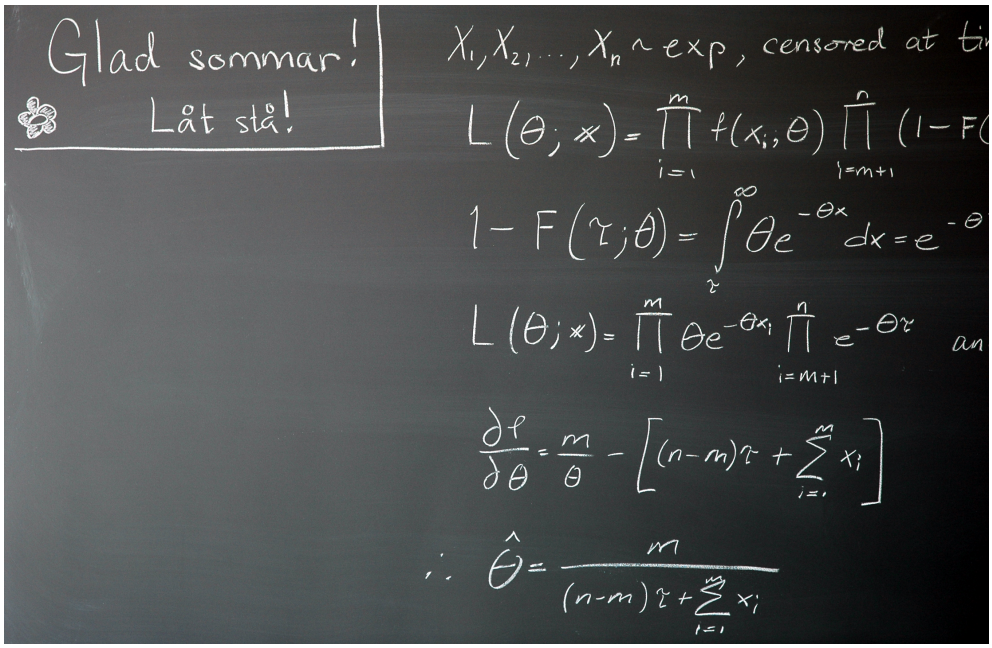
QVARTILEN

Svenska statistikersamfundets tidskrift

Produktionskostnad: 35 kronor

Årgång 22 • Nummer 2 • Juni 2007



Innehåll

		Telefon	E-post
Redaktör	Joakim Malmdin	090 - 786 55 30	qvartilen@gmail.com
	Jens Olofsson	019 - 30 34 60	qvartilen@gmail.com
Redaktion	Magnus Arnér	0520 - 48 30 02	magnus.arner@se.saab.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@astrazeneca.com
	Johan Lyhagen	018 - 471 28 44	johan.lyhagen@dis.uu.se

Qvartilen *ges ut av Svenska statistikersamfundet och publiceras fyra gånger per år – i mars, juni, september och december*

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 3: 15 augusti 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 LaTeX-format. Svenskspråkiga personer skriver på *svenska* och övriga på engelska. Kontakta redaktören för eventuella undantag.

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

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

Annonspriser

	Hel-	Halv-	Kvartssida
PLATSANNONSER/1 nr	4 000:-	2 000:-	1 000:-
ÖVRIGA ANNONSER/1 nr	6 000:-	3 000:-	1 500:-
ÖVRIGA ANNONSER/4 nr	12 000:-	6 000:-	3 000:-

Annonser bokas med Qvartilens redaktör. Medlemmar har 25% rabatt. *E-postutskick* och *adressetiketter* beställes 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. Moms om 25% tillkommer. E-postutskick kan enbart användas 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.

Ordföranden har ordet

Ett kvartal går fort, men det hindrar inte att det kan vara innehållsrikt. Det ser man kanske bäst när man stannar upp och blickar tillbaka. Mitt eget tänkande kring Statistikersamfundet ligger dock sällan bakåt utan mestadels framåt på vad som ska hända. Jag ringer många telefonsamtal, och jag gläder mig åt den positiva respons jag i allmänhet får när jag ber om synpunkter och hjälp.

När jag skriver detta pågår anmälningar till sommarskolan om data mining; när det läses har den redan varit. Det är bestämt att årsmötena för Samfundet och dess sektioner äger rum 15-16 oktober, och orten blir Stockholm för att många ska ha lätt att resa. Vi har efterfrågat nomineringar till Årets yngre statistiker. Statistiska föreningen har haft sitt årsmöte. Påverkar det Statistikersamfundet? Statistiska föreningen svarade ett klart ja på frågan om ett eventuellt samgående. Den väntar nu på Samfundets årsmöte.

Ett av syftena med ett samgående är att bli starkare. Det kan bli mer bredd, mer djup, mer debatt, mer dialog mellan statistiker och användare etcetera. En (ny) förening blir förstås vad medlemmarna vill och gör den till.

Som jag har skrivit förut i den här rutan är det viktigt för oss i styrelsen att veta vad medlemmarna tycker och önskar. Vi ger nu medlemmarna en extra uppmaning – besvara frågorna på sidan 4 i detta nummer av Qvartilen, så sammanställer vi svaren inför årsmötet.

En av många tankar är flera sektioner. Hur vore det med en sektion för finansiell statistik, en för miljöstatistik, en för matematisk statistik, en för statistiska beräkningar och programvara, en för konsultationer, en för undervisning, en för utbildning och en för forskning? Listan kan lätt göras mer omfattande. Förslagen i min uppräkning kan breddas, till exempel genom gruppering, om man så vill. Exempelvis kan utbildning föras samman med undervisning eller med forskning.

Jag varken menar eller tror att en lång rad sektioner ska startas precis nu. Däremot menar jag att det är önskvärt med fler sektioner så att sektionerna tillsammans täcker ett större intresse- och verksamhetsområde än idag. Detta gäller särskilt vid ett samgående, men inte enbart. Det skulle kunna vara ett långsiktigt mål att alla medlemmar är med i minst en sektion. Vad säger du, kära medlem? Du kan informera styrelsen även i detta avseende via frågorna på sidan 4.

Att starta en sektion kräver förstås entusiasm och en del arbete för att komma igång. Om det blir ett samgående, så kommer den nya föreningen att uppmuntra bildandet av sektioner, och ett startbidrag har föreslagits. Läs rapporten "Förslag till samgående mellan Statistiska föreningen och Svenska statistikersamfundet" även av detta skäl! Den kom i mars 2007 och finns på Samfundets hemsida `www.statistikersamfundet.se`.

Eva Elvers

Redaktören har ordet

Behövs statistik? Måste statistik "behövas"? Carl Friedrich Gauss kunde knappast ana att normalfördelningen skulle komma att användas som modell för en mängd olika fenomen inom natur, teknik och samhälle. Pionjärerna inom sannolikhetsteori var förmodligen ovetande om hur viktigt statistik, och därmed också den statistiska teorin, skulle komma att bli för framväxten av demokratiska samhällen. I vår strävan efter att förklara, förenkla, fördjupa, förutse och förstå den värld vi lever i är vetenskap, och däribland statistiken, ett utomordentligt verktyg.

Vare sig vi är engagerade i samhällsdebatten, tekniska landvinningar eller enbart statistik för dess skönhets skull, så är svaret på den första frågan "Ja". Men måste statistik behövas? Måste vetenskapen ses som ett verktyg vars nytta avgörs av dess användning? Med hänvisning till det första stycket är vi benägna att besvara den frågan med "Nej", även om vår horisont likväl som erfarenhet är begränsad.

En närbesläktad frågeställning är *varför* statistik behövs. En professor i nationalekonomi sade lite skämtsamt över en lunch att själva definitionen av att vara statistiker är att vara räknebiträde. Är det så vår omgivning ser på oss? Vi hoppas inte det. Är det så vi ser på oss själva? Vi tror inte det. Däremot ställer vi oss ibland frågan vad som är vår uppgift. Varför behövs statistik? Är det för att serva andra ämnen eller har vi som yrkeskår ett eget existensberättigande? Eftersom statistik är en vetenskap ägnad åt vetenskaplig metodik, behöver det ena inte utesluta det andra, men frågan kan vara nödvändig att ställa i ljuset av den utveckling vi ser idag. I syfte att göra besparingar blir det allt vanligare inom akademin att ämnen slås samman till större institutioner. Hur möter vårt ämne dessa förändringar? På många arbetsplatser finns det ofta ett behov av statistiker, men sällan finns där någon, och finns det en, är denne ofta ensam i sin yrkesutövning. Hur förhåller vi oss till detta faktum? Hur klarar vi av att hålla den statistiska fanan högt?

Vi statistiker tillhör ett litet skrå oavsett var vi är verksamma. Av denna anledning bör vi bli bättre på att samarbeta, visa vad vi kan och värna om vårt ämne. För om inte vi gör det, vem gör det då?

Detta nummer är fyllt av behövd statistik: samhällsdebatt, diskussion om statistikerns ansvar, statistikutbildning och mer teoretiska framsteg och förslag till lösningar. Håll till godo!

Till sist vill vi tacka KG Scherman för alla de intressanta artiklar om pensionsreformen som han har bidragit med under drygt ett års tid. Vi vill också passa på att tacka alla andra bidragsgivare till detta nummer, särskilt Adam Taube, för ett gott och stimulerande samarbete.

Glad sommar! *Låt stå!*

Joakim Malmdin och Jens Olofsson

Styrelsen

Styrelsebeslut tagna på sistone

Styrelsen har tillsatt en arbetsgrupp för att ta fram en yrkesetisk kod för statistiker.

Samfundet har successivt påbörjat övergången till ett webbaserat medlemsregister. I ett första skede har Samfundets register lyfts över till det nya systemet. I nästa steg kommer även sektionernas register att lyftas över. I framtiden kommer det även att vara möjligt att som enskild medlem gå in och redigera sina adressuppgifter. Det webbaserat medlemsregister kommer att underlätta både för styrelsen och den enskilda medlemmen.

Statistiska föreningens årsmöte gav klartecken till samarbetsgruppens förslag om ett eventuellt sammangående. Information om det eventuella sammangåendet finns förutom i detta nummer av Qvartilen på Samfundets hemsida.

Sommarskolan 2007 kommer att handla om *data mining* och huvudföreläsare är Paulo Vedicci. Sommarskolan startar den 12 juni och äger rum på Linköpings universitet.

Styrelsen har tagit fram förenklade rutiner inom Samfundet vad gäller administration, Qvartilen, e-postutskick samt löpande kostnader. Dessa är sammansatta för att underlätta styrelsens arbete och finns som bilagor till protokollen 2007-03-05 samt 2007-03-26.

Styrelsen har beslutat att stödja Surveysektionen med 5000 kr för ett baltiskt-nordiskt samarbete.

Årets årsmöte för Samfundet samt sektionerna kommer att äga rum i Stockholm den 15:e och 16:e oktober.

Styrelsen genom Marie Wiberg

Påminnelse

Du som ännu inte betalat medlemsavgiften på 130 kr för 2007 kan betala in på postgironummer 65 41 95-7. Ange ditt namn och medlemsnummer och år 2007 i samband med inbetalningen.

Frågor till Svenska statistikersamfundets medlemmar om nutid och framtid i verksamheten

Verksamheten i Svenska statistikersamfundet är till för medlemmarna. Vi i styrelsen vill veta mer än vi gör nu om vilken nytta som medlemmarna har av medlemskapet samt hur medlemmarna ser på Statistikersamfundets nutid och framtid. Frågorna har, naturligtvis, ett samband med det föreslagna samgåendet mellan Statistikersamfundet och Statistiska föreningen, men svaren är relevanta även om det inte blir något samgående. Vi ber därför dig som är medlem att besvara nedanstående fem frågor och att skicka ditt svar till den särskilda elektroniska brevlåda som anges nedan.

Svaren kommer att sammanställas av sekreteraren, Marie Wiberg. De kan komma att läsas av hela styrelsen. En sammanfattning av svaren kommer att presenteras på årsmötet den 16 oktober.

Enkäten är inte anonym – vi ber dig vänligen skriva under. Men, ingen individuellt identifierbar uppgift kommer att lämna styrelsen.

1. Vilka är de viktigaste skälen för dig att vara med i Svenska statistikersamfundet?

2. Ett samgående mellan Statistiska föreningen och Svenska statistikersamfundet har utretts och föreslagits. Hur ser du på det? *Motivera gärna.*

3. Vad skulle du vilja se i framtiden i Statistikersamfundet eller den eventuella nya föreningen? *Ange gärna typ av aktiviteter, exempel på innehåll etcetera. En ny sektion är ett exempel.*

4. Skulle du själv vilja medverka aktivt i verksamheten i Statistikersamfundet eller den eventuella nya föreningen? *På vilket sätt skulle du i så fall vilja göra det? Ange hur och med vad.*

5. Hur länge har du varit medlem i Statistikersamfundet?

Svara per e-post till `samfenkat@gmail.com` *senast* **15 augusti 2007**.

Styrelsen genom Eva Elvers, ordförande

Industriell statistik

Sektionen för Industriell statistik

Industriell statistik (IndStat) och *Svenska Förbundet för Kvalitet, Statistisk Metodik* (SFK-StaM) har under en tid pratat om ett samgående med gemensam styrelse och gemensam ekonomi. StaM startades 1990 men har under en längre tid inte haft någon styrelse eller verksamhet. SFK inser dock att det vore bra med en verksamhet inom statistisk metodik så därför har sektionen inte lagts ner. Planerna på samgående har dock lagts på is tills vidare och den främsta anledningen är det föreslagna samgåendet mellan Statistiska Föreningen och Statistikersamfundet. Kravet på juridisk person kommer ju till exempel att påverka hur stadgarna formuleras och det är ännu oklart hur medlemsformer och -avgifter kommer att utformas. Med ett samgående med StaM skulle IndStat få en större bas och ett naturligare samarbete med andra industriella sektioner inom SFK, till exempel fordonssektionen.

Styrelsen genom Ingemar Sjöström

Några spridda tankar

av Ingemar Sjöström, Sony Ericsson Mobile Communication

Efter några årtionden på Ericsson arbetar jag numera på Sony Ericsson. Jag anställdes inom kvalitetsorganisationen och de första två timmarna höll jag det där med statistisk analys lite hemligt ty "Human Resource" (HR) ville ha en "Quality Officer" och då behövdes det inte en statistiker! Nu är det dock ofta en skillnad på vad HR tror och vad företaget behöver, så nu eldar vi under pannorna för att få fyr på analysen. Sony Ericsson består mest av "Design Units" (DU) som konstruerar och utvecklar hårdvara och mjukvara för mobiltelefoner och tillbehör. Dessa avdelningar finns bland annat i Sverige, USA, Japan och Kina. Vi har stor kontakt med underleverantörer men inget att göra med den dagliga tillverkningen förutom saker som hänger ihop med designen (eller brist på design). Att arbeta på ett DU med dess välutbildade ingenjörer gör att man blir "jack-of-all-trades-master-of-none". Det betyder att man får en insikt i radioingenjörens kunskap eller hur en display är uppbyggd eller antennkonstruktörens trängda läge – *finnas men inte synas*. Man får lära sig främmande ord såsom *decibel*, *frekvens*, *lumen*, etcetera. Dessutom finns det andra besvärliga typer såsom teleoperatörer och inte minst Andersson som skall köpa, förstå och kunna använda det hon köpt.

På ett DU finns en mycket stor mängd möjligheter att applicera statistisk metodik och det är stimulerande att arbeta tillsammans med konstruktörerna. De är kunniga i matematik och datoranvändning och ett abstrakt tänkande och har ofta någon form av statistikutbildning bakom sig. Tyvärr brukar inte svensk industri stimulera sina anställda att använda den statistikkunskap de förvärvat tidigare så det blir ofta en Törnrosa-effekt. Men när vi möts över något problem och de känner igen gamla bekanta såsom Poisson eller binomial blir de entusiastiska över att det verkligen är användbart. Nu är ju inte entusiasm enbart av godo ty det kan bli lite väl fantasifulla konfidensintervall ibland. Vid något tillfälle fick en grupp människor i uppgift att samla in data till det klassiska födelsedagsproblemet, "$n = 23$", och efterföljande analys visade att utfallet var extremt osannolikt. När jag då meddelade gruppen att några måste ha fuskat, fyllt i formuläret utan att använda data, blev de djupt imponerade ty det visade sig att så var fallet. Jag försöker också få ingenjörerna att förstå att kunskap och *stats*makt är bra, det vill säga att ordentligt kunna sin teori då de till exempel möter leverantörer, operatörer eller andra motparter.

Ett vanligt dilemma för en konstruktör är ju att de nya produkterna inte finns eller finns i väldigt små prototypserier. Inte sällan möter vi dessutom nya leverantörer av vilka vi inte har någon erfarenhet. Då försöker vi granska deras kvalitets- och statistikarbete på annat sätt. De får till exempel visa vad de kan och gör och inte minst sättet att rita grafer är avslöjande. Om grafer presenteras i slentrianmässiga Excel-format, där μ och σ, n och $n-1$ blandats lite hur som helst, kan man fundera på tänkandet bakom.

Eftersom fältet av frågor och arbetsuppgifter är väldigt stort gäller det att skaffa nya kunskaper. I detta har jag fått ovärderlig hjälp av Lennart Nilsson, Umeå universitet, och ett ökat samarbete mellan industri och högskola och universitet vore önskvärt. ■

Föreningen för medicinsk statistik

Torsdagen den 19 april höll FMS sitt vårmöte vid Matematiska institutionen, Stockholms universitet, denna gång i samarbete med våra två systerorganisationer i Finland, SSL och FSB. Temat på mötet var "Epidemiologi och genetik" vilket lockade drygt 50 deltagare varav cirka tio deltagare från Finland.

Huvudtalare på mötet var Thomas Scheike från Köpenhamn som pratade om regressionsmodeller för överlevnadsdata. Förutom Thomas mycket intressanta föredrag hörde vi på en rad andra bra föredrag inom området av: Olof Bengtsson, Leif Groop, Ola Hössjer, Esa Läärä, Juni Palmgren och Janne Pitkäniemi. Jag tror de flesta deltagarna var mycket nöjda med workshopen, och att träffa våra kollegor från Finland var också givande.

Efter sommaren återkommer FMS med två nya arrangemang:

Första tillfället blir fredag 24 september i Göteborg då FMS anordnar en workshop om adaptiv design i Göteborg, som redan annonserats för Samfundets medlemmar. Föreläsare blir Christopher Jennison och Rob Hemmings. Mer detaljerad information finner du på `www.statistikersamfundet.se/fms/`.

Andra tillfället blir höst- och årsmötet som äger rum någonstans i Stockholm måndagen den 15 oktober. Statistikersamfundets årsmöte sker i Stockholm dagen därpå. Temat för höstmötet blir "Djur och statistik". Mer om detta senare, men vik dagen redan nu. Uppdaterad information finner du på `www.statistikersamfundet.se/fms/`.

