärr att det är svårt att uttala sig om huruvida ett resultat som avviker från ett observerat utfall verkligen är en signifikant avvikelse. Att uppskatta olika variabelers Monte Carlo-varians låter sig lätt göras genom att simulera modellen i ett antal replikat med varierande slumpföro, men att dessutom ta hänsyn till variansen med avseende på estimerade parametrar och samplingen av baspopulationen är mer komplicerat. Detta är ett område som behöver belysas ytterligare.

Några exempel på tillämpningar

Under sitt nu tioåriga liv har SESIM använts vid studier inom olika områden. I sin första version fokuserade modellen på utbildningsfrågor och användes bland annat för att göra beräkningar inför reformeringen av studiemedelssystemet, se [7]. Senare kom ålderspensioner att hamna i fokus och SESIM byggdes till stor del om för att anpassas till detta. Analyser av ålderspensionssystemet, till exempel utgiftsprognoser och analyser av systemets långsiktiga finansiella stabilitet, kommer också fortsättningsvis att vara en av de viktigaste fokusområdena för SESIM. Modellen har också använts i akademiska studier av pensionssystemet. I [8] beräknas ersättningsgraden, det vill säga i princip kvoten mellan inkomsterna efter och före pensioneringen, med hänsyn taget till såväl offentlig pension som avtalspension och privat pensionssparande.

Ett annat användningsområde är longitudinella inkomstfördelningsstudier. I [6] används SESIM för att vidga det traditionella ettårsperspektivet i inkomstfördelningsanalyser genom att istället se till fördelningen av livsinkomster, se även [9]. Genom den utjämning av temporära inkomstvariationer som åstadkoms då längre tidsperioder studeras erhålls en betydligt jämnare livsinkomstfördelning jämfört med en motsvarande årsinkomstfördelning, se Figur 1.



Figur 1: Inkomstfördelningar från ett enskilt år respektive över livet

Vidare beräknas i studien den omfördelning mellan individer, och inom en individs liv (över tiden), som uppstår genom transaktionerna mellan individerna och den offentliga sektorn, till exempel genom inbetalningar av skatter och utbetalningar av transfereringar. I studien konstateras att drygt 80% av de transfereringar och subventioner som individen får finansieras av egna skatter

och avgifter som betalats vid någon tidpunkt i livet.

I [3] används SESIM för att studera den offentliga sektorns omfördelning av resurser mellan olika generationer. Efter en omfattande kartläggning av olika generationers historiska och förväntade framtida in- och utbetalningar till välfärdssystemen beräknas individernas livsnetto gentemot den offentliga sektorn. En slutsats är att det förekommer omfördelningar mellan generationerna men att den är relativt liten. Resultaten talar dessutom mot den vanliga föreställningen om att 40-talisterna skulle vara särskilt gynnade. Det är i stället de egna inkomsterna förvärvade på arbets-, kapital- och bostadsmarknaden som förklarar vilka konsumtionsmöjligheter olika generationer har. Den omfördelning som sker mellan generationer är också mindre än den omfördelning som sker inom en generation.

Ytterligare en tillämpning av SESIM tillkom genom forskningsprojektet "The old Baby-Boomers", se [10], som syftade till att studera problematiken kring den åldrande befolkningen i Sverige. Resultaten presenteras i sin helhet i boken *Simulating an ageing population. A microsimulation approach applied to Sweden* [11] som under 2008 publiceras i Elseviers serie *Contributions to Economic Analysis*. Mer specifikt analyserades situationen för de stora efterkrigsgenerationerna som nu är på väg att lämna arbetsmarknaden för att pensioneras. Hur belastar detta de offentliga systemen i framtiden och vilken effekt får det för de aktuella kohorterna? I studien undersöks, genom simulering, ett antal frågeställningar kring bland annat inkomst- och förmögenhetsfördelning, bostadsmarknad, hälsofrågor och utnyttjande av sjukvård. Projektet är ett exempel på ett lyckat samarbete kring ett gemensamt modellverktyg. Genom att få tillgång till en redan existerande modell kunde man spara tid och resurser inom forskningsprojektet. Samtidigt kom SESIM genom samarbetet att utvecklas och förbättras vilket i sin tur gynnade Finansdepartementet. ■

Referenser

- [1] O'Donoghue, C. (2001) Dynamic Microsimulation: A Methodological Survey. *Brazilian Electronic Journal of Economics*. 4(2).
- [2] Edin, P.-A. and Fredriksson, P. (2000). *LINDA - Longitudinal Individual DATA for Sweden*. Working Pa-

per Series 2000:19. Uppsala University, Department of Economics.

- [3] Pettersson, T., Pettersson, T. and Westerberg, A. (2006). *Generationsanalyser - omfördelning mellan generationer i en växande välfärdsstat*. ESS, 6.
- [4] Klevmarken, N. A. (1998). *Statistical Inference in Micro Simulation Models: Incorporating external information*. Working Paper Series 1998:20. Uppsala University, Department of Economics.
- [5] Flood, L., Jansson, F., Pettersson, T., Pettersson, T., Sundberg, O. and Westerberg, A. (2005). *SESIM III - a Swedish dynamic microsimulation model*. Available at www.sesim.org.
- [6] Pettersson, T. och Tomas Pettersson, T. (2003) *Fördelning ur ett livscykelerspektiv*, bilaga 9 till Långtidsutredningen 2003/04. Fritzes, Stockholm. SOU 2003:110.
- [7] Ericson, P. och Hussénus, J. (2000). *Studiebidraget i det långa loppet*. Rapport till Expertgruppen för studier i offentlig ekonomi (ESO) Ds 2000:19.
- [8] Flood, L. (2007). Can We Afford the Future? An Evaluation of the New Swedish Pension System. Chapter 15 in Elseviers bookseries *International Symposia in Economic Theory and Econometrics*. Editors Harding, A. and Gupta, A.
- [9] Pettersson, T. and Pettersson, T. (2007). Lifetime Redistribution Through Taxes, Transfers and Non-Cash Benefits in Sweden. Chapter 8 in Elseviers bookseries *International Symposia in Economic Theory and Econometrics*. Editors Harding, A. and Gupta, A.
- [10] Klevmarken, N. A., Bolin, K., Eklöf, M., Flood, L., Fransson, U., Hallberg, D., Höjgård, S., Lindgren, B., Mitrut, A. and Lagergren, M. (2007). *Simulating the future of the Swedish baby-boom generations*. Working Paper Series 2007:26. Department of Economics, Uppsala University.
- [11] Klevmarken, N. A., Lindgren, B., Bolin, K., Lagergren, M., Flood, L., Eklöf, M., Hallberg, D. and Fransson, U. (kommande). *Simulating an ageing population. A microsimulation approach applied to Sweden*. Elsevier.

Presentation av ny sekreterare i Styrelsen – Lars Rönnegård

Jag bor i Djura utanför Leksand och är 37 år. Jag är tjänstledig från mitt jobb vid Högskolan Dalarna där jag är lektor i statistik. Min forskning är främst inriktad mot statistisk genetik och fram till april 2008 jobbar jag som forskare vid Centrum för Bioinformatik, Uppsala universitet, där jag ingår i ett forskarlag (www.computationalgenetics.se) som utvecklar metoder för att identifiera gener hos husdjur. Förutom statistisk

genetik har jag även bedrivit forskning med tillämpningar inom ekonometri och ekologi. Under de senaste två åren har jag också inriktad mig mot numerisk optimering där jag har ett samarbete med Avdelningen för teknisk databehandling vid Institutionen för informationsteknologi, Uppsala universitet. Jag ser fram emot en lärorik och spännande tid som sekreterare i Samfundets styrelse. ■

Teoretiska artiklar

Andreas Nordvall Lagerås tilldelades priset Årets yngre statistiker vid Samfundets årsmöte. Här presenterar Andreas kortfattat en del av sin avhandling.

Förnyelse teori för inversa subordinatorer – en genväg via Coxprocesser

av Andreas Nordvall Lagerås,
Centrum för teoretisk biologi och Institutionen för matematiska vetenskaper, Chalmers och Göteborgs universitet

Min avhandling som jag lade fram i våras bestod av fyra olika artiklar. På statistikersamfundets årsmöte presenterade jag de två senaste artiklarna som handlar om förgreningsprocesser och koalescensprocesser. Jag tänkte ta tillfället i akt och här presentera en annan artikel i avhandlingen. Artikeln heter "A renewal-process-type expression for the moments of inverse subordinators" och publicerades i *Journal of Applied Probability* [1].

Huvudidén är att vi kan förstå egenskaper för en kontinuerlig stokastisk process – här en invers subordinator – genom att studera egenskaper för en till denna associerad punktprocess – här en Coxprocess. Inversa subordinatorer är en sorts kontinuerlig motsvarighet till förnyelseprocesser, så vi förväntar oss egenskaper som liknar förnyelseprocessers.

Förnyelseprocesser

Förnyelseprocesser beskriver hur händelser inträffar i tiden med den speciella egenskapen att processen startar om då en händelse inträffar, det vill säga vid det ögonblicket beror den inte på sin historia och fördelningen för dess framtid är densamma som vid alla andra händelser. Hela processen karakteriseras av vad det är för fördelning för tiden T mellan händelserna. Om vi befinner oss mellan två händelser och vi vet att det var 20 minuter sedan förra händelsen och T exempelvis är likformig på $(0, 30)$ så vet vi att återstående tid är likformig på $(0, 30)$, det vill säga mellan två händelser kan historien spela roll för processens framtida utveckling.

Undantaget är Poissonprocesser, där tiden mellan händelser har en exponentialfördelning, och händelserna inträffar med samma intensitet λ hela tiden, oberoende av varandra.

Det är i många fall svårt att exakt ta fram fördelningen för N_t , antalet händelser fram till t , och att kvantifiera beroendet mellan hur många händelser som inträffar i disjunkta tidsintervall. Man kan dock med T 's fördelning uttrycka momenten för N och även simultana moment för till exempel N_s och $N_t - N_s$ då $0 < s < t$. Alla moment, även simultana, av högre ordning kan man ta fram om man känner till $U(t) := \mathbb{E}[N_t]$, den så kallade förnyelsefunktionen.

Specialfallet Poissonprocessen har förstås $N_t \sim \text{Po}(\lambda t)$ och momenten uttrycks enklast med *faktoriella moment* i stället för ordinära moment. Låt $x^{[k]} := x(x-1) \cdots (x-k+1)$. Vi deriverar den genererande funktionen för Poissonfördelningen och sätter argumentet till 1

och får

$$\mathbb{E}[N_t^{[k]}] = \frac{d^k}{ds^k} e^{\lambda t(s-1)} \Big|_{s=1} = (\lambda t)^k,$$

och i allmänhet

$$\mathbb{E} \left[\prod_{i=1}^n (N_{t_i} - N_{s_i})^{[k_i]} \right] = \prod_{i=1}^n (\lambda(t_i - s_i))^{k_i},$$

för disjunkta intervall $(s_i, t_i]$, $i = 1, \dots, n$.

För förnyelseprocesser i allmänhet ser det istället ut så här

$$\mathbb{E} \left[\prod_{i=1}^n (N_{t_i} - N_{s_i})^{[k_i]} \right] = \prod_{i=1}^n k_i! \int_C \prod_{j=1}^k U(dx_j - x_{j-1}),$$

där C är en särskild delmängd i $\mathbb{R}^k$, $k = k_1 + \dots + k_n$.

