

QVARTILEN

Svenska statistikersamfundets tidskrift

Produktionskostnad: 35 kronor

Årgång 22 • Nummer 3 • September 2007

Samfundet
♥
Statistiska föreningen
=
sant ?

Innehåll		Gästkolumnist	8
Ordföranden har ordet	3	Mikrosimulering	11
Redaktören har ordet	3	Debatt	17
Styrelsen	4	Bokanmälan	18
Industriell statistik	5	Kåseri om statistik som tillämpning	19
Föreningen för medicinsk statistik	6	Teoretiska artiklar	21
Surveysektionen	7	Rapporter	26
		Konferenskalendarium	28



		Telefon	E-post
Redaktör	Joakim Malmдин	090 - 786 55 30	qvartilen@gmail.com
	Jens Olofsson	019 - 30 34 60	qvartilen@gmail.com
Redaktion	Magnus Arnér	0520 - 48 30 02	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	08 - 553 269 80	marie.linder@socialstyrelsen.se
	Johan Lyhagen	018 - 471 28 44	johan.lyhagen@dis.uu.se

Ansvarig utgivare: Eva Elvers

Redaktör: Joakim Malmdin och Jens Olofsson

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

Adress: Qvartilens redaktion
c/o Joakim Malmdin
Statistiska institutionen, Umeå universitet
901 87 Umeå
qvartilen@gmail.com
alternativt
Qvartilens redaktion
c/o Jens Olofsson
ESI, Örebro universitet
701 82 Örebro
qvartilen@gmail.com

Riktlinjer för material

MANUSSTOPP NUMMER 4: 15 november 2007

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 L^AT_EX-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.

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.

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.

Annonsspriser

	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 drygt 1350 medlemmar verksamma inom industri, förvaltning, universitet och högskola.

Adress: Svenska statistikersamfundet
c/o Marie Wiberg
Statistiska institutionen, Umeå universitet
901 87 Umeå
sekrsamf@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.

Styrelsen

		Telefon	E-post
Ordförande	Eva Elvers	08 - 506 947 14	eva.elvers@scb.se
Vice ordförande	Björn Holmquist	046 - 222 89 26	bjorn.holmquist@stat.lu.se
Sekreterare	Marie Wiberg	090 - 786 95 24	sekrsamf@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
	Ann Lindqvist	08 - 553 835 68	ann.lindqvist@scania.com
	Boris Lorenc	019 - 17 67 83	sekreterare.survey@statistikersamfundet.se
	Anna Torráng	08 - 553 259 75	anna.torrang@astrazeneca.com

Ordföranden har ordet

När jag skriver denna ruta är det mitten av augusti. Arbetskorridoren som var stilla för två veckor sedan är tämligen fylld igen, och hemmavid har de flesta grannarna återvänt. Variationerna i svaren på den enkla frågan "Hur har du haft det i sommar?" är stora, förutom variationer i sysselsättning finns det variationer i väder – för att inte tala om variationer i tillfredsställelse med väder.

Ett stort och varmt tack till alla som har tagit sig tid att besvara enkäten om framtiden! Jag ägnar avsevärd tid åt att läsa och begrunda framförda synpunkter, reflektioner, åsikter och förslag av olika slag – avsevärd tid just för att det finns mycket i svaren som är viktigt och klokt att fundera runt. En första information om svaren finns i detta nummer av *Qvartilen*, och mer kommer på årsmötet.

Vilka slutsatser kan man dra av en sådan här undersökning? Vi vet vad de svarande har svarat. I övrigt vet vi inte vad medlemmarna tycker. Vilka valde att svara och vilka att avstå? Det lär vi inte få veta annat än för den lilla grupp som kommenterade detta. Det är även här en stor variation mellan svaren. Längden på medlemskap, som går från de helt nya till de mycket erfarna, är bara ett av många skäl.

Ett extra tack till alla som kan tänka sig att bidra till verksamheten på ett eller annat sätt, nu eller längre fram. Nya krafter behövs och är synnerligen välkomna. Om alla de goda tankar som framförs i enkätsvaren ska kunna realiseras krävs det betydande insatser av många.

Verksamhetsåret är nästan till ända noterar jag och därmed snart mitt andra år som ordförande. Det första året innehöll många upptäckter och lärdomar. Det andra året har känts litet mera vant – men bara litet. Ordförandeskapet är inte slentrianmässigt utan alltid spännande och innehållsrikt. Det eventuella samgåendet med Statistiska föreningen har satt sin prägel på båda åren. Det har krävt sin tid och det har betytt att verksamheten i övrigt har blivit mer förvaltande och mindre utvecklande än den annars skulle ha varit.

Jag tycker att man ska korta ordförandens startsträcka genom att planera så att ordföranden är med i styrelsen året före sitt första ordförandeår, lämpligen som vice ordförande. Det ger kunskap och mental förberedelse.

Ytterligare något som behövs är tid – hur mycket tid? Det kan vara bra att veta i förväg vad man ger sig in på. I efterhand finner jag det mindre viktigt. Jag har inte varit ordförande för min skull utan för verksamhetens. Det hindrar inte att det har gett mig själv mycket. Denna ideella verksamhet ligger mig varmt om hjärtat, och den gör det ännu mer nu än för två år sedan.

Eva Elvers

Redaktören har ordet

Så är sommaren över och hösten gör sig alltmer synlig i såväl arbete som natur. När detta nummer ligger i din brevlåda är det bara några få dagar kvar till Samfundets årsmöte, ett möte som har Samfundets framtid i sina händer och med det även den tidning som du nu har fått.

Vi är för en sammanslagning mellan Svenska statistiker-samfundet och Statistiska föreningen då vi bland annat vill tro att en större förening ger en ökad möjlighet till erfarenhetsutbyte mellan alla de olika delar som utgör vårt ämne. En större förening ger också större möjligheter att värna om och föra fram vad statistik är, både som ämne och som tillämpning.

Även som redaktör är vi för en sammanslagning och instämmer med samgåendegruppen i att en av de (många) praktiska frågorna som finns vid ett samgående är en ny medlemstidskrift. Idag har respektive förening varsin – Samfundet har *Qvartilen* och Statistiska föreningen har *Konfidensen*. Den rapport som har tagits fram av samgåendegruppen nämner inga detaljer om en medlemstidskrift. Vad som däremot sägs är att

En medlemstidskrift utgör ett viktigt forum för vetenskap, erfarenhetsutbyte, diskussion och debatt. Tidskriftens redaktion ska bedriva en bred verksamhet som främjar utbytet mellan medlemmar inom olika forsknings- och tillämpningsfält. Debatt och diskussion om statistiska synsätt och etiska frågor bör vara en viktig ingrediens.

Qvartilen anses av många vara en av Samfundets främsta tillgångar. Upplagan för det senaste numret var drygt 1 000 exemplar och vi vill gärna tro att det finns något för alla att läsa i varje nummer. Att den nya medlemstidskriften inte är, eller ska vara, *Qvartilen* tar vi för givet. Vad som däremot är angeläget, vid ett "Ja" till samgående, är att i detalj börja diskutera hur en medlemstidskrift för den nya föreningen bör utformas.

Anledningen till detta är att vi under året har lärt oss att det är bra med framförhållning, både vad gäller det tekniska arbetet med tidningen och, framför allt, när det gäller att vara i kontakt med presumtiva skribenter. Vi lever i en tid när tid är en bristvara för de allra flesta, trots att det är det enda vi alla har lika mycket av. Att få till den tidning du håller i din hand kräver en del arbete, inte bara av redaktör och redaktionsmedlem, utan även av dem som tar av sin tid för att ge tidningen sitt bidrag.

I detta nummer får du bland annat ta del av erfarenheter från en rad konferenser, sammanfattningar av nya doktorsavhandlingar, Kenneth Carlings syn på hur forskarutbildningen borde bedrivas samt bidrag av Anders Klevmarken och Erling Lundevaller Häggström, som ger oss två olika infallsvinklar till mikrosimuleringarnas förlovade land. Tack alla bidragsgivare för ett mycket gott och stimulerande samarbete!

Joakim Malmдин och Jens Olofsson

Styrelsen

Styrelsebeslut tagna på sistone

Sommarskolan 2007 genomfördes med 28 deltagare på Linköpings universitet.

Årets årsmöte för Samfundet samt sektionerna kommer att äga rum 15–16 oktober på Stockholms universitet, se kallelsen.

En enkät om hur Samfundets medlemmar ser på verksamheten och på det eventuella samgåendet med Statistiska föreningen har gått ut under sommaren. Vid tiden för Qvartilens manusstopp i augusti hade över 100 svar inkommit. De flesta av dessa svar ställer sig positiva till ett samgående. Det finns även några som är negativa samt flera som inte har någon åsikt. Ett flertal förslag på vad Samfundet ska arbeta med i framtiden har

inkommit, inklusive förslag på nya sektioner. Fullständig redovisning av enkäten kommer att ske på årsmötet i oktober.

En gemensam mall för sektionernas stadgar ska utformas och en arbetsgrupp är tillsatt för att genomföra detta.

Samfundet beslutade att ställa sig bakom en svensk ansökan om att genomföra ECAS summer school 2009.

Den första skrivningen om äganderätt till material i Qvartilen var tyvärr inte tillräckligt genomtänkt. Styrelsen har i samförstånd med redaktörerna tagit fram en ny text, som införs i detta nummer av Qvartilen (sidan 2).

Styrelsen genom Marie Wiberg och Eva Elvers

Program för Svenska statistikersamfundets årsmöte 2007

Tid: Tisdagen den 16 oktober 2007, ca kl 9.00 - 17.00.

Plats: Juristernas Hus, Frescativägen 18, Stockholms universitet.

Mat: För dig som anmält dig till middagen måndagen den 15 vill vi påminna om att den kommer att äga rum klockan 18.30 på Vårdshuset Kräftan, Roslagsvägen

101, Kräftriket 9, Stockholm.

Övrig information: Samfundets tre sektioner, Föreningen för Medicinsk Statistik, Industriell statistik och Surveysektionen, kommer att hålla sina respektive årsmöten under måndagen den 15 oktober, också vid Stockholms universitet. För mer information, kontakta respektive sektion. Varmt välkommen.

Styrelsen genom Marie Wiberg

Program

09.00 - 09.05	Välkommen
09.05 - 09.45	Professor Olle Häggström, Chalmers tekniska högskola och Göteborgs universitet. Föredrag <i>Vetenskap, pseudovetenskap och gråzonen däremellan</i>
09.45 - 10.10	Kaffepaus
10.10 - 11.35	Visionsblock
11.35 - 12.45	Lunch (på egen hand)
12.45 - 13.20	Katrin Kraus, Uppsala universitet. <i>Faktoranalys av ordinaldata med robust maximum likelihood: Har antalet svarsalternativ för de observerade variablerna någon betydelse för resultaten?</i>
13.20 - 14.00	Utdelning av priset <i>Årets yngre statistiker</i> samt föredrag av pristagaren
14.00 - 14.30	Kaffepaus
14.30 - 16.30	Årsmötesförhandlingar
16.30 -	Avslutning

Obs! Tidpunkterna för föredragen är preliminära, tidpunkten för årsmötesförhandlingarna är dock fastställd enligt ovan.

Industriell statistik

Thomas Svensson deltog i juli vid konferensen Mathematical Models in Reliability (MMR) och en workshop anordnad av ENBIS (European Network of Business and Industrial Statistics) i Glasgow. Här delger han oss sina intryck.

Konferens om matematiska tillförlitlighetsmodeller

av Thomas Svensson, SP Sveriges Tekniska Forskningsinstitut, Borås

Efter en terrorattack vid Glasgows flygplats på lördagen blev våra resor till staden ganska besvärliga. Inga allvarliga incidenter inträffade för konferensdeltagarna, men flygplatsen var stängd även på söndagen med svårigheter som följd. Själv hamnade jag i Newcastle för vidarebefordran med taxi till Glasgow, en tre timmars bilfärd. Jag kom ändå nästan i tid till ENBIS workshop på söndagseftermiddagen och vi höll vårt möte bland ett tiotal åhörare. Mötet innehöll en givande diskussion då två grupper, med ganska skilda utgångspunkter, möttes för presentationer med förberedda diskussioner. En grupp var från Italien. Deras område var tillämpning av tillförlitlighetsteori på cylindrar till båtmotorer. Den andra gruppen var från Göteborg och vi berättade om tillämpade metoder inom bil- och flygmotorindustri. Italienarna representerades av Fabrizio Ruggeri och Massimiliano Giorgio. Åke Lönnqvist från Volvo PV och undertecknad fick ensamma representera den svenska gruppen då Bo Bergman från kvalitetsteknik på Chalmers tyvärr var sjuk och tvingades stanna hemma.

På måndagen startade konferensen på allvar. Personligen är jag numera så inriktad på industriella tillämpningar att föredragen visade sig vara något bortom min horisont. Det var förstås helt i sin ordning då konferensens rubrik var *Mathematical Models in Reliability*. Jag bestämde mig ändå för att i mitt lyssnade utgå från frågorna "Är det användbart?" och "Har det i så fall kommit till nytta?" för att bedöma de olika föredragen under konferensen.

Resultat

Jag lyssnade på 29 föredrag. På frågan om användbarhet fick jag inga svar. Det är helt enkelt en fråga som i allmänhet inte ställs. Möjligen har man ibland en önskan om att det skall vara användbart, åtminstone inleds många föredrag med presentation av tänkbara tillämpningsområden (inspiration?). I de fall man överhuvudta-

get återknyter till tillämpningen är det emellertid oftast så att det saknas tillräckligt med data eller att det krävs ytterligare studier för att kunna använda metoderna.

Min Xie från Singapore, med svensk doktorsexamen, var en av de få som hade en uttalad ambition att lösa industriella problem och han sa under sitt föredrag ungefär så här:

Det är enkelt att publicera uppsatser inom det här området, det är svårare att göra något som är användbart och än svårare att få någon att använda det som ändå är användbart.

Den enda metod som presenterades bland de 29 presentationerna och verkligen har kommit till användning var en typ av komplettering av traditionellt fördelningspapper. Vasilij Krivtsov från Ford i USA presenterade idén att inkludera förkunskap i fördelningskattningar genom att skatta fördelningspapperets räta linje med regression och komplettera designmatrisen med förkunskapsrader. Detta är enkelt, tycks vara användbart, och har tydligen också kommit till nytta hos Ford.

Bo Bergman framförde i sin förberedda diskussion av Giorgios föredrag följande:

Essentially, predictions are based on three premises:

- 1 the phenomenon under study is predictable,*
- 2 some data from the phenomenon is available,*
- 3 a model with predictive capacity is found.*

Dessa förutsättningar skulle jag önska att fler av statistikerna funderade över inför sitt modellerande. Men, å andra sidan, konferensen hade inte alls något uttalat syfte att vara tillämpad så mina önskningar får väl stå för mitt eget intresse. Det kanske är så att jag hamnade på fel konferens. ■

Påminnelse

Du som ännu inte betalat avgiften på 130 kr för medlemskap i Svenska statistikersamfundet 2007 kan be-

tala in på postgironummer 65 41 95-7. Ange ditt namn och medlemsnummer och år 2007 i samband med inbetalningen.

Föreningen för medicinsk statistik

Program för FMS höst- och årsmöte

TEMA: Djur och statistik

TID: Måndagen den 15 oktober 2007

PLATS: Stockholms universitet, Reinholdsalen i Juristernas hus.

Sammanfattningar av föredragen finns på sidan www.statistikersamfundet.se/fms/.