Styrelsen genom Tom Britton

SSL Statisticians in the Finnish Pharmaceutical Industry
FSB The Finnish Society of Biostatistics

Annons

Surveysektionen

Utbildningskommittén har avlämnat sin slutrapport; den finns att tillgå på www.statistikersamfundet.se/survey/. Kommittén ger ett antal förslag för att främja utbildningen inom surveyområdet. Av naturliga skäl avser en del av förslagen universitetens verksamhet. Förslaget att inrätta tvååriga masterprogram inom surveyområdet har redan uppfyllts till stor del genom masterprogrammen i statistik vid Stockholms och Örebro universitet. Vidare föreslås ett utvecklat samarbete mellan de statistiska avdelningarna/institutionerna på olika universitet och att en gemensam nationell forskarskola med inriktning mot surveyområdet övervägs. Surveyinriktade kurser bör erbjudas som fristående kurser för kompetensutveckling för personer verksamma inom området. Därtill bör omfattningen av samfinansierade eller delade lärar-/forskartjänster mellan högskolan och myndigheter/företag ökas.

Kom ihåg att anmäla förtjänstfulla uppsatser inom surveyområdet till tävlingen om bästa uppsats 2006/07! Prissumman är på 5000 kr, och syftet med tävlingen är att stimulera intresset för samhällsstatistisk undersökningsmetodik bland studenter och handledare i grundutbildningen.

Surveysektionen deltar i ett informellt nätverk kring standarden ISO 20252 (Internationell standard för marknads-, opinions- och samhällsundersökningar). Standarden lades fast i fjol och framför allt brittiska aktörer har redan certifierat sig enligt denna. En svensk översättning kommer till början av hösten. I vilken utsträckning standarden kommer att tillämpas i Sverige avgörs av aktörerna på undersökningsmarknaden.

Styrelsen har hållit sammanträden per telefon den 28 februari, den 4 april och den 9 maj. Sektionens höstprogram är inte fastlagt, men vi hoppas bland annat kunna erbjuda en intresseväckande årskonferens den 15 oktober.

Styrelsen genom Jörgen Svensson

BLI MEDLEM I SURVEYSEKTIONEN!

Både personer och organisationer är välkomna som medlemmar i Surveysektionen. Läs om fördelarna med medlemskap på sektionens hemsida www.statistikersamfundet.se/survey/ under Medlemskap. Där går det också att anmäla intresse för medlemskap via ett webbformulär. Alternativt går det bra att kontakta sektionens sekreterare Boris Lorenc, e-post sekreteraren.survey@statistikersamfundet.se, telefon 019 - 17 67 83.

Framtidens pensioner

av KG Scherman, GD emeritus Riksförsäkringsverket

Pensionsreformen behöver ses över. Både den förra regeringen, genom dåvarande socialminister Berit Andnor, och den nuvarande, i den framlagda budgetpropositionen, har uttalat sin insikt om detta behov. Översynen behöver innefatta pensionerna och många andra områden. Det har med all önskvärd tydlighet framgått av tidigare artiklar i den här serien.

Ett reformprogram

Grunden för 1994 års historiska kompromiss om pensionerna står sig väl. Pensionssystemet måste vara anpassat till samhällsekonomiska förutsättningar, vilket betyder att pensionsåldern måste höjas och att kravet på arbete för en god pension också i övrigt måste öka när livslängden ökar. Men inom den ramen behöver den sociala utgångspunkten säkerställas.

Analysen behöver starta med en öppen redovisning av utvecklingen av *reformens förutsättningar*, så långt de nu kan överblickas. Resultatet av översynen måste få påverkas av hur förutsättningarna har utvecklas, till exempel beträffande arbetsmarknad, födelsetal och förväntad livslängd. En annan viktig utgångspunkt är beräkningar av vad pensionerna är och väntas bli jämfört med uppsatta mål. I det föregående har vi sett att redan för "modellpersoner" som jobbar länge och med jämn inkomst blir utfallet minst sagt magert. Detta kan utläsas till exempel i socialdepartementets rapport till Bryssel, den strategiska pensionsrapporten från 2005, och i SCBs rapport "Äldres välfärd" från 2006. Det finns en uppsjö av andra rapporter som visar ungefär motsvarande resultat, såsom OECDs översikt över pensionssystemen i världen och olika vetenskapliga rapporter i Sverige. Men kalkylerna blir ännu sämre om man gör studier på verkliga människor. Detta beror på att arbetsmönstren inte är så stabila som man tidigare trott. Dessutom inträder ungdomarna på arbetsmarknaden allt senare. Det nya pensionssystemet är konstruerat för att förmå folk att jobba längre. Det blir då en naturlig följd att pensionen vid 65 blir låg. Samma slutsats gäller beträffande konsekvenserna av ett senare inträdet på arbetsmarknaden.

En "normal" pensionsålder måste återinföras

Nu är den räknetekniska skillnaden mellan de oberoende källorna och de officiella inte så stor som den verkar. Skillnaden beror i stället till en betydande del på att den officiella informationen ganska konsekvent utgår från att människor "förhåller sig som de bör", det vill säga arbetar längre när livslängden ökar, respektive träder in på arbetsmarknaden relativt tidigt. De oberoende källorna sätter i stället fingret på följderna av att situationen för flertalet är annorlunda.

Det kan naturligtvis diskuteras vad som är det "riktiga" angreppssättet. Vill man diskutera vilka problem vi står inför och vad som krävs av ändrade beteenden och ändrade förutsättningar i övrigt så är en tydlig beskrivning av vart vi är på väg om beteendena inte ändras det bästa underlaget. Samma slutsats gäller för en diskussion av om reglerna är lämpligt utformade.

Det kan konstateras att redan idag är pensionsåldern, definierad som den ålder då man kan förvänta sig en hygglig pension, i praktiken klart högre än 65 år. Detta förhållande ligger väl i linje med vad som är samhällsekonomiskt nödvändigt. Samhällsekonomin fordrar att alla som kan arbeta också gör det, samhället kan inte stå neutralt i den frågan. Det är faktiskt ett missförstånd som ligger bakom den så kallade "aktuariska omräkningen" som görs i pensionssystemet för att beakta när människor väljer att ta ut sin pension. Samhället förlorar på att människor inte arbetar, även om deras pensioner minskas. Den högre pensionsåldern är också socialt motiverad. Inom ramen för ett solidariskt kontrakt mellan generationerna måste nämligen alla vara med och bära följderna av förändrade förutsättningar. Men det är nödvändigt att tala klartext om detta och att ta hänsyn till dem som faktiskt inte kan jobba längre.

För tydlighetens skull *måste en "normal" pensionsålder återinföras och höjas successivt* i linje med en ökad livslängd hos befolkningen, tillsammans med en motsvarande höjning av minimiåldern för när pensionen först får tas ut. Kring en sådan pensionsålder kan sedan individerna erbjudas ett eget val av när exakt pensioneringen skall inträffa, men justeringen av pensionens storlek måste ta hänsyn också till de samhällsekonomiska konsekvenserna av individens val. En sådan förändring skulle lägga grunden för en meningsfull dialog mellan politiker och väljare; alla skulle förstå vad det rör sig om, alla skulle förstå sitt eget ansvar och inte behöva leta i för de allra flesta obegriplig myndighetsinformation för att komma till samma resultat. Det skulle också placera reformarbetet där det hör hemma, nämligen i arbetsmarknaden och anställningsmöjligheterna. Politikernas ansvar blir också tydliggjort.

Arbetsmarknaden behöver reformeras

En normal pensionsålder är inte det samma som en pensionsålder som är lämplig för alla. Faktum är att det finns många som inte klarar att jobba till 65, än mindre att jobba längre. Andra får inget jobb, trots att de försöker. Människor över 65 behöver därför *ha tillgång till det sociala skyddsnätet, bland annat i form av sjukförsäkring, förtidspension och arbetslöshetsförsäkring, på samma villkor som de yngre*. Annars leder talet om *flexibilitet* och att "folk har rätt att själva bestämma när de

skall gå i pension" bara till ytterligare minskningar av pensionerna i framtiden. Med en tydligt angiven och realistiskt satt pensionsålder följer en lämplig åldersgräns för skyddsnätet helt naturligt: Den blir densamma som pensionsåldern.

Det finns också olika förutsättningar i olika branscher och för olika grupper av arbetstagare. Det är arbetsmarknadens sak att möta sådana skillnader, *arbetsmarknadens parter måste ta ett ökat ansvar*. Det gäller bland annat särlösningar för grupper som inte kan klara den generella pensionsåldern, behovet av sådana lösningar ökar när den generella pensionsåldern höjs. Men det gäller också behovet av att reformera arbetsmarknaden *för alla*. Detta är en stor och svår fråga. Just därför måste den stå i centrum. Det som till slut blir utformningen av pensionssystemet måste stå i samklang med vad som verkligen kan åstadkommas på arbetsmarknaden och med tydligt formulerade och brett accepterade värderingar i samhället.

Det måste bli möjligt och realistiskt att *fortsätta att arbeta efter uppnådda 65 års ålder*. Det livslånga lärandet måste bli en realitet, kompetensutveckling måste vara lika självklar för den som är 60 som för den som är 35. En något ökad ålder för när pensionen tas ut kan konstateras. Men vad ligger bakom denna utveckling? Och vad är det för jobb som erbjuds? Går utvecklingen parallellt med en uppvärdering av värdet av erfarenhet eller handlar det främst om *undantagssysslor*?

Dagens situation är oacceptabel

Tidpunkten för ungdomars *inträde* i arbetslivet behöver sänkas. En orsak är att de själva en gång i framtiden skall kunna få en vettig pension. Varje år i arbetslivet höjer pensionen, det är sant, men det är också det samma som att varje år utanför arbetsmarknaden sänker den. En annan orsak är att omfattningen av ungdomarnas arbetsinsatser påverkar pensionssystemets förmåga att betala pensionerna till dem som nu är gamla. Man måste skapa förutsättningar för den yngre generationen att börja arbeta tidigare än i dag, men även skapa fungerande incitament. Utvecklingen under det senast decenniet tyder på att endera eller bägge dessa faktorer brustit.

Dagens situation är att de unga i allt högre utsträckning står utanför arbetsmarknaden. En av de många olägenheter som följer med detta är att de får inga pensionspoäng. Hur väl avpassade är egentligen intjänandereglerna till de realiteter som möter dagens unga? Livsinkomstprincipen håller den, är den realistisk? Rättvis, socialt acceptabel? Vi står inför ett svårt val: Att se till att ungdomarna vill ha och får jobb eller förkorta kravet på intjänandetid eller ge beskedet: "Ok, så här är det, resten får du klara själv, gör som du kan och vill." Följdfrågan blir: Är det vettigt att göra alla starkt beroende av privata försäkringar? Vi har i en tidigare artikel sett på de stora svårigheter som möter den enskilde när hon skall försöka att själv ta hand om sitt försäkringsbehov.

Otvivelaktigt finns det berättigade krav att ställa på arbetsprestationer, både när det gäller äldre och yngre. Rimligen bör en diskussion landa i både radikala insatser för en bredare och mer inbjudande arbetsmarknad, i tydlighet när det gäller kraven på arbetsinsatser men också i en förändring av intjänandereglerna. *Dagens situation är oacceptabel.*

Taket behöver höjas – kraftigt

Vilka insatser som än görs när det gäller arbetsmarknad och intjänanderegler så måste vi förstå och acceptera att människors arbetskarriärer varierar kraftigt, mer än vi förstod när pensionsreformen förbereddes. Detta, liksom önskan att erbjuda effektiva pensionslösningar för alla, aktualiserar frågan om taket för den pensionsgrundande inkomsten. *Taket i det offentliga pensionssystemet bör höjas och höjas kraftigt*. Det nuvarande taket gör att pensionssystemet är varken heltäckande eller enbart ett grundsystem. Cirka 30% av svenskarna har inkomster över taket och är hänvisade till antingen avtalsförsäkringar eller, för egenföretagarna, att klara sig själva. Men alla med inkomster över taket drabbas av den beskattning som avgifterna på de inkomsterna innebär. De med varierande inkomster träffas hårt av detta. Vissa år har de låga inkomster, det ger låga pensionspoäng. Andra år är inkomsterna stora. Men pensionspoäng erhålles bara upp till ett ganska lågt tak och på inkomsterna däröver, som annars skulle grunda utrymme för avsättning till ett eget pensionssparande, betalas en extra skatt med över 10%. Egenföretagarna, nyckeln till utveckling av näringsliv och tillväxt, är en grupp som drabbas av denna ordning. Deras pensionsbehov borde tillgodoses av det offentliga pensionssystemet i långt högre omfattning än i dag. Vidare gäller att med dagens ordning lämnas de resursstarka utanför. Det offentliga pensionssystemet blir en bisak, som ändå ger så litet att det inte är någon anledning att intressera sig för dess utformning.

En allsidig diskussion måste också ta i frågan om att de äldre pensionärerna inte har fått del av den exceptionella standardstegring som inträffat under de senaste tio åren. Detta är en realitet som inte alls diskuteras. Men den kan inte i längden döljas med hänvisning till att motsvarande formellt sett skulle ha hänt även med det gamla pensionssystemet, då till och med i än större omfattning. Garantipensionens nivå är knuten till inflationen. När reallönerna ökar kommer garantipensionens betydelse successivt att minska. Det är vad som skett under de gångna tio åren. En sådan utveckling är inte acceptabel, och den saknar rättvisa och politisk trovärdighet. Garantipensionens relation till inkomsterna måste ses över och justeras från tid till annan. Det är en åtgärd som redan nu är högst motiverad. Men kravet att få del i den exceptionella standardstegring som inträffat sen 1996 gäller alla pensionärer, det vill säga även ATP-pensionerna. De bör rimligen få någon del av den starkt positiva utveckling av ekonomi och levnadsstandard som inträffat efter den djupa ekonomiska krisen under nittiotalets första hälft. Pensionärerna var med och bar bördan då, de bör få vara med också om den därefter följande utvecklingen. Självfallet bör en avvägning göras mot andra grupper som också har kommit efter. Men

att en sådan avvägning skulle kunna resultera i att ingen kompensation alls skulle ges, det är inte troligt.

AP-fonden *måste* få tillräckliga resurser för att kunna svar mot sin ursprungliga uppgift, som var att säkerställa att de redan 1994 befintliga åtagandena, till dem som då var pensionärer och dem som snart skulle gå i pension, kunde infrias. Att ta ytterligare medel därifrån är uteslutet. I stället bör behovet att återföra medel studeras.

I en allmän översyn av pensionssystemets detaljutformning bör också storleken på det årliga avdraget på $1,6\%$ vid indexeringen, den så kallade "normen", diskuteras. Att helt avskaffa denna norm skulle, om inga andra åtgärder vidtas, kosta i storleksordningen $2,5$ till 3% i höjd pensionsavgift. Icke desto mindre bör denna bestämmelse studeras i samband med andra insatser för att finjustera regelverket.

Alla partier måste engagera sig

Även premiepensionssystemet behöver förändras. Antalet fonder bör minskas radikalt, 700 fonder är en orimlighet. Vidare bör en minimiavkastning införas, i linje med vad som uttalades i 1994 års principer. En sådan ordning skulle ge en riskutjämning. En garanterad avkastning skulle också ge en realistisk utgångspunkt för prognoser över hur stora dessa pensioner skulle kunna förväntas bli, en observation som också redovisades av politikerna i underlaget för 1994 års beslut. Dagens optimistiska prognoser i de så kallade orangea kuverten kunde därmed få ett naturligt slut.

Omfattningen av pensionssystemets åtaganden är så stor att *diskussionen om de framtida pensionerna måste innefatta hela det politiska fältet*. Vi kan räkna med att ha gott om resurser de närmaste tio åren, då antalet aktiva fortsatt kommer att vara särskilt gynnsamt jämfört med antalet pensionärer. Det är därför angeläget att *spara* och göra *vettiga investeringar*. Därigenom skapas bästa möjliga produktionsförutsättningar för den tid då 40-talisterna når pensionsåldern. Men både *höjda avgifter till pensionssystemet och höjda skatter* måste få vara en del i en framtidsdiskussion. Med detta påstående vidgas intresset till att avse alla grupper i samhället. Och det är just vad som behövs, man kan inte ta en grupp i taget och ställa andra utanför. Det är en *vettig* och *rättvis* balans, mellan äldre och yngre, mellan barnfamiljer och pensionärer, mellan arbetslösa och dem som har ett arbete etcetera som måste eftersträvas. Det måste också vara en balans mellan samhällets olika insatser för de äldre, såsom sjukvård och hemtjänst, och utfallet av pensionssystemet. Balansen skall skapas i en *öppen demokratisk process* där politiker och myndigheter vidgår problemen och talar tydligt om krav och svårigheter, likaväl som om möjligheter och fördelar. Denna process kan inte ersättas av ett automatiskt verkande system, den så kallade automatiska balanseringen, "bromsen", bör därför avskaffas. Däremot bör de intressanta och för det svenska pensionssystemet unika beräkningsmetoder som utvecklats i det sammanhanget införlivas i ett sofistikerat varningssystem, som signalerar när något håller på att gå snett. Dessa beräkningsmetoder skulle därmed på ett naturligt och konstruktivt sätt kunna införlivas med en utformning av regelverket som ligger i linje med vad som alltid varit centrala värden i den svenska välfärdsmodellen. ■

Databas med kvinnliga matematiker

Matematiska institutionen vid Uppsala universitet är värd för den nystartade databasen *Kvinnliga matematiker*. Det är en öppen databas där kvinnliga matematiker med anknytning till Sverige har möjlighet att registrera sig. Alla som har en utbildning i matematik eller närbesläktat område, från examensarbete till professorsnivå, är välkomna.

Syftet med databasen är att få en överblick över de kvinnor som är aktiva inom matematik i Sverige. Tanken är att den som söker kandidater till exempelvis betygskommittéer eller utlysta tjänster lätt ska se vilka kvinnor som finns tillgängliga. Det är enkelt att söka i databasen på bland annat forskningsområde och examen.

De mer etablerade kvinnliga matematikerna i Sverige är förstås redan välkända, men när man söker kandidater för exempelvis doktorandtjänster kan det vara svårt att veta vilka kvinnor som kan vara intresserade. Därför är det viktigt att kvinnor som gör examensarbete i matematik eller närliggande områden får hjälp med att registrera sig. Då kan databasen bli ett bra hjälpmedel för rekrytering.

Det är naturligtvis helt frivilligt att vara med i databasen och man kan när som helst avregistrera sig. För att databasen ska innehålla aktuella uppgifter blir de registrerade en gång per år uppmanade per e-post att uppdatera sina uppgifter. Den som inte följer länken i e-brevet blir då automatiskt raderad.