Inversa subordinatorer

Egenskapen att processen startar om vid vissa tillfällen kan man utvidga även till en process $(\tau_t : t \geq 0)$ som inte räknar händelser som förnyelseprocessen, utan istället mäter någon kvantitet som ökar kontinuerligt. Vi kan tänka oss en reservoar som då och då fylls på med vatten, och vid varje ny, infinitesimalt liten, droppe som tillkommer så glöms historien bort och de framtida tillskotten har samma fördelning som efter varje annan vattendroppe. Det kan dock finnas perioder utan något inflöde och dess längder följer någon fördelning.

Processen $(\tau_t : t \geq 0)$, som kallas en *invers subordinator*, karakteriseras av tiden det tar från en omstart tills dess den har ökat en enhet, till exempel tiden till en ytterligare kubikmeter har strömmat in i reservoaren, vilket direkt kan jämföras med förnyelseprocesser som också karakteriseras av en sådan fördelning, med skillnaden att tillskotten för dem alltid kommer en och en.

Min artikel handlar om hur man kan ta fram egenskaper för inversa subordinatorer som har motsvarigheter i teorin för förnyelseprocesser. Exempelvis så finns det under vissa förutsättningar stationära fördelningar för processerna om man fördröjer dem, men här tänker jag fokusera på hur man beräknar moment för inversa subordinatorer. Det är inte uppenbart hur man kan få fram sådana resultat eftersom vanlig förnyelse teori utnyttjar att N_t är heltalsvärd.

Coxprocesser

Förnyelseprocesser kan ses som en generalisering av Poissonprocesser genom att man tillåter andra fördelningar än exponentialfördelningen för tider mellan händelser. Ett annat sätt att generalisera Poissonprocesser är att låta intensiteten variera med tiden så att $N_t \sim \text{Po}(L(t))$, där $(L(t) : t \geq 0)$ är någon icke-avtagande deterministisk funktion. Man talar då om en icke-homogen Poissonprocess. Denna har kvar oberoendet, men tiderna mellan händelser är inte längre likafördelade i allmänhet. Fortfarande gäller $\mathbb{E}[N_t^{[k]}] = L(t)^k$.

Om man låter $(\Lambda(t) : t \geq 0)$ vara en växande stokastisk process och givet ett utfall $\Lambda = L$ låter händelserna inträffa enligt en icke-homogen Poissonprocess med denna kumulativa intensitetsfunktion så får man en så kallad Coxprocess. Då gäller

$$\mathbb{E}[N_t^{[k]}] = \mathbb{E}[\Lambda(t)^k],$$

och i allmänhet

$$\mathbb{E}\left[\prod_{i=1}^n (N_{t_i} - N_{s_i})^{[k_i]}\right] = \mathbb{E}\left[\prod_{i=1}^n (\Lambda(t_i) - \Lambda(s_i))^{k_i}\right].$$

Observera att faktoriella moment för N motsvaras av ordinära moment för Λ .

Det kan finnas beroende över disjunkta tidsintervall på grund av beroenden i fördelningen för $(\Lambda(t) : t \geq 0)$:

$$\begin{aligned}\mathbb{E}[N_s(N_t - N_s)] &= \mathbb{E}[\Lambda(s)(\Lambda(t) - \Lambda(s))] \\ &\neq \mathbb{E}[\Lambda(s)]\mathbb{E}[\Lambda(t) - \Lambda(s)] \\ &= \mathbb{E}[N_s]\mathbb{E}[N_t - N_s].\end{aligned}$$

Coxprocesser är i allmänhet inte alls relaterade till förnyelseprocesser eftersom de båda typerna av processer generaliserar Poissonprocessen på helt olika sätt.

Genvägen

Läsaren frågar sig säkert nu vad Coxprocesser har med inversa subordinatorer att göra. Det är känt, men kanske

inte allom bekant, att en Coxprocess är en förnyelseprocess om och endast om $(\Lambda(t) : t \geq 0)$ är en invers subordinator. Vi kan nu använda detta resultat på följande sätt. Vad vi egentligen vill ta reda på är något uttryck för momenten för en invers subordinator $(\tau_t : t \geq 0)$. Vi tänker oss nu att $(\tau_t : t \geq 0)$ agerar slumpmässig intensitet i en Coxprocess $(N_t : t \geq 0)$, det vill säga τ fungerar som Λ ovan. Då har vi

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^n (\tau_{t_i} - \tau_{s_i})^{k_i}\right] &= \mathbb{E}\left[\prod_{i=1}^n (N_{t_i} - N_{s_i})^{[k_i]}\right] \\ &= \prod_{i=1}^n k_i! \int_C \prod_{j=1}^k U(dx_j - x_{j-1}).\end{aligned}$$

där första likheten gäller eftersom $(N_t : t \geq 0)$ är en Coxprocess med $(\tau_t : t \geq 0)$ som kumulativ intensitet, och den andra likheten gäller eftersom $(N_t : t \geq 0)$ också är en förnyelseprocess då $(\tau_t : t \geq 0)$ är en invers subordinator. Vidare så är $U(t) = \mathbb{E}[N_t] = \mathbb{E}[\tau_t]$.

Summa summarum kan vi alltså använda exakt samma uttryck när vi räknar på ordinära moment för inversa subordinatorer som när vi räknar på faktoriella moment för förnyelseprocesser. Detta resultat är nytt.

Andra, tidigare kända, resultat för inversa subordinatorer, till exempel att de kan fördröjas för att få en stationär fördelning, kan också visas genom att ta denna genväg över Coxprocesser som just med denna konstruktion också är förnyelseprocesser. Slutligen bör också nämnas att inversa subordinatorer faktiskt används i tillämpningar, exempelvis finansmatematik, dock inte i tillräckligt stor utsträckning som förnyelseprocesser. Det är emellertid så att inversa subordinatorer naturligt dyker upp på olika ställen i sannolikhetsteorin, särskilt som så kallad "lokal tid" för Markovprocesser, tack vare själva förnyelseegenskapen. ■

Referenser

- [1] Lagerås, A.N. (2005). A renewal-process-type expression for the moments of inverse subordinators. *Journal of Applied Probability* **42**(4).

Presentation av ny ledamot i Styrelsen – Jesper Rydén

Jag är universitetslektor i matematisk statistik vid matematiska institutionen, Uppsala universitet. Efter disputation i Lund våren 2002 tillbringade jag postdocperioder dels vid forskningsinstitutet Ifremer i Brest (se Qvartilen 18(3), sidan 14–15), dels vid Malmö högskola.

Mina forskningsintressen rör oftast tillämpningar inom oceanografi, meteorologi och liknande vetenskaper. En mängd problem från dessa områden kan formuleras och lösas med stokastiska modeller och metoder, hämtade från områden som stokastiska processer och fält samt extremvärdesanalys. På sistone har jag arbetat en del

med skattningsproblematik av så kallade återkomstvärden (100-årsvärden etcetera), ofta av intresse att beräkna för storheter som våghöjd och vindstyrka.

Jag är road av att undervisa och i Uppsala undervisar jag inom matematisk statistik ofta på kurser som är inriktade mer åt det statistiska hållet. En utmaning är att på bästa sätt introducera statistisk programvara.

Avslutningsvis ser jag fram emot en stimulerande, spännande och kreativ tid i styrelsen med anledning av samgåendet våren 2008. ■

Debatt

Fritt forum för medlemmar samt gästskribenter. En text får ej innehålla angrepp på enskilda individer. Styrelse och redaktion räknas ej som enskilda individer i sina respektive roller. I övrigt gäller att ansvarig utgivare ej skall kunna åtalas för innehållet.

Styrelsens svar på debattinlägg i 22-3

Debattinslag är välkomna i Qvartilen. De kan avse problem och frågor av olika slag: mer statistiskt orienterade frågeställningar (exempelvis metod- och modellval i vissa situationer), utbildningsfrågor eller styrelsearbete och dess former. Serien med debattinlägg av Mikael Möller är ett tydligt exempel på debatt av sistnämnda slag, och den visar att det under nuvarande former verkligen går att kommentera och kritisera styrelsen.

På omslaget till Qvartilen står att den är Svenska statistikersamfundets tidskrift. Ansvarig utgivare är Statistikersamfundets ordförande. Det är styrelsens uppfattning att

styrelsen naturligt är löpande ansvarig för Qvartilen, liksom för Samfundets övriga verksamhet. Detta utesluter dock inte att kritik mot styrelsearbetet kan föras fram i tidskriften. Det är redaktören och redaktionen som framställer Qvartilen. Styrelsen anser att såväl styrelse som medlemstidskrift är till för medlemmarna, och styrelsen ser dem inte som motsatser.

En ny formulering om äganderätt infördes i föregående nummer av Qvartilen (se nummer 22-3, sidan 2) oberoende av Mikael Möllers debattinlägg.

Styrelsen

Det behövs en ny högskoleförordning

av Per-Åke Lindblom

En engelsk språkvetare har skrivit att det på 1950-talet var omöjligt att förutse den dominanta ställning som engelskan skulle få i början av 2000-talet. Engelskan är idag Internets och datavärldens huvudspråk; engelskan har också en mycket stark ställning vad gäller film, bokutgivning, tidningar, flygplatser och luftfartstrafik, internationell handel och akademiska konferenser, vetenskap och teknik, medicin, sport, populärmusik och reklam. Denna expansion har varit särskilt snabb de två senaste decennierna, vilket framför allt hänger samman med datateknologins och Internets framväxt, den så kallade informationsrevolutionen. Det avgörande med denna är att man kan spara nästan obegränsade mängder information, som dessutom i stor utsträckning är sökbar, och som kan förmedlas sekundsnabbt över hela världen. Detta har ofelbart gynnat det engelska språket, eftersom datateknologin, både hård- och mjukvara, liksom Internet först utvecklades i USA, som dessutom är världens starkaste ekonomi. De som först insåg datateknologins och Internets praktiska betydelse var naturligtvis universitetet, så även i Norden. De nordiska länderna har tagits på sängen av engelskans expansion inom högskolan, inte minst i Sverige, där alla regeringar, oavsett partifärg, hittills har tillämpat en låt-gå-politik i fråga om svenska språkets ställning. Under 1990-talet inleddes diskussionerna om de nordiska språkens relation till engelskan och språksituationen överhuvudtaget i respektive land – den svenska "Mål i mun2"-utredningen tillsattes exempelvis 2000 och avlämnade sitt betänkande 2002. Men så sent som 2004 omnämnde den socialdemokratiska regeringens högskoleproposition, "Ny värld – ny högskola" 2004, inte med ett ord svenskans ställning på landets

högskolor. De domänförluster som de facto redan skett inom högskolan i Sverige har alltså inte föranlett några motåtgärder från högsta politiskt ansvarigt håll. Anglifieringen av den svenska högskolan understöds dessutom av Bologna-processen.

I Danmark inledde det så kallade rektorkollegiet diskussionen om språksituationen på högskolan 2003 genom publicering av en rapport med rekommendationer. Därefter har ett antal universitet och högskolor i Norden antagit eller utarbetat förslag till språkplaner. I Norge har till och med en arbetsgrupp tillsatt av norska högskoleverket lagt fram förslag till nationell språkpolitik för universitets- och högskolesektorn. Här i Sverige har Teknisk-naturvetenskapliga fakultetsnämnden vid Uppsala universitet, Göteborgs universitet och Kungliga Konsthögskolan i Stockholm antagit språkplaner – detta i motsats till övriga drygt 40 universitet och högskolor.