Styrelsen genom Tom Britton

Program (föredragets titel indikerar språk)

09.30-10.00	Registrering, kaffe
10.00-10.05	Tom Britton, ordförande FMS: Välkomsthälsning
10.05-10.35	Lars Rönnegård, Linnécentret för Bioinformatik, Uppsala Universitet: Husdjur öppnar vägen till ökad förståelse inom genetiken
10.40-11.10	Niclas Sjögren, AstraZeneca R&D Södertälje: Djur i läkemedelsutveckling
11.10-11.20	Bensträckare
11.20-11.50	Ulf Emanuelson, Institutionen för Kliniska Vetenskaper, Sveriges Lantbruksuniversitet: Vad finns i verktygslådan hos en veterinärmedicinsk epidemiolog?
11.55-12.25	Lennart Norell, Tillämpad statistik och matematik, Sveriges Lantbruksuniversitet: Djur och statistik vid SLU
12.25-13.40	Lunch
13.40-14.10	Howard Needham, European Centre for Disease Control and prevention: Public health aspects of Influenza A/H5N1 in the EU
14.15-14.45	Åke Pettersson, Statistiker och fågelinventerare, SCB och Sveriges Ornitologiska Förening: Att räkna fåglar
14.45-15.20	Kaffe med bulle
15.20-15.50	Dragi Anevski, Matematisk statistik, Chalmers: Tider mellan händelser i en räkneprocess och fågelskådning
16.00-17.00	Årsmötesförhandlingar
17.00-18.00	Eventuellt mingel
ca 18.30	Samfundets middag

Nya medlemmar

Samfundets styrelse hälsar följande medlemmar välkomna:

Shubila Balaile, Dan Gustafsson, Anna Norin, Ulf

Boman, Katarina Lundman, Tom Andersson, Johan Lindqvist, Carolina Thulin, Peter Pernemalm, Peter Zbornik, Tuna Zeliha, Daniel Wikström, Johan Nilsson, Erik Gustafsson, Amanda Waleh, Nicklas Pettersson, Tommy Nyberg och Frida Hultgren.

Surveysektionen

Surveysektionen har ansökt om medlemskap i den internationella motsvarigheten IASS (International Association of Survey Statisticians). IASS har funnits sedan 1973 och är öppet både för enskilda medlemmar och institutionsmedlemmar. Gordon Brackstone, Kanada, är ordförande, och Lilli Japac, Sverige, är vetenskaplig sekreterare. Medlemmarna är 1000–1100 till antalet och kommer från ett 130-tal olika länder. IASS tidning heter *Survey Statistician*. Sekretariatet för IASS finns i Libourne i Frankrike. Tidigare ansökningar har fastnat i byråkratins kvarnar, men vi hoppas att det går bättre den här gången.

Styrelsen har hållit sammanträden den 29 maj i Stockholm och den 22 augusti per telefon. En viktig fråga för Surveysektionen är att reda ut det juridiska förhållandet till moderföreningen, dvs. Svenska statistikersamfundet (i framtiden kanske Svenska statistikföreningen). Sektionen utgör i nuläget inte en egen juridisk person, men agerar i många avseenden som om den vore det.

Egentligen borde samfundet besluta om firmatecknare och ha det yttersta ansvaret för ekonomin för sektionen. I praktiken har dock sektionen uppträtt helt självständigt. I samråd med övriga sektioner inom samfundet arbetar nu Surveysektionen med att lägga fram ett förslag till nya stadgar till årsmötet, vilka gör det möjligt för sektionen att bli en egen juridisk person och ansöka om organisationsnummer för ideell förening. Ett steg i den riktningen kan vara att byta namn från sektion till förening, varpå vi också hamnar i paritet med Föreningen för medicinsk statistik.

Välkomna till årskonferensen den 15 oktober i Stockholm! Deltagandet i konferensen är gratis för alla. Endast årsmötet, som äger rum sist på eftermiddagen, är begränsat till medlemmarna. På konferensen bjuds det bland annat på ett panelsamtal om läget för den svenska utbildningen i surveystatistik.

Styrelsen genom Jörgen Svensson

Program för Surveysektionens årsmöte

10.00-10.05 Välkommen och praktisk information, Jörgen Svensson, SCB
10.05-10.40 *Hur når man mobilabonnenter?* Ulf Isander och Åsa Jaktlund, SIFO Research International.
10.40-11.15 *Hur mäter man inflationen?* Martin Ribe, SCB
11.15-11.25 Bensträckare
11.25-12.00 *ISO-standard för marknads-, opinions- och Samhällsundersökningar* Stig Holmer, SMIF
12.00-13.00 Lunch på egen hand
13.00-13.50 För 50 år sedan: *Det stora språnget i surveystatistik* Gunnar Kulldorff, Umeå universitet

13.55-14.00 Utdelning av Uppsatspriset 2006/07
14.00-14.20 Vinnarens presentation av uppsatsen
14.20-14.40 Kaffe
14.40-15.10 *Analys av kvasi-experimentella undersökningsdata inom arbetstidsförkortning och hälsa* Mats Nyfjäll, Statisticon
15.10-15.55 Paneldiskussion om utbildningsläget inom Surveyområdet Eva Leander, Thomas Laitila, Daniel Thorburn, Hans Nyquist
16.00-ca.17.00 Årsmötesförhandlingar
ca.17.10-17.50 Mingel
ca.18.30 Middag i Samfundets regi

Tredje internationella konferensen om företagsundersökningar

Den 18–21 juni hölls den tredje ICES-konferensen (International Conference on Establishment Surveys), som behandlade statistiska metodfrågor vid undersökningar av företag, lantbruk och andra organisationer. Tidigare ICES-konferenser har hållits i Buffalo 1993 och 2000. Årets konferens hölls i Montréal i provinsen Québec i Kanada på hotellet Hyatt Regency. Av konferensens cirka 410 registrerade deltagare kom 14 från Sverige, alla från Statistiska centralbyrån.

Det som kanske var mest slående med konferensen för min egen del var att jag fick se många av de kända surveyforskare som jag tidigare mest sett till namnet i referenslistor. Det blev också tillfällen till erfarenhetsutbyte med kolleger från Kanada, USA med flera länder.

Överhuvudtaget var det nordamerikanska inslaget starkt i konferensen, och jag fick incitament att bättra på mina engelska språkfärdigheter.

Föredragen spände över ett brett fält av metodfrågor, från varians estimation och säsongrensning till frågekonstruktion och elektroniska datainsamlingsmetoder. Även kvaliteten på föredragshållarna visade stor variation. Den avslutande middagssessionen bjöd på ett panelsamtal om möjligheter och problem med att använda administrativa datakällor för statistiska ändamål – ett hett ämne i tider med krav på minskad uppgiftslämnarbörd och minskade kostnader.

Jörgen Svensson

Gästkolumnist

I förra numret (22-2) av Qvartilen skrev Robert Lundqvist, Luleå tekniska universitet om hur han skulle vilja bedriva undervisning på grundnivå. Till detta nummer har vi ställt frågan: Hur skulle du vilja bedriva undervisning i ämnet (matematisk) statistik på forskarnivå? till Kenneth Carling, Högskolan Dalarna.

Vi behöver mål för forskarutbildning i statistik

av Kenneth Carling, Högskolan Dalarna

Jag har varit opponert på tre doktorsavhandlingar i statistik de senaste tre åren, och därtill läst, granskat och opponerat på ett ganska stort antal avhandlingar i andra ämnen. Dessa läsningar har väckt min nyfikenhet kring vad svenska avhandlingar i statistik handlar om idag och hur de är uppbyggda. Och vilka underliggande mål som är vägledande för dem.

Erfarenhet av Bologna-mål för kandidatexamen

Mina lösa funderingar och ostrukturerade tankar kring detta fick ett större mått av skärpa under hösten 2006 när vi i ämnesgruppen i statistik vid Högskolan Dalarna ägnade några intensiva månader att formulera reviderade kursplaner med anledning av att Bologna skulle implementeras vid högskolan. Jag hade aldrig tidigare befunnit mig i ett sammanhang där statistikutbildningens mål diskuterats med samma intensitet och skärpa som då. Och jag tycker att Bolognas tre målgrupper – *Kunskap och Förståelse, Färdighet och Förmåga* samt *Värderingsförmåga och Förhållningssätt* – är en betydande hjälp i att strukturera utformningen av kursplaner och därmed kursernas genomförande [1].

Det stod tidigt klart i denna diskussion att vi borde börja med att uttrycka målen för en statistikexamen (och vi valde att göra det för kandidatnivå). Ett skäl till att börja med examen snarare än en enskild kurs var att det är tämligen konkret att diskutera vad som ska känneteckna en färdigutbildad statistikstudent. En sådan diskussion kan till exempel föras genom att man tar ett arbetsgivarperspektiv.

Det visade sig dock vara besvärligt att relatera målen till en kandidatexamen. Bättre gick det när vi försökte relatera mål till examensarbetet (alltså C-uppsatsen). Vi uppnådde ganska snart en konsensus kring målen för *Färdighet och Förmåga*. Målen för *Värderingsförmåga och Förhållningssätt* såsom angivna i Högskolelagen ansåg vi vara i det närmaste direkt tillämpbara för statistik. Den längsta, och vingligaste, diskussionen rörde målen om *Kunskap och Förståelse*. Läsaren inser säkert snabbt att en grupp akademiska statistiker har svårt att nå enighet om frågor som: Ska en statistiker med kandidatexamen ha kännedom om momentgenererande funktioner? Behöver hon känna till vem Adolphe Quetelet var?

Den stora fördelen med att ha målen för examen formulerade är att arbetet med enskilda kursplaner följer en rent deduktiv logik. I stort sett reduceras arbetet till att bestämma vilka delmål som behöver sättas upp för re-

spektive kurs för att studenten ska nå de slutliga målen och därefter fastställa när under grundutbildningen studenterna ska arbeta mot delmålen.

Ett ytterligare steg tillkommer och det är att fastställa behörighetskraven till varje kurs, vilket är ganska lätt. Det kräver en uppfattning om vad studenten kan tillgodogöra sig avseende kunskap och färdighet under en stipulerad kurs tid. En ämnesgrupp med erfarenhet av statistikundervisning torde inte ha något större problem att skatta denna mängd. En enkel skattning är att utgå från vad tidigare studenter klarat av att lära sig på redan existerande kurser. Behörighetskraven bör vara kursens delmål minus den mängd som studenten skattats kunna tillgodogöra sig.

Vägen till en fullständig utbildningsplan på grundnivå torde nu vara uppenbara. Med början i examensarbetets mål stegar man sig nedåt i utbildningen till det att alla kursplaner upprättats och då har man som slutprodukt också erhållit behörighetskraven för att påbörja högskolestudier i statistik.

Låter det enkelt? Efter att ha deltagit i förra höstens arbete vill jag påstå att det är nästan lika enkelt som det låter.

Dagens mål för doktorsexamen

Högskolelagen anger generella mål för kandidatexamen och arbetet som beskrevs ovan gick ut på att specificera dessa mål för examen i statistik. Högskolelagen anger också generella mål för doktorsexamen och med tillgång till specifika mål för doktorsexamen i statistik så kunde studieplanen för forskarutbildningen härledas på samma sätt som ovan. Så låt oss undersöka de specifika mål som anges i studieplanerna till forskarutbildningen i statistik. Jag har gått igenom huvuddelen av landets lärosäten som erbjuder sådan utbildning och låter Stockholms universitet tjäna som exempel. Statistiska institutionen vid Stockholms universitet fick ett mycket gott omnämnande i samband med Högskoleverkets utvärdering av ämnet i landet och borde kunna tjäna som förebild för andra lärosäten.

Här är formuleringen av målen för forskarutbildningen i Stockholm enligt deras studieplan:

Forskarutbildningen syftar till att ge breda ämneskunskaper inom statistikens centrala delar, fördjupade kunskaper inom något statistiskt specialområde, förmåga att kunna tillämpa de statistiska metoderna på praktiska problem samt förmåga att följa den vetenskapliga litte-

raturen i ämnet och bedriva egen vetenskaplig forskning.

Jag gör två reflektioner när jag läser denna formulering. Den första är att vissa mål i Högskolelagen är utelämnade. Ett exempel är att de generella målen för *Färdighet och Förmåga* ställer krav på att doktoranden ska kunna utföra kvalificerade uppgifter inom givna tidsramar och ett annat är att doktoranden ska visa förmåga att i internationella sammanhang muntligt och skriftligt presentera och diskutera forskning och forskningsresultat.

Den andra reflektionen är att målen är så allmänt hållna att de inte ger någon vidare vägledning till hur forskarutbildningens ska utformas. Ta till exempel målet om breda ämneskunskaper inom statistikens centrala delar. Betyder det att den nyblivne doktorn ska vara väl förtrogen med litteraturen om survey-metodik och känna till alla för närvarande existerande procedurer och makron i SAS?

Jag blev ombedd att skriva denna artikel och angripa följande fråga "Hur skulle du vilja bedriva undervisning i ämnet statistik/matematisk statistik på forskarnivå?". Till synes ägnar jag all tid åt att undvika den, men vad jag vill ha sagt med denna artikel är att frågan inte låter sig besvaras om målen med utbildningen är otydliga. Så låt oss återgå till att diskutera målen.

Handledargruppen i Stockholm har säkert en betydande samsyn kring hur dessa, för mig vaga, formuleringar ska uttolkas och de har nog en vana att tillämpa dem i bedömningen av faktiska fall. När jag har varit med i betygsnämnder så har diskussion ofta handlat om målen med forskarutbildningen (och hur de olika målen ska vägas samman i ett slutomdöme) och ledamöterna har utbytt erfarenheter och uppfattningar. Problemet är att sådana diskussioner inte tecknas ned och sammanställs vilket försvårar en mer allmän diskussion om målen bland akademiska statistiker. Och framtida arbetsgivare för doktoranden (förutom högskolorna och universiteten) har usla möjligheter att påverka de faktiska målen eftersom företrädare för arbetsgivare mycket sällan är behöriga att delta i handledarmöten eller betygsnämnder.

En annan faktor som förmodligen påverkar de faktiska målen är vad tidskrifterna väljer att publicera. Jag har aldrig hört en diskussion som har handlat om vilka tidskrifter som är acceptabla och vilka som inte är det, vilket får mig att tro att det finns en samsyn där alla statistiska tidskrifter betraktas som acceptabla. Men hur är det med metodinriktade tidskrifter i andra discipliner?

Jag är benägen att tro att många handledare ser tidskriftsnormen som mål. Pondera att en doktorand lägger fram en sammanläggningsavhandling bestående av tre ensamförfattade arbeten på samma tema där samtliga är accepterade för publikation i en och samma statistiska tidskrift. Den doktoranden kunde nog vara helt säker på att få doktorsexamen. Men hur vet man att hon har "breda ämneskunskaper"? Eller är det så att "breda ämneskunskaper" inte är nödvändiga om hon besitter ett betydande djup i ett delområde? Så vitt jag förstår

ska Högskolelagens generella mål ses som nödvändiga villkor vart och ett av dem, vilket skulle innebära att en doktorand inte kan få examen om hon fallerar med avseende på något av målen. Och doktoranden i exemplet kanske inte borde godkännas.

Jag vet efter min granskning av studieplaner i landet att dokumenterade mål för forskarutbildningen i enlighet med Högskolelagens intentioner saknas. Jag tror dock att mindre grupper såsom handledargrupper har en viss samsyn kring målen. Dessa grupper bör fatta pennan och skrivna ner de mål som de tillämpar!

Några mål som jag önskar vore uppfyllda

I väntan på att handledare i statistik ska dela med sig av sina uppfattningar om specifika mål så tar jag upp några saker som jag tycker att en statistiker med doktorsexamen borde ha kunskap om eller ha visat förmåga till.