Matematiska institutionen i Uppsala hoppas att den här databasen ska bli ett användbart verktyg för att öka jämställdheten inom matematik och ett bra sätt för de kvinnor som är aktiva inom matematik att synas.

`www.math.uu.se/kvinnligamatematiker/`

Helen Avelin

Gästkolumnist

I förra numret (22-1) av Qvartilen skrev Harry Khamis, Wright State University, USA, om skillnader mellan Sverige och USA med avseende på undervisning. Till detta nummer har vi ställt frågan: Hur skulle du vilja bedriva undervisning i ämnet (matematisk) statistik? *till Robert Lundqvist, Luleå tekniska universitet.*

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

av Robert Lundquist, Luleå tekniska universitet

Att få utrymme att utgjuta sig om egna käpphästar i undervisningen är förstås en chans som är svår att inte försöka ta. Om det går att säga något väsentligt må väl vara osagt, men jag ska göra ett försök.

Först en del avgränsningar. Frågan gällde "undervisning i ämnet statistik/matematisk statistik på grundnivå". Till att börja med kan det vara bra att begränsa resonemanget till de kurser som ges inom högskolan. Det finns annat, både i grundskola och gymnasium, och det finns även en del riktade kurser till yrkesgrupper eller organisationer. Denna undervisning torde vara nog så intressant, men den lämnas därhän här och nu.

När det gäller högskolekurser finns det några olika vinklar:

- Den första kursen för dem som redan från början är inriktade på statistikstudier.
- Den första och kanske enda kursen i statistik/matstat för studerande med andra inriktningar: ekonomer, ingenjörer med flera.
- Den statistiska delen i en kurs av bredare slag, ofta benämnd "forskningsmetodik", "kvantitativa och kvalitativa metoder", en kurs där statistiken alltså bara blir en del.

Alla spåren är att betrakta som grundkurser, men de har förstås olika perspektiv och omfattning. Det är kanske ofta ingen större skillnad mellan de två första varianterna, för ofta kan "Statistik A1" vara en första ingång till senare statistikstudier liksom många inom matstat är teknologer från början. Dessa kurser är det också oftast vi själva som ger.

Det tredje spåret fungerar ofta inte på samma sätt. Det är inte alltid dessa kurser ges av statistiker, det är kanske till och med regel att den ges av undervisare inom den aktuella utbildningen: lärarutbildare, vårdlärare, sociologer eller liknande. Eftersom statistiker sällan är inblandade kanske detta spår inte är det vi tänker på mest, men ser vi till antalet personer som under sin högskoleutbildning går igenom statistikmoment torde dessa kanske utgöra en majoritet.

Statistik med praktisk vinkel

Spåren är givetvis olika, men det jag tänkte ta upp här är likheter och aspekter jag tror vi kan behöva stärka inom alla tre. Där är min egen första käpphäst att statistiken bör ha en *praktisk* vinkel med betoning på vår förmåga att kommunicera. Vi behöver verktyg för att bearbeta verkligheten, göra den begriplig för oss själva och för andra. Med den utgångspunkten blir det också självklart med en bas i data. Det blir alltså viktigt att utveckla studenternas förmåga att på ett smart och systematiskt sätt samla in, bearbeta och presentera data på ett begripligt sätt.

Ja, men detta gör vi väl? Jodå, det brukar finnas med ett avsnitt om beskrivande statistik i kurserna. Studenterna har också arbetat med just beskrivande statistik på tidigare stadier, så detta kan de väl? Mitt intryck är att det ofta finns en hel del brister. Dels är det inte ovanligt med studenter som kanske kan avläsa ett histogram, men som stupar på att konstruera ett själv. Dels har de inga färdigheter för att arbeta med metoder för beskrivning annat än för hand, och med tanke på de ofta stora och komplexa material vi ställs inför är det inget alternativ. Ett annat exempel är de studenter som både kan uttala ordet heteroskedasticitet och ge en förklaring, men som inte kan ge vettiga tolkningar av skattningarna av modellparametrarna.

Dessutom är det såvitt jag kan se detta med beskrivning och kommunikation "praktikerna" ägnar sig åt. I stålbranschen där jag jobbade ett kort tag var det en massa statistik, men då i form av grafiska beskrivningar av tidsserier, tabeller, medelvärden och annat som metodmässigt inte var särskilt krävande. I kommunens undersökning av kommunmedborgarnas fritidsvanor är det ofta inte relevant med mer än tabellsammanställningar och stapeldiagram.

Är då inte detta för enkelt? Tekniskt sett är det förstås inte särskilt svårt, men det finns några aspekter som gör det värt att lägga ner en hel del tid och möda på den här delen:

- Våra studenter kanske kan sätta ihop ett histogram utifrån ett givet material när de får i uppgift att göra just ett histogram, men kan de se vilken graf som passar ett visst material? Det brukar krävas en del handfasta erfarenheter av olika slags material för att kunna välja metod.
- Ska man idag kunna hantera data på ett meningsfullt sätt, och det även när det gäller enklare sammanställningar, måste man kunna göra det med programvara. Till och med korta enkäter ger så pass stora och komplexa material och många möjligheter att det kan te sig överväldigande utan hjälpmedel. Arbete med sådana hjälpmedel kräver en del träning.
- Metoder för bearbetning må vara enkla, men det är i regel inte lätt att utforma ett bra mätinstrument. Ett

exempel är enkäter. Konstruktion av sådana är kanske inte i första hand en statistisk fråga, det ställer krav på andra färdigheter. Det är ändå en del i vanliga statistiska uppgifter. Detsamma gäller sammanställning av data från tekniska mätningar. Hur ska data organiseras för att det ska vara användbart?

- Vad är ett bra urval? Hur utformar man ett bra experiment? I kurserna ingår jämförelser mellan två stickprov eller stickprov i par, men utrustas verkligen studenterna med en förmåga att randomisera, göra slumpurval och göra smarta försöksuppställningar?
- Oavsett om det handlar om enklare beskrivningar eller användning av mer avancerade metoder är det viktigt att statistiken uppfattas som *relevant* för de studerande. För egen del tror jag det bästa sättet att åstadkomma detta är att de får arbeta själva med egna material eller möjligen material som har sin grund inom det aktuella området för utbildningen. Det räcker inte att vi som lärare tar fram *konkreta* material och frågeställningar, för det verkar inte hjälpa vad vi som lärare säger. Konkret och realistiskt blir det först när de studerande själva gjort materialen "till sina".

Vad blir då konsekvenserna av denna betoning på kommunikation och beskrivande? Det finns säkert flera möjligheter, men själv skulle jag vilja göra mer av följande oavsett om det handlar om blivande statistiker, teknologer eller sjuksköterskor:

- Uppgifter som går ut på att studenterna samlar in egna material. Variablerna får gärna vara av olika slag, kategoriska och kvantitativa, i olika skalor. Datamaterialen kanske inte alltid kan vara realistiska i meningen att de hämtas från en tänkt vardag för den färdiga ingenjören eller ekonomen, men i vissa fall går det att komma ganska nära och i andra kanske man får nöja sig med material som ligger nära vår gemensamma vardag.
- Öppna uppgifter som ger möjligheter att välja olika ansatser och helst även ta fel metod, allt för att studenterna ska träna på inte bara de tekniska delarna med att använda en given metod utan även på att avgöra vilken metod som passar för en viss typ av data. En beskrivning av hur vi använder detta i några av statistikkurserna vid Luleå tekniska universitet finns i [6].
- Grafisk presentation är inte alltid så lätt. Många praktiker blir rätt drivna på egen hand, men vi borde kunna ge dem mer. Ska vi ta med [3] eller liknande som obligatorisk kurslitteratur för att stärka ett statistiskt moment vi kan utgå från att alla kommer att arbeta med?
- Användning av programvara, helst både med statistiskt inriktad programvara och standardprogram som Excel. Trots allt, i det kommande arbetslivet torde få ha tillgång till annat än Excel, och nog är det väl bättre att man kan göra några vettiga bearbetningar med Excel än att man inte gör dem alls? Eller kanske ännu värre, att man med Excel gör dåliga bearbetningar?
- Vi behöver ägna mer utrymme åt *dataproduktionen*: urval, experiment, försöksuppställningar, hur data organiseras och liknande frågor.
- Material bör samlas in så att det går att undersöka eventuella *samband*: korstabeller, sambandsplottar eller boxplottar uppdelade på undergrupper. Förutom att detta är vanliga uppgifter verkar vi vara utrustade med en fenomenal förmåga att söka mönster, även där det inte finns några. Dessutom är en möjlig effekt att det helt enkelt blir roligare.

En del av detta är i linje med de allmänt läsvärda rekommendationer som tagits fram av en grupp inom *American Statistical Association* (ASA), se exempelvis inledningen i [4].

Utveckla förmågan att hantera osäkerhet

Ett annat spår jag skulle vilja arbeta med mer är sannolikhetsteorin, och då särskilt att försöka utveckla vardagsförmågan att hantera osäkerheter och sannolikheter. Räkneregler och matematik är förstås på sin plats, men det är inte alltid självklart att man utifrån sådana färdigheter blir bättre på att hantera situationer präglade av osäkerheter.

En ingång är simulering. Det brukar i många kursböcker finnas övningar som går ut på att kasta tärning eller häftstift, och det kan förstås ge en viss förståelse. Nackdelen är dock att detta kan belysa det som upplevs som självklart, och därmed inte tillföra så mycket när det gäller förståelse. Det finns dock en rad mer komplicerade fenomen som till och med kan vara svåra att räkna på där simulering kan komma in. Exempel på detta finns i [5], en bok skriven för en "icke–matematiskt inriktad publik". För egen del använder jag gärna slumptalstabell snarare än programvara. Det går förstås snabbare med programvara, men det kan bli mer handfast med harvande i slumptalstabellen. Blir man mer förtrogen med simuleringens principer kan det förstås vara en lämplig fortsättning att överge slumptalstabellerna och istället arbeta med programvara.

En annan tråd är de allestädes närvarande betingade sannolikheterna. Det sätt vi oftast arbetar med dessa är utifrån en del definitioner och slutligen räknande med Bayes sats. Studenterna tycker ofta det är svårt, trots att det är en påtagligt relevant och närliggande fråga att fundera över vad det positiva resultatet i testet egentligen innebär.

Sannolikheter eller frekvenser?

Jag *tror* att det kunde vara en poäng att både ägna detta mer utrymme än vi normalt gör, och att vi kanske också ska börja använda andra metoder. Argument för detta ges i [1] eller [2]. Utifrån studier av både förmågan att göra korrekta bedömningar och av hur hårt kunskaperna sitter hävdar de att det är bättre att använda "frekvensresonemang" istället för betingade sannolikheter. Kort uttryckt, i stället för att formulera problem i termer av sedvanliga sannolikheter kan vi använda fiktiva populationer där vi räknar antal med givna egenskaper. Två

exempel:

- Sannolikheten att en kvinna utan symptom har bröstcancer är 0,8%. Om en kvinna har bröstcancer är sannolikheten 90% att mammografi ger positivt utslag. För en kvinna som inte har bröstcancer är motsvarande sannolikhet 7%. Om en kvinna fått positivt svar i mammografi, hur stor är då sannolikheten att hon verkligen har bröstcancer?
- Åtta av 1000 kvinnor utan symptom har bröstcancer. Av dessa 8 kommer 7 mammografi att ge ett positivt svar. Av de övriga 992 kvinnorna som inte har bröstcancer kommer mammografi att ge positivt utslag i omkring 70 fall. Tänk dig ett urval av kvinnor med positivt utslag. Hur många av dessa har bröstcancer?

Gigerenzer hävdar att den senare formen är överlägsen. Inte nog med att man klarar av att göra korrekta bedömningar i mycket högre grad, dessutom glöms tillvägagångssättet inte bort. Det ter sig rätt rimligt. Visst är vi flera som har sett studenter på eget bevåg lösa uppgifter av detta slag med hjälp av fiktiva populationer på ett liknande sätt?

Argumentet bygger på att vi nog inte har kommit så långt när det gäller förståelse av de rätt nyligt uppfunna sannolikheterna. Däremot har vi en inbyggd förmåga att hantera antal, antagligen utvecklad sedan tusentals år tillbaks. Hur gamla är sannolikheter i den form vi använder dem idag? Resonemanget får såvitt jag kan se också stöd från annat håll, exempelvis [7] som hävdar att vi som varelser inte föds ”nollställda” utan är utrustade med en medfödd förmåga att hantera just antal.

Hittills har jag för egen del betraktat dessa frekvenser som mindre lyckade varianter, för bäst hade det ju ändå varit med regelrätt hantering och användning av Bayes sats. Här kan det dock bli kommande förändringar i just detta avsnitt. Dessutom är det en förhoppning att kunna använda detta i fler grupper. I regel har betingade sannolikheter inte tagits upp i kortare kurser, men det finns kanske både möjligheter och poänger att ändra på detta. Detta med osäkra svar från medicinska tester är ett exempel på en högst relevant och förhoppningsvis intresseväckande fråga i kurser för vårdfolk.

Mer deskription, mindre inferens

Dessa kommentarer rör alltså sådant jag skulle vilja ha med mer av i grundkurserna. Om det nu inte går att få med detta utan att stryka något annat, vad ska då bort? Den frågan är besvärligare. En liten betraktelse dock: jag har sett få “praktiker” bestämma ett konfidensintervall eller utföra ett hypotestest, trots att detta ofta upptar en betydande del av kursen. Det förekommer i allehanda akademiska sammanhang, men sällan ute på stålverket eller på vårdcentralen. Varför gör de inte det? Ja, ett skäl kan förstås vara att berörda inte klarar av det. Ett annat skäl kan vara att det kanske inte är relevant. Varför granska *ett* bland alla möjliga väntevärden? Variablerna är många, och det handlar sällan om flera observationer av samma sak som i kursbokens exempel.

Är inferensen verkligen så viktig som man kan tro om vi ser till hur många sidor och hur mycket tid som ägnas åt det? Jag föreslår givetvis inte att vi kan stryka detta avsnitt, men det känns som om vi borde kunna göra det på något annat sätt. Med dagens gängse form missar vi att ta upp en rad faktiska svårigheter som våra studenter garanterat kommer att stöta på i sitt framtida arbetsliv. Samtidigt arbetar vi rätt mycket med stoff de knappast någonsin kommer att använda. Är detta rimligt? Eller uttryckt på ett lite annorlunda sätt: om vi idag tycker att det finns brister i en allmän statistisk förmåga, kan vi som statistiker då möta detta genom att göra mer av det vi gör idag? Om inte, hur ska morgondagens grundkurs se ut? ■

Referenser

[1] Gigerenzer, G. (2003). *Reckoning with Risk – Learning to Live with Uncertainty*, Penguin Books, London.

[2] Gigerenzer, G. and Edwards, A. (2003). Simple Tools for Understanding Risks: from Innumeracy to Insight. *BMJ* **327**, 741–744.

[3] Wallgren, A., Wallgren, B., Persson, R., Jorner, U. och Haaland, J-A. (1996). *Statistikens bilder – att skapa diagram*, Publica, Stockholm.

[4] Moore, D. S. (2003). *Basic Practice of Statistics*, WH Freeman, New York.

[5] Moore, D. S. (2003). *Statistics – Concepts and Controversies*, WH Freeman, New York.

[6] Lundqvist, R. (1999). Critical Thinking and the Art of Making Good Mistakes. *Teaching in Higher Education* **4**(4), 523–531.

[7] Butterworth, B. (1999). *Den matematiska människan*, Wahlström & Widstrand, Stockholm.

Nya medlemmar

Vi hälsar följande medlemmar välkomna:

John Gustafsson	Oscar Dunge
Olga Kostavelis	Anna Lindam
Rodrigo Espinoza	Dan Andersson
Elisabeth Mörk	Charlotte Jansson
Gustaf Rydevik	Ulf Albrechtsson
Joacim Rocklöv	Jonas Bergström
Anders Christianson	Anders Berglund
Joakim Englund	Ulrika Johansson
Anna Liljenberg	Mikael Gerhard
Karolina Carlsson	Malin Albing
Qiuang Gao	

Styrelsen

OPPORTUNITIES TO COMBINE PROBABILITY AND STATISTICS WITH CLIMATE RESEARCH AND MARINE SAFETY AT THE POSTDOC (ER) AND POSTGRADUATE (ESR) LEVEL

SEAMOCS, APPLIED STOCHASTIC MODELS FOR OCEAN ENGINEERING, CLIMATE, AND SAFE TRANSPORTATION, web page: `www.maths.lth.se/seamocs/`.

The EU Marie Curie Research Training Network SEAMOCS offers ***five positions for experienced (ER) researchers*** combining probability and statistics with marine sciences, hydraulics and coastal engineering, climatology and marine transportation safety. It also offers research training for ESR researchers for shorter and longer periods, from three months and up. The positions at two or more participating centres can be combined in order to get interdisciplinary training. The positions are restricted to candidates from EU or associated countries, and require mobility to a new host country.

Post Doctoral positions will be available for research on:

- New methods for uncertainty analysis of extremes from deterministic marine science/meteorological models, at Probability & Statistics, the University of Sheffield, UK (12 months).
- Locally stationary models for random waves – estimation and properties, at Laboratoire des Statistiques et Probabilités, Toulouse, France (12 months).
- Regional scale scenarios for coastal zone impact studies in future climate, at Hydraulics Laboratory, Katholieke Universiteit Leuven, Belgium, (12 months).
- Global scale extremes in future climate, at Royal Dutch Meteorological Institute, De Bilt, The Netherlands (12 months).
- "To be announced", at Naval architecture and Ocean engineering, Chalmers, Sweden (12 months).
- Wave modelling and uncertainties in wave forecasts including the uncertainties in the meteorological forcing, at Swedish Meteorological and Hydrological Institute, Norrköping (12 months).

Short or longer periods for Post Graduate researchers (PhD students) will be available at:

- Mathematical statistics, Lund University, Sweden – for Computational tools for marine applications.
- Probability & Statistics, the University of Sheffield, UK – for Statistical analysis of extremes in marine science and meteorology and in methods for the assessment of model uncertainty.
- Institute of Cybernetics, Tallinn Technical University, Estonia – to be announced.
- Royal Dutch Meteorological Institute, De Bilt, The Netherlands – for Intercomparison of modelled and observed climate extremes.
- Det Norske Veritas, Hövik, Norway – to be announced.

For more details on the positions and exact dates for application, see the network web page or contact the co-ordinator Georg Lindgren, `georg@maths.lth.se`. The research topics at Leuven and De Bilt are closely linked and can be combined to a 24 month period.

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.

Några kommentarer till Mikael Möllers debattinlägg i förra numret av Qvartilen

av Styrelsen

Qvartilen är betydelsefull för Svenska statistikersamfundet, och därmed blir dess redaktör en viktig person. Styrelsen och säkert många med oss uppskattar det stora och förtjänstfulla arbete som Mikael och även tidigare redaktörer har gjort för Qvartilen. Det var en tråkig nyhet när Mikael i augusti 2006 meddelade styrelsen att han från årsskiftet inte längre skulle ha tillräckligt med tid för att fortsätta som redaktör.