Högskolans särskilda ansvar

Försvaret av svenska språket vilar på tre huvudargument

1. Alla språk har ett inneboende egenvärde, det vill säga även svenskan, som är majoritetsspråket i Sverige.
2. Det är en mänsklig rättighet att kunna använda sitt eget språk.
3. Det funktionellt-demokratiska argumentet, det vill om det sker ett språkbyte, i det här fallet inom högskolan, kommer det att uppstå en klyfta mellan den så kallade bildade eliten och svenska folket, förmedling av vetenskapliga rön och resultat kommer att försvåras, vilket ytterst sett kommer att få negativa konsekvenser för demokratin.

Universitet och högskolor har ett särskilt ansvar att upprätthålla svenska som ett vetenskapligt språk, i att förmedla relevanta vetenskapliga resultat till samhället på begriplig svenska och med en heltäckande terminologi. Samtidigt består högskolans uppgift i att ge studenterna bästa möjliga utbildning, både för deras egen skull och till nytta för landet. Därför måste också språkfrågan brytas ner i ett antal delaspekter, vilket sker i de flesta av de nämnda språkplanerna, i fråga om undervisnings-, examinations- och avhandlingsspråk liksom kurslitteratur. Dessutom bör man utgå från ett rättighetsperspektiv.

Enspråkighet, parallellspråkighet eller mångspråkighet

Det finns vissa högskoleledningar i Norden, som uttryckligen har ställt som mål att högskolan inom en viss tidsram helt skall gå över till engelska som undervisningsspråk. En av de ledande förespråkarna för denna linje i Sverige är Anders Flodström, numera chef för Högskoleverket, som i en intervju i Campus säger:

Jag tror att det är en naturlig utveckling, men att tidsperspektivet är relativt långt. Förmodligen leder globaliseringen till att vi har undervisning helt på engelska om tio, femton år. [...] Vi blir också bättre och bättre på att undervisa på engelska. Om tre, fyra år är skillnaden mellan att gå en kurs i London eller Stockholm försumbar.

Bortsett från de negativa konsekvenser som en övergång till engelska som undervisningsspråk skulle ha för svenskan som samhällsbärande språk, vilken dessutom på sikt skulle skada Sveriges konkurrenskraft snarare än stärka den, uttrycker Flodströms uttalande en viss naivitet. Lärare som inte är perfekt tvåspråkiga från födsel eller uppväxt kommer alltid att tala sämre engelska än de som har engelska som modersmål.

Parallellspråkighet anges ofta som ett alternativ till en hotande engelsk enspråkighet på högskolan. Parallellspråkighet innebär att två eller flera språk – läs svenska och engelska – ska kunna användas sida vid sida med varandra. Men om man önskar att högskolans personal och studenterna i Sverige ska behärska fler främmande språk än engelska, exempelvis tyska, franska, spanska, kinesiska, ryska, arabiska eller kiswahili, bör man hellre tala om mångspråkighet, för att låna ett argument från den norska professorn i pedagogik, Birgit Brock-Utne.

Begreppet mångspråkighet har alltså en klarare – och mer offensiv – innebörd. Parallellspråkigheten har för övrigt redan gått förlorad vad gäller avhandlingar och forskningspublikationer inom den naturvetenskapliga sektorn.

Parallellspråkighet kan lätt missförstås som att det enbart gäller att reglera förhållandet mellan de båda dominerande språken inom den svenska högskolan, engelskan och svenskan. Kungliga Konsthögskolans i Stockholm språkpolicy fokuserar till exempel endast på förhållandet mellan svenska och engelska – alla andra språk är som bortblåsta. Det är viktigt att inse att andra främ-

mande språk, inte bara tyska, franska och spanska, kan tjänstgöra som motvikter till engelskan. EU:s målsättning är att alla EU-medborgare skall behärska två främmande språk förutom sitt modersmål. Svenska handelshögskolan i Helsingfors kan däremot tjäna som föredöme. Där måste alla studenter avlägga kurser i de två inhemska språken och i två främmande språk. Nu när Bologna-reformen är införd, så handlar det om sex studiepoäng per språk (24 sp) på kandidatnivån. Dessutom måste alla läsa minst sex studiepoäng till i något valbart språk på magisternivå. Totalt blir det alltså 30 studiepoäng i språk för en magister, och poängen fördelas på minst fyra språk.

I sitt svar den 30/8 2006 på Språkförsvarets rundbrev skriver företrädare för Örebro universitet:

Enligt Europarådets direktiv bör varje EU-medborgare behärska minst två språk utöver det egna, vilket innebär ytterligare ett främmande språk utöver det obligatoriska språket engelska. Mot bakgrund av detta och den ökade internationaliseringen, inte minst inom utbildningen, borde det vara en självklarhet för varje högskola/universitet att befrämja språkutbildningarna, en utbildning som redan på grund- och gymnasienivå borde premieras. Språkkunskaper är ett nödvändigt verktyg även vid studier av andra ämnen än språk och fyller således ett inom- och tvärvetenskapligt behov.

Några åtgärder

2006 uppskattades antalet kurser på engelska på svenska universitet och högskolor till 5000, eller 12,5% av alla kurser, en andel som säkerligen har ökat sedan dess. Både sunda förnuftet och all seriös och etablerad forskning säger att användning av modersmålet ger det bästa inlärningsresultatet. Därför bör undervisningsspråket inom högskolan normalt vara svenska och studenter med svenska som modersmål ska alltid ha rätt till undervisning på svenska (utom i rena språkämnen), särskilt på grundkursnivå. Ingen kurs ska kunna erbjudas på engelska, om den inte först har erbjudits på svenska. Givetvis kan undervisning också ske på engelska för gäst- och utbytesstudenter och på forskarnivå, om det senare är det mest ändamålsenliga.

En svenskspråkig student ska alltid ha rätt att examineras på sitt eget språk – utom i rena språkämnen. Utländska studenter ska ha rätt att examineras på engelska eller annat språk, om behöriga examinatorer finns tillgängliga. Vetenskapliga artiklar och avhandlingar kan skrivas på svenska, danska, norska, engelska, eller annat främmande språk, främst tyska, franska och spanska. En avhandling som skrivs på svenska eller annat jämbördigt språk enligt ovan förutom engelska, ska förses med en fullständig sammanfattning på engelska. En avhandling som skrivs på engelska eller annat främmande språk ska likaledes förses med en fullständig sammanfattning på svenska. Det enda som ska ha betydelse för val av avhandlingsspråk är om en avhandling kommer att kunna utsättas för en vetenskaplig granskning inom ämnesområdet el-

ler inte.

Tyvärr existerar det ingen sammanställning på nationell nivå över den engelskspråkiga kurslitteraturens andel kontra den svenskspråkiga inom den svenska högskolan. Samma argument som är giltiga för val av undervisningsspråk är dock giltiga för val av kurslitteratur. En norsk undersökning utförd av *Norsk institutt for studier av forskning og utdanning* (NIFU) visade att andelen engelskspråkig kurslitteratur på grundutbildningsnivå sjönk från 49% 1979 till 47% 2000, medan andelen norskspråkig kurslitteratur ökade från 38% till 49% under samma tid. Detta visar att andelen engelskspråkig kurslitteratur går att påverka genom en medveten språkpolitik vid högskolorna liksom också genom stödåtgärder från statsmakternas sida.

Det måste utarbetas en ny högskoleförordning, som re-

glerar frågan om arbets- och kommunikationsspråk, undervisningsspråk och examinations- och avhandlingsspråk. En ny högskoleförordning bör också präglas av ett rättighets- och mångspråkighetsperspektiv med ett huvudspråk som grund. I väntan på en ny högskoleförordning bör de lokala universiteten och högskolorna uppmuntras att utarbeta lokala språkplaner. Detta är ett utmärkt sätt för att medvetandegöra såväl högskoleledningar som undervisande personal och studenter om språkfrågans betydelse: bättre en dålig språkplan än ingen alls. Den som vill spela på den europeiska eller globala arenan kan inte göra detta förutan grundlig debatt och eftertanke. ■

Inlägget bygger på en utförligare text, *Förslag till språkpolitik för den svenska högskolan*, som finns utlagd på Språkförsvarets webbplats – www.sprakforsvaret.se/sf/

Kåseri om statistik som tillämpning

Vilka vill ha statistisk kvalitetsredovisning?

av Olle Sjöström

De så kallade totalsfelsmodellerna för statistiska surveyundersökningar var en viktig del av Tore Dalenius pedagogiska budskap som jag tog till mig. De ställde inte endast krav på formaliserade matematiska modeller utan kanske främst på den statistiska praktiken, både planeringen av en undersökning och implementeringen av statistikens resultat. Fel skulle helst vara mätbara och de skulle deklarerats, formaliserade krav på "optimal allokering" skulle helst vara uppfyllda och utvärdering (evalvering) var en dygd för att förbättra kommande undersökningar. *Detta var närmast omöjliga mål att realisera men var en stödjande tankemodell för det praktiska arbetet.* Detta verktyg fick för mig ett mer lödigt och realistiskt innehåll genom att Deming förde in "medspelaren"/beställaren i bilden (se kåseri nummer 22-3). Beställaren, även kallad uppdragsgivaren, hade ju ett ansvar för att godkänna "det sanna värdet" (Y) och att ta sitt ansvar för att kräva besked om fel i skattningar, vilket vanligen betyder extra kostnader! Statistikern fick en tydligare roll, och som Deming uttryckte det [1], en arkitektroll, och skulle ta ansvar i hela "designprocessen".

Beställarens ansvar

Mina erfarenheter av praktiken vid utredningsinstitutet vid Statistiska centralbyrån (SCB) blev dock att det var svårt att få en "beställare" att ta ansvar. Den som betalade ville inte alltid ta sådant ansvar utan delegerade uppdraget nedåt och det ledde inte alltid till ett lyckat samarbete. Då det gällde uppdrag från en statlig utredning att genomföra en undersökning kunde det vara ännu svårare att hitta någon som tog ansvar. Det var märkligt nog så att statistikern – av beställarens bekvämlighets-skäl – ibland förväntades ta beslut i frågor som hörde

till beställarens ansvar. Ett illustrativt exempel var en situation vid planering av en rökaneundersökning i början av 1960-talet vid utredningsinstitutet. Frågan uppstod om det var nödvändigt att skilja mellan halsbloss, gombloss respektive enkel inhalering av röken – eller bara ha två kategorier, halsbloss och inhalering. Självklart var det den medicinska expertisens ansvar att svara på frågan om "gombloss" var en relevant definition för den statistiska analysen. Men inblandade läkare kunde inte komma överens om ett svar på frågan utan hänsköt beslutet till statistikern, som dock med rätta avvisade erbjudandet. Läkarna fick till slut lov att bestämma sig! En annan erfarenhet vid utredningsinstitutets undersökningar gällde kvalitetsredovisning. Långt ifrån alla beställare var intresserade av att konfidensintervall publicerades eller föreföll bry sig om bortfall och mätfel. Men våra ambitioner var höga, satta av den beslutsteori som logiskt motiverar ovan nämnda modeller för "totalfelet" och av det "överjag" som Tore Dalenius inplanterat i oss, och vi hade ingen tanke på att sänka nivåerna. Det gällde tydligen bortfallet, som sällan översteg tio procent och i enstaka fall låg under en procent. Det fick kosta och vid utredningsinstitutet erbjöd vi inte heller sänkta kostnader genom att sänka ambitionerna.