Jag bad Rickard Stridbeck identifiera de 20 senaste avhandlingarna och rekvirera dem. Han lyckades i februari 2007 ha fått tag på de flesta med undantag för avhandlingar framlagda vid Handelshögskolan i Stockholm (som ville ha betalt för att skicka dem). Med detta material samt Högskoleverkets rapporter om Statistik (och Demografi) och Matematisk Statistik som grund vill jag lämna några reflektioner.

Jag skulle vilja föreslå ett kunskapsmål: förtrogenhet med tillämpningsområdet och ett aktivt bidrag till dess kunskapsutveckling.

En doktor i statistik torde normalt ägna en stor del av sin tid åt att samarbeta med forskare eller utredare med bakgrund i andra discipliner. Med tanke på det så tycker jag att det är en intressant fråga hur närvarande andra discipliner är i avhandlingarna i statistik. Det första jag noterar efter att ha gått igenom materialet är att i huvuddelen av avhandlingarna ingår någon medförfattare från en annan disciplin – oftast i form av ett samarbete kring en av avhandlingens uppsatser. Något förvånande är det då att finna att avhandlingssuppsatserna aldrig publiceras i icke-statistiska tidskrifter. Ett lysande undantag är Claudia Libisellers avhandling [2]. Det kan naturligtvis hända att så sker efter att avhandlingen lagts fram. Hursomhelst infinner sig frågan: Anser doktoranderna att statistiska tidskrifter har högre meritvärde och därför väljer dem i första hand eller beror det på att uppsatserna som godkänts som goda statistiska arbeten betraktas som ointressanta eller undermåliga i den disciplin uppsatsen berör?

Ytterligare förvånande är att referenserna uteslutande är till statistiska tidskrifter trots att uppsatserna behandlar ett flerdisciplinärt område och återigen är Claudia Libisellers avhandling ett undantag liksom David Bocks avhandling [3]. Jag undrar vad medförfattarna från en annan disciplin fyller för roll när det verkar som att den disciplinen saknar relevanta referenser och inte heller har intressanta tidskrifter för publicering. En mer oroande tanke är att de faktiska målen med doktorsexamen inte innefattar målet om förtrogenhet med teori och lit-

teratur i den disciplin (förutom statistik) avhandlingen berör.

Jag skulle vilja föreslå ett färdighetsmål: visa förmåga att samla in data och bearbeta dessa inför en statistisk analys.

I de flesta introduktionsböcker i statistik beskrivs ämnet att handla om att samla in och att analysera data. Det skulle man inte ha gissat om man läser svenska, nutida avhandlingar i statistik. Ingen av de jag har gått igenom bygger på data insamlade av doktoranden (Peter Werners avhandling [4] visar dock på god förtrogenhet med datainsamlingsarbetet). Jag hoppas att handledarna betraktar förmågan att samla in och bearbeta data som ett faktiskt mål för examen, men efter denna genomgång så är jag tveksam om så är fallet.

Till slut skulle jag vilja ta upp frågan: vilken utbildningsbakgrund har doktorander i statistik? En reflektion man kan göra av det material jag har gått i genom är att det är tunt med avhandlingar i en del områden som har stor användning av statistisk metodik. Nämnas kan, som exempel, Demografi, Psykologi, Sociologi och mera allmänt teknik och naturvetenskap (fast jag har inte studerat avhandlingar i matematisk statistik). Är detta en följd av en ensidig rekrytering av doktorander män? Och har vi som akademiska statistiker en skyldighet att verka för

att inga subdiscipliner i statistik lämnas utforskade?

Den senare frågan ställde jag till Roland Granqvist som är professor i Företagsekonomi för att höra hur han såg på den i sitt ämne. Han menade att handledare i Företagsekonomi har ett kollektivt ansvar att alla delar i ämnet utforskas. Jag är benägen att ha samma åsikt när det gäller statistik. ■

Referenser

- [1] Svensk författningssamling (2006) *Högskolelagens nivåbeskrivningar och generella examensbeskrivningar*, SFS 2006:173.
- [2] Libiseller, C. (2003). *Considering Meteorological Variation in Assessment of Environmental Quality Trends*. Doctoral thesis. Matematiska institutionen. Studies in Statistics No. 3, Linköpings universitet.
- [3] Bocks, D. (2004). *Statistical surveillance. Optimal decision times in economics*. Doctoral thesis. Skriftserie - Publications, No. 31, Statistiska forskningsenheten, Göteborgs universitet.
- [4] Werner, P. (2005). *On the cost-efficiency of probability sampling based mail surveys with a web response option*. Doctoral thesis. Linköping studies in statistics, 1651-1700 ; 6, Matematiska institutionen, Linköpings universitet.



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

Demografisk mikrosimulering

av Erling Häggström Lundevaller, Umeå universitet

Den demografiska utvecklingen i västvärlden präglas av minskad fertilitet och mortalitet. I många områden närmar sig fertiliteten ett barn per kvinna vilket leder till en snabb och kraftig befolkningsminskning om man bortser från effekten av migration. Kombinerat med att vi lever allt längre innebär det stora förändringar av befolkningssammansättningen med konsekvenser för bland annat ekonomi och socialförsäkringssystem. Allt färre kan komma att få försörja allt fler. Sverige har varit relativt lite påverkad av detta på grund av höga fertilitetstal (1,77; 2005) och stor immigration (96 000 personer; 2006). Prognoser från SCB visar en ökning från 9 miljoner invånare 2004 till 10,5 2050. Även om Sverige som helhet inte kommer att påverkas i en nära framtid kan det finnas problem på regional nivå. Olika delar av landet befinner sig i olika stadier av den demografiska omvandlingen till en äldre befolkning. Norrlands inland ligger före i och med att det där redan finns en relativt gammal befolkning som ger en liten bas av unga som kan förväntas få egna barn. Att studera sådana områden kan ge en bild av hur utvecklingen kan komma att bli i andra delar av landet.

De stora effekter den demografiska utvecklingen har på samhället gör det viktigt att ta fram verktyg för att prediktera utvecklingen samt studera olika scenarior baserade på olika utveckling av fertilitet, mortalitet och migration. Det kan hjälpa till att på ett realistiskt sätt dimensionera samhällsservice och social försäkringssystem i förhållande till befolkningens behov och skatteunderlag. Inte bara tiden utan också rummet har betydelse i och med att personerna kan flytta vilket också måste kunna hanteras.

Denna artikel syftar till att ge en beskrivning av hur och varför ett mikrosimuleringssystem kan användas för att belysa sådana processer och ge ett enkelt exempel i form av en simulering av den demografiska utvecklingen i Västerbotten.

Varför mikrosimulering?

Demografiska prediktioner görs ofta på aggregerade data och sådana prognoser finns att tillgå från till exempel SCB. Så varför göra sig det extra besväret med att göra en mikrosimuleringsmodell? För våra syften här är ett uppenbart skäl att vi vill kunna göra prognoser på valfri nivå av rumslig aggregering. Databasen vi utgår från lokaliserar individerna i 100×100 meters rutor vilket gör att det går att följa den demografiska utvecklingen på detaljerad nivå genom mikrosimuleringsansatsen. Genom sin utgångspunkt i individernas beteende ger mikrosimulering ett verktyg för experiment av hur fertilitet, mortalitet och migration påverkar den regionala utvecklingen. Det är oftast av naturliga skäl omöjligt att

genomföra verkliga experiment av denna typ. Ett annat skäl är att ansatsen gör det möjligt att använda individspecifika socioekonomiska variabler i modellerna vilket skulle kunna förbättra prognoserna.

Ett konkret exempel kan vara universitetets effekt på demografin i Umeå. Invånarantalet i Umeå är runt 111 000 varav 29 000 är studenter. Majoriteten av dessa är kvinnor och kvinnliga studenter står således för en stor del av de kvinnor som är i barnafödande ålder. Det är dock troligt att många kommer att skjuta upp barnafödandet till efter studiernas avslutande och sedan bidra med sin fertilitet någon annanstans än i Umeå. Är det så borde en mikrosimuleringsmodell som tar hänsyn till individernas socioekonomiska situation och flyttmönster kunna ge bättre prognoser för den demografiska utvecklingen speciellt på regional nivå.

Intressanta möjligheter öppnar sig genom att individerna även kan ha en familjetillhörighet i simuleringarna. Det gör det möjligt att inkludera familjeeffekter för till exempel fertilitet och migration i modellerna samt att följa hur den demografiska utvecklingen och flyttmönster påverkar familjestrukturen. Den demografiska utvecklingen för olika delpopulationer, till exempel baserat på socioekonomisk- eller etnisk bakgrund, går också att följa.

Mikrosimulering i Umeå

Kulturgeografiska institutionen Umeå har länge haft behov av verktyg för att kunna studera befolkningens utveckling. Institutionens forskningsintresse kan sammanfattas i begreppet populations geografi, men inom den ramen studeras speciellt regional flyttning, internationell flyttning, transportgeografi, regional utveckling och turism. I mycket av detta arbete används mikrosimulering som hjälpmedel. Både tid och rum är av stort intresse varför stor energi lagts på att utveckla mikrosimuleringsprogram som kan hantera både tids- och rumsdimensionen.

Olika mikrosimuleringsprogram har utvecklats men det som använts mest kallas SVERIGE. Förutom de rent demografiska modulerna som kommer att användas här finns även moduler som hanterar familjetillhörighet, arbetsplats, bransch, arbetsdeltagande och ekonomiska ersättningar. Tidigare bedrevs det huvudsakliga arbetet vid SMC (Spatial Modelling Centre) i Kiruna som en filial till Kulturgeografiska institutionen i Umeå. Nu har SMC flyttat merparten av sin verksamhet till Umeå. Ett exempel på tillämpning som gjorts med hjälp av SVERIGE är en studie om platser attraktion och arbetskraftutbud, där sambandet mellan dessa studerats. Ett annat exempel är en studie av de långsiktiga socioekonomiska effekterna av stora investeringar på små och medelstora

orter. Den studerar speciellt effekterna av att investera i en anläggning för slutförvar av kärnbränsle. SVERIGE skrevs ursprungligen i C++ men nu används huvudsakligen C#. Programmen R och Python har också använts i simuleringsarbetet.

För att skatta ekvationer för simuleringsmodellerna och som basår för simuleringarna behövs bra data. I Sverige har vi unika data på individnivå som kan användas för detta och en databas som går under namnet ASTRID har byggts upp vid institutionen som utnyttjar detta. Den innehåller detaljerade socioekonomiska data för hela Sveriges befolkning samt alla arbetsplatser och företag.

Simulering av Västerbotten

Som en illustration av hur en demografisk mikrosimulering kan göras och vad man kan få för information från en sådan visas här en simulering för Västerbotten.

Strukturen i ett simuleringsprogram är enkel. Först läses data in från databas. Sedan skattas modellerna som ska simuleras om detta inte redan är gjort i ett annat program. Den centrala delen är looperna över åren som innebär att individernas värden för variablerna uppdateras en gång per år. Årsdata sparas antingen som rådata eller i mer bearbetad form.

Data för år 2002 har används både för estimation och som start år för simuleringarna. Olika statistiska tekniker har använts för att skatta de olika modellerna som ingår. Individernas dödsrisk har skattats med en logistisk regression liksom även sannolikheten att flytta till ett annat land eller att flytta till en annan kommun i Sverige. Antalet barn en kvinna kan få är 0, 1 eller 2 och modellen har skattats med multinominell logistisk regression och endast kvinnor 16–44 år gamla får barn i modellen. Förklaringsvariablerna som använts i detta enkla exempel är ålder, åldersgrupp (0–5, 6–16, 17–28, 29–43, 44–58, 59–78, 79+), kön och födelseland i fem grupper (1. *Sverige* 2. *Resten av Norden* 3. *Västeuropa plus Australien, Kanada, Japan, Nya Zeeland och USA* 4. *Östeuropa* och 5. *Resten av världen*).

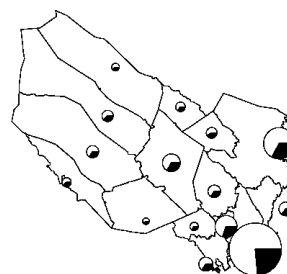
För sannolikheten att få barn har även ålder i kvadrat använts. För inflyttningen till kommunerna har en helt annan strategi valts. Observerade inflyttningar till kommunerna 2002 bildar en pool som inflyttare senare år slumpas ifrån (det vill säga inte de individerna utan nya individer med en del egenskaper satta till samma som de i poolen). Antalet immigranter bestäms av en kvot; antalet immigranter 2002 dividerat med antalet invånare i kommunen 2002, som multipliceras med invånarantalet aktuellt år för att ge ett realistiskt antal immigranter det året.

Simuleringsresultat

Figur 1 och 2 visar simuleringsresultat för den demografiska utvecklingen i Västerbotten åren 2005 och 2050. Det är viktigt att komma ihåg att detta bara är ett resultat av endast en körning, nya körningar med nya frön för slumpalen ger nya resultat även om skillnaderna är

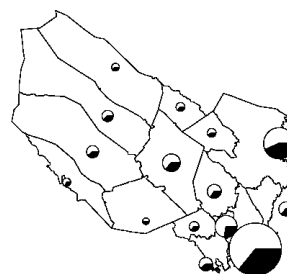
mycket små. Totalt minskar befolkningen i Västerbotten något under den här tiden. Ett av de kanske överraskande resultaten är att proportionen över 55 år ganska stabil i inlandet (här definierat som de fem västligaste kommunerna: Storuman, Sorsele, Dorotea, Vilhelmina och Åsele) men ökar kraftigt i Umeå.

Västerbotten 2005



Figur 1: Simuleringsresultat Västerbotten 2005

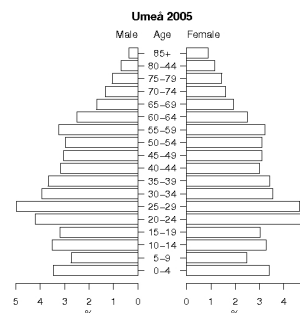
Västerbotten 2050



Figur 2: Simuleringsresultat Västerbotten 2050

Anmärkning Figur 1 och 2: Cirkulernas storlek är proportionella mot befolkningsstorleken och andelen över 55 års ålder motsvaras av den svarta delen.

För att förstå detta kan vi titta på befolkningspyramiderna för dessa år. Låt oss först titta på befolkningspyramiden för Umeå 2005 i Figur 3 som är pyramidformad från åldern 25 och uppåt. Detta kan till stor del förklaras av Umeås historia och funktion. Att Umeå är en universitetsort kan delvis förklara den stora befolkningen mellan 20 och 30. Kommunen har växt kraftigt under många decennier på grund av dess roll som regionalt centrum för administration, utbildning och sjukvård. Med bakgrund av detta är det naturligt att Umeå har en relativt liten andel äldre som kommer att öka över tiden.



Figur 3: Befolkningspyramid Umeå 2005

För att behålla proportionerna skulle ett större och större antal unga behöva tillkomma för att kompensera det ökande antalet äldre. Som kontrast till situationen i Umeå kan vi titta på befolkningspyramiden för inlandet i Figur 4. Pyramiden har en smal midja beroende på en liten befolkning mellan 20 och 30 år då många i den gruppen flyttat till andra kommuner. Kategorin 10–20 år är dock stor och beteendet hos dessa kommer i hög grad att påverka inlandets framtid. En slutsats som man kan dra är att flyttningsbeteendet hos de unga vuxna är mycket viktigt för den regionala demografiska utvecklingen. Det är även den kanske svåraste delen att prediktera. Regionen har redan en relativt gammal befolkning och eftersom dödligheten är högre hos äldre hålls proportionen äldre ner. Detta förklarar att andelen äldre hålls ganska konstant över tiden i denna region.