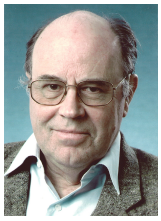
Årsmötet väljer en styrelse att sköta verksamheten under ett år – fram till nästa årsmöte då styrelsen redovisar och står till svars för vad den har gjort och inte gjort. Det åligger styrelsen att hantera löpande frågor; i annat fall blir det förlamning i verksamheten mellan årsmötena. Qvartilen, som ska komma ut fyra gånger om året, hör hit. Styrelsens ledamöter väljs på ett år, och det är vanligt att en individ sitter i två år. Det är, med denna stora omsättning i åtanke, praktiskt att ha nedskrivna rutiner. Styrelsen har under det senaste året skrivit ned flera sådana rutiner som stöd för sitt och framtida styrelsers arbete. Qvartilen är ett exempel.

Ett av syftena med rutinerna för Qvartilen är att precisera ansvarsfördelningen, det vill säga tydligare ange vem som utser vem och vem som ytterst är ansvarig. Denna så kallade beslutsordning anger att det är styrelsen som utser redaktör och redaktionen. Mikael argumenterar i sitt inlägg för att årsmötet bör utse redaktör och att denne skall vara ansvarig endast inför årsmötet. I dagsläget är det omöjligt, eftersom ett sådant förfarande inte finns i Samfundets stadgar. Formellt är det endast styrelsen som kan hållas ansvarig och som beviljas ansvarsfrihet vid årsmötet. Vi anser därför att styrelsen har det formella ansvaret även för Qvartilens utgivning. Därmed förbehåller styrelsen sig rätten att formellt utse och som en sista utväg möjligheten att avsätta en redaktör som missbrukar sin position – även om ingen väntar sig att det ska inträffa. I annat fall kan det hävdas att styrelsen inte sköter sitt åtagande.

Vi håller med Mikael om att demokrati är svårt och det är svårt att på förhand förutse alla eventualiteter. Vi är säkert överens också om att arbetet i en ideell förening bygger på förtroende, överenskommelser och ibland kompromisser. Att styrelsen till exempel tagit på sig att formellt utse redaktionsmedlemmar (efter samråd med redaktören som det står i beslutsordningen) innebär inte att vi vill göra det över huvudet på någon. Styrelsen och Qvartilen är båda till för medlemmarna. Det finns således ingen motsats dem emellan, snarare tvärtom. Redaktören är förstås fri att kritisera styrelsen för dess styrelsearbete. Skulle styrelsen försöka påverka innehållet i någon annan riktning skulle vi nog väldigt snart stå utan redaktör och tidskrift.

Den som vill läsa beslutsordningen hittar den via länken `www.statistikersamfundet.se/protokoll/Styrelse060817.pdf` där den finns som bilaga till styrelseprotokollet. Den var med på årsmötet på en punkt som behandlade den då aktuella frågan med ny redaktör för Qvartilen. Alla medlemmar är välkomna att i Qvartilen och Forum framföra synpunkter på dokumentet och styrelsens arbete i övrigt och att göra det öppet. För stunden ser vi inte tillräckligt starka skäl för att ändra rutinerna vad gäller Qvartilen. ■

En personlig intervju



Till detta nummer har vi gjort en intervju med en person som har en lång och gedigen erfarenhet inom vårt ämne. Personen är professor Adam Taube, Institutionen för informationsvetenskap, Uppsala universitet. Han disputerade i statistik vid Uppsala universitet 1969.

På de statistiska barrikaderna

av Redaktören

Varför valde du att läsa statistik och matematik efter din studentexamen?

Jag valde att läsa statistik och matematik på grund av nyfikenhet och nöjeslystnad. Det var en lärare på gymnasiet som väckte mitt intresse för matematiken. Matematiken blev en naturlig start. En föreläsning av Herman Wold, Uppsala universitet, lockade in mig på statistik. Kanske skulle man kunna kalla det en lättsinnig "random walk".

Hur kommer det sig att en statistiker blev UNESCO-expert i Etiopien?

Det var så att UNESCO stödde ett East-African Statistical Training Center. Det fanns även ett samarbete med universitetet i Addis Abeba där jag hamnade. Projektet pågick alltså redan när jag kom dit och jag var inte den ende svensken. Det fanns redan en "junior assistant" från Göteborg.

Du har erfarenhet av samarbete mellan läkare och statistiker såväl i Sverige som i USA. Är skillnaden stor och vad skulle vi i Sverige kunna förbättra oss på?

Att säga att jag har erfarenhet från USA är en viss överdrift. Jag har gjort många studieresor och sett på nära håll hur samarbetet går till, speciellt vid Dana-Farber i Boston. Statistikerna där kommer in i projekten på ett mycket tidigare stadium än i Sverige. Det är mera av "från ax till limpa". Här i Sverige är det inte så mycket ax utan mer limpa. Det är med beundran jag såg hur samarbetet var ordnat.

I Sverige skulle det behövas en ökad medvetenhet hos dem som använder oss, att de har nytta av oss redan på planeringsstadiet. Inom medicin samlar man ofta in data själv och går till statistikern när det behöver räknas. Statistikerna måste kunna "tala för varan" och för oss själva.

Vad behöver vi statistiker tänka på när vi samarbetar med "praktiker"?

Att statistiker har kommunikationsproblem. Vi statistiker måste kunna framföra meningar med både subjekt och predikat och helst utan formler.

Jag har också tänkt på att det är viktigt att vi sätter på pränt vad vi har sagt exempelvis vid ett möte. På detta sätt kan vi konkret visa på vad vi bidragit med, vilket är svårt om det stannar endast vid ett samtal. Det är också lättare att värdera bidraget om det finns på pränt. Då har man något att peka på.

I din bok, *Mest medvind*, för du fram att du anser att inom medicin sätts medförfattare ofta upp på lösa boliner. Har du någon uppfattning om hur det är inom vårt ämne?

Jag vet inte riktigt hur det är inom andra discipliner, men det är speciellt illa utvecklat inom medicin. Det händer att man får höra att de inte kan betala för arbetet men man kan få stå med på "peket". Det borde framgå vem som gjort vad, precis som inom film. Visst, det kan vara svårt för vissa projekt, men för andra är det absolut möjligt. De som har gjort något ska stå med. Har man inte gjort något, ska man inte stå med. Så enkelt är det.

Kanske kan man se det som Typ I-fel och Typ II-fel; de som har gjort något står inte med och de som inte har gjort något står med. Båda felen behöver minimeras eller snarare risken för dem båda behöver minimeras. Detta gäller naturligtvis inte bara medicin utan jag hade en diskussion med en annan forskare, Kalle Norling, som visade en uppsats inom fysik vilken hade 40 olika författare!

Själv har jag inte velat stå med när jag inte har gjort något och velat stå med när jag däremot *har* gjort något och det har ibland lett till kontroverser. Det händer att man inom medicin frågar en statistiker en fråga av enklare karaktär bara för att i förordet kunna tacka denne. Därigenom antyds att statistikern även tittat på det övriga. Man måste vara på sin vakt!

Mitt intryck är att det inte skulle vara samma djungel inom vårt ämne. Kanske beror det på att jag inte skrivit så mycket tillsammans med andra statistiker.

Du är en engagerad person i ditt yrkesutövande. Bör vi statistiker ta en aktivare roll i debatten än vad vi gör idag?

Ser man någonting som man anser är tokigt, bör man gå in och skriva en artikel om detta, vilket är vad jag har gjort, men eldunderstöd har saknats. Vid en konfe-

rens i Göteborg härområdet ansåg somliga av deltagarna att statistiker inte ska delta i debatten utan vi ska endast servera. Jag är förbluffad över denna ståndpunkt.

Det är betydelsefullt för ämnet att någon för dess talan. Visst, det är inte möjligt att ge sig på alla mål utan man får nischa in sig, vilket jag har försökt att göra.

Förr var statistiken ett känt ämne. Nu sitter vi och mumlar formler i ett hörn och folk har inte riktigt klart för sig vad det är vi gör. (Ibland har vi inte det själva heller.)

Vad anser du är statistikerns roll och ansvar i yrkesutövandet?

Är man inblandad i en bearbetning bör man kunna försvara den med gott samvete. Man måste också kunna tala i klartext och försöka begripa de problem man har att lösa, förstå den saklogiska frågan. Självklart är det inte bara vi som måste tala i klartext utan även de som vi arbetar tillsammans med. Det finns projekt som jag har varit tvungen att avsluta då jag inte har lyckats förstå grundproblemet.

Ett exempel på kommunikationsproblem är att samma ord kan ha olika betydelser inom olika discipliner. *Parameter* är ett sådant exempel, men det finns fler.

I dina ögon, vad kännetecknar en god statistiker?

En god statistiker ska ha känsla för numeriska storleksordningar, en känsla för kvantiteter. Det måste finnas en känsla för om svaret är rimligt. Vidare måste man som statistiker ha respekt för de saklogiska experterna. Vi kan inte komma med näsan i vädret, med modeller och tro att vi kan allt. Fingertoppskänslan som forskare har inom sitt eget ämne ska inte underskattas utan den ska respekteras, även om de kanske genomfört klantiga statistiska beräkningar.

Du har säkert, som vi, mötts av Disarelis uttryck "Lögn, förbannad lögn och statistik". Hur har du valt att bemöta detta?

Jag har för det mesta valt att inte bemöta detta utan blivit uppgiven och trött. Det förvånar mig att folk berättar detta för mig som om det vore en nyhet trots att de känner till att jag varit verksam inom ämnet i decennier. Folk vet inte vad statistik egentligen är. Genom att vi berättar, skriver och bråkar får de kännedom. Ett ordentligt åskväder kan ibland rensa luften.

Hur ser du på statistikämnets framtid?

Jag är pessimistisk. Jag är rädd för att vi får ett allt större programfixering. Är vi inte uppe på de statistiska barrikaderna finns en fara för att statistik som ämne kommer att spela en allt mindre roll. Folk letar efter program istället för statistiker. Vi har inte kunnat föra ut, saluföra vad vi är bra på. Ett ökat deltagande i debatten är en nödvändighet. Vi måste synas mer. Skriva artiklar och delta i debatten i vår roll som statistiker.

Vi måste också värna om ämnet. Stora bitar av tårtan tas annars över av andra discipliner. Vi ska väl inte alla bli ekonometriska räknebiträden?

Vilka för- respektive nackdelar ser du med att vårt ämne är uppdelat i statistik och matematisk statistik?

Svår fråga. Några fördelar kan jag direkt inte se. Däremot när det gäller nackdelar förbluffas jag över det bristande samarbetet de två grenarna emellan. Vi skulle åstadkomma mycket mer med gemensamma krafter.

Jag upplever det som om man inom matematisk statistik inte har samma problem som inom den samhällsvetenskapliga grenen där ämnet håller på att ätas upp. Kan det tänkas vara så att de är finare på att skriva formler? Det finns risk för att den samhällsvetenskapliga grenen av ämnet försvinner då våra studenter i regel har sämre matematikkunskaper med sig, det finns färre timmar att fördela medan lärostoffet i mångt och mycket är detsamma. Det ska också sägas att jag de senaste åren hållit på med mitt och har kanske inte så god överblick längre.

Vi fick tidigt lära oss att inte berätta på fester eller vid andra sociala sammanhang vad vi arbetar med. Det skulle vara en "festdödare". Hur upplever du att du blivit bemött av personer utanför vårt ämne när du har berättat vad du arbetar med?

För det första vet inte folk vad statistik är men de tror att det är vansinnigt tråkigt och förstår inte hur man kan syssla med det. Jag har lärt mig att det inte lönar sig att berätta i festsammanhang. Det som är viktigt är att man inte tar det personligt, för då riskerar man att gå under.

Till och med min egen hustru somnar när det pratas statistik. Däremot kan jag istället säkerligen ha"dödat" fester genom att spela gitarr – för mycket och för länge.

Skulle du rekommendera en avgångselev från gymnasiet att bli statistiker idag och i så fall varför?

Absolut! Med förtjusning om de är nöjeslystna. Det är ju ett sådant fascinerande ämne.

Om du skulle ge ett råd till alla statistiker, oerfarna som erfarna som läser detta, vad skulle det vara?

Träna dig i att formulera dig i såväl tal som skrift förståeligt. Lär dig så mycket matematik som möjligt. Själv önskar jag att jag inte hade avslutat mina egna matematikstudier så tidigt. Jag har två betyg, vilket motsvarar 40 poäng. Jag skulle ha behövt mycket mer.

Vilken eller vilka frågor anser du vara viktigast att vi statistiker kliver upp på de statistiska barrikaderna och kämpar för?

Vi bör kämpa för medvetenheten att varje statistisk analys baseras på en teoretisk modell eller en teoretisk struktur och så häpnar jag även över attityden till bortfall.

Kan jag inte här få passa på att få flika in ett råd, som jag ofta ger till medicinare, då de frågar "om man får göra si eller så"

> – Du får göra hur du vill bara du talar om hur du har gjort.

Bearbetningen ska vara transparent så att folk förstår vad som har genomförts. ■

Teoretiska artiklar

Denna artikel är en sammanfattning av författarens doktorsavhandling Estimation and Inference for Quantile Regression of Longitudinal Data - With Applications in Biostatistics *som lades fram vid Avdelningen för statistik, Institutionen för informationsvetenskap, Uppsala universitet, den 10 november 2006. Kappan finns tillgänglig på nätet* [†].

Kvantilregressioner för longitudinella data

av Andreas Karlsson, Centrum för klinisk forskning Västerås, Uppsala universitet och Landstinget Västmanland

Ett datamaterial som har samlats in över en viss tidsperiod genom att vid upprepade tillfällen mäta en viss egenskap hos en grupp individer kallas för longitudinella data. Det som utmärker longitudinella data är sålunda att varje individ har flera observationer för varje enskild variabel, och att dessa observationer är mätta vid olika tidpunkter. Vanligtvis antas observationerna från samma individ vara korrelerade, medan observationer från olika individer antas vara oberoende. Logitudinella data kan något förenklat ses som en serie av tvärsnittsdata, där varje tvärsnitt är tagen vid olika tidpunkter. Jämfört med vanliga tidsserier, som vanligtvis består av en enda lång tidsserie, består longitudinella data vanligtvis av många korta tidsserier. Vid analysen av longitudinella data är man oftast främst intresserad av att studera hur en viss variabel förändras över tid. Den främsta fördelen med longitudinella data är att de gör det möjligt att följa en enskild individ över tid och att man vid inferens kan dra nytta av fler observationer både mellan och över individer, och därmed få säkrare resultat.

Det finns en mängd olika metoder och modeller för att analysera longitudinella data. Ofta används regressionsanalys, med tid som en oberoende variabel och en variabel som man mätt över tid som beroende variabel. Dessa regressionsmetoder är baserade på medelvärdet som centralmått, men som är väl känt är medelvärdet ett dåligt centralmått för skeva fördelningar, vilka är vanligt förekommande i verkligheten. Ett alternativ till medelvärdet är medianen, vilken i regressionsfallet ger upphov till medianregression. Medianen är i sin tur bara ett specialfall av kvantiler, och medianregression därför endast ett specialfall av kvantilregression.

Genom att beräkna regressionslinjer för flera olika kvantiler av ett longitudinellt dataset får man en fördelning av regressionslinjer som visar den beroende variabelns fördelning för varje enskild tidpunkt. Detta gör det möjligt att studera hur variabelns fördelning förändras över tid. Genom att titta på en enskild individs placering i fördelningen vid varje tidpunkt är det också lätt att följa hur denna individ utvecklas över tid i relation till övriga individer. Detta ger en mycket mer fullständig bild av datamaterialet än vad som uppnås när endast medelvärdesregression används.

En nackdel med att använda median- eller kvantilregression istället för medelvärdesregression är att det saknas analytiska formler för att skatta regressionsparametrarna, varför man hänvisas till att använda sig av numeriska metoder. Detta leder i sin tur till svårigheter att härleda de statistiska egenskaper som är nödvändiga för inferens. För longitudinella data, där observationer från samma individer är korrelerade, är det ännu svårare att härleda de statistiska egenskaperna. Bootstrap är en metod som lämpar sig väl att använda vid inferens för komplicerade dataset där de statistiska egenskaperna är svåra att härleda, och är därför en naturlig metod att använda vid inferens för kvantilregressioner för longitudinella data.

Skattning av kvantilregressioner

Låt git beteckna observation $t = 1, ..., T_i$ för individ $i = 1, ..., n_g$ i grupp $g = 1, ..., G$ mätt vid tidpunkt τ_{git}, y_{git} den beroende variabeln för observation git, x_{gitk} den oberoende variabeln $k = 1, ..., K$ för y_{git}, ε_{git} en felterm, och β_{gh}, $h = 1, ..., H$, parameter h för grupp g i funktionen $f(\cdot)$. Vidare, låt $\mathbf{x}_{git}$ beteckna en $K \times 1$-vektor med de K oberoende variablerna för y_{git}, $\mathbf{y}_{gi}$ en $T_i \times 1$-vektor med de T_i beroende variablerna för individ i i grupp g, ε_{gi} motsvarande $T_i \times 1$-vektor med de T_i feltermerna ε_{git}, β_g en $H \times 1$-vektor med de H parametrarna β_{gh} för grupp g, och $\mathbf{X}_{gi}$ en $T_i \times K$-matris där raderna utgörs av de transponerade vektorerna $\mathbf{x}'_{git}$. Slutligen, låt $f(\mathbf{x}_{git}, \beta_g)$, förkortad $f(\star)$, beteckna en känd funktion av $\mathbf{x}_{git}$ och β_g och $\mathbf{f}(\mathbf{X}_{gi}, \beta_g)$, förkortad $\mathbf{f}(\star)$, en $T_i \times 1$-vektor där raderna utgörs av funktionerna $f(\star)$. En regressionsmodell för longitudinella data kan sedan skrivas som

$$y_{git} = f(\star) + \varepsilon_{git},$$

eller, i matrisform,

$$\mathbf{y}_{gi} = \mathbf{f}(\star) + \varepsilon_{gi}.$$

Eftersom feltermerna $\varepsilon_{gi1}, ..., \varepsilon_{giT_i}$ för individ i är korrelerade är det naturligt att vid skattningen av β_g ta med denna korrelation i estimationsprocessen. Låt w_{git} beteckna vikten för observation git och $\mathbf{W}_{gi}$ en symmetrisk $T_i \times T_i$-viktmatris för individ i i grupp g med elementen $w_{gitt'}$, $t, t' = 1, ..., T_i$, där $w_{gitt'}$ är vikten för

[†] www.diva-portal.org/diva/getDocument?urn_nbn_se_uu_diva-7186-1__fulltext.pdf

observation git. Vidare, låt $I\{\bullet\}$ beteckna indikatorfunktionen

$$I\{y_{git} - f(\star) < 0\} = \begin{cases} 1 \text{ om } y_{git} - f(\star) < 0 \\ 0 \text{ om } y_{git} - f(\star) \geq 0 \end{cases},$$

och låt för en kvantil q, där $0 < q < 1$, $\rho_q(y_{git} - f(\star))$ beteckna kontrollfunktionen

$$\rho_q(y_{git} - f(\star)) = (y_{git} - f(\star))(q - I\{\bullet\}) \quad (1)$$

och $\rho_q(\mathbf{y}_{gi} - \mathbf{f}(\star))$ en $T_i \times 1$-vektor där raderna utgörs av kontrollfunktionerna (1). Slutligen, låt $\beta_g(q)$ beteckna parametervektorn β_g för kvantil q, $\widehat{\beta}_g(q)$ en vektor med skattade värden för $\beta_g(q)$ och $\mathbf{1}_{T_i} = (1, ..., 1)'$ en $T_i \times 1$-vektor med 1:or. En vägd kvantilregressionsskattning för longitudinella data ges då av

$$\widehat{\beta}_g(q) = \min_{\beta_g} \sum_{i=1}^{n_g} \sum_{t=1}^{T_i} w_{git} \rho_q(y_{git} - f(\star)), \quad (2)$$

eller i motsvarande matrisfall

$$\begin{aligned} \widehat{\beta}_g(q) &= \min_{\beta_g} \sum_{i=1}^{n_g} \mathbf{1}'_{T_i} \mathbf{W}_{gi} \rho_q(\mathbf{y}_{gi} - \mathbf{f}(\star)) \quad (3) \\ &= \min_{\beta_g} \sum_{i=1}^{n_g} \sum_{t=1}^{T_i} \sum_{t'=1}^{T_i} w_{gitt'} \rho_q(y_{git} - f(\star)). \end{aligned}$$

Som framgår är ekvation (2) ett specialfall av ekvation (3) där alla $\mathbf{W}_{gi}$ är diagonalmatriser med diagonalelementen $w_{git} = w_{gitt'}$ och $\widehat{\beta}_g(0.5)$ ett specialfall som skattar en medianregression.

