

Qvintensen

SVENSKA STATISTIKFRÄMJANDET

1

Innehåll

Ordföranden har ordet	3
Redaktören har ordet	3
Styrelsen informerar	4
Föreningen industriell statistiks styrelse	5
Surveyföreningens styrelse	7
Främjandets höstmöte	9
Jubileum för den officiella statistiken	12
Student's t 100 år	16
Debatt: Statistiker behöver utbildning i kommunikation	19
Medlemstombolan	20
Visualisering på Internet	21
Rapporter från konferenser och seminarier	23
Hänt och hört	27
Teoretiska artiklar	28
Konferenskalendarium	32

Ansvarig utgivare Bernhard Huitfeldt
Redaktör Joakim Malmdin 090-786 55 30
Biträdande redaktör Helen Lindkvist 019-17 62 63
Redaktion Magnus Arnér, Industriell statistik
Alf Fyhrlund, Främjandet
Anders Holmberg, Surveyföreningen
Marie Linder, FMS
Johan Lyhagen, Främjandet

Adress Qvintensen
c/o Joakim Malmdin
Statistiska institutionen
Umeå universitet
901 87 Umeå
E-post qvintensen@gmail.com

Annonspriser

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

Annonser bokas med redaktören. *Annonssutskick per e-post* bokas med Främjandets sekreterare. Priset är 5 000 kronor för institutionsmedlem eller motsvarande, och 7 500 kronor för övriga annonsörer. På alla priser tillkommer 25 % moms.

Svenska statistikfrämjandet

Ordförande	Bernhard Huitfeldt	08-552 456 34
Vice ordförande	Åsa Greijer	018- 18 22 49
Sekreterare	Jesper Rydén	018-471 32 88
Skattmästare	Annika Lindblom	019- 17 60 86
Klubbmästare	Staffan Lindqvist	070-200 21 43
Surveyföreningen	Anders Holmberg	019- 17 69 05
FMS	Mats Rudholm	031-703 73 72
Industriell statistik	Ingemar Sjöström	08-553 835 68
Övrig ledamot	Xavier de Luna	090-786 55 59
Övrig ledamot	Christina Kvarnström	08- 30 33 17

Adress Svenska statistikfrämjandet
Box 480
751 06 Uppsala
E-post sekfram@gmail.com
Webbplats www1.internetplatsen.se/statistiker

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

Svenska statistikfrämjandet äger upphovsrätten till publicerat material då det är framställt av redaktionsmedlem, styrelsemedlem, någon till Främjandet knuten förenings styrelse eller person som erhållit arvode för sitt bidrag.
© Svenska statistikfrämjandet 2008.

Ordföranden har ordet

Det har nu gått åtta månader sedan Svenska statistikfrämjandet bildades. Vi i styrelsen har lärt känna varandra riktigt bra och hunnit bli ganska varma i kläderna. Antalet personer som aktivt tagit ställning till medlemskap i den nya föreningen ligger fortfarande klart under det antal som utgörs av de tidigare medlemmarna i Samfundet respektive Statistiska föreningen. Detta oroar oss lite. Vi håller på att kartlägga varifrån bortfallet kommer, om det har administrativa orsaker eller om tidigare medlemmar av någon anledning inte vill vara med i den nya föreningen. Vi kommer också att satsa på rekrytering av helt nya medlemmar. Jag är klart medveten om att medlemskap i vår förening är helt beroende av om vi kan tillgodose de intressen som man förväntar sig. Vi hoppas att kunna göra det genom våra olika aktiviteter, som jag kortfattat beskriver nedan.

Att regelbundet kunna sätta en tryckt tidskrift med intressant läsning i händerna på våra medlemmar är en av våra viktigaste uppgifter. Detta vill vi prioritera. Vår nuvarande redaktör Joakim Malmdin kommer att avsluta sitt uppdrag i och med detta nummer av Qvintensen för att slutföra sin forskarutbildning. Vi håller nu på att rekrytera en ny redaktör bland våra medlemmar för att säkerställa en fortsatt utgivning. För att hålla kontakten med medlemmarna lite oftare skickar vi också ut nyhetsbrev. Det senaste kom i början av december.

Vår höstkongress hölls i Stockholm den 6 november med temat hälsosam statistik. Det blev en mycket intressant kongress som väl belyste hur statistik samverkar med olika medicinska sakområden för att underlätta komplicerade beslut. Ett fylligare referat från kongressen finns publicerad på annan plats i detta nummer av Qvintensen. Presentationerna kan man återfinna på vår webbplats. Vår klubbkommitté är nu på plats med Staffan Lindqvist som ordförande. Vi räknar med att komma igång med regelbundna medlemsmöten kring aktuella statistiska frågor i mitten av januari.