Hur dålig är officiell statistik?

Men bristen på intresse för kvalitetsredovisning (eller felredovisning) gällde också själva ämbetsverket, Statistiska centralbyrån. Det var i alla fall min uppfattning och sedan jag slutat fick jag en artikel publicerad i *Dagens Nyheter* under rubriken *Hur dålig är officiell statistik?* Den kritiska frågan gällde det faktum att användare av den officiella statistiken inte fick veta något om fel och

därför saknade möjlighet att ta ställning till olika slag av fel i undersökningar. Det var en plädering för kvalitetsredovisning och därmed för så kallade evalveringsundersökningar (utvärderande undersökningar) som möjliggjorde sådan redovisning. Det blev ett antal diskussionsinlägg, både för och emot, i tidningen. Det var givetvis en besvikelse att statistikerns intresse för fel, eller kvalitetskomponenter i statistiska data, inte var besvarat. Det föreföll mig som om risker i beslut eller slutledningar, betingade av fel, inte verkade bekymra beställare eller användare av statistikens resultat. Till detta kom en annan besvikelse. I början av 1960-talet hade jag sammanställt en formelsamling som underlag för programmering som speciellt gällde skattningar av konfidensintervall. Det var ett generellt program för olika flerstegsurval och olika estimatorer. Med den tidens datorkraft blev det en mycket svår uppgift för IBM, trots att de inledningsvis hade talat om att det skulle bli en "övningsuppgift". Programmet blev inte klart förrän 4–5 år senare då jag själv hade slutat vid utredningsinstitutet. Men jag var nyfiken och i början på 1970-talet ringde jag till utredningsinstitutet och frågade om programmet användes. Nej, det hade blivit dyrt att programmera och därför lades en särskild avgift på användningen av programmet, som avskrivning på investeringen, och det ville inte användarna i regel betala för. Detta uppfattade jag, med den policy som vi tidigare hade haft, som ansvarslost av utredningsinstitutet.

Intresset för fel i statistikens skattningar blir äntligen besvarat!

Ett omfattande uppdrag fick jag av Vägverket i slutet av 1960-talet. Uppdraget var att utforma ett system för trafikmätningar för allmänna vägnätet som skulle ge skattningar av trafikflöden (medeltal per dygn) på ett stort antal olika vägvägnät (cirka 20 000). Samtidigt skul-

le en stor mängd trafikarbeten (transportarbeten) skattas med indelningar efter vägtyper och regioner. Det innebar först ett omfattande arbete med att klargöra vad trafikskattningarna skulle användas till och de ungefärliga kraven på precision i skattningarna för olika användning. Den statistiska plan som utformades innebar relativt små stickprov för varje skattning av trafikflöde och en estimator som utnyttjade hjälpinformation om trafikens variation under hela året. Skattningar av konfidensintervall var möjliga för trafikflöden men skulle kräva åtskilligt av datorprogramarbete för att kunna realiseras. Till min överraskning och glädje backade Vägverket inte ut från dessa kostnader och då de första skattningarna av trafikflöden publicerades 1969 redovisades konfidensintervall för varje skattat trafikflöde.

Emellertid blev det en proteststorm från många håll inom Vägverket mot den redovisade osäkerheten: Det hade man inte varit med om tidigare. Vägverkets ledning stod dock fast vid sina ambitioner och till nästa publicering av trafikflöden, tre år senare, hade dessa protester ebbat ut. Kvalitetsredovisningen i form av konfidensintervall blev kvar och är det ännu i dag.

Hur kunde det bli så lyckligt för statistikern? Svaret är att beslutsteori var ett aktuellt ämne även vid tekniska högskolor och den dåvarande överingenjören, sedermera generaldirektören Carl Olof Ternryd, hade övertygelsen att beslut under fiktiv säkerhet, det vill säga att inte (låtsas) veta något om fel, var föga effektivt. Det hade även visats i en förstudie hur avsevärda samhällskostnader uppstod på grund av sådan omedvetenhet om fel. Denna medvetenhet om beslut och risker hade också min närmaste kontaktman vid Vägverket, Börje Thunberg. ■

Referenser

- [1] Deming, W.E. (1960). *Sample Design in Business Research*. Wiley, New York.

Hänt och hört

ANDERS BJÖRKSTRÖM, Matematisk statistik vid Stockholms universitet, disputerade den 28 september, på en avhandling med titeln *Regression methods in multidimensional prediction and estimation*. Fakultetsopponent var professor Philip J Brown från University of Kent.

YINGFU XIE, Biostokastikum, SLU, försvarade den 9 november sin doktorsavhandling *Maximum Likelihood Estimation and Forecasting for GARCH, Markov Switching, and Locally Stationary Wavelet Processes*. Fakultetsopponent var Professor Zongwu Cai, University of

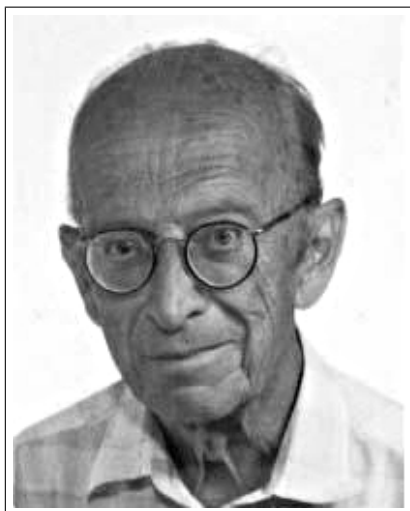
North Carolina at Charlotte, USA.

TERMEH SHAFIE, Statistiska institutionen vid Umeå universitet, försvarade den 16 november sin licenciatahandling *On-site Sampling in Economic Valuation Studies*. Opponent var Annica Isaksson, Linköpings universitet.

XAVIER DE LUNA, Statistiska institutionen vid Umeå universitet, utnämndes den 1 december till professor.

ESBJÖRN OHLSSON är antagen som docent i matematisk statistik vid Stockholms universitet.

Till minne av Bertil Matérn



Professorn emeritus Bertil Matérn har avlidit i en ålder av 90 år; hans närmaste är hustrun Carin samt döttrarna Barbro och Gunhild med familjer.

Bertil Matérn föddes 1917 i Göteborg. Han tog studenten vid Norra Latin i Stockholm 1935, studerade vid Stockholms högskola och blev fil. kand. 1938, pol. mag. 1941, fil. lic. 1947 samt fil.dr och docent 1960. Från 1941 var han hos professor Harald Cramér vid högskolans institut för försäkringsmatematik och matematisk statistik. 1945 blev han handplockad till Statens skogsforskningsinstitut (SFI) av professorn Manfred Näslund för att arbeta med frågor kring Riksskogstaxeringen.

Han fick i uppdrag att utreda noggrannheten i linje- och provytetaxeringar, vilket han redovisade redan 1947 i sin lic-avhandling – en klassiker inom skogstaxering. Under det året gifte han sig med fil. kand. Carin Berglund, som också hade läst vid Stockholms högskola.

Efterhand blev Bertil allmänt engagerad inom SFI, och 1949 inrättades ett kontor för matematisk statistik med honom som chef. Han var ett år stipendiat vid universitet i Iowa och North Carolina, och förordnades 1951 som speciallärare vid Skogshögskolan.

1962 sammanslogs SFI och Skogshögskolan, och en institution i skoglig matematisk statistik med professur tillkom, som Bertil uppehöll tills han efter kallelse utnämndes på den 1963. Under 1970-talet utlokaliserades högskolan till Umeå, Uppsala och Garpenberg och Bertil fick en ambulerande tillvaro fram till pensioneringen 1981.

Lic-arbetet från 1947 är ett pionjärbete inom spatial statistik, och dess betydelse kan inte överskattas. Som exempel kan nämnas att Peter Whittle från Nya Zeeland lärde sig svenska genom att med hjälp av ett lex-

ikon tillgodogöra sig Bertils idéer inom den spatiala statistiken. Doktorsavhandlingen från 1960 är en milstolpe inom den spatiala statistiken och kom i nytryck genom Springer-Verlag 1986. Den är flitigt citerad och finner alltför tillämpningar. Ett avsnitt benämns i facklitteraturen som “Matérn processer”. Spatial statistik har länge varit under snabb utveckling. Då är det inte alltid som upphovsmännen är fortsatt centrala inom området, men det är Bertil!

Bertil var med i flera internationella organ och i Skogs- och lantbruksakademien. Han hade lätt för språk och kunde flera. Han var anlitad både hemma och utomlands för utrednings- och sakkunniguppdrag, inte bara skogliga. Bland annat hade han sådana i Österrike, Mexico och Östtyskland.

Föreningen Skogen tilldelade Bertil sin Guldkvist 1982 med motiveringen att han “får finna sig i att vara en slags institution bland majoriteten av svenska jägmästare, av honom har de nämligen fått de kunskaper de har om skoglig statistik. ...snabbt går han på de enkla lösningarna till stundom mycket invecklade problem. ... varje medarbetare som kommer den store statistikern något närmare in på livet finner en finstämd, något självironisk humorist. ...”

Må det vara tillåtet att här återge en episod från SFI som visar lite av Bertils jag. Professor Näslund hade då blivit chef för institutet, och han styrde det med fast hand, bland annat genom cirkulärskrivelser som utfärdades av hans medarbetare, byrådirektören Alf Allard. Vid personalens sedvanliga julfest var ett ABC skrivet av Bertil och Carl Tersmeden ett återkommande och uppskattat inslag. Ett år löd versen på C: “Cirkulären från Allard uppgår nu till en miljard”. Bertil och Carl blev snabbt uppkallade till Näslund, som förehöll dem att det där var verkligen inget att skämta om! – Nästa år löd versen: “Cirkulären från Allard uppgår ej till en miljard”. Någon uppkallning blev det inte den gången!

Bertil läste skönlitteratur, samlade på mynt och började under senare år släktforska. Han tyckte om att röra på sig, var med i gymnastik, tog joggingrundor och promenader. Semestertid blev det ibland resor med hela eller delar av familjen och övernattning i tält, även turer till fjällen några gånger. Familjens sommarviste är annars huset på Ingarölandet.

Namnet Bertil Matérn kommer att för framtiden vara inskrivet i den matematiska statistikens utvecklingshistoria. Och Bertils många forna elever, kolleger och vänner bevarar det ljusa minnet av en god lärare, en vördad kollega och en avhållen vän.

Bo Ranney och Gustaf von Segebaden

40-årsjubileum för Vinterkonferensen



Jubileumskonferens i Borgafjäll

SÖNDAGEN DEN 2 MARS – TORSDAGEN DEN 6 MARS 2008

Statistiska institutionen och institutionen för matematik och matematisk statistik vid Umeå universitet inbjuder till sin 40:e vinterkonferens.

Statistics and the Law:	David Balding, Imperial College, London, UK
Statistics and Disease Surveillance:	Martin Kulldorff, Harvard Medical School, Boston, USA
Statistics and Brain Imaging:	Keith Worsley, McGill University, Montreal, Canada

Plats: Hotell Borgafjäll, Box 46, 917 04 Borgafjäll. Telefon 0942-421 00, fax 0942-421 07. Ytterligare information: www.borgafjall.se

Mat och logi: Boende i flerbäddsrums inklusive frukost, lunchpaket/lunchbuffé, middagsbuffé 3417 kr/person (3995 kr inklusive moms)

Konferensavgift:

- 2000 kr (2500 kr inkl moms) för deltagare verksamma inom högskolan.
- 1700 kr (2125 kr inkl moms) för doktorander som presenterar poster.
- 5000 kr (6250 kr inkl moms) för övriga deltagare.