Bootstrap av kvantilregressioner

Eftersom ett longitudinellt dataset har en speciell korrelationsstruktur är det naturligt att använda en bootstrapmetod som bevarar denna struktur. Detta kan göras genom att dra med återläggning från de n_g paren $(\mathbf{y}_{gi}, \mathbf{X}_{gi})$ med alla beroende och oberoende observationer för en enskild individ. Låt $\widehat{\beta}^*_{gb}(q) = \left(\widehat{\beta}^*_{g1}(q), \ldots, \widehat{\beta}^*_{gh}(q), \ldots, \widehat{\beta}^*_{gH}(q)\right)'$ beteckna bootstrapskattning $b = 1, \ldots, B$ av $\beta_g(q)$. Efter att ha genererat bootstrapstickprovet $\left(\mathbf{y}^*_{g1}, \mathbf{X}^*_{g1}\right), ..., \left(\mathbf{y}^*_{gm_g}, \mathbf{X}^*_{gm_g}\right)$ och konstruerat en viktmatris $\mathbf{W}^*_{gj}$ för varje par $\left(\mathbf{y}^*_{gj}, \mathbf{X}^*_{gj}\right)$, $j = 1, ..., m_g$, beräknas $\widehat{\beta}^*_{gb}(q)$ med formeln

$$\widehat{\beta}^*_{gb}(q) = \min_{\beta_g} \sum_{j=1}^{m_g} \mathbf{1}'_{T_j} \mathbf{W}^*_{gj} \rho_q\left(\mathbf{y}^*_{gj} - \mathbf{f}\left(\mathbf{X}^*_{gj}, \beta_g\right)\right).$$

Val av vikt

För att kunna använda ekvationerna (2) eller (3) måste först en vikt w_{git} eller $\mathbf{W}_{gi}$ konstrueras. Vikterna kan baseras på t.ex. residualerna e_{git} från en skattning av $\widehat{\beta}_g(0.5)$ med $w_{git} = 1$. I en omfattande simulationsstudie av modellen

$$y_{git} = \beta_{g1} + \frac{\beta_{g2} - \beta_{g1}}{1 + \exp(\beta_{g4}(x_{git} - \beta_{g3}))} + \varepsilon_{git} \quad (4)$$

där feltermerna följer AR(1)-modellen

$$\varepsilon_{git} = \rho \varepsilon_{gi,t-1} + u_{git}$$

vilka sedan standardiserades med avseende på median och varians, jämfördes sex olika residualbaserade vikter samt vikten $w_{git} = 1$ med avseende på genomsnittligt absolut procentuellt fel (MAPE) och väntevärdesriktighet. Jämförelsen gjordes för kvantilerna $q = 0.5$, 0.75 och 0.9, autokorrelationerna $\rho = 0$, 0.5 och 0.95, olika antal individer och observationer samt olika fördelningar för u_{git}. En jämförelse gjordes också mellan medianregressionen $\widehat{\beta}_g(0.5)$ och en vägd medelvärdesregression som skattades med minstakvadratmetoden.

Resultatet av studien ger vid handen att skillnaderna mellan de sju vikterna är små, men vikterna $w_{git} = \frac{1}{\sqrt{s^2_{gi}}}$ och $w_{git} = \frac{1}{u_{gi}}$, där

$$s^2_{gi} = \frac{1}{T_i} \sum_{t=1}^{T_i} (e_{git} - \overline{e}_{gi})^2$$

$$u_{gi} = \frac{1}{T_i} \sum_{t=1}^{T_i} |e_{git} - \overline{e}_{gi}|$$

och

$$\overline{e}_{gi} = \frac{1}{T_i} \sum_{t=1}^{T_i} e_{git}$$

där $i = 1, \ldots, n$, gav något bättre resultat än övriga vikter. Väntevärdesriktigheten är god, särskilt för medianregressionen. Vidare fann vi att en ökad korrelation ρ medför mer precisa skattningar, men ju längre ifrån medianen som q är desto sämre är precisionen. Vilken fördelning som används för u_{git} spelar däremot ingen större roll. I jämförelsen med den vägda medelvärdesregressionen fann vi att den vägda medianregressionen är mer robust.

Bootstrap för biaskorrigering och beräkning av konfidensintervall

En simuleringsstudie av modell (4) genomfördes för att undersöka användandet av bootstrap för att (a) korrigera bristande väntevärdesriktighet för $\widehat{\beta}_g(q)$ och (b) beräkna konfidensintervall för $\beta_g(q)$. Studien visar att bootstrap fungerar väl för att korrigera bristande väntevärdesriktighet och för att beräkna konfidensintervall. Men metoder som konstruerar konfidensintervall som samtidigt korrigerar för bristande väntevärdesriktighet fungerar dåligt, liksom att använda parameterskattningar som är

korrigerade för bristande väntevärdesriktighet tillsammans med ett konfidensintervall. Låt $\widehat{\beta}^*_{gh(b)}(q)$, $h = 1, ..., H$, beteckna ordningsstatistikan b av bootstrapreplikaten $\widehat{\beta}^*_{gh(1)}(q) \leq \cdots \leq \widehat{\beta}^*_{gh(B)}(q)$ för $\widehat{\beta}_{gh}(q)$ och $\widehat{\beta}^*_{g(b)}(q)$ en $H \times 1$-vektor där raderna består av ordningsstatistikorna $\widehat{\beta}^*_{gh(b)}(q)$. Den metod som ger bäst täckningsgrad för ett $100(1-2\alpha)\%$ konfidensintervall och samtidigt ger det smalaste konfidensintervallet är vikten $w_{git} = 1$ och konstruktionen

$$\widehat{\beta}^*_{g(\alpha(B+1))}(q) \leq \beta_g(q) \leq \widehat{\beta}^*_{g((1-\alpha)(B+1))}(q). \quad (5)$$

Resultaten av simuleringsstudien visar vidare att ju längre ifrån medianen som q är desto bredare är konfidensintervallet, och ju högre korrelation ρ desto smalare konfidensintervall.

Bootstrap-baserade hypotestest

Ytterligare en simuleringsstudie genomfördes av modell (4) för att undersöka användandet av bootstrap för att testa hypotesen om likhet mellan en viss kvantils regressionsparametrar för två grupper. Studien delades in i två delar, där likhet testades mellan (a) två parametrar $\beta_{1h}(q)$ och $\beta_{2h}(q)$, med enkel- eller dubbelsidig mothypotes, och (b) två grupper med samtliga H parametrar $\beta_{1h}(q)$ och $\beta_{2h}(q)$, med dubbelsidig mothypotes. Resultatet av studien var att för (a) fungerade det bäst att använda vikt $w_{git} = 1$ med teststatistikan

$$Z^* = \frac{\left(\widehat{\beta}_{1h}(q) - \widehat{\beta}_{2h}(q)\right) - \left(\beta_{1h}(q) - \beta_{2h}(q)\right)}{\sqrt{\widehat{V}^*\left(\widehat{\beta}_{1h}(q)\right) + \widehat{V}^*\left(\widehat{\beta}_{2h}(q)\right)}}, \quad (6)$$

där $\beta_{1h}(q) - \beta_{2h}(q) = 0$ under nollhypotesen, bootstrapestimatorn $\widehat{V}^*\left(\widehat{\beta}_{gh}(q)\right)$ ges av

$$\widehat{V}^*\left(\widehat{\beta}_{gh}(q)\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\beta}^*_{ghb}(q) - \frac{1}{B}\sum_{b=1}^{B}\widehat{\beta}^*_{ghb}(q)\right)^2,$$

där $g = 1, 2$, samt kritiska värden från normalfördelningen används. För (b) fungerade det bäst att använda vikten $w_{git} = 1$ tillsammans med teststatistikan

$$W_p^{c*} = \left[\mathbf{R}\widehat{\beta}^c_0(q)\right]' \left[\mathbf{C}\widehat{\mathbf{V}}^*\left(\widehat{\beta}_{\mathrm{o}}(q)\right)\mathbf{C}'\right]^{-1} \times \left[\mathbf{R}\widehat{\beta}^c_0(q)\right], \quad (7)$$

där $\mathbf{R}$ är en $H \times 2H$-matris med restriktioner för regressionsparametrarna,

$$\mathbf{C} = \left.\frac{\partial \mathbf{R}\beta_0(q)}{\partial \beta_0'(q)}\right|_{\widehat{\beta}_0(q)},$$

$\widehat{\beta}^c_0(q)$ och $\widehat{\beta}_0(q)$ är $2H \times 1$-vektorer med parameterskattningar korrigerade respektive okorrigerade för bristande väntevärdesriktighet, och $\widehat{\mathbf{V}}^*\left(\widehat{\beta}_{\mathrm{o}}(q)\right)$ är en bootstrapskattad varianskovariansmatris för $\widehat{\beta}_0(q)$. De kritiska värdena för (7) fås från de B ordningsstatistikorna av den bootstrapbaserade teststatistikan

$$W_b^{c*} = \left[\mathbf{R}\widehat{\beta}^*_{0b}(q) - \mathbf{R}\widehat{\beta}^c_0(q)\right]' \left[\mathbf{C}\widehat{\mathbf{V}}^*\left(\widehat{\beta}_{\mathrm{o}}(q)\right)\mathbf{C}'\right]^{-1} \times \left[\mathbf{R}\widehat{\beta}^*_{0b}(q) - \mathbf{R}\widehat{\beta}^c_0(q)\right],$$

där $\widehat{\beta}^*_{0b}(q)$ är en $2H \times 1$-vektor med bootstrapskattade parametrar.

Slutsatser

De huvudsakliga slutsatserna från denna undersökning är:

- Valet av vikt spelar liten roll för precisionen i parameterskattningarna.
- Vid beräkning av konfidensintervall, använd vikten $w_{git} = 1$ för parameterskattningarna, undvik korrigering för bristande förväntningsriktighet och beräkna konfidensintervallen med formel (5).
- Vid hypotestest av likhet mellan två olika grupper, använd $w_{git} = 1$ med (6) vid test av två parametrar och $w_{git} = 1$ med (7) vid samtidigt test av samtliga parametrar.

■

Erik Wallgren, Institutionen för informationsvetenskap, Uppsala universitet, försvarade den 27 april sin doktorsavhandling Essays on Capability Indices for Autocorrelated Data*. Fakultetsopponent var professor Jan Wretman, Stockholms universitet.*

Kapabilitetsindex - ett duglighetsindex eller ett odugligt index?

av Erik Wallgren, Örebro universitet

Statistisk kvalitetsstyrning har blivit ett stort forskningsområde sedan starten som väl kan sägas inträffa när Walter A. Shewhart på 1920-talet kontrollerade (styrde) kvaliteten på tillverkade enheter med hjälp av styrdiagram av olika slag. Han såg den oundvikliga variabiliteten i en tillverkningsprocess som en kärnpunkt och delade upp

variabiliteten i två grupper: genuin variation och variation som kan tänkas bero på identifierbara och åtgärdbara orsaker. Shewhart använde uttrycket "common and special causes". Gränsdragningen mellan grupperna är naturligtvis diffus och ju bättre du lär känna din process, desto fler komponenter bland "common causes" blir möjliga att åtgärda. Jag skulle vilja säga att mitt bidrag till området är att identifiera en viktig variationskälla, seriellt beroende, som hittills bakats in i "common causes" och att ge förslag på hur detta beroende ska hanteras.

Tre olika index

Kapabilitetsindex fanns som idé i Japan på 60-talet men först på 70-talet introducerades begreppet capability ratio, eller kapabilitetsindex, som kvoten mellan tillåten och verklig variabilitet i en process, det vill säga

$$\frac{\text{tillåten variabiliet}}{\text{processvariabilitet}}$$

där den tillåtna variabiliteten skrevs som skillnaden mellan högsta och lägsta tillåtna nivå för en speciell kvalitetskarakteristika, gränser som normalt bestäms av produktens användbarhet och definieras i samarbete med användaren.

Med traditionella beteckningar skrivs ett sådant index

$$C_p = \frac{\text{USL - LSL}}{6\sigma}$$

där täljaren är skillnaden mellan "Upper (and Lower) Specification Limit" och nämnarens σ är processens standardavvikelse. Med 6σ avses den totala variationsvidden som processen klarar av. Ju duktigare man är att hitta, och åtgärda, "special causes", desto lägre blir σ, och ju högre blir indextalet.

Svagheten med C_p är naturligtvis att måttet inte reagerar för en eventuell avvikelse från ett målvärde. Bara om processen har låg variabilitet eller ej. Lösningen på detta problem blev ett nytt index,C_{pk}, använt tidigare men publicerat av Kane (1986)

$$\begin{aligned} C_{pk} &= \frac{\min(\text{USL - } \mu, \mu \text{ - LSL})}{3\sigma} \\ &= \frac{d - |\mu - M|}{3\sigma} \end{aligned}$$

där μ är processens medelvärde, d är längden på intervallet USL, LSL med mittpunkten M.

Ytterligare ett sätt att ta hänsyn till en eventuell avvikelse från målvärdet är

$$\begin{aligned} C_{pm} &= \frac{\text{USL - LSL}}{6\sqrt{\sigma^2 + (\mu - T)^2}} \\ &= \frac{\text{USL - LSL}}{6\sqrt{MSE}} \end{aligned}$$

där T står för målvärdet, oftast samma som M.

Alltså tre grundläggande varianter av ett duglighetsindex. Men vad står bokstäverna C, p, k och m för? Att C står för Capability och p för process är väl klart liksom att m-et antyder kopplingen till MSE. Men k-et? Den förklaringen jag tror mest på är att k står för "katayori", japanska för "avvikelse", och som i sig definieras som $|M - \mu|/d$. Det är ju just detta som skiljer C_{pk} från C_p.

Efter dessa indexförslag kom störtskuren ("The avalange" enl. Kotz et al.) av förslag på allt mer sofistikerade indexvarianter. Varianter som kombinerar egenskaper hos de tre nämnda varianterna och varianter som värderar de två felkällorna $|M - \mu|$ och σ olika. Men de tre indexen, framför allt C_{pk}, har hunnit bli etablerade och håller ännu ställningarna i industriella sammanhang.

$C_{pk,}$, som är det ännu dominerande indexet i industrin, och C_{pm} kräver alltså att processens faktiska medelvärde och standardavvikelse är kända men normalt skattas dessa från ett urval från processen. Ett urval som definitionsmässigt är ett obundet slumpmässigt urval, *OSU*, men som i praktiken allt som oftast är systematiskt. Med urvalet kommer ju hela estimationsproblematiken in i bilden och många författare har bidragit med variansskattningar och konfidensgränser för de olika indexen och Kotz et al. ger en utförlig sammanfattning av detta.

Men utgångspunkten för estimationen har varit att urvalet består av observationer på lika fördelade, oberoende och normalfördelade variabler, vårt vanliga *olf*-antagande som i sådana här sammanhang ofta är orealistiskt. Normalfördelningsantagandet har ifrågasatts och behandlats av många författare men det är oberoendeantagandet jag har intresserat mig för. Just förhållandet att ett urval inte är slumpmässigt utan systematiskt, urvalet består av till exempel var 10:e exemplar eller att ett prov tas var 10:e minut, gör att vi måste utgå från att det kan finnas ett seriellt beroende mellan successiva observationer.

För en grundlig genomgång av de olika indexvarianterna se [1] och för en översikt om vad som skrivits på området se exempelvis [2].

Mätningar på autokorrelerade variabler

Målsättningen för avhandlingen [3] var att ta fram konfidensgränser för C_{pk} och C_{pm} i situationer där successiva observationer är mätningar på autokorrelerade variabler. I fyra artiklar har C_{pk} och C_{pm} studerats där processen antagits kunna beskrivas med antingen en $AR(1)$ eller en $MA(1)$-process. Men i en femte artikel har båda indexvarianterna behandlats med antagande om en allmän $ARMA(p, q)$-process. Då $AR(1)$ och $MA(1)$ båda är specialfall av den allmänna $ARMA(p, q)$-processen, kan resultaten från de fyra första artiklarna tas fram som specialfall av resultaten i den femte artikeln.

För C_{pk} har jag utnyttjat en Taylorlinjärisering av $\hat{C}_{pk}$ och för stora urval kan en $100(1 - \alpha)\%$ nedre konfidensgräns tecknas:

$$\hat{C}_{pk} - z_{1-\alpha}\sqrt{\frac{\hat{C}_{pk}^2}{2n}\sum_{h=-(H+3)}^{h=(H+3)}\hat{\rho}_h^2 + \frac{1}{9n}\sum_{h=-(H+3)}^{h=(H+3)}\hat{\rho}_h}$$

Här är $\hat{\rho}_h$ den skattade autokorrelationen för lag h och båda summorna avser en summation av de H första signifikant skattade termerna i korrelationsfunktionen, med positiv och negativ lag, samt ytterligare 3 för att minska risken att utesluta signifikanta skattningar som kan dyka upp senare i funktionen.