Som också framgår på annan plats finns det ett initiativ att bilda en ny statistisk förening med inriktning mot forskning och utbildning. Tanken är att den skall kunna bli antagen som ny sektion inom Främjandet. Detta är glädjande och kommer att intressera många medlemmar, inte bara de som har sin verksamhet vid universitet och högskolor. Vi välkomnar andra initiativ att bilda nya sektioner inom Främjandets ram och vi kan ge både administrativt och ekonomiskt stöd till detta från moderföreningen.

Ett helt nytt initiativ är att inrätta en årlig utmärkelse kallad Årets statistikfrämjare. På annan plats finns ett upprop till nominering av kandidater till denna utmärkelse. Jag hoppas att vi får in många förslag och att kriterierna skall återspegla den nya föreningens inriktning.

Hoppas att långhelgen har blivit vederkvickande på alla sätt och jag ser fram emot ett intressant år med Främjandet. Väl mött på våra aktiviteter.

Bernhard Huitfeldt

Redaktören har ordet

Debatten som under året har förts i media angående tillsatser i vår mat har satt fingret på en för oss statistiker välbekant problematik: inferens och kausalitet. Hur gör vi för att dra rimliga slutsatser utifrån en observationsstudie? På senare tid har forskare inom det medicinska området tagit bladet från munnen och uttryckt en farhåga för att det inte kan uteslutas att användningen av vissa tillsatser kan ligga till grund för den explosionsartade ökning av allergier vi sett de senaste decennierna. Men hur bevisa det utan kontrollerade experiment? Idag, den 14 december, publiceras en debattartikel i DN till tillsatsernas försvar. Problematiken känns igen från den långdragna, och stundtals infekterade, debatten om rökning och lungcancer. Med den skillnaden att här rör det sig om något som är många gånger mer komplext.

Av alla tidningsartiklar jag läst i ämnet har inte en enda statistiker uttalat sig, och förmodligen inte heller blivit tillfrågad. Det här är ett typexempel på där vår förening skulle kunna bidra med ett konsultationsråd i likhet med hur den engelska motsvarigheten till Främjandet aktivt deltar i samhällsdebatten i frågor som rör tolkning av statistik. Rådet skulle också kunna bidra med vissa semantiska klargöranden. Vad betyder det till exempel att något är "vetenskapligt bevisat"?

Jag drar mig till minnes att jag för flera decennier sedan läste en bok om hur vissa former av cancer ökat lavinartat i Västeuropa sedan andra världskrigets slut, medan samma cancerformer i stora delar av Afrika var helt okända. Författaren pekade på skillnaden i matvanor. I de delar av Afrika där man inte upptäckt ett enda fall av tjocktarmscancer åt befolkningen inte raffinerat vitt mjöl och sex gånger mer fibrer än vad européer gjorde. Rådet som gavs var att man skulle äta mer av frukt och grönsaker och bröd innehållandes mycket fibrer. Men var det vetenskapligt bevisat?

När jag nu söker på Internet är råden inte längre lika entydiga. En läkare säger till en tidning att det har visats att det inte finns något samband mellan fiberintag och tjocktarmscancer. På en annan sida läser jag att ett lägre pH-värde i tjocktarmen, vilket man funnit att människor i Afrika har, leder till att man inte utvecklar cancer i tjocktarmen. I samma text står det att ett större fiberintag sänker pH-värdet. När ska man överge beviskraven och tillämpa försiktighetsprincipen?

I detta nummer bjuds vi på flera intressanta artiklar. Vi har referat från Främjandets höstkongress och SCB:s jubileumsdag, en artikel om den officiella statistikens 150-åriga historia och en som uppmärksammar att det var 100 år sedan William Sealy Gosset publicerade sin banbrytande "The probable error of the mean" i Biometrika. Därtill bjuds vi på artiklar om visualiseringsverktyg av statistik på Internet, en intervju med ännu en slumpvald medlem, rapporter och avhandlingsreferat.

Avslutningsvis vill jag rikta ett stort tack till alla bidragsgivare.

Joakim Malmdin

Styrelsen informerar

Mötesverksamhet i form av lunch med anslutande föredrag var ett uppskattat inslag i Statistiska föreningen, och ambitionen är att fortsätta med sådan verksamhet i Svenska statistikfrämjandet. En klubbkommitté, under ledning av adjungerad klubbmästare Staffan Lindqvist, har utsetts och ett första lunchmöte väntas äga rum i januari. Mötena är avgiftsfria för medlemmar i Svenska statistikfrämjandet.