Anmälan: Via elektronisk anmälan på hemsidan www.math.umu.se eller till fax nr 090-786 5222 senast den 10 januari 2008.

Konferenskommitté:

- Xavier de Luna, 090-786 5559, xavier.deluna@stat.umu.se
- Lennart Bondesson, 090-786 6529, lennart.bondesson@math.umu.se
- Marie Wiberg, 090-786 9524, marie.wiberg@stat.umu.se
- Ingrid Westerberg Eriksson, 090-786 5225, iw@math.umu.se

Sponsras av:



HANDELSHÖGSKOLAN • UMEÅ UNIVERSITET
U·S·B·E
UMEÅ SCHOOL OF BUSINESS • SWEDEN

Välkomna!

Livet som statistiker

Livet som statistiker på EU-kommissionens generaldirektorat för översättning

av Martin Karlberg, DGT, Luxemburg

Europeiska kommissionen har sammanlagt cirka 25 000 anställda. Av dessa arbetar ungefär var tionde på generaldirektoratet för översättning (DG Translation, DGT). Därmed är DGT världens största översättningstjänst och man översätter texter för EU-kommissionens räkning till och från EU:s samtliga (för närvarande 23) officiella språk. Sätet finns i Luxemburg, men eftersom de flesta beställarna finns i Bryssel är ungefär hälften av personalen stationerad där. Enkel kombinatorik ger vid handen att det enbart vid översättning mellan de olika officiella språken rör sig om 506 möjliga (ordnade) språkpar. Alla språkkombinationer är dock inte lika vanliga och merparten av översättningen sker antingen till eller från endera engelska eller franska. Ungefär 1,5 miljoner sidor (1 sida motsvarar cirka 1500 tecken) översätts årligen, varav cirka 80 procent översätts av DGT:s egen personal, resten av frilansöversättare. Analys- och utvärderingsenheten är ansvarig för DGT:s statistik; statistikfunktionen är bemannad av två personer inklusive undertecknad.

Från Eurostat till DGT

Min första post på EU-kommissionen var föga förvånande Eurostat där jag arbetade med kvalitets- och utvärderingsfrågor. Jag började strax efter den så kallade Eurostatkrisen då situationen var ganska turbulent. I dessa strömvirvar hamnade jag så att säga i bakvattnet, eftersom myndighetens fokus med nödvändighet snarare låg på kärnverksamheten än på stödverksamhet som kvalitet och utvärdering. Jag kom dessutom ganska snabbt att sakna det praktiska statistikarbetet. Det är roligare att själv producera statistik än att utvärdera hur andra gör det! Det gyllene tillfället uppenbarade sig i form av en intern platsannons då DGT sökte en statistiker i Luxemburg.

Till dukat bord

Det faktum att ingen tidigare arbetat som statistiker vid DGT betyder inte att ingen statistik producerats tidigare – tvärtom. Till skillnad från många andra verksamheter inom offentlig förvaltning så har översättningstjänsten en mycket påtaglig produktion registrerad på individnivå. Detta är alltså en av få verksamheter där man kan mäta akademikers produktivitet! Av denna anledning, så har produktions- och produktivitetsstatistik sammanställts under årtal, under överinseende av en permanent arbetsgrupp för statistikfrågor bestående av bland annat översättare, chefer och IT-expert. Eftersom verksamheten har ett starkt IT-stöd för dokument- och uppdragshandling, skapas och lagras en mängd processdata automatiskt – ett Eldorado för en statistiker! För att garantera intern konsistens, så finns dessutom ett sofistikerat datalager innehållande relevanta processdata, och ett helautomatiserat system för produktion av en mängd

regelbundna statistikrapporter.

Produktions- och efterfrågestatistik

Den statistik som produceras är huvudsakligen kvantitativt beslutsstöd (till exempel olika nyckeltal) för DGT:s ledning och för EU-kommissionären för flerspråkighet (för närvarande rumänen Leonard Orban) samt diagram och tabeller för presentationer av verksamheten i olika interna och externa sammanhang. Till en stor del handlar det om deskriptiv produktions- och efterfrågestatistik, där den metodologiska utmaningen kanske inte är särskilt stor, men där man ändå som statistiker kan bidra till kvaliteten genom till exempel adekvata och översiktliga, men informationstäta, diagram utan överflödiga



DGT:s hela statistikfunktion på en och samma bild. David till vänster och Martin till höger. Fotograf: J Janssens

3D-effekter och amputerade y-axlar. För den deskriptiva statistiken har jag stor hjälp av min belgiske assistent David, som tidigare arbetade med statistik vid EU-kommissionens generaldirektorat för fiske och havsfrågor. Utöver statistikproduktion får jag ibland redigera de kvantitativa inslagen i diverse strategidokument och hjälpa till med urvalsdragning för och analys av utvärderingar. Dessutom gör jag vissa metodutredningar (där "metod" ofta kan stå för IT-rutiner eller administrativa procedurer snarare än statistisk metod). Jag arbetar även en del med metadatafrågor både när det gäller definitioner av primärdata och när det gäller beskrivningar av statistikrapporter. Jag arbetar däremot inte alls med datorlingvistik som lär vara en statistikintensiv disciplin.

Hierarkiskt ordnade

En utmaning när det gäller översättningsstatistik är att produktionsdata är *hierarkiskt ordnade*. Vi kan som exempel ta en översättning av ett original, som delvis är på franska och delvis på engelska, till 20 andra språk. Detta är en beställning som blir till 20 uppdrag (ett för varje språk som det ska översättas till) som ges till språkavdelningarnas översättningsenheter. Vid varje enhet delas uppdragen upp i (minst) två bearbetningar (en för översättning från engelska, en för översättning från

franska), totalt får vi alltså (minst) 40 bearbetningar. Om originalet är långt, så kan bearbetningen vid vissa enheter fördelas på flera individer genom att den delas upp i flera uppgifter. Om en enhet har hårt belastade medarbetare, så kan en bearbetning läggas ut på en frilansöversättare eller lämnas över till en annan enhet inom samma språkavdelning.

För att ytterligare komplicera det hela, så har vi dessutom *tidsaspekten* att ta hänsyn till. Efterfrågan kommer hela tiden före produktionen – exempelvis kan en beställning i januari leda till en leverans i februari – och olika nivåer av produktionsdata kommer in vid olika tider. De flesta uppdrag inom en beställning kanske är klara i april, men hela beställningen är klar först i maj. Situationen kompliceras ytterligare när man vill göra detaljerade produktivetsberäkningar eftersom en översättares produktion registreras punktvis när bearbetningen är klar, medan antalet arbetade timmar registreras kontinuerligt i tidsredovisningssystemet. Detta fenomen är givetvis extra påtagligt när en översättare har ett långt dokument att översätta under en lång tidsperiod.

Administrativa processdata av hierarkisk natur

Eftersom jag inte har någon direkt erfarenhet av liknande problem från mina tidigare verksamhetsområden, så är det svårt att kunna knyta min verksamhet till någon teoretisk grund. Om du känner till tidigare arbeten som skulle kunna vara av intresse i sammanhanget (“administrativa processdata av hierarkisk natur”?) skulle jag vara mycket intresserad av att ta del av dessa. I avvaktan på detta, så får jag lita till fantasi och sunt förnuft, något som jag har försökt tillämpa på problemet som jag beskriver nedan.

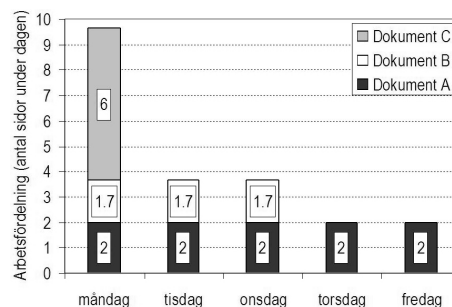
Ett problem

Ett exempel på de problem som jag ställts inför är prognostiska uppskattningar av arbetsbelastning. Eftersom jag inte känner till några tidigare resultat inom området, så har jag klurat lite på min egen kammare. Nedan kommer jag att beskriva problemet genom ett exempel och mitt förslag till lösning. Eftersom problemet inte är direkt statistiskt (jag behandlar varken prognos- eller mätfelsaspekter), så har lösningen säkert redan identifierats inom någon helt annan disciplin. Om någon Qvartilenläsare har något tips om detta, så emottages det tacksamt!

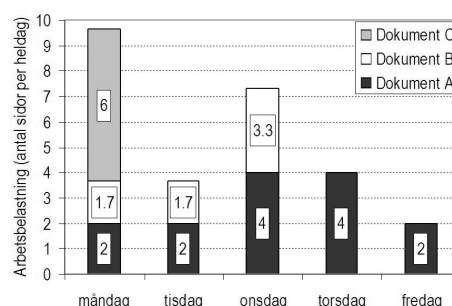
Uppgift: På måndag morgon får en översättare 3 dokument att översätta. Dokument A (10 sidor långt) ska vara klart på fredag kväll, Dokument B (5 sidor) ska vara klart på onsdag kväll, och Dokument C (6 sidor) ska vara klart redan på kvällen samma dag. Översättaren arbetar halvdag onsdag och torsdag. För enkelhets skull gör vi (det helt orealistiska) antagandet att inga nya dokument kommer in under veckan, och struntar i att viss frånvaro (till exempel vid akut sjukdom) är oförutsägbar. Givet detta, hur beräknar man bäst översättarens arbetsbelastning för de olika dagarna?

Enkel lösning: Varje dokument fördelas jämnt över dagarna mellan beställning och leverans (se Figur 1). För

varje arbetsdag ges arbetsbelastningen av kvoten mellan antalet sidor och antalet persondagar som översättaren arbetar (se Figur 2). I vårt exempel innebär detta enbart att “halvdagarnas” arbetsbelastning blir dubbelt så stor som antalet sidor som ska översättas denna dag.



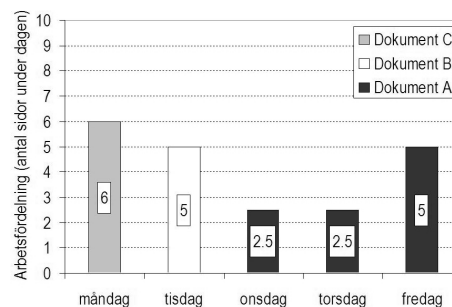
Figur 1. Arbetsfördelning (enkel lösning)



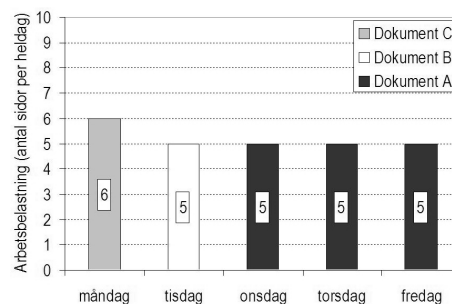
Figur 2. Arbetsbelastning (enkel lösning)

Problem: Redan detta extremt enkla exempel visar de huvudsakliga problemen med den enkla ansatsen:

- orealistiskt hög arbetsbelastning under en av “halvdagarna” (onsdag),
- orealistiskt hög arbetsbelastning under början av perioden (måndag).