Figur 4: Befolkningspyramid Västerbottens inland 2005

Avslutningsvis hoppas jag att läsaren av denna artikel sett att mikrosimulering är ett bra verktyg för att studera demografiska förlopp och genom exemplet fått förståelse för hur det kan genomföras i praktiken. ■

Statistisk inferens i mikrosimuleringsmodeller

av N. Anders Klevmarken, professor i ekonometri, Nationalekonomiska institutionen, Uppsala universitet

Simulering är en teknik som användes för många olika ändamål inom olika discipliner. "Mikrosimulering" och "mikrosimuleringsmodeller" är en terminologi som kommit att användas inom samhällsvetenskaperna och framför allt i nationalekonomi. En mikrosimuleringsmodell kan ses som en uppsättning regler som opererar på ett urval av mikroenheter såsom enskilda individer, hushåll eller företag. Varje mikroenhet karaktäriseras av vissa egenskaper (variabler) och när modellen simuleras uppdateras dessa egenskaper.

Modellen kan till exempel endast bestå av de deterministiska regler som utgör vårt inkomstskattesystem och som opererar på ett urval av skattebetalare för vilka man känner inkomster, avdrag och andra förhållanden som behövs för att beräkna skatterna för varje individ (en så kallad tax-benefit modell). Med en sådan modell kan man till exempel beräkna inkomst efter skatt för varje individ och disponibel inkomst för varje hushåll med olika antagande om skattesatser, avdragsregler etcetera. Man kan således se hur förändringar i skattesystemet slår både på de totala skatteintäkterna och vilka individer (hushåll) som vinner respektive förlorar på en skattereform.

Simuleringsmodellen kan emellertid vara betydligt mer komplex och även innehålla stokastiska relationer som avses beskriva hur individer, hushåll och företag förändrar sitt beteende. Avsikten med till exempel en skattereform är ju ofta att försöka påverka folks beteende såsom att stimulera till ökat förvärvsarbete eller senare pension. I så kallade dynamiska mikrosimuleringsmodeller finns relationer som skriver fram själva undersökningspopulationen, det vill säga man använder modeller för att simulera födelser av nya mikroenheter och dödsfall bland gamla.

Mikrosimuleringstekniken har sitt ursprung i den ameri-

kanska ekonomen Guy Orcutts ideer om att man i ekonomisk analys borde försöka efterlikna naturliga experiment, vilka resulterade i den dynamiska mikrosimuleringsmodellen DYNASIM [1], [2], [3]. Mikrosimulering fick senare under 1980- och 1990-talen förnyad betydelse när många västlänningar ville reformera sina skattesystem och behövde instrument för att analysera konsekvenserna av nya skatteförslag. I dag finns både statiska tax-benefit modeller och dynamiska mikrosimuleringsmodeller i många länder. I Sverige finns vad som kanske är den äldsta ännu i praktiskt bruk varande mikrosimuleringsmodellen vid Försäkringskassan. Det är den modell som ursprungligen utvecklades av Tor Eriksen för att kunna göra framskrivningar av ATP-systemet [4], [5]. Hos Statistiska centralbyrån finns den statiska tax-benefit modellen FASIT, och finansdepartementet har en stor dynamisk mikrosimuleringsmodell SESIM.

De flesta mikrosimuleringsmodeller simulerar betingade fördelningar, det vill säga man drar slumpmässigt från olika betingade fördelningar, varvid betingningen bestäms av respektive individs egenskaper. Ofta finner man till exempel probit- eller logitmodeller, Tobitmodeller och regressionsmodeller i dessa mikrosimuleringsmodeller. En probitmodell kan till exempel bestämma sannolikheten för att en individ ska vara sysselsatt givet individens kön, ålder, utbildning etcetera. Däremot brukar dessa modeller inte vara särskilt detaljerade i sin beskrivning av hur olika beslutsprocesser går till, till exempel hur en individ söker efter arbete hos olika arbetsgivare och så småningom når en överenskommelse om arbete och lön. Utvecklingen av sådana modeller, som mer i detalj beskriver interaktionen mellan olika agenter och deras väg till beslut, har dock påbörjats under senare år och dessa modeller brukar kallas agentbaserade mo-

deller. De har dock knappast fått någon större praktisk betydelse ännu.

Inferens i en statisk tax-benefit modell

Inferensproblemet i de så kallade statiska tax-benefit modellerna är ganska begränsat. Ett urval av mikroenheter dras från en population enligt någon väldefinierad urvalsplan. För dessa mikroenheter observeras vissa egenskaper som modellen i simuleringen ändrar. Någon ny stokastisk osäkerhet introduceras dock inte i simuleringen. Den stokastiska osäkerheten kommer i detta fall endast från urvalsdragningen. De nya värdena användes för att med givna urvalsvikter uppskatta storheter som till exempel total inbetald skatt, percentiler för disponibel inkomst eller ginikoefficienten för disponibel inkomst. Inferensen går till den (ändliga) population ur vilken urvalet dragits.

Verkligheten är dock inte alltid så enkel. I praktiken är man ofta tvungen att använda ett urval som är ett eller ett par år gammalt, medan man vill göra inferens till den aktuella populationen av skattebetalare. Urvalsvikterna ger dock inte en sådan inferens. Sedan urvalet drogs kan populationen ha genomgått väsentliga förändringar. Åldersfördelningen och inkomstfördelningen kan exempelvis ha förändrats. Ofta försöker man komma tillrätta med detta problem genom att justera vikterna genom någon form av kalibreringsförfarande [6], [7] och [8].

I regel är man primärt intresserad av att analysera effekterna av skatteförändringar, det vill säga vilka förändringar till exempel i inkomstpercentiler eller ginikoefficient som en skatteförändring medför, och man kan ju nära förhoppningen att dessa förändringar inte samvarierar starkt med populationsförändringarna. I så fall behöver en mindre lyckad kalibrering kanske inte spela så stor roll.

Inferens i en dynamisk mikrosimuleringsmodell

I en dynamisk mikrosimuleringsmodell finns det ingen konstant population till vilken man kan dra en inferens, då modellen i sig definierar hur populationen förändras både till sin storlek och sammansättning. Det kan till och med vara så att de policyförändringar man är intresserad av att studera påverkar befolkningens storlek och sammansättning. I denna typ av mikrosimulering är det därför bara meningsfullt att diskutera inferens till den "superpopulation" som definieras av modellen. Det utesluter dock inte att man (i efterhand) kan utvärdera modellens realism genom att jämföra simuleringar med verkligheten.

Inferensproblematiken i en dynamisk mikrosimuleringsmodell har mycket gemensamt med den statistiska inferensen i ekonometrin. En principiell skillnad är möjligheten att i mikrosimulering är huvudsyftet att designa och estimerar en modell som så väl som möjligt kan simulera fördelningen för de nyckelvariabler man är intresserad av, medan fokus i traditionell ekonometri ofta är på de betingade fördelningarnas medeltal, det vill säga

man söker konsistenta och effektiva skattningar av modellens parametrar. Dessa parametrar har dessutom ofta en ekonomisk tolkning. I mikrosimuleringstillämpningen är tolkningen av enskilda modellparametrar inte lika viktig, medan det i stället blir centralt att kunna replikera hela fördelningen av nyckelvariablerna väl, även dess svansar. Ett exempel på detta är studier av fattigdom, då man måste kunna simulera de vänstra svansarna i inkomst- och förmögenhetsfördelningarna. Det kan således finnas anledning att fundera över vilka estimationskriterier man bör använda i mikrosimulering. Med hänsyn till syftet bör man inte ha ett kriterium som enbart eller huvudsakligen straffar avvikelser från medelvärdet, utan ett kriterium som även straffar avvikelser från högre moment. En GMM-ansats skulle då kunna vara ett naturligt val. Komplexiteten, icke-linjära relationer och förekomsten av diskreta hopp i en typisk mikrosimuleringsmodell kan emellertid göra det svårt att evalvera momentvillkoren. En lösning på detta problem kan vara att utnyttja det faktum att vi har och göra med en simuleringsmodell och ersätta GMM med den simulerade momentmetoden (the simulated methods of moment).

Antag följande enkla modell,

$$y_t = g(x_t, \epsilon_t, \theta) \quad (1)$$

x_t är en exogen variabel, ϵ_t en slumpvariabel med känd täthetsfunktion och θ en okänd parameter.

$$E(y_t|x_t) = E(g(x_t, \epsilon_t, \theta_0|x_t)) = k(x_t, \theta_0) \quad (2)$$

Vi antar att $k(x_t, \theta)$ inte kan skrivas på sluten form.

Den grundläggande tanken är således att vi ska estimerar θ så att vi får en fördelning av simulerade y -värden, y_t^s , med egenskaper som så nära som möjligt överensstämmer med dem som täthetsfunktionen för y_t har. Det skulle möjligen kunna förefalla naturligt att välja så att följande uttryck minimeras,

$$\sum_{t=1}^n (y_t - y_t^s(\hat{\theta}))^2 \quad (3)$$

Det har emellertid visats i [9, sidan 20] att denna så kallade "path calibrated estimator" inte nödvändigtvis är konsistent. Om man tillämpar denna estimator på följande enkla modell,

$$y = x\beta + \sigma\epsilon \quad \text{där } \epsilon \sim IID(0, 1) \quad (4)$$

blir skattningen av β konsistent medan $\text{plim } \hat{\sigma} = 0$, se [10]. Denna estimator råder oss således att ignorera slump termen ϵ när vi simulerar och endast använda genomsnittsprediktionen. Detta förfarande överensstämmer uppenbarligen inte med mikrosimuleringens syfte.

En alternativ ansats är att använda vad som har kallats en "moment calibrated estimator", som minimerar avståndet mellan observerade och simulerade moment. Denna

estimator tillåter i princip inte bara kalibrering mot moment i tvärsnittsfördelningar utan även mot övergångsfrekvenser och intertemporala korrelationer, vilket kan bli betydelsefullt i dynamiska modeller. Låt θ vara en vektor av storleken p och x_t en vektor av storleken r . Låt vidare $K(y_t, x_t)$ vara en vektorfunktion med storleken q , och

$$E(K(y_t, x_t)|x_t, \theta_0) = k(x_t, \theta_0) \quad (5)$$

K skulle till exempel kunna vara identitetsfunktionen och kvadraten av y_t . Definiera också en $r \times q$ matris $Z_t = I_{q \times q} \otimes x_t$. Från antagandet att x_t är exogen följer att,

$$E[Z_t(K(y_t, x_t) - k(x_t, \theta_0))] = 0 \quad (6)$$

Då $k(x_t, \theta)$ inte har någon sluten form definierar vi en väntevärdesriktig simulator för k ,

$$\tilde{k}(x_t, \epsilon^s, \theta) = \frac{1}{S} \sum_{s=1}^S K(g(x_t, \epsilon^s, \theta), x_t) \quad (7)$$

där ϵ^s är en vektor av S oberoende slumpstermer ϵ_t^s dragna från p.d.f. för ϵ_t . En simulerad GMM estimator erhålles då som,

$$\hat{\theta} = \arg \min_{\theta} (\sum_{t=1}^n Z_t [A - B]) \times \Omega (\sum_{t=1}^n Z_t [A - B]) \quad (8)$$

där $A = K(y_t, x_t)$, $B = \tilde{k}(x_t, \epsilon^s, \theta)$ och Ω är en $r \times r$ symmetrisk positiv semidefinit matris. I [9] visas att denna estimator är konsistent. Estimatorns kovariansmatris har två komponenter, en som är kovariansmatrisen för den vanliga GMM-estimatorn, och en som beror på hur väl k simulerats. Det optimala valet av Ω beror på den okända fördelningen för y_t , men en simulerad estimator av det optimala Ω har anvisats i [9, sidan 32].

Det bör observeras att antalet momentvillkor i (6) får inte vara mindre än antalet okända parametrar, ty eljest blir modellen inte identifierad.

Det kvadratiske uttrycket i (8) kan minimeras med vanliga gradientbaserade metoder om första och andra ordningens derivator med avseende på om θ existerar. Om modellen innehåller diskontinuiteter med avseende på θ måste andra optimeringsmetoder användas. Mikrosimuleringsmodeller som innehåller skatteregler och regler för olika former av bidrag är typiskt diskontinuerliga i variabler, vilket möjligen kan medföra att de även blir diskontinuerliga i parametrar.

Varför inte använda maximum likelihood metoden? I praktiken är det ofta inte möjligt. Det finns som regel inga teoretiska skäl att välja någon viss parametrisk familj av fördelningar och med tanke på mikrosimulerings syfte kan det vara olämpligt att göra starka antaganden i detta avseende. I stället kan det vara bättre att skatta en struktur för betingade medelvärden och sedan i

simuleringen sampla från den empiriska residualfördelningen. I vissa fall, när det till exempel gäller att simulera inkomst och förmögenhetsfördelningen, måste vi arbeta med fördelningar som är kraftigt skeva, med stark excess och som inte följer kända parametriska fördelningar. Slutligen, en stor mikrosimuleringsmodell innehåller en ganska komplex blandning av olika modeller, funktionsformer och beroendeantaganden så det kan vara svårt att sätta upp en likelihoodfunktion.

Storleken på en typisk mikrosimuleringsmodell, blandningen av olika modelltyper och funktionsformer samt nödvändigheten att använda olika datakällor gör att man i praktiken sällan skattar modellens alla parametrar simultant. I stället brukar man skatta varje delmodell för sig och sedan sammanföra dem till en stor simuleringsmodell. Om detta är ett acceptabelt förfarande eller ej beror på modellstrukturen. Om modellen är hierarkisk och dess stokastiska egenskaper sådana att man har en rekursiv eller blockrekursiv struktur kan ett styckevis estimationsförfarande försvaras med att man får konsistenta skattningar. Finns det däremot ett stokastiskt beroende mellan delmodellerna ger en styckevis estimation inkonsistenta skattningar. Jämför den gamla diskussionen om rekursiva respektive interdependenta modeller. Se även avsnitt 4.2 i [10].

Ett stort problem vid varje större mikrosimuleringsprojekt är bristen på goda data. I allmänhet är det inte möjligt att finna en enda datamängd som kan användas både för estimation och som bas för simuleringarna. Ofta saknas nyckelvariabler i det urval på vilket simuleringarna görs och man måste då dels imputera startvärden på dessa variabler i datamängden, och dels skatta den modell, som ska skriva fram dessa startvärden, från en extern datakälla. Förutsättningen för att detta över huvud taget ska gå är givetvis att de variabler som "förklarar" den saknade variabeln alla ingår i simuleringens modellens datamängd och skrivs fram av modellen. I total avsaknad av externa data har forskare använt sig av "guestimates" som kalibrerats mot kända "benchmarks". Kalibrering är i detta sammanhang inget annat än försök att anpassa dessa "guestimates" så att modellen simulerar fördelningarna för nyckelvariabler av intresse på ett hyggligt sätt. I detta avseende har mikrosimulering släktskap med så kallade allmänna jämviktsmodeller. I båda fallen förlitar man sig för mycket på kalibreringstekniken. Hansen och Heckman kritiserade denna ansats därför att de ansåg att man visste för litet om egenskaperna hos estimaten [11]. I själva verket är egenskaperna okända och därmed gäller detsamma för simuleringarna. Kalibreringstekniken tenderar också att gömma undan ett allvarigare problem, nämligen att kalibreringen vanligen sker mot en eller några få benchmarks eller datapunkter, och därför inte nödvändigtvis identifierar alla parametrar på ett entydigt sätt.