För C_{pm} har jag utnyttjat att kvoten $C_{pm}^2/\hat{C}_{pm}^2$ för oberoende observationer är icke-centralt χ^2-fördelad med en skalfaktor som beror av urvalsstorleken och icke-centraliteten. Från simuleringar och studier av momentgenererande funktioner har jag dragit slutsatsen att detta gäller även approximativt för beroende observationer, och efter en transformation till central χ^2, har jag kommit fram till följande $100(1-\alpha)\%$ nedre konfidensgräns för C_{pm}

$$\hat{C}_{pmr} \cdot \left(\chi_\alpha^2\left(\hat{\nu}\right)/\hat{\nu}\right)^{1/2}$$

Här är $\hat{\nu}$ är det skattade antalet frihetsgrader för en central χ^2-variabel

$$\hat{\nu} = \frac{n(1+\hat{\lambda})^2}{\sum\hat{\rho}_h^2 + 2\hat{\lambda}\sum\hat{\rho}_h}$$

och $\hat{\lambda}$ är en skattad icke-centralitetsparameter

$$\hat{\lambda} = \frac{(\hat{\mu}-T)^2}{\hat{\sigma}^2}$$

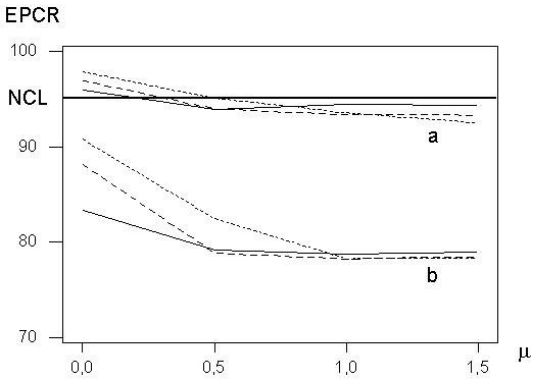
De två korrelationssummorna i uttrycket för frihetsgraderna bestäms på samma sätt som för C_{pk}.

Den empiriska täckningsgraden

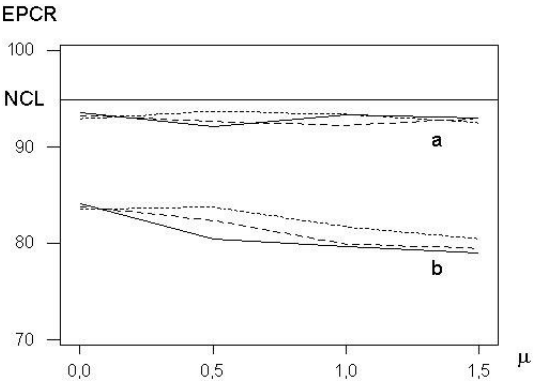
I mängder av Monte Carlosimuleringar har jag skattat C_{pk} och C_{pm} och bestämt nedre konfidensgränser för dessa. För olika urvalsstorlekar, olika stora avvikelser från målvärdet, olika brusvarianser och framför allt, för olika nivåer på autokorrelationen.

Simuleringarna har sammanfattats i figurer som beskriver hur den empiriska täckningsgraden, *EPCR*, påverkats av korrelationens storlek, för varje index och för varje modellantagande. Simuleringarna omfattar också vissa känslighetsanalyser: hur påverkas till exempel täckningsgraden om vi gör ett felaktigt antagande om tidsseriemodell?

I den avslutande artikeln har jag utgått från en *ARMA(1,2)* och en *ARMA(2,1)* process med realistiska parametrar och en nominell konfidensgrad 95% För den förstnämnda kan följande figurer redovisas. I varje figur ges *EPCR* för tre olika nivåer på brusvariansen och om autokorrelationen tas med i skattningen eller ej. En figur för C_{pk} och en för C_{pm}.



Figur 1: EPCR för den undre konfidensgränsen för C_{pk}



Figur 2: EPCR för den undre konfidensgränsen för C_{pm}

Anmärkning: För olika värden på μ och σ_Z, ger de tre övre kurvorna, a, täckningsgraden då autokorrelationen tagits med i beräkningarna medan de tre nedre kurvorna, b, ger täckningsgraden då autokorrelationen exkluderats. För båda figurerna gäller: heldragen linje: $\sigma_Z = 0,5$; streckad linje: $\sigma_Z = 1$ och punktlinje: $\sigma_Z = 1,5$. Nominell täckningsgrad: 95%.

I båda fallen är den empiriska konfidensgraden nära den nominella när vi tar hänsyn till autokorrelationen jämfört med om den negligeras, och även om situationen var väl vald, är det realistiskt att tro att resultatet skulle bli likartat för andra *ARMA*-processer och andra realistiska val av parametrar.

Men att urval som används för skattning av för C_{pk} och C_{pm} ofta är systematiska innebär ytterligare ett problem som inte berörts tidigare. I en kemisk process till exempel görs kontrollmätningar med jämna mellanrum. Med ett antal mätningar som grund bestäms ett index, till exempel C_{pk}. Men här kan man säkert räkna med ett seriellt beroende och ju tätare observationerna görs, desto starkare kan man utgå från att autokorrelationen är, och desto lägre blir det skattade indexet. Man kan alltså i princip själv till en viss gräns bestämma nivån på index genom att göra observationerna med lämpligt tidsavstånd. I avhandlingen redovisas ett datamaterial från en kemisk industri där C_{pk} varierar från 2 till 2.3 beroende på samplingavståndet och C_{pm} varierar på samma sätt mellan 1.7 och 2.0. Följden av detta måste bli att om ett kapabilitetsindex används i processtyrningen och/eller i samband med kontrakt mellan en producent och en konsument, är det viktigt att man definierar vilket index som skall användas och med vilket samplingavstånd observationerna görs.

Urvalsstorleken spelar roll

Naturligtvis spelar också urvalsstorleken en stor roll och urval mindre än 50 kan inte rekommenderas när man kan misstänka att seriell korrelation finns i processen.

Ja, slutligen: är kapabilietsindex ett dugligt index? Många kritiska artiklar har skrivits, många mycket kritiska, och kritiken har riktats mot försöket att beskriva processdugligheten med ett univariat mätning. För papper till exempel är det en kombination av rätt vikt, rätt slitsstyrka, rätt tjocklek etc, som avgör processens duglighet. En multivariat duglighetsbestämning vore alltså att föredra.

Men viktigast är kritiken mot att sammanfatta en avvikelse från målvärdet och standard-avvikelsen i ett mått. Vad ska man tro om C_{pk} blir litet? Har vi hamnat långt från målvärdet eller har standardavvikelsen ökat? Mycket effektivare är att beskriva förändringen i medelvärde och standardavvikelse i styrdiagram.

Men kapabilitetsindex används i till exempel i avtal mellan tillverkare och konsumen; producenten skall leverera från en process med ett index större än till exempel 1,5 eller 2. Om man då lär sig att i stället för en punktskattning av ett kapabilitetsindex, formulera ett krav på en undre gräns för index, som tagit hänsyn till en rimlig seriell korrelation, kan man räkna med att leveranser från underleverantören blir bättre.

Men kan man komma överens om att styrka leveranserna med styrdiagram bör informationen bli ännu bättre. Man kan se från diagrammen om processen kan betraktas som stabil, det vill säga se om tillverkningsprocessen har en konstant genomsnittsnivå och en konstant varians. Den informationen kan man inte få från ett kapabilitetsindex som ju bara ger information från ett enda urval.

Sammanfattning

Kapabilitetsindex är ett mått som används och om det används bör man ta hänsyn till de faktorer som påverkar, till exempel det seriella beroendet. Men man kan i de flesta fall få bättre information från styrdiagram. Alltså, egentligen ett ganska odugligt index.

Detta är en kort sammanfattning av det vikitigaste i min avhandling i statistik. Arbetet började när Olle Carlsson från Umeå universitet kom till Örebro och han fick mig att skriva om kapabilitetsindex; ett arbete som avslutades först i år.

En personlig anekdot

Samuel Kotz's skugga vilar tungt över *Capability Indices*, och en strålande sommardag, jag vill minnas att det var midsommardagen, i slutet på 1990-talet passerade han Stockholm och hade gjort upp med Olle och mig att träffas för att diskutera gemensamma problem. Som sagt, söndagen var strålande och vi hämtade Kotz på Arlanda och körde upp till fåfängan på Söder för lunch. Vi hade först tänkt oss en skön stund med ett glittrande Stockholm vid våra fötter, men Kotz tänkte annorlunda; den mörkaste platsen längst in i en vrå, utan uns till utsikt. Där ville han sitta och dricka dekoffinerat kaffe och prata om våra problem. Något sådan kaffe kunde serveringen inte erbjuda så han fick en kopp hett vatten istället till ett hutlöst pris. Att så många människor var lediga och att så många människor var informellt klädda kunde han inte förstå. För honom tycktes det inte finnas något annat än arbete. Vi åt middag senare på dagen och vi kom att prata om svenska statistiker, som verkligen imponerade på honom. Det märkliga var att han nog i huvudet kunde räkna upp fler än Olle och jag tillsammans. ■

Referenser

[1] Kotz, S. and Lovelace, C. R. (1998). *Process Capability Indices in Theory and Practice*. Arnold, London.

[2] Kotz, S. and Johnson N.L. (2002). Process Capability Indices - A review, 1992-2000. *Journal of Quality Technology* **24**, 188–95.

[3] Wallgren, E. (2007). *Essays on Capability Indices for Autocorrelated Data*, Avhandling. Uppsala universitet.

Annica Dominicus försvarade den 24 mars 2006 sin avhandling Latent variable models for longitudinal twin data*. Opponent var professor Mike Kenward, London School of Hygiene and Tropical Medicine, University of London. Handledare var professor Juni Palmgren, Matematisk statistik, Stockholms universitet, och bihandledare var professor Nancy L. Pedersen, Institutionen för medicinsk epidemiologi och biostatistik, Karolinska Institutet.*

Modeller med latenta variabler för longitudinellla tvillingdata

av Annica Dominicus, AstraZeneca R&D

Longitudinella studier av tvillingar tillför viktig information i sökandet efter en förklaring till hur arv och miljö påverkar mänskliga egenskaper och hur denna påverkan förändras med åldern. I statistiska modeller för upprepade mätningar på tvillingar representeras gener och miljöfaktorer som inte är observerade av latenta variab-

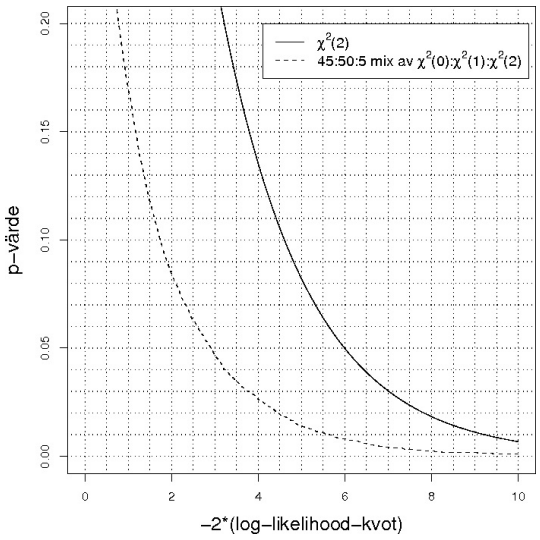
ler med en specificerad sannolikhetsfördelning. På så sätt kan variationen i en egenskap hos olika individer delas upp i komponenter som förklaras av gener och miljöfaktorer. Varianskomponenterna kan sedan beskrivas som en funktion av ålder.

Avhandling är en bred syntes av statistisk modellering av longitudinella tvillingdata och utvecklar denna genom att (i) utvärdera en tvåfas-modell för modellering av variation som en funktion av ålder baserat på longitudinella data, (ii) undersöka hur skattningar av varianskomponenter påverkas av bortfall från tvillingstudier, och (iii) presentera en metod för hypotestest av varianskomponenter. Statistisk metodik illustreras med analyser av empiriska data på kognitiv förmåga från en longitudinell studie kring åldrande, *The Swedish Adoption/Twin Study of Aging* (SATSA). Ovan nämnda bidrag återfinns i tre artiklar [1], [2], [3] som en del av avhandlingen.

I det första arbetet [1] jämförs olika modeller för att beskriva upprepade mätningar av kognitiv förmåga. Fokus ligger på en två-fas-modell där åldern vid övergång från den första till den andra fasen tillåts vara olika för olika individer. Modellen är icke-linjär i de latenta variabler som beskriver förändringen över tid, vilket komplicerar skattning av modellparametrar. En approximativ maximum likelihood-metod som är baserad på en linjärisering av modellen utvärderas och jämförs med en Bayesiansk metod som är baserad på Markov chain Monte Carlo simulering. Tvåfas-modellen visas vara mycket flexibel i att beskriva variationen som en funktion av ålder. Detta är till skillnad från den linjära och den kvadratiska modellen som ofta används för denna typ av modellering.

I den andra artikeln [2] undersöks hur skattningar av varianskomponenter påverkas när tvillingdata är ofullständiga, till exempel på grund av bortfall av deltagare i longitudinella studier. Den metod som ofta används för att anpassa modeller till ofullständiga tvillingdata är baserad på ett antagande om icke-informativt bortfall, det vill säga att bortfallsmekanismen helt kan förklaras av observerade data. Detta arbete utvecklar metodologin genom att formulera en modell för tvillingdata med informativt bortfall. Bortfallets påverkan visas bero på hur faktorerna som styr bortfallsmekanismen hänger ihop med de faktorer som influerar egenskapen i fråga. Mätningar av kognitiv förmåga från SATSA, gjorda med 13 års mellanrum, analyseras och en markant nedgång i skattad genetisk varians observeras. Baserat på en simuleringsstudie utifrån ett antal olika hypoteser om den underliggande bortfallsmekanismen, dras slutsatsen att nedgången inte helt kan förklaras av borfallet av deltagare i studien.

Tredje artikeln [3] presenterar en metod för hypotesprövning av varianskomponenter i statistiska modeller för tvillingdata. Eftersom varianser är icke-negativa placerar nollhypotesen parametrarna på randen av parameterrymden. Det medför att standardresultat inom maximum likelihood-teori som brukar användas ger felaktiga resultat. Som lösning till detta härleds den asymptotiska fördelningen för likelihoodkvot-statistikan för test av varianskomponenterna i klassiska tvillingmodeller. Resultatet är en mix av χ^2-fördelningar. Fördelningen har större sannolikhetsmassa för små värden på likelihoodkvot-statistikan än den χ^2-fördelning som vanligtvis används.



Figur 1: Jämförelse av p-värdeskurvorna

Anmärkning: P-värdeskurvorna är baserade på $\chi^2(2)$-fördelningen respektive den korrekta 45:50:5 mixen av $\chi^2(0)$, $\chi^2(1)$ och $\chi^2(2)$ för likelihood-kvot-test av E-modellen gentemot ACE-modellen.

Resultatet exemplifieras med likelihoodkvot-testet av den så kallade E-modellen gentemot den så kallade ACE-modellen. Detta motsvarar ett test av hypotesen att endast individspecifika miljöfaktorer (E) har en påverkan. I denna modell ingår alltså inte några varianskomponenter för faktorer som helt eller delvis delas av tvillingar, såsom gener (A) och gemensam miljö (C). Figur 1 demonstrerar effekten av att standardmässigt använda en $\chi^2(2)$-fördelning som referens istället för den korrekta mixen av χ^2-fördelningar, som härleds till en 45:50:5 mix av $\chi^2(0)$, $\chi^2(1)$ och $\chi^2(2)$. I figuren tydliggörs att tidigare tvillingstudier kan ha misslyckats att påvisa familjeeffekter på grund av att fel sannolikhetsfördelning har använts. ■

Referenser

[1] Dominicus, A., Ripatti, S., Pedersen, N. L. and Palmgren, J. (2006). *Modelling variability in longitudinal data using random change point models.* Research Report 2006:1, Matematisk statistik, Stockholms universitet.

[2] Dominicus, A., Palmgren, J. and Pedersen, N. L. (2006). Bias in variance components due to nonresponse in twin studies. *Twin Research and Human Genetics*, **9**(2), 185-193.

[3] Dominicus, A., Skrondal, A., Gjessing, H. K., Pedersen, N. L. and Palmgren, J. (2006). Likelihood ratio tests in behavioral genetics: Problems and solutions. *Behavior Genetics*, **36**(2), 331-340.

Lars Ängquist, Matematikcentrum, Avdelningen för matematisk statistik, Lunds universitet, försvarade 16 mars sin doktorsavhandling Pointwise and Genomewide Significance Calculations in Gene Mapping through Nonparametric Linkage Analysis: Theory, Algorithms and Applications. *Fakultetsopponent var docent Staffan Nilsson, Matematiska vetenskaper, Chalmers Tekniska Högskola, Göteborg.*

Pointwise and Genomewide Significance Calculations in Gene Mapping through Nonparametric Linkage Analysis: Theory, Algorithms and Applications

av Lars Ängquist, Matematikcentrum, Avdelningen för matematisk statistik, Lunds universitet

I *kopplingsanalys* (linkage analysis), eller i en något mera generell mening vid genletning, så söker man efter sjukdomsgener längs ett *genom*. Här kan man tolka ett genom som en mängd av hela, eller bitar av, olika kromosomer. Med avseende på en mängd sammanhängande flergenerationella släkter så observerar man då, längs genomets kromosombitar, markördata, det vill säga *genotyper* bestående av nedärvda anlag (alleler) från fädernet respektive mödernet. Dessa observationer analyseras sedan tillsammans med iakttagelser gällande individernas *fenotyper*, med vilket här avses aktuell sjukdomsstatus (sjuka/friska/status okänd). Summan av kardemumman är att man vill försöka lokalisera sjukdomsgener genom att finna onormalt starka *kopplingar* mellan nedärvningen av anlag vid vissa kromosompositioner (lokus) och fördelningen av fenotyper över släkternas inkluderade individer. Detta vill man åstadkomma med så god precision som möjligt. En nyckelobservation är då att en, i någon mening, *signifikant avvikelse*, med avseende på kopplingen mellan genotyper och fenotyper, från vad som kan förväntas under hypotesen om *slumpmässig nedärvning* statistiskt sett tyder på en genetisk komponent kopplad till motsvarande observationslokus. (Begreppet slumpmässig nedärvning härstammar från Gregor Mendel.) En intressant avvikelse består vanligtvis av att de olika fenotypgrupperna internt delar fler nedärvda alleler, i någon mening, än vad som kan anses vara rimligt vid slumpmässig nedärvning med avseende på motsvarande lokus.