Figur 3. Arbetsfördelning (“balanserad” ansats)



Figur 4. Arbetsbelastning (“balanserad” ansats)

Alternativ ("balanserad") arbetsfördelning: Den alternativa algoritmen som jag har utvecklat är ganska enkel att klä i ord (men något svårare att implementera generellt, se [1]):

- a) Sortera arbetsuppgifterna efter tidsfrist och lägg de akuta först.
- b) Givet tidsfristen, fördela arbetsbelastningen så jämnt så möjligt (låt arbetsbelastningen bestämma arbetsfördelningen, i stället för tvärtom).

Man får då en arbetsfördelning (se Figur 3) som är så "proportionell" mot resurserna som möjligt, vilket ger en arbetsbelastning som är så "jämn" som möjligt (se Figur 4). Här är det endast måndagen med som bryter mönstret (sex sidor som måste översättas just den dagen på grund av den korta tidsfristen), medan arbetsbelastningen är helt jämn (fem sidor per heldag) under resten av veckan.

Den "balanserade" ansatsen har alltså fördelen att arbetsbelastningen är så jämn som möjligt. En styrka med den "enkla" ansatsen är dock, förutom den okomplicerade beräkningsmetoden, att den indirekt, om än oavsiktligt, och helt okontrollerat, kompenserar för det faktum att det faktiskt kan komma in nya uppdrag med korta tidsfrister under tisdag till fredag, varför det kanske inte är så klyftigt att fördela om arbete till senare tider i veckan.

På jungfrulig mark

Efter många års arbete inom organisationer där statistikantingen har varit kärnverksamheten, eller en stödverksamhet tillräckligt stor för att motivera en egen avdelning, så hamnade jag alltså under förvären 2005 på EU-kommissionens översättningstjänst, vilket jag inte hade

föreställt mig i mina vildaste drömmar när jag började på Eurostat!

Jag antar att mina erfarenheter är desamma som hos många andra som arbetar som ensamma statistiker i en större organisation; å ena sidan förväntas man vara "expert" på allt från himmel till jord, å andra sidan måste man lära sig att argumentera för sådant som man har vant sig vid att betrakta som självklart. Man måste lära sig att hålla en konstruktiv dialog med statistikbeställare, och eftersom det finns en stor mängd begåvade kollegor med stort intresse för statistikfrågor och stor insikt i tillämpningsspecifika problem, så måste man i denna dialog ofta vara beredd att ödmjukt lyssna och lära. Ibland måste man också påminna sig själv om att de operativa databaserna faktiskt finns till för att översättningsverksamheten ska löpa effektivt, inte för att skapa översättningsstatistik.

På pluskontot står det faktum att jag får arbeta i en internationell miljö. Exempelvis är vi ett dussin medarbetare av åtta olika nationaliteter på vår egen enhet; kommunikationen internt och med samarbetspartners sker huvudsakligen på engelska och franska. Andra fördelar är givetvis att jag som första statistiker får arbeta inom ett intressant (och datarikt!) område, den låga andelen "administrativt nedfall" i förhållande till det praktiska/statistiska arbetet och den stora friheten, under ansvar, när det gäller utformningen av översättningsstatistiken. ■

Referenser

- [1] Karlberg, M. (2007). *Workload assessment at the world's largest translation service*. Paper presented at ISI, Lissabon.

Allt från mödradödlighet till företagsregister - om arbetet i ett statistikprojekt i Laos

av Hans Pettersson, National Statistics Centre, Laos

Hälsoministern, Dr Ponemek Dalaloy, tittar utöver församlingen och sammanfattar:

Nu har jag lyssnat på vad ni experter säger om antalet döda kvinnor i samband med barnafödande i Laos. Vi har mina experter här på ministeriet som säger att deras urvalsundersökningar visar att antalet är 530 per 100 000 födda levande barn. Sedan har vi WHO som på basis av globala modeller säger att antalet är 660 och vi har National Statistics Centre som säger att de har beräknat antalet till 740 baserat på data från senaste folkräkningen. Det här måste jag fundera på.

Vi är i slutet på ett möte på hälsoministeriet som – om man skall se positivt på det – har varit litet av ett seminarium om osäkerhet i statistik och statistiska felkällor. Det egentliga syftet med mötet är att hälsoministern skall bestämma vilka siffror från Laos som skall

presenteras i internationella databaser. Vi har diskuterat svårigheten att mäta trender i mödradödlighet (antalet döda kvinnor i samband med barnafödande per 100 000 födda levande barn) från hälsoministeriets urvalsundersökningar som dels är gjorda med olika design, dels är små. Vi har också diskuterat WHO:s modeller som är byggda enbart på data från länder med bra statistik; fattiga länder som Laos är inte med. Den lokale WHO-representanten erkänner att han inte förstår sig på modellerna. De är komplicerade och osäkerheten är stor (200 – 1600!) men WHO i Genève pressar på för att få skattningen accepterad av Laos.

Från National Statistics Centre har vi sagt att det i avsaknad av nationella register är svårt att mäta nivån särskilt väl, alla mätningar är osäkra. Man får nöja sig med att mycket grovt uppskatta trenden över lång tid. Det är ju trenden som är viktigast ur Laos perspektiv, inte om La-

os är något bättre eller sämre än grannländerna. Därför bör man lita på de mätningar av trenden som kan göras från folkräkningarna från åren 1995 och 2005. Det finns visserligen problem med underrapportering men räkningarna är gjorda med identisk design och materialet är mycket större än i urvalsundersökningarna. Slutet på diskussionen blev livlig; vårt betonde av osäkerheten i mätningarna gjorde att vi blev kallade "high-flying theoretical people" av en av deltagarna från hälsoministeriet som ville ha bestämda besked. Att bli kallade teoretiker kändes lite konstigt för oss statistiker som arbetar på golvet. Vi – min kollega Thirakha och jag – åkte tillbaka till National Statistics Centre med känslan att diskussionen skulle komma att fortsätta.

Reducering av mödradödligheten är ett av åtta *Millennium Development Goals*. Det är ett stort projekt som drivs av FN. Alla utvecklingsländer har föresatt sig att reducera mödradödligheten med 75% under en tjugofemårsperiod (1990 – 2015). Liknande mål finns för barnadödlighet, fattigdom, utbildning etcetera. Det finns ett stort nationellt och internationellt intresse för uppföljningen av målen; FN:s statistiska kontor har en särskild websajt för projektet. Uppföljningen ställer stora krav på det nationella statistiksystemet; många utvecklingsländer har liksom Laos problem med uppföljningen.

Det nationella statistiksystemet – ja, det är därför jag sitter här på kontoret i Vientiane i Laos. Jag är utsänd som projektledare för ett SCB projekt som syftar till att stärka landets statistikbyrå. Mitt kontor är ett spatiöst rum i statistikbyråns villalikhade byggnad ett stenkast från den mäktiga Mekongfloden. Närmaste grannar är ett Budisttempel och en billackeringsverkstad som ibland sprider illaluktande ångor över omgivningen.

Jag är den siste i raden av projektledare som jobbar med National Statistics Centre, i juni 2008 kommer projektet att avslutas. Historien började 1991 då SCB fick en förfrågan från National Statistics Centre för att bygga upp byrån.



Delar av personalen vid National Statistics Centre. Deputy Director General Ms Phonesaly står i mitten längst fram. Fotograf: NSC

Vid den tiden var det en statistikbyrå med synnerligen rudimentära funktioner så det var från början klart att det var frågan om en massiv och lång insats, en insats som skulle finansieras av Styrelsen för Internationellt Samarbete (Sida). Projektet lades upp som ett brett samarbete

som skulle täcka alla verksamheter vid en nationell statistikbyrå. Under åren har tio långtidskonsulter jobbat i projektet och över 100 korttidskonsultationer har genomförts (företrädesvis av SCB:are). Mina uppgifter är, förutom projektledning, huvudsakligen rådgivning och stöd i utvecklingsarbetet (förändringsprocesser) och utbildning dels i form av formell undervisning och dels i form av individ- eller grupphandledning.

Statistikbyrån i Laos

National Statistics Centre är en mycket liten statistikbyrå, totalt jobbar bara cirka 60 personer här. Det har varit det svåraste problemet för vårt projekt; byrån är för liten för att kunna fungera fullt ut som en nationell statistikbyrå. Ett kraftigt intag av unga personer under senare år har något förbättrat situationen. Det har gett en åldersfördelning liknande den på SCB för några år sedan med en grupp halvgamla veteraner och en grupp unga "noviser" och inte så många däremellan. De äldre på National Statistics Centre är så gott som alla utbildade i det forna östblocket. De yngre har oftast en ekonomexamen från Laos nationella universitet. I stort sett alla pratar en passabel engelska, i många fall som ett resultat av kurser i engelska som projektet har finansierat.

SCB:s projekt

Projektet syftar till att utveckla kapaciteten i det nationella statistiksystemet. Primärt gäller det att bygga upp statistikbyråns kapacitet att producera officiell statistik och koordinera det nationella statistiksystemet. Insatsen har två huvudelement: kunskapsöverföring och infrastrukturbyggande.

Kunskapsöverföring avser baskunskaper i alla aspekter av statistikproduktion men också utveckling av till exempel ledarskapskompetens. Formella kurser är en viktig del eftersom det i Laos inte finns någon institution för statistikutbildning. Än viktigare är nog ändå praktiska övningar och erfarenheter från den löpande produktionen. Statistikbyrån har med projektets stöd genomfört två folkräkningar, en jordbruksräkning och en företagsräkning och därtill ett antal nationella urvalsundersökningar både av hushåll och av företag. Undersökningarna har fungerat som stora laborationsövningar för personalen där alla delar av en undersökning har genomförts med stöd av konsulter.

Parallellt med kunskapsöverföringen har det skett en massiv uppbyggnad av byråns infrastruktur. Projektet har finansierat cirka 30 bilar och cirka 50 motorcyklar för fältarbetet. Investeringarna i IT-infrastruktur är också betydande, vi har köpt några hundra datorer, ett antal servrar, skannrar, skrivare med mera. National Statistics Centre har idag en modern IT-miljö med få motsvarigheter på andra håll i Laos.

Det finns många intressanta saker från projektet som jag skulle kunna ta upp men jag skall koncentrera mig på några stycken.

Rapportsystem kontra urvalsundersökningar

Laos orienterar sig gradvis mot marknadsekonomi men rutiner och tankemönster från tiden med centralstyrd planekonomi lever delvis kvar i förvaltningen. Detta märks också i statistikproduktionen som i många fall är baserad på (totalräknade) rapportsystem kopplade till planeringssystemen, med de kvalitetsproblem detta medför. Det stora problemet är förstås att de rapportrande enheterna har ett egenintresse att redovisa att planeringsmål har uppnåtts. Ett annat problem är att data aggregerats i distrikt och provinser på väg upp till nationell nivå vilket gör det omöjligt att bedöma och kontrollera kvaliteten i grunddata. Avsaknaden av kvalitetskontrollmöjligheter gör att man är hänvisad till bedömningar av statistikresultatens rimlighet när man bedömer kvaliteten i grunddata.

Jerker Johnsson, en av mina företrädare, har gjort en analys av rimligheten i den statistik som publicerats om produktivitet utveckling och real tillväxt under en tioårsperiod i jordbrukssektorn (statistik baserad på data från rapportsystemen). Han bedömde, på basis av data från jordbruksräkningen, att produktivitet utvecklingen och den reala tillväxten var markant överskattad i statistiken.