Även om alla parametrar har estimerats med standardmetoder förekommer det att man använder kalibrering av de simulerade variabelvärdena mot kända benchmarks för att modellen inte ska spåra ur eller därför att

det av ett eller annat skäl är önskvärt att modellen inte avviker från kända förhållanden eller officiell statistik. I till exempel Finansdepartementets SESIM kalibrerar man mot Statistiska centralbyråns befolkningsframskrivningar. Det förekommer även att man använder kalibrering för att reducera Monte Carlo osäkerheten i simuleringarna.

Om de simulerade fördelningarna avviker mer från de observerade än vad modellens stokastiska egenskaper motiverar är detta en klar indikation på att data antingen förkastar den uppsättning parameterskattningar som använts eller modellstrukturen som sådan. I så fall ska modellen inte kalibreras för att ge rimliga simuleringar utan respecificeras och estimeras om. Om emellertid data inte förkastar modellen, men simuleringarna ändå avviker mycket från kända förhållanden, kan kalibrering vara ett sätt att öka precisionen i simuleringarna. Ett alternativ till att justera de simulerade variabelvärdena är att justera parameterskattningarna. I själva verket kan man se kalibrering som estimation under bivillkor. Bivillkoret är att simuleringarna ska överensstämma med benchmarks. I det enkla fallet att man har en linjär regressionsmodell kan problemet formuleras som en traditionell estimation under bivillkor, såsom den behandlas i vanliga läroböcker i ekonometri. Förutom explicita uttryck för estimaten under bivillkoren kan man härleda motsvarande "justeringsfaktorer" för de simulerade värdena, se [10]. Från denna övning framkommer ett intressant resultat, nämligen att de effektiva justeringsfaktorerna inte är desamma som de enkla proportionella justeringar som brukar användas i praktiken och dessutom att det inte bara är de variabler som ingår i bivillkoren, som ska justeras, utan i princip alla endogena variabler i modellen.

Det är inte lika uppenbart hur kalibreringen ska gå till när modellen är icke linjär. Om man kan utgå från att alla preliminära parameterestimater, det vill säga estimaten innan kalibreringen, är konsistenta, är en ansats att finna nya, kalibrerade estimater, som avviker så lite som möjligt från de konsistenta preliminära estimaten men gör att modellen uppfyller kalibreringsvillkoren. Skriv mikrosimuleringsmodellen på följande form,

$$y_{it} = g(y_{i0}, \dots, y_{it-1}, x_{i0}, \dots, x_{it}; \theta) \quad (9)$$

där y_{it} är en vektor bestående av k endogena variabler för en individ i vid tidpunkten t , x_{it} en vektor av exogena variabler inklusive icke observerade slumpvariabler och θ en vektor av p parametrar. Låt Y_T vara en $k \times n$ matris av simulerade värden för en period T utanför den samplade tidsperioden och skriv kalibreringsvillkoret på följande sätt,

$$\bar{Y}_T = R(Y_T) \quad (10)$$

där $\bar{Y}_T$ är en matris bestående av benchmarks och med dimensioner som är mindre än dimensionerna för Y_T . R är en matrisfunktion. Låt antalet villkor som impliceras av (10) vara c . Typiskt gäller att $c < p$. Om till exempel Y_T är en vektor av totaler, då blir $R(Y_T) = Y_T JN/n$

där J är en $n \times 1$ vektor med enhetselement och N är populationens storlek. (Urvalet har antagits vara ett enkelt OSU.) Kalibrering kan nu ske genom att följande Lagrangeuttryck minimeras med avseende på θ ,

$$(\theta - \hat{\theta})' V^{-1} (\theta - \hat{\theta}) + \lambda' (\bar{Y}_T - R(Y_T)) \quad (11)$$

där V är kovariansmatrisen för $\hat{\theta}$. Detta uttryck kan minimeras med numeriska metoder. Genom att Taylorutveckla bivillkoret kring $\hat{\theta}$ kan man även härleda ett explicit uttryck för den kalibrerade estimatoren, se [12], där även ett test av bivillkoret finns redovisat. Resultaten i [12] bygger dock på att en första ordningens Taylorutveckling ger en god approximation, vilket inte behöver vara fallet om modellen är kraftigt icke-linjär. I så fall är en direkt minimering av (11) att föredra. (Den kvadratiske avståndsfunktionen är ett naturligtval, men det finns alternativ. I ett annat sammanhang, kalibrering av estimatorer i survey sampling, gav Deville och Särndal (1992) exempel på flera avståndsfunktioner.)

Fortsatt utveckling av goda metoder behövs

Mikrosimuleringsmekniken har under senare år blivit allt mer använd av myndigheter, departement och organisationer i deras policyarbete. Inferensproblemen har emellertid i stor utsträckning ignorerats. Inte ens i den akademiska litteraturen finns det någon närmare diskussion av dessa problem. Den stokastiska osäkerheten kommer från tre olika källor: urvalsförfarandet när den datamängd väljs på vilken simuleringarna utförs, skattningarna av modellens parametrar och från själva simuleringen. För att utvärdera egenskaperna hos simuleringarna och validera dem behöver man kunna ta hänsyn till alla dessa tre källor. Endast i undantagsfall för mycket små modeller kan man räkna med att kunna få explicita uttryck för de simulerade värdenas egenskaper, se till exempel [13], i allmänhet kommer man att vara hänvisad till numeriska replikationsmetoder. Den stora ökningen i datakraft och den förbättrade tillgången på mikrodataba har gjort mikrosimulering praktiskt möjlig, men inferensproblemen behöver utredas ytterligare så att man i praktiskt arbete kan använda sunda metoder med kända egenskaper. ■

Referenser

- [1] Orcutt, G.H. (1957). A new type of socio-economic system. *Review of Economics and Statistics*. **58**, 773–797.
- [2] Orcutt, G.H., Greenberger, M. Korbelt, J. and Rivlin, A. (1961). *Microanalysis of Socioeconomic Systems: A Simulation Study*. Harper and Row, New York.
- [3] Orcutt, G.H., Caldwell, S. and Wertheimer, R. (1976). *Policy Explorations Through Microanalytic Simulations*. The Urban Institute, Washington D.C.
- [4] Eriksen, T. (1973). *En prognosmodell för den allmänna tilläggs pensioneringen*. Riksförsäkringsverket, Stockholm.

- [5] Klevmarken, N.A. (1973). En ny modell för ATP systemet. *Statistisk Tidskrift*. **1973**(5), 403–443
- [6] Deville, J-C. and Särndal, C-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*. **87**(418), 376–382.
- [7] Lindström, K. (1997). *Utvärdering av framskrivningstekniken i FASIT-modellen*. Rapport från Statistiska centralbyrån, Stockholm.
- [8] Lundström, S. (1997). *Calibration as a standard method for treatment of nonresponse*, PhD-theis. Department of Statistics, Stockholm University.
- [9] Gouriéroux, C. and Monfort, A. (1996). *Simulation-Based Econometric Methods*. Oxford University press, Oxford.
- [10] Klevmarken, N.A. (2002). Statistical inference in Micro Simulation Models: Incorporating External Information. *Mathematics and Computers in Simulation*. **59**(1–3), 255–265.
- [11] Hansen, L.P. and Heckman, J.J. (1996). The Empirical foundations of Calibration. *Journal of Economic Perspectives*. **10**(1), 87–104.
- [12] Klevmarken, N. A. (2007). Dynamic Microsimulation for Policy Analysis – Problems and Solutions. Chapter 2 in *Simulating an ageing population. A microsimulation approach applied to Sweden*, eds. Klevmarken, N.A. and Lindgren, B. Kommer att publiceras 2008 o Elseviers bokserie *Contributions to Economic Analysis*.
- [13] Pudney, S. and Sutherland, H. (1996). Statistical Reliability in Microsimulation Models With Econometrically-Estimated Behavioural Responses. Chapter 21 in *Microsimulation and Public Policy*, ed. Harding, A. North-Holland Elsevier.

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.

Samfundets medlemmar bör utse redaktör för Qvartilen – del 2

av Mikael Möller, TeX-Försäljning AB

I nummer 1 av Qvartilen 2007 skrev jag ett debattinlägg som, baserat på demokratiska argument, förespråkar ett skiljande mellan styrelsen och Qvartilen. Jag påpekade där vad jag ansåg vara fel i den nuvarande styrelsens beslut rörande Qvartilen. Styrelsen har valt att inte gå i diskussion om sakförslaget. Se Qvartilen 22-2

När någon skriver “eftersom ett sådant förfarande inte finns i Samfundets stadgar” (Qvartilen 22-2, styrelsens inlägg) blir jag orolig ty detta ordval får jag ofta höra av icke-demokrater. Detta är inget hållbart argument av samma anledning som att det inte finns något hållbart argument för att kvinnor icke skall ha rösträtt. Stadgar skrivs av medlemmar och ändras av medlemmar. I Samfundets stadgar står:

d) Val av ordförande, vice ordförande, sekreterare, skattmästare (även kallad kassaförvaltare) och övriga styrelseledamöter.

som med 2/3-dels majoritet enkelt ändras till

d) Val av ordförande, vice ordförande, sekreterare, skattmästare (även kallad kassaförvaltare) och övriga styrelseledamöter **samt redaktör**.

Det förslag jag gav är baserat på mina erfarenheter av nuvarande och andra styrelser. Förslaget innehåller verktyg för att hålla en styrelse borta från det redaktionella arbetet samt verktyg för att se till att medlemmarna har kontroll över tidningen. Det är speciellt illavarslande när en styrelse använder sig av uttrycket “företse alla eventualiteter” ty detta kan man aldrig göra men uttryck-

et sätter hämsko på nödvändig debatt och inget hotar de aktuella beslutsfattarna.

Anledningen till att jag tog upp frågan om Qvartilen i nummer 1 innevarande år beror på samgåendet med Statistiska föreningen vilket kommer att skapa en stor förening med möjlighet till statistisk och samhällslig ’lobby’-verksamhet. Styrelser i stora föreningar har dock en otäck böjelse att gå emot sina medlemmars intressen och för att göra detta svårare behövs flera demokratiska ben. Sådana demokratiska ben är till exempel fristående och självständiga sektioner inom den, förhoppningsvis, kommande paraplyorganisationen.

Det viktigaste demokratiska benet i en förening är dess medlemstidning, men det krävs att tidningens redaktör utses av medlemmarna direkt, så att denne ges styrka att i kritiska lägen gå emot styrelsen det vill säga hålla debatten fri.

I sin strävan att lägga locket på hänvisar Samfundets styrelse till arbetet med att skapa rutiner för Samfundets olika verksamheter det vill säga vad kassören skall göra, vad sekreteraren skall göra och så vidare. Dessa rutiner är dock endast av praktisk natur. Det ställer sig helt an-

norlunda vad gäller Qvartilen och dess redaktör vilket visar sig i att vi har lagat om pressfrihet, yttrandefrihet och rätten att i tryck framföra åsikter. Det är dock ett vanligt retoriskt knep, att prata om annat än det som skall diskuteras, men jag tror att knepet faller platt i den välutbildade skara som Samfundets medlemmar utgör. Styrelsen använder sig även av andra retoriska knep för att föra diskussionen bort från sakfrågan: man berättar om hur styrelsen väljs, att man "tagit på sig att utse redaktionsmedlemmar", påstår indirekt att jag inte för en öppen diskussion med mera.

Jag hade öppna diskussioner med styrelsens sekreterare. Vidare närvarade jag vid ett styrelsemöte, som jag upplevde som ett försvarsmöte för den av mig förda arbetslinjen för Qvartilen. Den beslutsordning som styrelsen dock antog, för Qvartilen, fick varken redaktören eller redaktionen yttra sig om. Vem är det som för slutna diskussioner i slutna rum?

Självklart finns det en motsatsställning mellan en förenings styrelse och en förenings medlemstidning. Den senare är ett organ för föreningens medlemmar – inte dess styrelse. Detta kommer förr eller senare reta styrelsen och som, när detta inträffar, då försöker ta kontroll över tidningen vilket har, enligt min mening, inträffat i flera föreningar, till exempel Hyresgästföreningen, Stockholms Kooperativa Hyresgästförening, SACO och SKTF. I Samfundet gör man det i samband med att man tar fram rutiner för övrig verksamhet. Vill man dölja den egentliga orsaken: Att få kontroll över medlemmarnas tidning och därmed deras sinnen. Eller är det bara en fråga om en rad misstag?

Under de tre och ett halvt år jag arbetade som redaktör lärde jag mig mycket om hur man tar fram en tidning. Bland annat hur inblandning i tidningsarbetet, vilket stundtals ägt rum, indirekt eller direkt, faktiskt försvårar/förtråkar/omöjliggör detta arbete. Det är denna typ av inblandning som mitt förslag, till organisering av

Qvartilen, avser förhindra. I styrelsens svar saknar jag en diskussion om de olika punkterna i mitt förslag.

Under mitt sista år som redaktör såg jag vart det barade hän och försökte då stävja den nya inriktningen. Jag föreslog att redaktören skall väljas av medlemmarna samt att tidningen skall ha en egen ekonomi fri från styrelsens. Tyvärr har sedan dess flera styrelsebeslut skapat en demokratiförlust.

Styrelsen hänvisar till ett styrelsemöte men många fler har ägnats Qvartilen. Varför? Är inte redaktören och redaktionen betrodda att sköta Qvartilen? På mötet 070509 formulerades ett mycket olyckligt beslut vars konsekvens är att författarna fräntas rätten till sina egna artiklar om dessa publiceras i Qvartilen. Men är det inte något feltänkt här? Hur skulle Bo Rothstein ha ställt sig till att publicera i Qvartilen under dylika villkor? Hur skulle vi kunna ha publicerat Jan Hagbergs historiska artiklar? Varför blev KG Scherman mycket irriterad och ringde mig för att diskutera dessa nya villkor?

Jag har, efter diskussion, förstått att de givna formuleringarna stämmer illa med vad som avsetts och ger därför här förslag på formulering som bättre täcker den kunskap diskussionen gav mig:

De författare som medverkar i Qvartilen äger copyrighten till sina artiklar. Författaren ger Qvartilen rätten att trycka artikeln, elektroniskt och på papper, samt behålla den ersättning som kan uppstå härför. Om en originalartikel används i något annat sammanhang är Qvartilen tacksam om originalkällan anges. Artiklar som arvoderats äger Qvartilen och författaren gemensamt och var för sig copyrighten till.

Författare som så önskar kan erhålla sin artikel (särtryck) på elektronisk form att fritt förfoga över.

Bokanmälan

Robert Lundqvist, Bygga tabeller och diagram med standardprogramvara, eget förlag, 2007, (vi + 64 sidor).

Lundqvists bok är avsedd som en snabb och lättanvänd handledning i hur man utför enkla statistiska operationen med allmänt tillgänglig programvara. Beskrivningen baseras på användning av Microsoft Excel 2003 under Windows XP.

Boken är upplagd enligt principen "från ax till deg", det vill säga från kodning och inmatning av data till produktion av tabeller, diagram och enklare bearbetningar. Att baka limpan, det vill säga producera den färdiga rap-

porten/analyserna berörs inte. Några av Excels inbyggda funktioner för sannolikhetsberäkningar beskrivs. Boken orienterar också om alternativa programvaror, både kalkylprogram och statistiska program. Den innehåller också en kort diskussion om urval och kvalitet.

Främsta användningsområdet torde vara som självinstruerade studiematerial vid elementära statistikkurser, eller för den som utan tidigare erfarenhet behöver utföra enklare bearbetningar av mindre datamaterial.