Planeringsarbete inför årsmötet i Uppsala den 26 mars och sektionmötena den 27 mars har utförts. Lokaler har bokats på universitetsområdet Polacksbacken och en programkommitté arbetar med planeringen av årsmötet. Arbetet med verksamhetsberättelse och budget, vilka presenteras vid årsmötet, inleds efter nyår.

Styrelsen har beslutat att introducera en ny utmärkelse: *Årets statistikfrämjare*, en person som uppmärksammas för insatser under det föregående året. Kriterier för val av Årets statistikfrämjare och därtill hörande rutiner har formulerats av en arbetsgrupp.

Policy för e-post

Styrelsen har tagit beslut om följande policy för spridning av information via e-post:

- *Information från Främjandet inklusive kurser, konferenser och möten där Främjandet eller någon av föreningarna är huvudarrangör* distribueras gratis till alla.
- *Information om kurser, konferenser och möten inom vårt verksamhetsområde* distribueras till den som önskar sådan information till en administrativ kostnad av 1 000 kr (exklusive moms) för arrangören.
- *Jobbannonser* distribueras till dem som vill ha sådan information till en kostnad av 5 000 kr (exklusive moms) för medlemmar respektive 7 500 kr (exklusive moms) för icke-medlemmar.

Som allmän policy gäller att vara försiktig med reklam för produkter. En länk i ett nyhetsbrev kan vara acceptabelt. Styrelsen beslutade vidare att inte längre sälja adressetiketter baserade på medlemsregistret.

Styrelsen har också diskuterat framtidsfrågor, till exempel yrkesetisk kod, medlemstidning, utbildningsfrågor och eventuell ny förening, ekonomi och medlemsrekrytering, grafisk profil och webbplats, samt utåtriktad verksamhet.

Styrelsen genom Jesper Rydén, sekreterare

Nominering till Årets statistikfrämjare

Styrelsen har beslutat inrätta en årlig utmärkelse: *Årets statistikfrämjare*. Till Årets statistikfrämjare utses en eller flera personer bosatta i Sverige som

... gjort en slagkraftig insats på statistikområdet under det senaste året som givit publik uppmärksamhet. Insatsen kan utgöras av en innovativ metod, en smart analys, en effektiv presentation eller att på ett framgångsrikt sätt ha utmanat en etablerad sanning. Insatsen kan ligga inom det samhällsrelaterade, kommersiella eller vetenskapliga området.

Personen behöver inte nödvändigtvis vara medlem i Svenska statistikfrämjandet.

Belöningen består av ett prisföremål och ett diplom. Den utvalde blir inbjuden att hålla ett föredrag på årsmötet, blir presenterad i Qvintensen, och får själv skriva en artikel om sin insats i Qvintensen.

Vi inbjuder härmed medlemmarna att ge förslag på en eller flera personer som du tycker uppfyller ovanstående kriterier. Skicka förslaget med motivering till bernhard.huitfeldt@astrazeneca.com senast den 15 januari 2009. Utdelningen av priset äger rum första gången vid årsmötet den 26 mars 2009.

Föreningen industriell statistiks styrelse

Föreningen industriell statistik, tidigare Sektionen för industriell statistik inom Svenska statistikersamfundet, bildades i september 2000. Vi har för närvarande cirka 50 medlemmar och verkar för en spridning av statistiska metoder inom industrin, stimulerar utbyte av yrkeskunskaper mellan föreningens medlemmar, utgör ett forum för diskussioner samt främjar forskning och utbildning inom föreningens verksamhetsområde. Föreningen arrangerar två seminarier per år, ett i samband med årsmötet på våren och ett höstseminarium, och ger också kurser till personer verksamma inom industrin. Information om vad vi gjort hittills finns på www.matstat.umu.se/indstat/. En ny webbplats är under uppbyggnad.



Kerstin Vännman, ordförande

Jag är sedan 2004 professor i matematisk statistik med inriktning industriell statistik vid Luleå tekniska universitet. Jag har varit verksam i Luleå sedan 1975, då jag blev universitetslektor vid dåvarande Luleå tekniska högskola. Jag blev docent 1996 och biträdande professor 1998. Min nuvarande forskning är inom industriell statistik, speciellt duglighetsanalys och statistisk processtyrning.

Jag har också engagerat mig mycket i utbildningsfrågor. I många år var jag verksam i Utbildningskommittén inom Svenska statistikersamfundet, i International Statistical Institute Education Committee och International Association for Statistical Education. Jag har också suttit som internationell representant i styrelsen för American Statistical Association.