En stor utmaning för projektet har varit att introducera nationella urvalsundersökningar som ersättning för eller komplement till rapportsystemen. Vi fick faktiskt börja med att övertyga National Statistics Centre om urvalsundersökningarnas fördelar! De östblocksutbildade statistikerna hade väldigt lite erfarenhet av sådana. Nästa steg är att få beslutsfattare i ministerier att förstå och lita på resultat från urvalsundersökningar. Det är en pågående process, vi skulle behöva hålla kurser i statistiskt tänkande för ministerierna men det har vi inte måktat med ännu. Vi skulle också behöva höja kompetensen hos statistikerna vid ministerierna så att kvalitetsmedvetandet stärks. Trots allt måste en hel del sektorstatistik baseras på rapportsystem och vägen att förbättra systemen går via höjt kvalitetsmedvetande hos ministeriernas statistikenheter.

Företagsräkning och företagsregister

För mig, som hemma enbart har jobbat med hushålls- och individstatistik, har arbetet här med företagsstatistiken gett nya och omskakande erfarenheter av metodproblem. Det har hänt att jag har längtat tillbaka till hushållsstatistiken när jag jobbat med ram- och mätproblemen i företagsundersökningarna här.

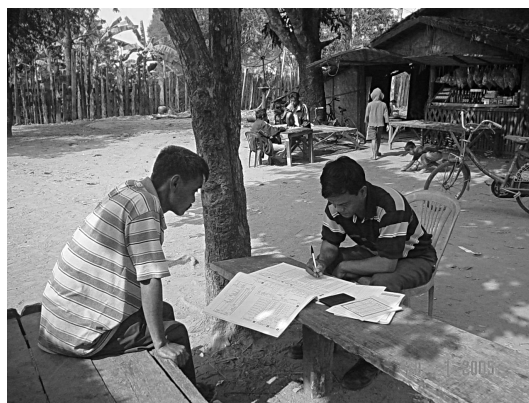
Skatteförvaltningen och industriministeriet har notoriskt dåliga register över företag. De urvalsundersökningar av företag som gjorts har varit svåra att räkna upp till totalnivåer med rimlig kvalitet på grund av täckningsproblemen. Detta har varit ett långvarigt problem för nationalräkenskaperna. För att råda bot på detta genomfördes med projektets stöd en företagsräkning under 2006. Fältpersonal besökte under två veckor alla de 10 000 byarna i landet (även stadsdelar benämns här byar) och lista-

de alla företag (arbetsställen) i byn. Listningen inbegrep också de små familjeföretag utan några anställda utanför familjen ("*subsistence business*") som utgör den stora volymen av företag. Två månader senare gjordes en intervjuundersökning av alla företag utom de företag som ligger i byar i väglöst land. Intervjun avsåg basuppgifter som näringsgren (ISIC), varukoder (CPC) antal anställda med mera. Granskade och korregerade data för 140 000 företag, eller egentligen arbetsställen, ligger nu i en databas. Nästa steg är att koppla arbetsställen till företag. Det ser ut att bli en knepig uppgift. I vissa fall har man vid intervjun kunnat ange skattenummer för företagets huvudkontor och då är det lätt. Men i regel har man inte skattenumret och då måste man koppla via uppgifter om huvudkontorets namn och adress. Adresser i egentlig mening finns inte under bynivå så det kommer att bli en del pusslande för att få ihop det.

Förhoppningen är att företagsräkningen ska kunna bilda grund för ett företagsregister. National Statistics Centre har diskussioner med de två andra institutionerna som har register över företag: ministeriet för industri och handel och skatteavdelningen på finansministeriet. Jag hoppas att samarbetsviljan finns hos alla parter så att institutionella hinder kan undanröjas.

Datainsamling – mätproblem

Det mesta av mitt arbete sker här på kontoret. Det blir därför intressanta avbrott i rutinen de gånger jag kan resa ut i fält och följa arbetet i intervjuundersökningarna. Intervjuarna och vi som följer med ses som representanter för statsförvaltningen och tas artigt emot i byarna. Denna artighet tar sig ibland lite väl påfrestande uttryck; att som representant för den centrala makten högtidligen bli serverad ljummen hembränd sprit från en plastdunk - mitt på blanka eftermiddagen i 35 graders värme - är sådant som man ibland måste genomlida för den goda sakens skull.



En intervju ute i en av byarna i Laos för *Expenditure and Consumption Survey*
Fotograf: NSC

Byns ordförande måste alltid formellt godkänna undersökningen. Efter att detta är gjort är alla utvalda hushåll eller företag i byn i princip tvungna att delta i undersökningen. Vi får i stort sett inget vägrarbortfall. Det är ju bra men säkert kommer en del undertryckt vägran istället ut som svarsfel. Att ställa upp för intervjun är en sak,

att ge korrekta svar en helt annan sak. Jag förstår bara litet av det som sägs vid intervjuerna men kroppsspråket hos vissa småföretagare som intervjuas om omställning och inkomster är tydligt nog. Det är nog ingen risk att inkomsterna överskattas. Den typen av medvetna svarsfel är nog vanligare i företagsundersökningarna än i hushållsundersökningarna. I hushållsundersökningarna härrör svarsfelen ofta från genuin okunskap om svaret eller att frågan inte förstås korrekt. Ett exempel på det förra är åldern på äldre personer; man ser tydlig *“age heaping”* på tiotal i åldersfördelningen från folkräkningen. Mätproblemen är allra störst i Laos *Expenditure and Consumption Survey* (LECS) som är den viktigaste av statistikbyråns hushållsundersökningar. Det är en utvidgad hushållsbudgetundersökning där man bestämt sig för att använda dagboksметод för att mäta hushållets utgifter och konsumtion under en månad. I ett land där läs-och skrivkunnigheten ligger kring 70% innebär detta särskilda problem. För att få tillräcklig kvalitet i dagboksdata måste intervjuaren assistera vissa hushåll så gott som dagligen med dagboksförandet. Det betyder att intervjuaren måste bo i byn under hela månaden. Kostnaderna för fältarbetet blir höga men det är priset för acceptabel kvalitet i grunddata.

Solnedgång över Mekong

Arbetsveckan avslutas i regel med en öl på Sunset Bar. Det är ett ganska slitet ställe vid Mekongfloden som

har en fantastisk atmosfär och som kan bjuda på spektakulära solnedgångar. Göthe Isacsson, en pensionerad SCB:are som bor här sen elva år kommer oftast hit. Det blir förstås en hel del prat om jobbet från veckan som gått. Så småningom, när den intensivt röda solen sänker sig över *“the mighty Mekong”*, glider samtalet över till andra saker; ofta diskussion om böcker (Göthe är en bokslukare) eller varför min golfsving är så miserabel trots all möda jag lägger ner på att förbättra den. Det kan vi prata länge om medan det tropiska mörkret sänker sig över Vientiane.

Hur det gick med mödradödligheten? Den historien är inte slut än. De två mätningarna på folkräkningsdata med 10 års mellanrum visar att antalet troligen är oförändrad (hög) under perioden. Hälsoministeriet tycker det är konstigt att antalet inte har minskat när man jobbat så mycket med förbättrad hälsovård. Ministeriet litar inte på mätningarna och menar att dödligheten måste ha gått ned. Mot det talar att den enskilt viktigaste påverkande faktorn – andelen födslar med läkare, sköterska eller barnmorska närvarande – har varit i stort sett oförändrad under perioden. Viktiga intressenter som UNFPA, UNICEF och WHO tror därför på noll-trend (men WHO vill naturligtvis ha sina nivåskattningar i redovisningen). ■

Rapporter

Statistikerträff i Göteborg

Torsdagen den 13 september gick den första statistikerträffen i Göteborg av stapeln. Denna anordnades av AstraZeneca och Statistik konsulterna och tanken var att under mer lättsam seminarieform skapa en plattform för statistiker och statistik användare att “nätverka” och uppmärksamma nya tillämpningsområden för statistiska metoder.

Staffan Nilsson inledde med att berätta om det nya masterprogrammet i biostatistik som drivs av matematiska vetenskaper på Chalmers. Staffan visade på ett engagerat sätt nyheterna i den biostatistiska grenen som utgörs av de fem inriktningarna: matematik, tillämpad matematik, finansmatematik, matematisk statistik och biostatistik.

Magnus Pettersson informerade om arbetet med att skapa en yrkesetisk kod för statistiker och gav förhoppningen om att resultatet av detta arbete skulle komma inom inte allt för avlägsen framtid.

Därmed var punkterna av mer informationskaraktär genomgångna och ordet gick till Martin Gellerstedt, Högskolan Väst, som talade om “Diagnostik - en tolkningsfråga?” Tanken var att presentera ett statistiskt område

som visserligen existerar men som kanske inte alltid får den uppmärksamhet som det förtjänar. Behoven inför framtiden är stora på statistiksidan! Martins presentation var tankeväckande, intressant och fullt begriplig för alla.

Magnus Pettersson, Statistik konsulterna, talade sedan om *Churn Management* (analys och upptäckt av kunder som lämnar ett företag) tillämpat inom telekommunikation och bilindustrin samt förklarade hur longitudinella respektive tvärsnittstudier snarare kompletterar än överlappar varandra. Statistisk processkontroll (SPC) tillämpat inom dessa branscher är relativt oprövat men definitivt en framkomlig väg med hjälp av den statistiska verktygslådan.

Karin Nelander, AstraZeneca, avslutade med en presentation om Trellisgrafik, ett hjälpmedel för visualisering av data i flera dimensioner. Hon kunde genom exempel och beskrivningar demonstrera användbarheten med detta grafiska användargränssnitt. Vi förstod också att detta är ett verktyg som kan bespara statistiker onödigt analysarbete längre fram. Presentation gav många uppslag och jag vet att flera laddade ner och testade redan dagen efter.

Efter föredragen avslutade vi på en restaurang ett stenkast från Chalmersområdet med mingel, mat och dryck under avspända former.

När man lyssnat på föredragen under träffen frågar man sig hur många andra tillämpningsområden, nya metodtillämpningar och smarta användargränssnitt man egentligen missat! Det bästa sättet att undvika detta är att skapa fler plattformar för samverkan och nätverk bland sta-

tistiker och statistikanvändare.

AstraZeneca och Statistikkonsulterna, vilka tagit initiativet till denna träff, vill gärna se en fortsättning nästa år vid ungefär samma tid. Har du, eller känner du någon kollega, som vill vara med och kanske är intresserad av att hålla ett av föredragen under en eftermiddag, så meddela gärna oss.

Mats Rudholm, Statistikkonsulterna

Heldagsseminarium på Frimurarehotellet i Linköping

av Peter Werner, SCB

Sedan 1995 har först *Linköpings universitet* och därefter *Sveriges Marknadsundersökare (SÖK)*, *European Society for Opinion and Marketing Research Sverige (ESOMAR)* och *Surveysektionen* i Svenska Statistikersamfundet regelbundet inbjudit till ett heldagsseminarium på Frimurarehotellet i Linköping om aktuella metodfrågor kring marknads- och opinionsundersökningar. Seminarierna syftar till att vara ett forum där akademiska forskare och företrädare för de privata och offentliga undersökningssektorerna kan träffas och diskutera gemensamma undersökningsproblem.

Årets seminarium, som gick av stapeln den 2 oktober, hade temat "Framtidens undersökningar: Kommer vi bara att kunna intervjua personer som är vidtalade i förväg?" och skulle ur olika synvinklar behandla den ständigt aktuella frågan hur vi ska handskas med de problem som orsakas av försämrade urvalsramar och minskande svarsbenägenhet.