Göran Arnoldsson

Kåseri om statistik som tillämpning

av Olle Sjöström

Det är min erfarenhet att det nästan helt saknas diskussion bland statistiker om våra erfarenheter i statistikens tillämpningar. Är det den starka dominansen av formell matematik som har tystat oss? Varför har inte den kommunikativa matematiken dominerat statistisk utbildning och forskning? Klassificeringen i formell respektive kommunikativ matematik finns i den matematisk-didaktiska forskningen i dag.

Det fanns på 1700-talet ett gammalt franskt uttryck, "statistique raisonnée", ungefär "statistik som något att resonera kring". Vid ett samtal med en matematisk forskare för några år sedan, vid institutet för matematisk didaktik i Bielefeld, uttryckte han sig ungefär så, att analyser av statistiska data gav upphov till "vandringsvägar" i huvudet genom funderingar eller i samtal. Detta är statistisk analys som kunskapsbildning, inte som en speciell sort av matematisk analys. Vad vet vi om vad som händer med sådana "processer i huvudet" i dag? Ger de statistiska redovisningarna i dagspress och tidskrifter lyft för tankar och idéer hos dem som läser budskapen – eller kan de begränsa perspektivet eller ge "desinformation"? Samtalar vi själva, som sta-

tistiker, om de frågor som statistiska redovisningar aktualiserar?

I brist på samtal om sådant finner jag att kåseriet är den mest passande formen för att odla en känsla av tanke- och yttrandefrihet i vårt ganska tysta "statistiksamhälle". Ingen känner sig nödd att svara på frågor som väcks i ett kåseri, men blir inte heller förbjuden att ta till orda. Det går att kåsera om stort och smått, från vad vi menar med statistiskt teori eller hur vi använder det statistiska språket till att berätta om händelser eller episoder i den statistiska praktiken eller hur vi valde statistikeryrket. Det får vara en tillräcklig motivering för kåseriformen. Här kommer nu det första kåseriet om hur jag valde statistik som yrke: hur Herman Wold (1908-1992) i Uppsala på 1950-talet och Tore Dalenius (1917-2002) i Stockholm på 1960-talet gav idéer, impulser och handledning. Motivet till denna början i det självbiografiska är att det ger en bakgrund till de kommande kåserierna, bland annat om kvalitetsredovisning i praktiken, hur vi som statistiker språkar med andra och om vad vi menar med statistisk metod. □

Mina erfarenheter av Herman Wold blev positiva, inte bara för att han gav mig "spets" i min allra första tentamen för två betyg (motsvarande 40 poäng i dag) utan även för att han var den första "statistiska entusiast" som jag mötte. På så vis blir det att jag skriver ett försvar för Wold mot skvaller av mindre fördelaktigt slag som har hörts från lärare och forskare vid statistiska institutionen i Uppsala. Mitt perspektiv är studentens mer anspråkslösa och naiva men jag vill beskriva den uppfattning om statistik som jag fick under mina första år som student i Uppsala på 1950-talet.

Då var statistik ett ämne som studenten bara kunde börja läsa på vårterminen i Uppsala. De som studerade ämnet var i huvudsak de med politiska karriärambitioner, speciellt hade många valt utbildning till politices magister. I den var statistik var ett obligatoriskt ämne tillsammans med statskunskap och nationalekonomi. Själv upptäckte jag och valde statistikämnet genom att Erland Hofsten i sin bok *Praktisk statistik* [1] förde en del intressanta resonemang, bland annat om bevisföring ("resonerande statistik"!)). Att även jag valde utbildning till politices magister var mer av nyfikenhet på ämnen, som inte hade funnits med i gymnasiekurserna, än politiskt intresse.

Det första goda intrycket av Wold kom på introduktionsföreläsningen där han övertygade om ämnets betydelse och komplexa innehåll, med både allmänna orienteringar om viktiga statistiska redovisningar och intressanta tekniker med olika grader av matematisk underbyggnad.

Just vid den tiden förde Wold in en ny variant i grundutbildningen, nämligen en matematisk-statistisk sådan som alternativ under andra terminens studier.

Wold såg betydelsen av en allmän yrkesutbildning till statistiker. Det fanns då sedan länge, före den drastiska centraliseringen av statistiska verksamheter till Statistiska centralbyrån efter 1960, en utvecklad statistisk byråtradition. Den erbjöd en genomtänkt praktik med en början som amanuens och med reglerad befordringsgång till aktuarie- och byråchefstjänster. De statistiska byråerna var fler än ett 25-tal och de var produktiva genom olika slag av uppsatser inom de områden som varje byrå hade ett statistiskt ansvar för. Sådana uppsatser var en del av den statistiska kurslitteraturen i Uppsala och förberedde studenten på kommande möjligheter i en karriär som statistiker. Annonser om lediga platser vid byråer med skilda inriktningar, olika ämnesområden, gav oss statistikstudenter löften om kommande praktik med varierande innehåll. Praktiken inom byråerna kan sägas motsvara juristens eller läkarens praktiktjänstgöringar. Wolds intresse för den statistiska praktiken yttrade sig även i att han åtog sig ordförandeskapet i statistiska föreningen under senare delen av 1950-talet.

Ytterligare ett plus i min uppfattning om Herman Wold var hans intresse för statistikens historia som, när jag studerade, yttrade sig i vissa orienterande föreläsningar men som senare ledde till böckerna om *Mål och metod i tillämpad statistik*, delen *Orientering i det statis-*

tiska arbetsfältet [2]. Han närmade sig det område av den historiska statistiken som jag själv senare forskade i, nämligen den så kallade statskunskapstraditionen med Hermann Conring som startpunkt och med en utveckling av den byråstatistiska traditionen i Centraleuropa under 1800-talet. Det var den senare som realiserades i Sverige genom uppbyggnaden av de statistiska byråerna i början av 1900-talet. Jag har velat ta fram en bild av en klok och empatisk statistikprofessor som inspirerade till valet av ett yrke som statistiker. "Den briljante Wold" med internationellt rykte hör hemma i de tillämpningar där hans matematiska och statistiska kompetens skapade framgångar och internationell berömmelse. Ekonometrin, psykometrin och kemin är de områden som kommer tydligast fram i sökningar på Internet. Skvallret talar om Wolds besatthet av sin egen forskning och om en påstådd dröm om Riksbankens ekonomipris i Nobels anda. Många säger sig ha lidit av Wolds så kallade flagg-tablå, men den blev jag inte besvärad av, inte ens påtrugad. Mitt sista möte med Wold var vid ett seminarium om grundundervisning i statistik som Bengt Molander, filosofen, och jag hade i Uppsala i mitten av 1980-talet i projektet *Utbildning för tillämpning av statistik – kunskap och kunskapssyn*. Hans intresse var inte att ta miste på och det framgick inte minst av samtalen efter seminariet.

Efter fullbordad pol.mag.-utbildning blev min första statistiska praktik på Konjunkturinstitutet (1958) och ett år senare på Socialstyrelsen, på den statistiska byrå där Erland Hofsten var byråchef. Hofsten var en utmärkt handledare som ställde krav på att uppdrag skulle slutföras, i mitt fall en utredning om rabatternas möjliga betydelse för konsumentindex. Eftersom jag ville bredda mina statistiska erfarenheter sökte jag en tjänst vid Statistiska centralbyråns utredningsinstitut. Det blev en längre och lärorik period med praktik med urvalsundersökningar, både stora (till exempel arbetskraftsundersökningar) och små (till exempel kundundersökning för ett bensinföretag). Jag läste den nyutkomna boken av W. Edwards Deming, *Sample Design in Business Research* [3], och fick avsevärd hjälp av hans diskussion om fördelningen av ansvar mellan uppdragsgivare och statistiker, såväl i planering som genomförande av en undersökning. Som tidigare engagerad schackspelare tilltalade det mig att hitta in i en meningsfull och mer social spelsituation! Men framför allt blev denna femåriga praktikperiod en närmare kontakt med Tore Dalenius, då docent vid Statistiska institutionen i Stockholm. Han hade inspirerat till inrättandet av utredningsinstitutet 1953 och genomfört det statistiska urvalet av områden ("basurvalet").

"Some would call him an idealist", skriver Carl-Erik Särndal om Tore Dalenius, som en kontrast till hans rykte som tämligen arg kravställare och "renhållare", bland annat som extraopponent vid disputationer. "Dåliga undersökningar hör hemma i papperskorgen", "skit in, skit ut", "ännu ett stort slag i luften" var exempel på invektiv han strödd omkring sig. Han var emellertid en strikt föreläsare och var, liksom Wold, en inspirerande

entusiast för statistik. Det är nu med den tidige praktikers begränsade perspektiv som jag vill karakterisera honom. Genom att "basurvalet" av områden vid utredningsinstitutet var moget för en revidering, bland annat på grund av ny hjälpinformation, fick jag överraskande detta revisionsuppdrag - med Dalenius som handledare. De blev en mycket intressant period som till min egen överraskning ledde till en licentiatavhandling. För en översiktlig redovisning, se Utredningsinstitutets nya basurval, *Statistisk tidskrift* 1967:1. Efter genomgångna kurser i samplingteori och -metodik fick jag också uppdrag att vara lärare i dessa kurser i mitten av 1960-talet och kom att bokstavligen att bli granne med Tore på statistiska institutionen. Med erfarenheterna från utredningsinstitutet som bakgrund kom den stora läroboken av Hansen, Hurwitz och Madow, med Tores föreläsningar som komplettering, att bli ett slutsteg i min utbildning till statistiker med surveyundersökningar som huvudinriktning. Till licentiatutbildningen hörde också en bredare matematiskt orienterad studiekurs för Sten Malmquist, då professor vid institutionen. Det var inte svårt att samarbeta med Tore Dalenius och min uppfattning var att han i grunden, trots sin ibland "benhårda yta", inte var auktoritär utan både önskade och accepterade motsägelser och diskussion. Han tyckte om kamposporter, till exempel boxning, och jag fick då och då matcha honom i ett parti blixtschack.

Min tjänst på SCB:s utredningsinstitut lämnade jag 1964 för att övergå till det som jag önskat, en statistisk konsultpraktik. Något senare blev jag erbjuden en forskarsistenttjänst vid statistiska institutionen i Stockholm. De kommande 20 åren skulle jag dela mina statistiska aktiviteter mellan konsultpraktik, cirka 60 %, statistisk lärarverksamhet och forskning, cirka 40 %, forskning om den samhällsstatistiska tillämpningens villkor i Sverige. Det som jag nu sökte var nya och bredare erfarenheter av tillämpningar; "att komma närmare problem och problemställare", brukade jag säga och att söka formulera en teori om statistik som tillämpning. Statistisk teori var för mig inte statistikens matematik utan ställde frågor kring kunskapsbildning genom statistik. Viktiga frågor var för min del statistikerrollen, statistikens och statistikerns organisatoriska villkor och utbildningen. Jag var stimulerad av de första årens yrkeserfarenheter och av diskussioner med arbetskamrater vid utredningsinstitutet, speciellt Tor Eilert Eriksen, Gunnar Jacobsson och Lars Lindgren. ■

I nästa nummer: Vilka vill ha statistisk kvalitetsredovisning?

Referenser

- [1] Hofsten, E. (1963). *Praktisk statistik. Hur man planerar och genomför en statistisk undersökning*. Norstedts förlag.
- [2] Wold, H. (1963). *Orientering i det statistiska arbetsfältet*. Biblioteksförlaget, Stockholm.
- [3] Deming, W.E. (1960). *Sample Design in Business Research*. Wiley, New York.

Teoretiska artiklar

Den 7 maj i år disputerade Niklas Norén vid Stockholms Universitet på en avhandling i matematisk statistik med titeln Statistical methods for knowledge discovery in adverse drug reaction surveillance. Av fakulteten utsedd opponent var professor Heikki Mannila vid Helsinki University of Technology. Forskarstudierna bedrevs inom ramen för Världshälsoorganisationens (WHO) program för internationell biverkningsövervakning, med professor Rolf Sundberg vid Stockholms Universitet som huvudhandledare och professor Ralph Edwards vid WHO Uppsala Monitoring Centre som biträdande handledare.

Statistiska metoder för förbättrad biverkningsövervakning

av G. Niklas Norén, Matematiska institutionen, avdelningen för matematisk statistik, Stockholms universitet

Where an existing body of data is relevant to a question of policy, those data are going to be used whether we like it or not. If statisticians refuse, others will attempt inference. [1]

Det inledande citatet är taget ur professor David Finneys installationstal som ordförande för Royal Statistical Society, 1973. Det knyter direkt an till min avhandling genom att de problematiska data det refererar till är samma biverkningsrapporter som jag själv ägnar mig åt i min avhandling. Som Finney beskriver i sitt tal och som jag hoppas kunna ge viss inblick i nedan finns stora utmaningar vad gäller att analysera dessa biverkningsrapporter. Samtidigt innehåller de potentiellt väldigt värdefull information och kan möjliggöra tidig upptäckt av viktiga läkemedelsbiverkningar som inte identifierats i kliniska prövningar. Citatet ovan understryker att vi statistiker som yrkesgrupp är bäst rustade för att vägleda analys av data – även när dessa data är av väldigt problematisk natur. Förutsättningarna för mitt doktorandprojekt var sådana att det hade kunnat utföras inom ett annat ämnesområde än matematisk statistik, men jag är för egen del övertygad om att forskarutbildningen i matematisk statistik gett mig bästa möjliga grund för de utmaningar jag ställts inför.

Biverkningsrapporter

Trots att nya läkemedel genomgår omfattande kliniska prövningar innan de godkänns för utbredd användning måste arbetet med att kartlägga och utvärdera ett läkemedels biverkningsprofil fortsätta under hela dess livscykel. Inget läkemedel är utan risk för oönskade effekter, och ett nytt läkemedels fullständiga säkerhetsprofil är aldrig känt då det godkänns för allmän användning, dels omfattar kliniska prövningar sällan tillräckligt många patienter för att tillförlitligt kunna upptäcka väldigt ovanliga eller idiosynkratiska biverkningar, dels är kliniska prövningar noggrant övervakade medan läkemedel när de kommer i verklig användning inte nödvändigtvis förskrivs i angiven dos, till rätt typ av patienter och under rätt tidsrymd.

En viktig källa till information om tidigare okända biverkningar efter att ett nytt läkemedel godkänts för allmän användning är rapporter av misstänkta biverknings-

fall i vanlig klinisk verksamhet. Sådana rapporter samlas in världen över i system som till allra största del bygger på frivillig rapportering. Ursprungligen tilläts bara läkare rapportera men nu godtar vissa länder även rapporter från sjuksköterskor, apotekspersonal och ibland även direkt från patienter. WHO hanterar världens största databas av biverkningsrapporter med nära fyra miljoner rapporter från 83 länder mellan 1967 och 2007. En grupp bestående av ett fyrtiotal medicinska experter världen över granskar rutinmässigt ett urval av alla inkommande rapporter. Målet är att på så tidigt stadium som möjligt identifiera misstänkta fall av okända biverkningar. Beräkningseffektiva metoder för filtrering och mönsterigenkänning är ett viktigt stöd i detta arbete. Med 200 000 inkommande rapporter varje år behöver läkarna hjälp att fokusera på rätt grupper av rapporter. Dessutom används statistiska metoder som stöd för den medicinska granskningen för att karakterisera större grupper av rapporter och identifiera möjliga störvariabler eller skevheter i rapporteringen som kan förklara förhöjda rapporteringsfrekvenser.