En försvinnande kort bakgrund

I introduktionen till min avhandling [1] så beskrivs de grundläggande och nödvändiga genetiska begreppen från den statistisk-genetiska disciplinen. Dessutom introduceras basala begrepp som, till exempel, *nedärvningsprocessen* av genetiska anlag, den genetiska *sjukdomsmodellstrukturen* som statistiskt beskriver ramarna för möjliga kopplingar mellan fenotyper och genotyper samt hur utbredd sjukdomen kan tänkas bli, och ibland även var aktuella sjukdomslokus är lokaliserade. Ytterligare begrepp som diskuteras är *datamaterialet* bestående av observerade släkter, *nedärvningsvektorn* som fullständigt beskriver hur nedärvningen av anlag har gått till i en specifik släkt och vidare olika sätt att beskriva mängden av tillgänglig *genetisk information*.

Efter detta så ges en introduktion till så kallad *enlokus icke-parametrisk kopplingsanalys*, där fokus ligger på *signifikansberäkningar* för en viss typ av teststatistika kallad för *NPL score* (Nonparametric linkage score). Begreppet icke-parametrisk syftar till att inget explicit antagande om strukturen av den genetiska modellen görs; enlokusanalys är ett uttryck för att man letar efter ett sjukdomslokus i taget längs det aktuella genomet. Vidare så utförs, vagt uttryckt, signifikansberäkningar i syfte att kvantifiera huruvida, vid analysen funna, intressanta resultat avviker, i en statistisk mening, tillräckligt mycket från det normala för att man skall våga tro på att man har hittat något sjukdomsrelaterat genomområde. Sådana resultat är naturligtvis kopplade till ett specifikt lokus men metoden kan i princip, med trovärdighet, endast ange att man statistiskt sett har funnit ett sjukdomslokus inom ett område som omringar detta lokus. Vidare, om man letar efter sjukdomar som är kopplade till nedärvningen med avseende på två stycken sjukdomslokus så utför man en *tvålokusanalys*. Även vissa generaliseringar till detta utvidgade fall, samt kopplingar och skillnader till den alternativa analysmetoden *parametrisk* kopplingsanalys, ingår i introduktionen. Vilket kanske kan förstås från relaterad definition ovan så antas vid parametrisk analys kunskap (antagande) om underliggande sjukdomsmodell.

I den tredje delen av introduktionen så beskrivs översiktligt vissa angränsande och/eller alternativa samt kompletterande forskningsfält inom ramen för den statistisk-genetiska kontexten. Slutligen så sammanfattas innehållet i de i avhandlingen fyra olika inkluderade artiklarna. En fullständig avhandlingsintroduktion återfinns via `www.maths.lth.se/matstat/staff/larsa/`.

Vad vi har gjort (eller försökt att göra)

Allmänt kan sägas att om man letar efter gener över substantiellt stora kromosomområden så ger detta upphov till signifikansmässiga tolkningsproblem på grund av så kallad *multipel testning*. Huvudinriktningen för avhandlingens fyra papper är att på ett rimligt sätt utföra signifikansberäkningar (även i form av statistiska styrkeberäkningar) i olika situationer relaterade till såväl enlokus- som tvålokus icke-parametrisk kopplingsanalys i samband med genomvid, i ovanstående mening, multipel testning.

I de två första artiklarna behandlas enlokusanalys vari det första (med Ola Hössjer) förbättrar och utvidgar vissa existerande *analytiska approximationer* för att utföra relaterade signifikansberäkningar [2, 3, 4]. Med analytiska approximationer så menas här att man härleder formler (slutna uttryck) så att man däri kan sätta in aktuella värden på inkluderade, fixa eller skattade, para-

metrar och således direkt få fram numeriska approximationsvärden. Själva uttrycken bygger här på *extremvärdesteori* för sannolikheten att överskrida höga trösklar med avseende på Ornstein-Uhlenbeck-lika stokastiska processer. Vår metod bygger, till exempel, på en kvantilkoppling mellan standard normalfördelningen och en kontinuerlig version av den aktuella marginella NPL scorefördelningen. I slutändan leder detta till en korrigeringsmetod med avseende på normalavvikelser för den senare fördelningen.

Artikel nummer två (med Ola Hössjer) behandlar samma problematik, men här är approximationerna baserade på så kallade *Monte Carlo simuleringar* istället för beräknade med hjälp av fasta analytiska approximationsformler. Denna typ av simuleringar innebär, löst uttryckt, att man slumpmässigt (exakt eller approximativt) genererar (simulerar) fram förlopp eller processer av den typ man är intresserad av och sedan analyserar utfallet av dessa förlopp. I det traditionella fallet med Monte Carlo simuleringar med avseende på genomvid icke-parametrisk kopplingsanalys så uppkommer i vissa fall en beräkningsmässig problematik då det tar, i någon mening, för lång tid att generera tillräckligt många förlopp som är analysmässigt intressanta. Detta beror på att den teststatistika, det vill säga den slumpvariabel som stoppas in i de analytiska eller simuleringsbaserade approximationsformlerna, då generellt sett alltför ofta antar för låga värden under vårt analysscenario, vår nollhypotes. För att lösa detta så inför vi, för att uttrycka det enkelt, ett slumpmässigt placerat artificiellt sjukdomslokus som då i allmänhet, vid simuleringar, leder till högre värden på teststatistikan i närheten av detta lokus. För att få en korrekt probabilistisk tolkning så korrigerar vi för införandet av denna procedur genom att på ett visst sätt väga samman de olika förloppens resultat med avseende på approximationsformeln. Detta utförs med hjälp av så kallad *vägd simulering* (importance sampling). Här jämförs resultaten med avseende på beräkningstid för fix varians (cost-adjusted relative efficiency), med gott resultat, med den relaterade metoden beskriven i [5].

De två avslutande artiklarna riktar in sig på tvålokusanalys. Den första av dessa (med Dragi Anevski och Holger Luthman) behandlar så kallad *obetingad* tvålokusanalys, vilket innebär att man simultant (samtidigt) letar efter två olika sjukdomsgener. I vårt fall består här mängden av släkter enbart av så kallade sjuka syskonpar (affected sib-pairs); vi har med andra ord oss tillhanda ett homogent familjematerial vari varje familj består av ett par föräldrar och ett par affekterade barn till dessa föräldrar. En generell grundkontext målas upp med begreppsapparat samt diskussion av olika angreppssätt och olika typer av relaterade signifikansberäkningar (signifikansnivåer och styrka) med avseende på diverse möjliga situationer. Bland annat så tittar vi på sammansatta nollhypoteser i den meningen att vi tillåter inkluderande av (högst) ett sjukdomslokus vid motsvarande tvålokusdefinitioner. Vi härleder sedan, till exempel, en konservativ analytisk approximationsformel med avseende på *minst gynnsamma fördelning* (least favourable distribution) samt metoder för att uppskatta statistisk signifikans genom att parameterskatta aktuell nollhypotes. Detta innebär, mer explicit, att bestämma instans med avseende på sammansättningen, det vill säga att skatta sjukdomsmodellen under nollhypotesen.

Slutligen, i den sista artikeln (med Ola Hössjer och Leif Groop), utvecklas ett generellt angreppssätt för signifikansberäkningar etcetera gällande så kallad *betingad* tvålokusanalys. Den betingade analysen kan ses som en hybrid mellan enlokus- och tvålokusanalys där man betingar med avseende på någon typ av *information* från ett första *betingningslokus* innan man sedan letar efter ett andra lokus. Här kan betingningslokusen vara givna apriori eller skattade utifrån en initial enlokusanalys. Vidare så kan informationen man betingar på vara enlokusresultat, från teststatistikan, i vårt fall i form av NPL scorer, eller motsvarande underliggande nedärvningsvektorer. Här har det första alternativet beskrivits i [6] medan det andra är en utvidgning utförd i denna artikel. Detta ger alltså upphov till sekventiella snarare än simultana tvålokusmetoder. Man kan också notera att av central betydelse i detta sammanhang är begreppet *icke-centralitetsparameter* (noncentrality parameter), vilket här enkelt uttryckt är ekvivalent med väntevärdet av aktuell teststatistika under en väldefinierad sjukdomsmodell (alternativhypotes). I artikeln härleds optimala versioner av NPL scoret med avseende på icke-centralitetsparametern. Man kan också notera att i detta sammanhang så är denna parameter nära besläktad med styrkefunktionen. ■

Referenser

[1] Ängquist, L. (2007). *Pointwise and Genomewide Significance Calculations in Gene Mapping through Nonparametric Linkage Analysis: Theory, Algorithms and Applications*. Doctoral thesis 2006:15, Department of Mathematical Statistics, Lund University.

[2] Feingold, E., Brown, P. O. and Siegmund, D. (1993). Gaussian Models for Genetic Linkage Analysis Using Complete High-Resolution Maps of Identity by Descent. *American Journal of Human Genetics* **53**, 234-251.

[3] Lander, E. S. and Kruglyak, L. (1995). Genetic Dissection of Complex Traits: Guidelines for Interpreting and Reporting Linkage Results. *Nature Genetics* **11**, 241-247.

[4] Tang, H. K. and Siegmund, D. (2001). Mapping Quantitative Trait Loci in Oligogenic Models. *Biostatistics* **2**, 147-162.

[5] Malley, J. D., Naiman, D. Q. and Bailey-Wilson, J. E. (2002). A Comprehensive Method for Genome Scans. *Human Heredity* **54**, 174-185.

[6] Cox, N. J., Frigge, M., Nicolae, D. L., Concannon, P., Hanis, C. L., Bell, G. I. and Kong, A. (1999). Loci on Chromosomes 2 (NIDDM1) and 15 Interact to Increase Susceptibility to Diabetes in Mexican Americans. *Nature Genetics* **21**, 213-215.

Rapporter

På seminarium med Jan Wretman

av Dan Hedlin, SCB

I efterhand kan man se att någon liten, slumpmässig händelse betydde mycket för ens fortsatta liv. Jag gick på en kurs i urvalsundersökningar och så småningom kom jag att ägna mig åt survey sampling, eller urvalsundersökningar, som yrke. Eftersom Jan var min första lärare i survey sampling ville jag gärna bevista hans "avskedsseminarium" på Statistiska institutionen vid Stockholms universitet.

Seminariet började med att Jan citerade Nationalencyklopedin. "Det här seminariet kommer att handla om 'name-dropping'", sa han, "det vill säga att 'låta uppräkning av kända personers namn få (alltför) stor plats i framställning, konversation eller dylikt, vanligen i syfte att vinna prestige genom att antyda bekantskap med dessa personer eller deras verk.'"

Jan föddes alldeles i slutet på 30-talet i Medelpad. Skoltiden tillbringades i Norrköping, dit familjen flyttade. Det gick hyfsat i skolan, men han hade svårt att klara skrivningarna i matte, eftersom han var för långsam. Matematik kändes omöjligt, och han valde att gå latinlinjen i gymnasiet, vilket gick betydligt lättare.

Efter gymnasiet kom ett år i lumpen, livets absoluta bottennivå i välbefinnande. När det var genomlidet tänkte Jan att han skulle bli samhällsvetare, mycket för att slippa matten. Han började läsa till en politices magisterexamen i Uppsala 1959 i vilken statistik ingick som obligatoriskt ämne. "Bäst att klara av det först, eftersom det verkar tråkigast" tänkte Jan. Om man inte hade gått reallinjen på gymnasiet, så skulle man först gå en förberedande kurs i matte. Matteläraren var den mycket unge och mycket entusiastiske filosofie kandidaten K. G. Jöreskog. Det var kanske den bästa kurs Jan gått på överhuvudtaget – och han tänkte om beträffande matten.

Det började med statistik

Så kom då en termins statistikstudier. Det fanns inga delkurser och ingen plan för undervisningen. Åtminstone fanns ingen plan som var känd för studenterna. Man visste aldrig vilken föreläsare som skulle dyka upp, eller vad han skulle prata om. Det var två eller tre dubbeltimmar föreläsning varje vecka. Trots allt var terminens sammanlagda kursinnehåll rätt lik dagens grundkurs. Det fanns inga deltentor, utan hela terminens kurs tenterades på en tvådagars skriftlig tenta vid terminens slut. De som blev godkända på den skriftliga tentan fick sedan gå upp i en muntlig tenta för professorn själv, Herman Wold.

Första terminen gick lätt, så Jan fortsatte med statistiken en termin till, och det gick också lätt. Bland lärarna fanns en som han minns med särskild värme, Ejnar Lyttkens. Även andra terminen avslutades med en tvådagars skriftlig tenta följd av en munta.

Efter fler ämnen – nationalekonomi, statskunskap och matematik – blev det en tredje termins statistik. På den nivån fanns det ingen som helst undervisning, inte ens någon fastställd litteraturlista. Vid ett snabbt möte med Wold skrev denne för hand ner en lista med litteratur, och uppmanade Jan att komma igen senare för muntlig tentamen. Där hade frågorna till en början ingen som helst anknytning till kurslitteraturen. Men så kom slutligen frågor på några av Wolds egna artiklar, som ingick i kurslitteraturen. Jan hade förstås lusläst artiklarna och Wold blev helt nöjd.

Det var lätt att få undervisning på lösa timmar i Uppsala och Jan hade massor av kurser för olika beteendevetare på Olle Vejdes bok *Hur man räknar statistik*. Några andra unga personer (sedermera välkända i statistiksammanhang) i samma situation var till exempel Bernhard Huitfeldt, Ulf Jorner och Bengt Swensson.

Det var ingen mening att stanna kvar i Uppsala om man inte ville hålla på med Wolds specialområde – ekonometri. Det fanns visserligen några personer som kunde arbeta vid sidan om Herman Wold med andra ämnen, men det verkade kräva enorm självständighet. De som fanns där i skuggan var bland andra K.G. Jöreskog (faktoranalys) och Paul Seeger (variansanalys).

Jan bestämde sig i alla fall för att börja på SCB men på hösten detta år hände en märklig sak.

Under en militär repmånad kom ett motorcykelbud ut i terrängen och meddelade att Jan måste följa med till den provisoriska telefonstationen någonstans i skogen. Det var mycket viktigt. Den som sökte honom var Gunnar Kulldorff. Han erbjöd ett jobb i Umeå. Jan tackade ja på stående fot och började i januari 1967.

Den nystartade institutionen i Umeå var mer lik en modern universitetsinstitution. Bredd uppfattades som något positivt. En grupp som studerade teori för urvalsundersökningar kom igång. Där ingick bl.a. Claes Cassel och Carl-Erik Särndal. Det var en av Jans bästa perioder i den akademiska världen.

Andra rundan

Efter licentiatexamen i Umeå började Jan för andra gången på Statistiska centralbyrån. Året var 1972. Det fanns då en enhet som kallades UI eller utredningsinstitutet. Där var kunskaperna i urvalsfrågor något bristfälliga. Debiteringssystemet var detsamma som nu på SCB. För uppdrag fick man en viss summa pengar och metodarbetet fick därför inte kosta för mycket. Arbetsuppgifterna var till exempel att härleda variansformler. Jan gick så småningom över till enheten för statistiska me-

toder, STM. Man betraktades med misstänksamhet av övriga anställda inom SCB. Det blev mer och mer sammanträden och mindre och mindre nytta. Kanske det är likadant nu på SCB, undrade Jan och tittade på Dan i publiken som blev svarslös.

Efter många år lämnade Jan sålunda SCB och gick till Stockholms universitet. Undervisningen var överraskande lik hur den varit 40 år tidigare i Uppsala. Kanske har datorerna gjort det möjligt att gå igenom mer stoff nu.

Onyttig, men inte ensam

En tillbakablick leder till obehagliga frågor av typen: "Vad har du gjort av ditt liv?" "Har du varit till någon nytta?" "Nej, jag tror knappast det", menade Jan. "Men det skäms jag inte för eftersom jag knappast är ensam om att vara onyttig", sa Jan med en gest ut mot den fullsatta seminarielokalen. "Däremot är jag lite skamsen över att inte ha arbetat så mycket med statistiska tillämpningar utan i stället sett statistiken mer som intellektuell förströelse."

Jan har av någon anledning aldrig fått gå på någon pedagogisk kurs. Numera är det ett krav för universitetslärare. Ouppnåeliga förebilder som föreläsare har varit sådana som K.G. Jöreskog, Jan Lanke, Bengt Rosén och Carl-Erik Särndal. Själv har Jan, med en för honom lika karakteristisk som otrolig undervärdering, aldrig känt sig särskilt lyckad som lärare. Undervisningen på grundnivå idag går mest ut på att förmedla kokbokskunskaper. Man får inte förklara hur saker hänger ihop då kurser och kursmaterial inte är upplagda på det sättet. Ibland har det också känts lite konstigt att undervisa om metoder som av förre institutionschefen ansetts mest lämpade för dumbommar.

Institutionschefen Hans Nyquist tackade Jan för seminariet och livsinsatsen hittills. "Pension är ju bara ett ord för att ens lön betalas ut genom andra administrativa kanaler, och att man får större frihet att arbeta med det man vill", tyckte Hans, möjligen med den inställning man kan ha när man själv har många år kvar.

Min egen första kontakt med Jan var en kurs i urvalsundersökningar som gavs på SCB i samarbete med matematisk statistik i Stockholm. Hälften av oss var matstattare och andra hälften av kursdeltagarna var anställda på SCB. De pratade om "stubbsen" och "mubbsen" utan att inse att de som inte jobbar på SCB kanske inte förstår vad de pratar om. När jag flera år senare började läsa i den bok som numera är internationellt känd som *Gula boken*, hörde jag nästan Jan prata. Det var precis hans stil och uttryck. Gula boken – *Model Assisted Survey Sampling*, av Carl-Erik Särndal, Bengt Swensson och Jan Wretman (1992) – är idag huvudboken i lite mer avancerad urvalsteori i stora delar av världen. Jag tror att en av de böcker jag själv läst mest i, av alla böcker överhuvudtaget, måste vara Gula boken.

En annan mycket välkänd bok som Jan medverkat till är den mer teoretiska *Foundations of Inference in Survey Sampling*, av Claes Cassel, Carl-Erik Särndal och Jan Wretman (1977). Jag har träffat åtskilliga personer som säger att det är den bästa boken som finns i urvalsteori.

Efter seminariet frågade någon Jan varför han inte sa något om sina böcker. Han såg förvånad ut och sa:

– Jag tänkte nog inte på det.

■

Workshop om att undersöka rörliga populationer

av Anders Holmberg, SCB

Den 28 mars arrangerade Surveysektionen en workshop i Stockholm på temat rörliga populationer. Talare var Kjell Wallin från Svensk Naturförvaltning AB och Institutionen för växt och miljövetenskaper vid Göteborg universitet, Maj Eriksson Gothe och Jonas Jonsson SCB samt Åke Pettersson Sveriges Ornitologiska Förening och SCB.