Dagens första talare var Bengt Larsson, TNS Gallup, tillika ordförande i Sveriges Marknadsundersökningsföretag (SMIF). TNS Gallup har nyligen gått över från telefon- till webbaserad omnibus, främst på grund av ökade svårigheter att nå folk per telefon. Bengt redogjorde för en studie där man kört de båda insamlingsmetoderna Computer Assisted Telephone Interviewing (CATI) och Computer Assisted Web Interviewing (CAWI) parallellt i en och samma omnibus. Frågorna handlade om varumärkeskännedom och resultaten visade att metoderna i vissa avseenden genererade olika svarsmönster. CAWI gav exempelvis:

- en större andel respondenter som nämnde åtminstone något märke,
- fler märken per respondent och
- färre svårstavad märken, vilka dessutom mer sällan nämndes först.

För att hantera problemet med jämförelser över tid har TNS Gallup valt att transformera historiska CATI-data för jämförbarhet med kommande CAWI-resultat. Kopplat till nämnda studie pratade Bengt också allmänt om urvals- och generaliserbarhetsproblematiken vid webbaserade undersökningar.

Därefter berättade Ulf Isander, SIFO Research International, om en studie som han och kollegan Åsa Jaktlund genomfört i syfte att jämföra den grupp av personer som endast har mobil telefon med den grupp som har fast telefon. Som nästkommande föredragshållare, Nicklas Källebring från Synovate Temo, senare skulle avslöja har uppskattningsvis hela 7,6 % av Sveriges befolkning (16-75 år) endast mobil telefon, medan 89,1 % har fast telefon (varav 3,8 % endast fast). I SIFO RI:s veckovisa omnibus ingår i dagsläget bara personer som har fast telefon. Studien visade att det var mycket svårt att skapa en acceptabel urvalsram av personer med endast mobil telefon. Dessutom indikerade resultaten att det fanns skillnader mellan denna grupp och gruppen med fast telefon vad gällde vissa attitydfrågor. Även om SIFO RI:s omnibus för närvarande inte befinner sig i riskzonen föreslår Ulf och Åsa bland annat att man skapar en roterande panel av personer som endast har mobil telefon, vilken vid behov kan utnyttjas för att komplettera omnibusurvalet.

Efter en god lunch och ett kort föredrag av Thomas Floreteng (konsult, tidigare vid Post- och Telestyrelsen) om konvergensen mellan fast och mobil telefoni fick seminariet ett annat ansikte, då de tre kommande föredragen var av övervägande teoretisk karaktär. Huruvida det var denna innehållsliga förändring eller vanlig efter-lunch-trötthet som fick en och annan åhörare att tappa fokus ska jag inte spekulera i.

Först ut var Stig Danielsson, biträdande professor vid Linköpings universitet, som värmdde upp auditoriet med en pedagogisk genomgång av den grundläggande teorin för propensity score-viktning. Därefter fortsatte Boris Lorenc, SCB, på den inslagna vägen och presenterade en ny, icke-parametrisk estimator för användning vid icke-sannolikhetsurval. Mer om denna estimator och dess egenskaper finns att läsa i Boris doktorsavhandling från Stockholms universitet, 2006.

Det tredje teorispäckade föredraget hölls av Peter Lundquist, SCB, som i likhet med Boris presenterade utdrag ur sin doktorsavhandling, också denna från Stockholms universitet, 2006. Peter har fokuserat på hur man kan skatta intervjuarvarians, det vill säga den del av den tota-

la variansen som enligt en viss mätfelsmodell uppkommer på grund av att olika intervjuare påverkar en och samma uppgiftslämnare på olika sätt. I sitt föredrag diskuterade han mot teoretisk bakgrund hur man kan designa en intervjuarvariansstudie, där syftet är att skatta intervjuarvariansen på bästa möjliga sätt.

Dagens sista framträdande gjordes av en grupp studenter från Linköpings universitet som följt upp svenska privatägda undersökningsföretag gällande tillämpningen av Surveysektionens *Standard för bortfallsberäkning*.

Resultaten var överlag ganska nedslående. Exempelvis saknade hela 14 av 35 svarande företag kännedom om att det finns en standard och 8 företag ansåg att en sådan inte behövs. Detta bör betyda att Surveysektionen har en del missionsarbete framför sig.

Sammanfattningsvis var det en trevlig dag med flera intressanta föredrag. Däremot var det ett taktiskt misstag att lägga teoriblocket efter lunch, vilket jag tror att en majoritet av det dryga hundratalet deltagare håller med om. ■

Bokanmälan

All statistikundervisning ifrågasatt

“Alla svanar är vita” var ett uttalande som länge ansågs sant, ända tills man i Australien upptäckte en svart svan. Sådana oväntade och, om man så vill, extrema observationer, som ställer tidigare teorier på huvudet, är ämnet för en bok med titeln *The Black Swan* av Nassim Nicholas Taleb (Random House, New York 2007). Bokens undertitel är *The Impact of the Highly Improbable*. Författaren är född i Libanon, uppvuxen i Frankrike och nu bosatt i USA. Han har arbetat i ett finansföretag, men är nu “Dean’s Professor in the Sciences of Uncertainty” vid University of Massachusetts i Amhurst.

Författaren vidhåller att det är de extrema och därför oväntade observationerna som har avgörande inverkan på händelseförloppet både i ekonomiska och andra sammanhang. Antalet dödsfall per månad i Thailand varierade förmodligen på ett tämligen förutsägbart sätt tills tsunamikatastrofen åstadkom en synnerligen extrem observation. Det var en typisk Svart svan. Det är nödvändigt att utforma teorier och modeller så att de tar hänsyn till möjligheten av att det uppträder en Svart svan, för det är de som svarar för de stora förändringarna.

Med denna bakgrund är boken en vidräkning med praktiskt taget allt vad statistisk och ekonomisk teori heter. Författaren är särskilt allergisk mot normalfördelningen, som han anser har alldeles för smala svansar. Det är de riktigt extrema observationerna som påverkar skeendet så starkt att vi i själva verket aldrig kan göra några prognoser. Vi lever i *Extremistan*, inte i *Mediocristan*. Med fördömandet av normalfördelningen följer också en stark skepticism mot regressionsanalys och flertalet andra statistiska metoder. Författaren uppmanar de läsare som haft otur att bli utsatta för undervisning i statistik att glömma allt som lärdes ut där!

Det är inte heller många ekonomer som får godkänt av författaren. En av dem är Keynes, vars *Treatise on Probability* från 1920 anses värdefull, en annan från senare tid är Kahneman, som fick “Nobelpriset” i ekonomi 2002.

I övrigt får den svenska vetenskapsakademins nobelkommitté för ekonomipriset skarp kritik för att med få undantag, varav då Kahneman är ett, ha belönat ekonomer som baserat sina teorier på normalfördelningsantaganden! Naturligtvis påpekar Taleb hur Scholes och Merton, som 1977 fick ekonomipriset för sin finansmarknadsteori, bara ett år senare fick se sitt finansbolag gå i konkurs som följd av den finansiella krisen – en typisk Svart svan, en extrem observation.

Man kan inte beskylla författaren för att inte ha följt den vetenskapliga diskussionen runt de problem han tar upp. Bokens bibliografi upptar 28 sidor och täcker ett mycket vitt fält.

Jag blev ändå irriterad när jag läste den här boken. Det var sällan jag kunde hålla med författaren i hans slutsatser, men hans resonemang var oftast så luddigt att jag inte kunde sätta fingret på någon punkt och säga: Här har du fel! Det skulle vara lätt att ignorera en opinionsyttring av det här slaget, som så uppenbart går till överdrift i sin kritik, men eftersom *The Economist* uppmärksammade boken i vintras – det var på det sättet jag observerade den – kanske vi ska förbereda oss på att den kan få ett visst negativt inflytande på den allmänna synen på statistik och statistiker.

Och kanske vi ändå ska ta åt oss något av kritiken. Visst är vi väl ganska fixerade vid normalfördelningen! Det är inte ofta man ser någon annan fördelning användas vid beräkning av en maximum likelihood-skattning. Och visst är det svårt att göra prognoser! Jag har gjort många långtidsprognoser i mina dar, och jag har levt tillräckligt länge för att uppleva hur dåligt de slog in. Ofta kan man hänföra felet till att modellen inte tagit hänsyn till någon ny företeelse, men antagligen skulle man kunna se det så att det är en Svart svan som dykt upp. Men vi kan väl inte låta bli att göra prognoser?

Ladda upp dig med argument till försvar för normalfördelningens berättigande!

Erik Ruist

Konferenskalendarium

Här anges först och främst nordiska och europeiska konferenser samt vissa kurser. För fler konferenser hänvisas till de länkar som finns på Svenska statistikersamfundets hemsida – www.statistikersamfundet.se. Kalendariet baseras på de önskemål och förslag som medlemmarna lämnar till Qvartilen.

Datum	Konferens	Plats	Hemsida
2–6 mar 2008	Vinterkonferens – 40-årsjubileum: Statistics and the Law, Statistics and Disease Surveillance, Statistics and Brain Imaging	Borgafjäll	www.matstat.umu.se/aktuellt/vinterkonf/v-konf-08/index.asp
4–7 mar 2008	8th German Open Conference on Probability and Statistics	Aachen, Tyskland	gocps2008.rwth-aachen.de
21–25 apr 2008	The International Conference on Trends and Perspectives in Linear Statistical Inference LINSTAT'2008 in Celebration of Tadeusz Calinski's 80th Birthday	Bedlewo, Poland	linstat08.au.poznan.pl/
16–18 may 2008	International Conference on Interdisciplinary Mathematical and Statistical Techniques, IMST 2008 / FIM XVI	Memphis, USA	www.msci.memphis.edu/IMST2008-FIMXVI/
8–11 jun 2008	International Workshop on Recent Advances in Time Series Analysis	Protaras, Cyprus	www.ucy.ac.cy/~rats2008/
22–25 jun 2008	The 28th Annual International Symposium on Forecasting	Nice, France	forecasters.org/isf/
30–2 juli 2008	The 1st International Symposium on Biopharmaceutical Statistics	Shanghai, China	www.isbiostat.org/
13–18 jul 2008	XXIVth International Biometric Conference 2008	University College Dublin, Dublin, Ireland	www.conferencepartners.ie/ibcdublin2008/
14–19 jul 2008	The seventh World Congress in Probability and Statistics	Singapore, Singapore	www.ims.nus.edu.sg/Programs/wc2008/index.htm
23–26 jul 2008	The 17th International Workshop on Matrices and Statistics, IWMS-08, in Honor of Professor T.W. Anderson's 90th Birthday	Tomar, Portugal	www.ipt.pt/iwms08
3–7 aug 2008	2008 Joint statistical meeting	Denver, US	www.amstat.org/meetings/jsm/2008/
17–21 aug 2008	29th Annual Conference of ISCB - the International Society for Clinical Biostatistics	Copenhagen, Denmark	www.iscb2008.info/
24–29 aug 2008	COMPSTAT 2008: International Conference on Computational Statistics	Porto, Portugal	www.fep.up.pt/compstat08
27–31 aug 2008	The 2008 European Meeting of the Econometric Society	Milan, Italy	www.eea-esem2008.org/
22–24 sep 2008	ENBIS	Aten, Grekland	www.enbis.org/