Allmänna förutsättningar

Analysen av biverkningsrapporter är problematisk i flera avseenden: rapporterna utgör inget slumpmässigt urval, de ger bara information om sådana förskrivningar som misstänks ha orsakat en biverkning och kan i allmänhet inte antas ha inkommit oberoende av varandra. De utgör därmed ingen god grund för traditionell inferens. Däremot är biverkningsrapporter en ovärderlig källa för att generera hypoteser om okända biverkningar *för vidare uppföljning*. Det är viktigt att komma ihåg att bakom varje rapport ligger en klinisk bedömning att händelsen i fråga var tillräckligt oväntad eller avvikande för att läkaren skulle bemöda sig att fylla i och skicka in en rapport. Statistikerns roll är dock en annan än vad vi kanske är vana vid. Förutom att vi förväntas utveckla metoder till stöd för den fortlöpande filtreringen och medicinska granskningen av inkommande rapporter är en viktig uppgift att hjälpa medicinska experter till välgrundade prioriteringar och undvika mindre välgrundade sådana.

Bedömningen av huruvida ett misstänkt läkemedel-biverkningssamband förtjänar mer noggrann utredning görs alltid av medicinsk expertis. Därför finns starka

argument för att så långt det är möjligt utnyttja relativt enkla statistiska metoder. Risken är att upplevt ogenomträngliga metoder resulterar i statistiskt bländverk som avleder de medicinska experterna från underliggande problem med data som de i grunden känner väl till. Ofta kan kombinationen av flera, relativt enkla metoder och modeller ge ett mer balanserat och lättillgängligt underlag för beslut än en enskild, ytligt sett mer sofistikerad analys som riskerar ha förvanskats av datakvalitetsproblem. Min avhandling omfattar utveckling eller vidareutveckling av ett antal metoder för mer effektiv analys av stora samlingar biverkningsrapporter: en ny metod för att upptäcka dubletter i databasen [2], vidareutveckling av en robust metod för att identifiera avvikande rapporteringsmönster i stora samlingar biverkningsrapporter [3], [4] och en ny metod för att upptäcka misstänkt läkemedelsinteraktion [5]. Dessutom föreslår vi en ny metod för att skatta osäkerheten vid prediktion med så kallade Bayes-klassificerare [6].

Dubletter

En faktor som försvårar effektiv analys av stora samlingar biverkningsrapporter är att enskilda biverkningsfall ibland genererar flera till synes oberoende rapporter. Olika läkare kan ha träffat samma patient och skickat in varsin rapport, eller en och samma läkare kan ha skickat rapporter dels direkt till ett läkemedelsverk och dels till ett eller flera berörda läkemedelsföretag (som i sin tur vidarebefordrar sina rapporter till ett eller flera läkemedelsverk). Dubletter i databasen kan också uppstå till följd av att uppföljningsrapporter med mer utförlig information om tidigare rapporterade fall felaktigt behandlas som nya rapporter. I runda tal fem procent av alla biverkningsrapporter antas vara dubletter men siffran är osäker och vår erfarenhet är att dubletterna dessutom är ojämnt spridda i databasen, vilket gör att de kan vilseleda både statistiska metoder och medicinsk granskning. Effektiva metoder för att identifiera misstänkta dubletter för vidare uppföljning kan därför vara till mycket stor nytta.

Problemet vad gäller att identifiera misstänkta dubletter är att separata biverkningsrapporter relaterade till samma fall endast i undantagsfall innehåller exakt identisk information. Dessutom är rapporterna i WHO-databasen anonymiserade och innehåller aldrig någon patientinformation utöver ålder och kön. Kvaliteten på rapporterna är vidare sådan att inget rapportfält i allmänhet kan anses vara fullt tillförlitligt, vilket gör regelbaserade metoder för att identifiera misstänkta dubletter otillräckliga för ändamålet. Vi valde istället att utvidga och anpassa till vår tillämpning Copas och Hiltons så kallade träff-bom-modell (*hit-miss model*) för hur avvikelser mellan relaterade databasposter uppstår [7]. Med träff-bom-modellen används likelihoodkvoter för modellen att två rapporter härrör från samma misstänkta biverkningsfall jämfört med modellen att de är helt oberoende, som mått för att avgöra om två rapporter är misstänkt lika eller ej.

Fördelen med detta angreppssätt är att det ger en allmän

metod som fungerar för många typer av data. Vi föreslår en ny utvidgning till numeriska data som även kan användas för löpande text tillsammans med valfritt mått för textavstånd [2]. Trots att poängsättningen för observerade rapportpar är relativt sofistikerad är det enkelt att se varför ett visst rapportpar markerats som misstänkta dubletter. Träff-bom-modellen gör vidare relativt starka antaganden om den underliggande modellen vilket har fördelen att många av dess parametrar kan skattas baserat på allmänna egenskaper hos databasen i fråga. Detta är en fördel i vår tillämpning där väldigt få bekräftade dubletter finns att tillgå.

Vår metod för att identifiera misstänkta dubletter är i rutinmässig användning i WHO-databasen sedan 2005 och resultaten hittills är mycket lovande. I en tidig utvärdering baserad på knappt 2000 norska rapporter från 2003 identifierades utöver ett antal redan kända dubletter en dittills okänd dublett. Det mest intressanta med detta rapportpar var att det vid efterföljande granskning inte omedelbart framstod som en uppenbar dublett, bland annat eftersom patientens ålder och datumet för den misstänkta biverkningen skiljde sig något åt mellan de två rapporterna. Anledningen att träff-bom-algoritmen ändå markerat rapportparet som misstänkta dubletter visade sig vara att de två rapporterna hade fem läkemedel gemensamma som dessutom sällan samförskrivs. I detta fall visade sig träff-bom-modellen vara mer tillförlitlig än vår egen bedömning vid manuell granskning. I en större separat studie för ett annat lands räkning identifierade algoritmen fler än 150 dittills okända dubletter, trots att landet ifråga redan bedriver egen dubblettkontroll bland annat baserad på lokalt lagrad information om patientens födelsedatum och initialer.

Avvikande rapporteringsfrekvenser

Ett viktigt led i granskningen av biverkningsrapporter är den första filtreringen av inkommande rapporter efter oväntat höga relativa rapporteringsfrekvenser. Vårt allmänna angreppssätt är att jämföra det observerade antalet rapporter av en viss biverkning för ett visst läkemedel med ett förväntat antal under en enkel modell. Eftersom vi saknar tillförlitlig information om användningen av olika läkemedel på global nivå använder vi oss av databasen själv som referens. I den enklaste modellen utgår det förväntade antalet rapporter från den övergripande relativa rapporteringsfrekvensen för biverkningen i fråga och det totala antalet rapporter på läkemedlet. För att prioritera stora avvikelser mellan observerat och förväntat antal rapporter utan att ge allt för stor känslighet för låga förväntade antal rapporter använder vi en krympt observerat-förväntat-kvot:

$$\log_2 \frac{Obs + 1/2}{Exp + 1/2} \quad (12)$$

Den kan härledas bayesianskt som logaritmen av maximum aposteriori-skattningen av intensitetsparametern λ för en Poisson-fördelning $Po(\lambda \cdot Exp)$ med a priori-fördelning Gamma $G(1/2, 1/2)$ för λ . Ur ett frekventistiskt perspektiv kan den helt enkelt betraktas som en mer

robust version av logaritmen av den rena observerat-förväntat-kvoten, med reducerad varians och väntevärdesavvikelse mot 0.

Ofta har avvikande rapporteringsfrekvenser andra orsaker än ett verkligt samband mellan det ingående läkemedlet och biverkningen. Läkemedlet i fråga kan till exempel tendera att samförskrivas med ett annat läkemedel som orsakar den misstänkta biverkningen. Observerat-förväntat-kvoten kan justeras för att eliminera eventuellt inflytande av misstänkta störvariabler, genom att stratifiera databasen med avseende på den misstänkta störvariabeln och beräkna ett viktat medelvärde av de stratumspecifika observerat-förväntat-kvoterna. Bortsett från krympningen visar sig detta vara ekvivalent med att ersätta det förväntade antalet rapporter i hela databasen med summan av det förväntade antalet rapporter i varje stratum [3]. Trots likheten med Mantel-Haenszel-skattade oddskvoter är dock den justerade observerat-förväntat-kvoten mer känslig för förekomsten av enstaka väldigt små strata än den justerade oddskvoten [4].

Misstänkt läkemedelsinteraktion

Läkemedelsinteraktion kan leda till förhöjd risk för biverkningar vid samtida intag av två eller fler läkemedel. Olika läkemedel kan till exempel konkurrera om samma biologiska receptor. En annan möjlighet är att det ena läkemedlet hämmar ett enzym som bryter ned det andra och därigenom orsakar en oavsiktlig överdos. Att kunna identifiera fall av misstänkt läkemedelsinteraktion är viktigt dels för att kunna undvika olämpliga läkemedelskombinationer i framtiden men också för att läkemedel som annars skulle ha dragits tillbaka ska kunna kvarstå på marknaden med varningar gällande samförskrivning.

Eftersom samförskrivna läkemedel ofta undviks vid kliniska prövningar utgör biverkningsrapporter ett särskilt viktigt komplement för att identifiera misstänkt läkemedelsinteraktion. Ett antal metoder har föreslagits för detta ändamål, men resultaten har inte varit lovande. Vi tror att en möjlig förklaring till det är att tidigare metoder utgått från grundmodeller där avsaknad av interak-

tion motsvarat ungefär multiplikativa effekter av samförrekommande riskfaktorer. Exempel ur WHO-databasen såväl som tidigare diskussion i den epidemiologiska litteraturen antyder att detta är ett olyckligt val. Vår grundmodell utgår istället från att riskbidraget från olika läkemedel är ungefär additivt i avsaknad av interaktion. Givet denna enkla grundmodell använder vi oss av en krympt observerat-förväntat-kvot för att identifiera förhöjda relativa rapporteringsfrekvenser som kan vara orsakade av misstänkt läkemedelsinteraktion [5]. ■

Referenser

- [1] Finney, D. J. (1974). Problems, data, and inference. *Journal of the Royal Statistical Society: Series A* **137**, 1–23.
- [2] Norén, G. N., Orre, R., Bate, A. and Edwards, I. R. (2007). Duplicate detection in adverse drug reaction surveillance. *Data Mining and Knowledge Discovery*, 305–328.
- [3] Norén, G. N., Bate, A., Orre, R. and Edwards, I. R. (2006). Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *Statistics in Medicine* **25**(21), 3740–3757.
- [4] Hopstadius, J., Norén, G. N., Bate, A. and Edwards, I. R. (2007). Adjustment for potential confounders in adverse drug reaction surveillance. Submitted for publication.
- [5] Norén, G. N., Sundberg, R., Bate, A. and Edwards, I. R. (2007). *A statistical methodology for drug-drug interaction surveillance*. Research Report 2007:6, Mathematical Statistics, Stockholm University.
- [6] Norén, G. N. and Orre, R. (2005). Case based imprecision estimates for Bayes classifiers with the Bayesian bootstrap. *Machine Learning* **58**(1), 79–94.
- [7] Copas, J. and Hilton, F. (1990). Record linkage: statistical models for matching computer records. *Journal of the Royal Statistical Society: Series A* **153**(3), 287–320.

Magnus Karlsson har skrivit avhandlingen Load Modelling for Fatigue Assessment of Vehicles - a Statistical Approach. Avhandlingen lades fram den 23 februari 2007 vid Institutione för matematisk statistik på Chalmers.

Lastmodellering inom fordonsindustrin - ett statistiskt problem

av Magnus Karlsson, Matematisk statistik, Chalmers

Lastbilar konstrueras för att utföra hårt arbete under mycket varierande förhållande. De skall klara gruvdrift i arktisk kyla, timmertransport i otillgängliga skogar och containerfrakt på kuperade alpvägar. Samtidigt har man begränsad fordonsvikt så varje extra kilo i konstrukti-

onen medför motsvarande minskning i lastförmåga, men även om man vill hålla nere vikten för att kunna lasta maximalt är det nödvändigt att garantera livslängden och tillförlitligheten hos komponenterna.

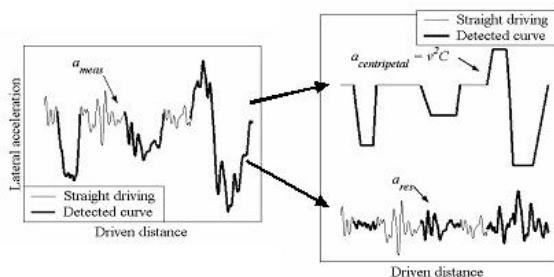
För att kunna konstruera komponenterna så att de upp-

fyller dessa mål är det av stort intresse att ha information om vilka belastningar som komponenterna utsätts för. Det finns en stor variation i dessa belastningar. Variationen beror bland annat på vilken miljö fordonet körs i, hur föraren kör, på vilket sätt man använder fordonet och fordonets specifika dynamiska egenskaper. På grund av den inneboende variationen är det naturligt att behandla lastmodeller utifrån statistiskt perspektiv.

Ett ytterligare problem är att man tidigt i designprocessen vill kunna ta hänsyn till de belastningar som framtidens fordon kommer att utsättas för. Det är därför nödvändigt att ha en beskrivning av populationen av kundlaster som gäller generellt för alla fordon.

Lastmodeller

Traditionellt har man inom fordonsindustrin enbart gjort mätningar på "tuffa" kunder. De uppmätta lasterna har man sedan på olika sätt försökt återskapa på provbanor. Detta har gjort att verifieringen av bilar har blivit starkt kopplad till det enskilda företags provbana och uppfattningen om den verkliga kundpopulationen har varit mycket vag. För att få en bättre förståelse för kundpopulationen finns metoder för att dela in kundernas användning i olika klasser till exempel beroende på vilken vägtyp som kunderna kör på. Det finns också modeller för att beskriva vägprofiler som ger upphov till vertikala laster i bilarna när man kör över vägen.



Figur 1: Kurvkörning

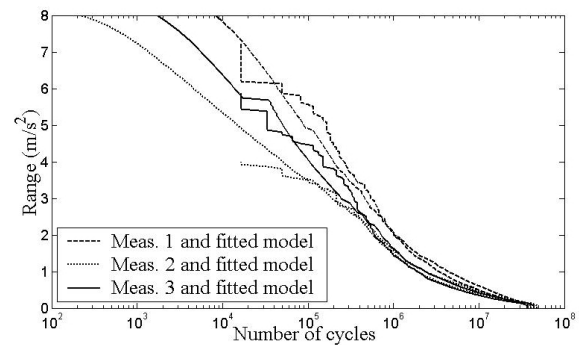
I mitt arbete har jag fokuserat på att få fram en modell för att beskriva laster som kommer in i bilarna från sidan. Dessa laster kan delas in i två delar: en del som kan förklaras av laster från kurvkörning och en residualdel som kommer från fordonets dynamik och förarens ratt-rörelser (se figur nedan). Lasterna från kurvkörning kan i sin tur förklaras av kurvaturen i den enskilda kurvan och vilken hastighet som föraren väljer att köra med.

Parametrisk modell

Genom en parametrisk modell som beskriver fördelningen av kurvaturer och ett regressionssamband mellan hastighet och kurvatur kan man få fram en modell som förklarar lasten på grund av kurvkörning. Residualdelen modelleras som en stationär transformerad Gaussisk process. De parametrar som ingår i modellen skattas från varje mätning.

För att verifiera att modellen ger en tillräckligt god beskrivning av de verkliga uppmätta lasterna har jag härlett

approximativa uttryck för den förväntade ackumulerade delskadan.