En tydlig slutsats som kan dras från workshopen är att rörliga populationer måste vara en både rolig och tuff utmaning för den som arbetar praktiskt med surveyundersökningar. Någon självklar undersökningsdesign eller manual för att genomföra undersökningar existerar inte, utan befintlig teori måste alltid anpassas till de specifika förutsättningar som gäller för målpopulationen. Detta framkom framför allt av Kjell Wallins presentation. Kjells titel var "Att få en uppfattning av det man inte kan se – stickprovstagning och skattning i frilevande populationer". Varianterna på att räkna djur och växter är många: De kan fångas och märkas och återfångas, man kan fånga och döda individer, man kan studera kända ansamlingar, man kan genomföra flyginventeringar, man kan observera spår och lämningar i form av avtryck, ljud, skadeverkningar och exkrementer och ibland använder man flera av dessa varianter samtidigt. I samband med estimationen så är modeller regel snarare än undantag.

Efter Kjells presentation fortsatte Maj Eriksson Gothe och Jonas Jonsson med att berätta om förutsättningarna och genomförandet av en turistundersökning (IBIS) under rubriken "Att räkna turister – undersökning av inkommande besökare i Sverige". Flera av de metoder som Kjell berättat om för djur och växter går naturligtvis inte att applicera när man ska räkna turister. Redan i definitionen av begreppet turism uppkommer svårighe-

ter med populationsavgränsning som sedan löper vidare till själva mätningsförfarandet. Maj och Jonas berättade om historiken bakom turismundersökningar i Sverige, designen av IBIS-undersökningen (en del av den gjordes i form av gränsstudier och en annan del gjordes vid turistanläggningar), dess datainsamling, kort om skattningsförfarandet samt de erfarenheter som gjordes. Två svårigheter som fanns var dels språket, (trots att intervjuformuläret fanns översatt till flera språk uppstod språkförbistringar) och dels logistik, där framför allt Sveriges landgränser med otaliga gränsövergångar visade sig svåra att hantera.

Tyvärr så är någon uppföljningsundersökning av IBIS, som gjordes mellan 2000 och 2003, ännu inte planerad. Den främsta orsaken är att kostnaderna för den här typen av undersökningar är förhållandevis höga och att det beställarkonsortium som stod bakom IBIS inte längre finns.

Fågelräkningar

Åke Pettersson avslutade workshopen med att beskriva metodik och resultat från fågelinventeringar i Sverige. Titeln på Åkes föredrag var "Att räkna fåglar – Svensk Fågeltaxering". Naturvårdsverket ansvarar för den nationella fågelövervakningen som består av fyra delar:

- Svensk häckfågeltaxering och vinterfågelräkning
- Fångst av flyttande fåglar i Ottenby
- Räkning av dagsträckande fåglar (i Falsterbo)
- (Inter)nationella sjöfågelräkningar

Mycket av informationsinhämtningen har dock genom åren byggt på ett stort inslag av ideell verksamhet.

Med tonvikten lagd på första delen ovan så beskrev Åke historiken och metodiken bakom fågelräkningar i Sverige. Den svenska häckfågeltaxeringen och vinterfågelräkningen startade 1969 och metodiken bestod då av vad man kallar revirkarteringar. 1975 började man med att de som skulle räkna fåglarna (under häckningstid) fritt fick välja tid och plats för rutter som de gick. De fria rutterna blev fler och fler över åren och hade de nackdelarna att de inte var habitat-representativa, de hade ojämn geografisk spridning och vissa rutter kunde försvinna/avslutas om biotopen försämrades. Under 1996 införde man istället standardrutter, där Sverige nu är indelat i 724 rutter med geografisk spridning över hela landet och metodiken är betydligt mera standardiserad. Vidare berättade Åke om verksamheten och statistik från Ottenby fågelstation samt metodiken som används när man räknar dagsträckande fåglar i Falsterbo. Tiden räckte inte riktigt till men han hann även nämna något om de sjöfågelräkningar som görs. Mer information om fågeltaxeringen och fågelräkning allmänt kan man få på, `www.biol.lu.se/zooekologi/birdmonitoring` och/eller `www.sofnet.org`.

Debatter som uppkommer om exempelvis rovdjursbestånd, fågelinfluensan och turismens betydelse för ekonomin visar att informationsbehov finns och snabbt kan aktualiseras. Undertecknad noterade under workshopen att det, trots att det uppenbarligen finns ett förhållandevis stort behov av beslutsunderlag inom dessa områden (djurliv och turism), så verkar en konsekvent ansats och samhällsplan för att genomföra undersökningar inte existera i Sverige. I stället styrs insatserna av tillfälliga, inte sällan lokala/regionala behov och mer eller mindre ideella insatser. Att det finns olika intressen inblandade medför även att införandet av standardiserad metodik försvåras. Vad gäller undersökningar av vilt djurliv tycks föregångslandet i dessa sammanhang vara USA. Där har lagstiftning och myndighetsambitioner medfört att man har mycket större kunskap och systematik i sin statistik. Därmed vet man också mer om tillståndet i sina vilt levande populationer. I Sverige verkar det däremot vara eftersatt område i samhällsplaneringen.

Delar av materialet från workshopen kommer att finnas tillgängligt på Surveysektionens hemsida `www.statistikersamfundet.se/survey`, under Arkiv, Workshopar. ■

Rapport från Marknadsundersökningens dag

av Boris Lorenc, SCB

För andra året i rad har Marknadsundersökningens dag ägt rum, i år den 19:e april. Målgruppen var marknadsundersökare, köpare av deras tjänster och andra allmänt intresserade av kundundersökningar. Ett rekordhögt antal – 140 personer – deltog i evenemanget där syftet var att stärka branschen genom bildning i form av presentation av utvecklingsnyheter och aktiv diskussion om bland annat kvalitet i undersökningar.

En givande föreläsning hölls av Peter Laybourne (Neuroco), vars företag är det första i världen med att använda elektroencefalogram (EEG) i marknadsundersökningssyfte. Enligt honom kan de på ett direkt sätt mäta kunders känslor, motivation och handlingsberedskap (eller brist på dessa) med hjälp av EEG. Hårdvaran har hunnit bli så effektiv att man kan spela in upp till 30 timmars aktivitet i verkliga sammanhang, till exempel i ett köpcentrum, utan att det stör den huvudsakliga aktiviteten. Elektroderna kan gömmas under en mössa eller dylikt och inspelningen sker på en pryl som är lika stor som en VHS kassett. Svårigheten är förstås att tolka det inspelade materialet. Experterna kan till exempel ta fram ett "engagement index" och ett "emotion index". Ett nästa

steg som utlovas är att koppla ihop EEG-ansatsen med ögonavläsning.

Webbundersökningsföretag säger sig ha bekymmer med vissa panelmedlemmar, vilka kategoriserades enligt

- Hyperaktiva (de är med i många paneler) – inte bevisat att det är ett problem, men de är under uppsikt.
- Förfalskande – svåra att upptäcka men fällor för att avslöja dessa bör läggas ut, till exempel Benfords lag.
- Ouppmärksamma – dessa klarar man genom att förbättra frågeblanketter, samt eventuellt exkludera respondenter som visar sig ouppmärksamma upprepade gånger.
- Betingade (de som lär sig hur de ska svara för att få belöning) – samarbete mellan paneler förespråkades.

Vi fick också höra om tillämpning av etnografi inom marknadsundersökningar på läkemedelsområdet, med syftet att bättre förstå kunder och deras behov. Tillvägagångssättet innefattar många nedslag i målpopulationers vardag.

Som åhörare gick det att välja mellan tre 2-timmars pooler med korta presentationer och gruppdiskussioner. Undertecknad valde poolen "Tveksamma undersökningar". Gunilla Ericson, nyhetschef på Dagens Arbete, lade fram orsaker till att statistik i media presenteras på ett relativt okunnigt sätt: löpandebandsprincipen i redaktionsarbetet, många vikarier, journalisters matteskräck. John Wigzell från Storbritannien pratade om hur man säljer bra undersökningar, vilka han definierade som "de undersökningar som vi gör", medan dåliga undersökningar "görs av de andra". Under diskussionen nämndes idén att ta fram mall för en minimal, standardiserad teknisk beskrivning av undersökningar med syfte att därifrån lätt kunna bedöma undersökningskvaliteten.

På utställarsidan kunde man lägga märke till ett flertal webbundersökningsföretag som erbjöd tjänster av sina paneler med villiga respondenter – inte ett ord om inferens nämndes i sammanhanget. Svenska spel var med och arbetade hårt för att respondenter ska belönas med trisslotter.

Dagen avslutades med en paneldebatt om kvalitet i undersökningar. Föreningen Marknadsundersökningens dag, med Henrik Hall i spetsen, stod som organisatörer för denna intressanta dag som hölls på Skansen i Stockholm. ■

Konferens i Genéve: datorintensiva metoder inom finansiell ekonometri

av Suad Elezović, Statistiska institutionen, Umeå universitet

Mellan den 20:e och 23:e april i år ägde *International Workshop on Computational and Financial Econometrics* (CFE07) och *Kick-off meeting of the ERCIM Working Group on "Computing and Statistics"* rum i Genéve, Schweiz. Huvudarrangör var ekonometriska institutionen vid Genéve universitet. Det främsta ändamålet var att främja vidareutveckling av datorintensiva metoder inom ekonomi, finans och statistik i allmänhet. Konferensen startade med registrering av samtliga deltagare, redan på fredagsmorgonen och fortsatte i full fart fram till eftermiddagen på söndagen. Mer än 200 presentationer, 45 reguljära sessioner och omkring 250 deltagare lovade ett lyckat evenemang. Konferensen hade fem huvudföreläsare, John Mulvey, Peter Rousseeuw, Dimitri Bertsekas, Claudio Albanese och James G. MacKinnon. Huvudföreläsningar handlade om stokastisk programmering, algoritmer för robusta multivariata metoder, stora system av linjära ekvationer, operatörsmetoder och bootstrapmetoder. Enligt min uppfattning verkade det som om den sistnämnde, James MacKinnon, Queen's University (Kanada), som talade om sin 25-åriga forskningserfarenhet kring bootstrapmetoder, gav den mest uppskattade föreläsningen under hela konferensen.

På konferensen kunde man möta folk från alla världens hörn men de flesta var nog doktorander från länder runtomkring Schweiz. Från Sverige deltog bara tre doktorander; två från Linköping och jag. Större grupper utanför Europa kom från USA, Japan, Kanada och Kina. Ämnen har varierat från traditionella statistiska metoder till mer inriktade metoder för analys av finansiella och ekonomiska tidserier. Samtliga deltagare erbjöds möjlighet att skicka in sina opublicerade artiklar till tidskriften *Computational Statistics & Data Analysis*. De artiklar som blir antagna kommer att publiceras i en speciell utgåva som handlar just om datorintensiva ekonometriska och statistiska metoder.

I övrigt kan det konstateras att organisatören gjorde ett otroligt bra jobb med hänsyn till antal deltagare och omfattning av konferensen. Redan från registreringsögonblicket fick man utförlig information i form av ett häfte med kartor, turistbroschyrer, ritning av konferenslokaler med föreläsningsschema och dylikt. Fanns lite tid över så kunde man njuta av staden med otaliga mysiga kaféer, restauranger och en hel del turistattraktioner kring Genéve sjön. Addera sommarliknande väder i april så blev upplevelsen fullkomlig.

På hemsidan `www.csdassn.org/europe/CFE07/` finns det mer utförliga information om konferensen och relaterade ämnen. ■

Hänt och hört

Disputationer

JIMMY OLSSON, Matematisk statistik, Lunds universitet, försvarade den 26 januari sin doktorsavhandling *On Bounds and Asymptotics of Sequential Monte Carlo Methods for Filtering, Smoothing, and Maximum Likelihood Estimation in State Space Models*. Fakultetsopponent var professor Ya'acov Ritov, The Hebrew University of Jerusalem, Jerusalem, Israel.

AZRA KURBASIC, Matematisk statistik, Lunds universitet, försvarade den 2 februari sin doktorsavhandling *Topics in human gene mapping*. Fakultetsopponent var professor Dr. Thomas F. Wienker, Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn.

ANNA BORNEFALK HERMANSSON, Institutionen för informationsvetenskap, Uppsala universitet, försvarade den 16 februari sin doktorsavhandling *Resampling Evaluation of Signal Detection and Classification: With Special Reference to Breast Cancer, Computer-Aided Detection and the Free-Response Approach* . Fakultetsopponent var professor Dankmar Böhning, Applied Statistics, School of Biological Sciences, University of Reading.

ANDREAS NORDVALL LAGERÅS, Matematisk statistik vid Stockholms universitet, försvarade den 14 mars sin doktorsavhandling med titeln *Markov Chains, Renewal, Branching and Coalescent Processes: Four Topics in Probability Theory*. Fakultetsopponent var professor Ingemar Kaj från Uppsala.

OLLE ERIKSSON, Matematiska institutionen, Linköpings universitet, försvarade den 16 mars sin doktorsavhandling *Sensitivity and Uncertainty Analysis Methods, with Applications to a Road Traffic Emission Mode* Fakultetsopponent var professor Rolf Sundberg, Stockholms universitet.

LARS ÄNGQUIST, Matematikcentrum, Avdelningen för matematisk statistik, Lunds universitet, försvarade den 16 mars sin doktorsavhandling *Pointwise and Genomewide Significance Calculations in Gene Mapping through Nonparametric Linkage Analysis: Theory, Algorithms and Applications*. Fakultetsopponent var docent Staffan Nilsson, Matematiska vetenskaper, Chalmers Tekniska Högskola, Göteborg.

SARA LARSSON, Matematisk statistik, Lunds universitet, försvarade den 13 april sin doktorsavhandling *Statistical Modelling of Cell Cycle Dynamics* . Fakultetsopponent var professor Olle Nerman, Matematiska vetenskaper, Chalmers Tekniska Högskola, Göteborg.

SOFIA ÅBERG, Matematisk statistik, Lunds universitet, försvarade den 25 april sin doktorsavhandling *Applications of Rice's formula in oceanographic and environmental problems*. Fakultetsopponent var professor Jean Marc Azqaïs, Laboratory of Statistics and Probability, University Paul Sabatier, Toulose.

ERIK WALLGREN, Institutionen för informationsvetenskap, Uppsala universitet, försvarade den 27 april sin doktorsavhandling *Essays on Capability Indices for Autocorrelated Data*. Fakultetsopponent var professor Jan Wretman, Stockholms universitet.

NIKLAS NORÉN, Matematiska institutionen, Stockholms universitet, försvarade den 7 maj sin doktorsavhandling *Statistical methods for knowledge discovery in adverse drug reaction surveillance*. Fakultetsopponent var professor Heikki Mannila, Helsinki University of Technology, Helsingfors.

INGRID SVENSSON vid Institutionen för matematik och matematisk statistik, Umeå universitet, försvarade den 16 maj sin doktorsavhandling *Estimation of Wood Fibre Length Distributions from Censored Mixture Data*. Fakultetsopponent var docent Aila Särkkä, Chalmers Tekniska Högskola, Göteborg.

Licentiatseminarier

MARTIN OHLSON, Tekniska högskolan vid Linköpings universitet, presenterade den 21 februari 2007 sin licentiatavhandling *Testing Spatial Independence using a Separable Covariance Matrix*. Opponent var professor Dietrich von Rosen Institutionen för biometri och teknik, SLU/Ultuna.

IRIS J. Y. WANG, Högskolan Dalarna, presenterade den 29 mars sin licentiatavhandling *Non-parametric statistics in the context of program evaluation*. Opponent var Fil. dr. Andreas Karlsson, Centrum för Klinisk Forskning, Västerås. Seminariet hölls vid Ekonomikum, Uppsala universitet.

Utmärkelse till Karl Gustav Jöreskog

American Psychological Association (APA) har utsett Karl Gustav Jöreskog, professor emeritus i statistik vid Uppsala universiet, till en av två pristagare till *Distinguished Scientific Award for the Applications of Psychology*. Professor Jöreskog delar priset tillsammans med professor Peter M. Bentler, Psychology and Statistics, University of California, Los Angeles.

Jöreskog och Bentler får priset för deras bidrag inom det psykometriska området där de varit med om att utveckla strukturella ekvationsmodeller, *Structural Equation Modeling*, (SEM).

Under 60- och 70-talen arbetade Jöreskog med modeller, procedurer och datorprogrammering (LISBEL) för att kunna använda observationsdata till att testa pskykologiska teorier.

Mer information finns på hemsidan `www.apa.org/science/psa/apr07dis.html`.

Avsändare/Sender:
Svenska statistikersamfundet
c/o Marie Wiberg
Statistiska institutionen, Umeå universitet
901 87 Umeå

B

Sverige
Porto betalt
Port payé

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**
25–29 juni	Workshop: New direction in Monte Carlo methods	Fleurance, Frankrike	`www.adapmc07.enst.fr/`
1 juli	Case Study Workshop on Reliability	Glasgow, Skottland	`www.enbis.org/`
1–4 juli	Mathematical Methods in Reliability: Methodology and Practice	Glasgow, Skottland	`www.mmr2007.org/`
4–6 juli	Workshop on statistical metadata	Vienna, Österrike	`www.unece.org/stats/documents/2007.07.metis.htm`
9-11 juli	MCP 2007 Vienna	Vienna, Österrike	`www.mcp-conference.org/`
9–11 juli	14th Applied Probability INFORMS Conference	Eindhoven, Holland	`appliedprob.society.informs.org/`
16–20 juli	ICIAM 07	Zurich, Schweiz	`www.iciam07.ch/`
16–20 juli	RSS International Conference - Statistics & Public Policy-Making	University of York, Storbritannien	`www.rss.org.uk/rss2007`
29 juli–2 aug	The 28th Annual Conference of the International Society for Clinical Biostatistics	Alexandroupoli, Grekland	`www.iscb2007.gr/`
16–20 aug	TIES 2007, 18th annual meeting of the International Environmetrics Society.	Mikulov, Tjeckien	`www.math.muni.cz/ties2007/`
18–20 aug	International Symposium On Business and Industrial Statistics 2007	Azorerna, Portugal	`www.isbis2007.uac.pt/`
22–29 aug	International Statistical Institute, 56th Biennial Session	Lissabon, Portugal	`www.isi2007.com.pt/isi2007/`
30 aug–1 sep	International Conference on Statistics for Data Mining, Learning and Knowledge Extraction	Aveiro, Portugal	`www2.mat.ua.pt/iasc07/`
24–26 sep	7th annual conference of ENBIS	Dortmund, Tyskland	`www.enbis.org/`
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/`

ENBIS: European Network for Business and Industrial Statistics
MCP: Multiple Comparison Procedures
INFORMS: The Institute for Operations Research and the Management Sciences
ICIAM: International Council for Industrial and Applied Mathematics

I nästa nummer

Bidrag på temat *mikrosimulering*

ISSN 0283 - 3654