Figur 2: Omfångsspektrum

Dessa uttryck är baserat på etablerad utmattningsteori för delskada samt så kallad *rainflow cycle count* som är ett sätt att räkna ekvivalenta lastcykler från tidsignaler. Dessa cykler kan också kopplas till signalens uppkorsningar av intervall. Det är även möjligt att få fram ett uttryck för förväntat omfångsspektrum för den parametriska modellen. Detta visar antalet cykler som har omfång större än en viss nivå. Se Figur 2 för ett exempel.

Med hjälp av denna teknik kan man alltså visa att lastmodellen ger en tillräckligt god beskrivning för den enskilda mätningen. Förhoppningen är dock att kunna beskriva kunder mer generellt. Då parametrarna fångar in beteendet i en mätning kan man se dem som en slags lastvariabler som beskriver hur svår lasten är. Genom att studera hur parametrarna i modellen varierar från mätning till mätning beroende på olika faktorer kan man få en förståelse för belastningarna och dess orsaker. Faktorer som kan vara intressanta att ta hänsyn till är till exempel förare, marknad och typ av väg.

Ett sätt att beskriva kunder är att dela in användningen efter hur mycket kunderna kör i olika vägtyper. Körningen inom varje vägtyp är liknande, men det finns en variation mellan vägtyperna. Variationen mellan olika kunder består då i att de kör olika mycket i de olika vägtyperna. Denna beskrivning har en fördel i att vägtyperna i så fall skulle kunna beskrivas globalt, det vill säga de ser mer eller mindre lika ut oavsett marknad. Se Tabell 1.

För att se hur väl denna idé om indelning stämmer kan parametrarnas variation studeras genom att använda variansanalys. Eftersom den modell som föreslagits ovan bygger på att det finns en viss variation på grund av vägtyp så är förhoppningen att andra effekter inom vägtypen som till exempel de från marknader ska vara små. För varje parameter kan en faktoreffektmodell anpassas som till exempel kan ha följande utseende:

$$y_{ijkl}^{(m)} = \mu^{(m)} + \tau_i^{(m)} + \beta_j^{(m)} + \gamma_{k(j)}^{(m)} + (\tau\beta)_{ij}^{(m)} + (\tau\gamma)_{ik(j)}^{(m)} + \epsilon_{(ijk)l}^{(m)}, \quad m = 1, \dots, p \quad (13)$$

där $\mu^{(m)}$ är medelnivån för parametern. Effekterna $\tau_i^{(m)}$ och $\beta_j^{(m)}$ kommer från den i :te vägtypen respektive den

Tabell 1: Kunder och vägtyper

	Klass 1	Klass 2	Klass 3	Klass 4	Klass 5
Kund 1	$A\%$	$B\%$	$C\%$	$D\%$	$E\%$
Kund 2	$J\%$	$K\%$	$L\%$	$M\%$	$N\%$
...					

j :te marknaden. Samverkanseffekterna mellan vägtyp och marknad representeras av $(\tau\beta)_{ij}^{(m)}$. Föroreffeften inom marknaden finns med som $\lambda_{k(j)}^{(m)}$ och i den mån det finns en samverkan mellan förare och vägtyp kommer den med som $(\tau\lambda)_{ik(j)}^{(m)}$. Slutligen är $\epsilon_{(ijk)l}^{(m)}$ den slumpmässiga variationen på grund av den enskilda mätningen. Denna modell har alltså med en rad faktorer som förhoppningsvis är relativt små.

Komplikation av korrelation

En komplikation i att studera parametrarna en och en är att det finns en korrelationsstruktur att ta hänsyn till. Ett typiskt exempel på detta är att vägar med många kurvor ofta har snäva kurvor. Parametern som styr antalet kurvor är alltså korrelerad med den som styr kurvaturfördelningen.

För att komma till runt detta problem har jag använt mig av en teknik som ortogonaliserar parametrarna. Idén bakom denna är att bara några få av parametrarna verkligen påverkar delskadan väsentligt. Ortogonaliseringen är därför knuten till hur mycket de olika parametrarna påverkar delskadan på så vis att effekten av den parameter som påverkas mest tas bort från de övriga. Därefter tas effekten av den som påverkas näst mest bort från de övriga och så vidare. Resultatet är att det mesta av delskadan kan förklaras med någon eller några av parametrarna.

För att så se om detta är en framkomlig väg att beskriva

lasterna har jag studerat en stor mätserie som genomfördes innan projektet startade. Eftersom mätserien inte var planerad med analysmetoden i åtanke finns det problem med obalans och att urvalet inte gjorts slumpmässigt utan efter olika andra kriterier. En förenklad modell för varje vägtyp för sig har därför anpassats till dessa data. Detta ger en bild av hur stora variationerna är även om det är svårt att tolka en del av effekterna. Man kan dock se att indelningen i vägtyp tycks relativt bra då till exempel marknadseffekten är liten.

Det är dock svårt att dra några handfasta slutsatser från dessa data, men en ny försöksplanering föreslås som bör kunna råda bot på en del av de problem som finns med den tidigare datainsamlingstekniken. Och till detta syfte är de gamla mätningarna mycket användbara.

Slutsats

Sammanfattningsvis kan det konstateras att modellering av laster är mycket användbar i fordonsindustrin då man behöver ta hänsyn till laster i de flesta steg i utvecklingsprocessen. Genom ett statistiskt angreppssätt kan man enklare kvantifiera vad det är som påverkar lasternas variation.

Mitt arbete visar att det är fullt möjligt att modellera lasterna med relativt enkla parametriska modeller, men det visar också på ett stort behov av försöksplanering så att man kan optimera de ofta kostsamma och komplicerade mätningarna. ■

I oktober får vi svaret på frågan: Statistiska föreningen + Samfundet = SANT?

Runt årsskiftet 2006/07 arbetade en gemensam arbetsgrupp med att ta fram ett förslag för hur ett samgående mellan Statistiska föreningen och Svenska statistikersamfundet skulle kunna genomföras. Arbetet var intensivt och ambitionen var att ta fram ett så heltäckande förslag som möjligt. Förslaget skulle alltså innehålla ett förslag på namn, strukturen på underenheter, de ekonomiska frågorna (både vid själva övergången och på sikt), stadgar och hemsidor/medlemstidskrift.

I början av mars år 2007 fanns så ett färdigt förslag som sändes till samtliga medlemmar i Statis-

tiska föreningen inför årsmötet den 22 mars. Rapporten finns tillgänglig på www.statistikersamfundet.se/downloads/rapport_sf_ssf_070306.pdf.

Statistiska föreningens årsmöte sa ja till ett samgående med Svenska statistikersamfundet i enlighet med det förslag som tagits fram. Innan samgåendet kan bli verklighet måste dock motsvarande beslut fattas av Samfundets medlemmar. Frågan tas upp på Samfundets ordinarie årsmöte den 16 oktober år 2007.

Samgåendegruppen genom Åsa Greijer

Rapporter

Sommartider - konferenstider

Under sommaren erbjöds det många konferenser, både de med lång tradition och de mera färska, stora och små, specialinriktade och generella. Med detta smörgåsbord av konferenser är det inte lätt att välja. Vi bestämde oss att åka på en av nykomlingarna, European Survey Research Association, som anordnade sin andra konferens i Prag. Den pågick direkt efter midsommar, 25–29 juni, och lockade så många som 400 registrerade deltagare från 38 länder. Deltagarna hade mycket olika bakgrund, allt från matematiker vid universitet till praktiker som gör enkätundersökningar. Detta ledde till en stor variation bland de 300 presenterade artiklarna. Konferensen hade delat upp artiklarna i sex huvudgrupper "Comparative Research", "Quality of Survey Questions", "Social, Political and Administrative Behavioral Research", "Data Analysis", "Sampling and Nonresponse" och "Mode of Data Collection and Data Enhancement" förutom en slaskgrupp där de som inte platsade i de övriga hamnade. Bland föredragen fanns allt från artiklar om den brittiska kriminalstatistiken, över metoder att få respondenter att svara på telefonintervjuer, till mycket statistiskt teoretiska arbeten, exempelvis om stokastiska differentialekvationer. Våra artiklar (Katrins om effekten av antalet svarsalternativ vid faktorianalys och Johans om medelfel inom strukturella ekvationsmodeller) tillhörde de mer teoretiska bidragen.

Av de praktiska bidragen fick vi värdefull och ibland

oväntad kunskap, till exempel att "Picknick participation" är en bra indikator på socialt kapital, att trädgårdsarbete är mer väldefinierat än hushållsarbete, att man behöver en telefon för att delta i en telefonintervju och att man måste komma ihåg att ange i enkäten i vilken valuta (£ eller €) som brittiska företag ska redovisa sina vinster. En stor andel av konferensens bidrag var relaterad till *European Social Survey*, som är en undersökning som ska kunna användas för att jämföra europeiska länder. Tyvärr var det många inom detta område som behandlade ordinaldata som om det var intervalldata man arbetade med. Och som vanligt var det inte speciellt populärt att ge förslag på adekvata metoder som kan användas istället. Uppenbart finns det ett stort behov av oss statistiker på mer tillämpade konferenser!

Vi kan inte avsluta utan att nämna en av de mest intressanta presentationerna, som gavs av Marika Wenemark från Linköpings universitet. Hon berättade om hur man kan öka kvalitén på enkätsvar genom att fokusera på respondenten, exempelvis genom att använda färre frågor, öppna svarsalternativ, ett introducerande brev innan enkäten skickas ut och endast en påminnelse.

Katrin Kraus och Johan Lyhagen, Uppsala universitet

P.S. För de som aldrig har varit i Prag så kan en resa dit varmt rekommenderas! DS.

Internationell sommarskola vid Örebro universitet

Den 15-17 augusti hölls för sjätte året i rad en sommarskola arrangerad i samarbete mellan Örebro universitet och Statistiska centralbyrån. Ämnet för året var "Urvalssamordning" och huvudtalare var professor Richard Valliant, University of Michigan och the Joint Program for Survey Methodology at the University of Maryland. Övriga talare var Lennart Nordberg och Annika Lindblom, båda från Statistiska centralbyrån.

Deltagarna kom från en rad olika länder. De baltiska staterna och länderna på Balkan var välrepresenterade. Av de drygt 50 deltagarna var cirka 20 från Statistiska centralbyrån. Kursen startade med att Richard Valliant på ett pedagogiskt sätt gick igenom grundläggande urvalsteori för att i slutet av den första dagen komma fram till hur man konstruerar bland annat permanenta slumpetal som kan användas vid samordning av urval. Han belyste flera praktiska problem som uppstår vid urvalsundersökningar.

Dag två började med fortsatt genomgång av urvalsmetoder och metoder för samordning av urval. Både positiv



Richard Valliant, Lennart Nordberg och Annika Lindblom föreläste under sommarskolan om samordnade urval. Foto: Thomas Laitila

och negativ samordning togs upp och dessutom hur man går till väga praktiskt vid roterande urval. Eftermiddagen ägnades åt SAMU som är ett system för samordning av urval som Statistiska centralbyrån har byggt upp. Lennart Nordberg började med att presentera Före-

tagsdatabasen som är grunden för SAMU-systemet och Annika Lindblom fortsatte med att prata om hur själva systemet fungerar i praktiken. Dagen avslutades med en genomgång av hur man skattar variansen för förändringsskattningar då man har samordnade urval. Det var lite svårt att ta till sig på så kort tid, men det kanske klarar efter läsning i [1] som det hänvisades till.

Den andra dagen avslutades för dem som ville med en trevlig och mycket god middag i Naturens hus vid Rynningeviken som är ett naturreservat i närheten av Hjälmaren. Vi fick också en guidad tur i området med dess växtlighet och fågelliv.

Sista dagen gick Richard Valliant igenom urvalsalloke-

ring och estimation och knöt ihop säcken på ett bra sätt genom en sammanfattning av de tre dagarna. Jag tyckte det var en bra kurs med mycket koppling till praktiska problem, och presentationerna var bra och gripbara.

Richard Valliants presentation och en referenslista finns på www.oru.se/esi/summerschool07.

Marina Jansson, SCB

Referenser

- [1] Nordberg, L. (2000). On variance estimation for measures of change when samples are coordinated by the use of permanent random numbers. *Journal of Official Statistics* . **16**, 363–78.

Second Baltic-Nordic Conference on Survey Sampling

Gunnar Kulldorf initierade 1997, vad som har kommit att bli, en serie årliga workshops med syfte att ge ökad kunskap inom området survey sampling i de baltiska staterna, alltså Estland, Lettland och Litauen. Föreläsare hämtades huvudsakligen från Sverige och Finland. Från de baltiska staterna deltog ett antal yngre statistiker på forskarutbildningsnivå från universitet och centralbyråer. Deltagarna i dessa workshops uppmuntrades att presentera egna rapporter vilka sedan diskuterades. Hela tiden har kunskapsnivån höjts och 2002 var tiden mogen för att ha en regelrätt konferens, "First Baltic-Nordic Conference on Survey Sampling", som ägde rum i Ammarnäs. Verksamheten med årliga workshops har sedan fortsatt och i år var det dags för den andra konferensen, "Second Baltic-Nordic Conference on Survey Sampling", den här gången i Kuusamo, Finland.



Carl-Erik Särndal och Harvey Goldstein Foto: Gunnar Kulldorf

Konferensen gick av stapeln den 2-7 juni i de glesbefolkade men skogrika trakterna i nordöstra Finland. Ett uppskattat och för många exotiskt inslag var ett besök i Oulanka nationalpark. Drygt 70 deltagare kom från många olika håll. Förutom från de baltiska och nordiska länderna var det deltagare från Frankrike, Holland, Itali-

en, Kanada, Polen, Slovenien, Storbritannien, Tjeckien, Tyskland och Ukraina. Många uppskattade att det här var en konferens med inriktning på just survey sampling. Detta i samhället så viktiga, men inom universitetsvärlden tämligen försummade ämnesområde.

Konferensens huvudområden var (1) *design och analys av komplexa undersökningar* och (2) *användning av hjälpinformation*. Två speciellt inbjudna föreläsare gav så kallade "keynote lectures". Det var Harvey Goldstein, University of Bristol och Carl-Erik Särndal, Université de Montréal. Goldsteins namn brukar förknippas med flernivåmodellering och mycket riktigt handlade hans föreläsningar om detta. Särndals namn förknippas med GREG-skattning och kalibrering, och hans föreläsningar handlade om tänkandet bakom dessa begrepp, med speciell tillämpning på domänskattningar och justering för bortfallsskevhet.

Dessutom höll deltagarna egna föredrag (i några fall mera "invited" än "contributed"), sammanlagt ett femtiotal. Presentationerna handlade mycket om hur man försöker lösa olika surveyproblem under de speciella förutsättningar som råder i olika tillämpningssituationer i olika länder.

Ett subjektivt urval av problemområden som berördes var skattningar av små domäner, estimation när man har bortfall, kalibrering och modellassisterad (ibland modellbaserad) estimation i mer komplicerade fall, longitudinella studier, icke-sannolikhetsurval, arbetsmiljöundersökningar, nya urvalsförfaranden för olika tillämpningar, urvals- och estimationsproblem inom företagsstatistik, jordbruksstatistik, skoglig statistik, arbetskraftsundersökningar och inkomstundersökningar. Sammanfattningsvis en välordnad och matnyttig konferens med många tillfällen till kontakter med kollegor inom surveyområdet.

Jan Wretman

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
22–24 okt	ICAS4 – Advancing Statistical Integration and Analysis	Beijing, Kina	www.stats.gov.cn/english/icas
4–7 mar 2008	8th German Open Conference on Probability and Statistics	Aachen, Tyskland	gocps2008.rwth-aachen.de
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/

I nästa nummer

Bidrag på temat *visualisering av kvantitativ information*
Olle Sjöström kåserar vidare