Thomas Svensson, sekreterare

Jag har under halvannat decennium arbetat med tillämpning av matematisk statistik inom området mekanisk utmattningshållfasthet och disputerade i ämnet 1996 på Chalmers. Under större delen av denna tid har jag varit anställd på Sveriges tekniska forskningsinstitut i Borås, under några år avbrutet av en tjänst hos Fraunhofer-Chalmers centrum (FCC) för industrimatematik i Göteborg.

För närvarande är jag mest engagerad i tillförlitlighet av mekaniska strukturer och arbetar då främst med industriella problem i nära samarbete med FCC och Matematisk statistik på Chalmers. Ett annat stort intresseområde är modellering av mätosäkerhet. Inom detta område ger jag kurser för industriverksamma ingenjörer och har också skrivit en handbok i ämnet.

Lennart Nilsson, kassör

Jag är en av dessa som aldrig gått ut ur skolan utan fortfarande är verksam inom Umeå universitet sedan 36 år tillbaka, nu som universitetslektor och utbildningsledare för Tekniska högskolan. Under resans gång har jag provat på många olika uppdrag. De flesta av akademisk

eller administrativ karaktär, men även några roliga som till exempel sekreterare i Svenska statistikersamfundet. Under den perioden startade vi sektionen för industriell statistik för vilken jag varit kassör hela tiden.

Magnus Arnér

Jag är sedan ett år verksam vid Tetra Pak i Lund där jag främst arbetar med robust konstruktion och försöksplanering. Jag gick min utbildning i Linköping och har tidigare arbetat i bilindustrin.

Bo Bergman

Jag är innehavare av SKF-professuren i kvalitetsutveckling vid Chalmers tekniska högskola i Göteborg. Dessförinnan var jag under femton år professor i kvalitets teknik vid Tekniska högskolan i Linköping och ännu tidigare adjungerad professor i tillförlitlighetsteknik vid KTH. Jag arbetade 15 år vid Saabs flygdivision som tillförlitlighetsexpert och internkonsult avseende statistisk problemlösningsmetodik. Under denna tid disputerade jag i matematisk statistik vid Lunds universitet. Jag har en lång publikationslista inkluderande läroböcker inom kvalitetsområdet och har handlett ett stort antal doktorsavhandlingar och licentiatavhandlingar. Jag är medlem i bland annat International Statistical Institute och i International Academy for Quality.

Ann Lindqvist

Jag har jobbat på Scania i sju år och har då varit verksam inom två analysområden, nämligen fordonsreparationer och variabler loggade under fordonets drift. I båda fallen är målet att huvudsakligen skapa underlag för beslut om konstruktionsförändringar. Innan det jobbade jag på Partek (Rockwool), men inte som statistiker.

Min statistikutbildning skaffade jag mig sent i livet, under förra lågkonjunkturen för byggindustrin (början av 90-talet). Jag följde utbildningsprogrammet Computational Statistics på Skövde högskola parallellt med mitt arbete. Det var både roligt och givande att plugga och arbeta samtidigt då jag kunde använda mina förvärvade kunskaper direkt.

Magnus Pettersson

Jag arbetar sedan drygt tio år på Statistikkonsulterna i Göteborg med statistiska analyser och projektledning. Jag utbildade mig på matematikerlinjen i Stockholm och skuttade över till Göteborg för forskarutbildning i biostatistik som jag avslutade med en licentiatexamen 1998. Främst är jag intresserad av området industriell statistik.

Jag har intresserat mig en hel del för Six Sigma, som jag menar har haft det goda med sig att paketera och marknadsföra statistik bland tillämparna i näringslivet, speciellt ingenjörer och ekonomer. Pedagogik tycker jag är viktigt eftersom en statistikers arbete ofta handlar om att förstå och göra sig förstådd.

Ingemar Sjöström

Att arbeta världsomspännande som industristatistiker på Sony Ericsson är fascinerande. Man blir en "jack-of-all-trades-master-of-none" genom att umgås med välutbildade ingenjörer inom till exempel antennteknik, display, kamera, mekanik, radiosignalering och mjukvara. Man lär sig också om vad operatörerna och kunderna på olika marknader vill ha och behöver.

Alltifrån starkare högtalare i larmande städer till mättade färger för kunder i Asien. De problem och utvecklingsprojekt man blir inblandad i är av skiftande karaktär och svårighetsgrad. Jag har ringa kontakt med fabriker, men vi tillverkar ofta förserier och då är utvecklingsavdelningarna involverade. □

Reflektioner av en industrianställd civilingenjör

Svensk industri har, precis som industrin i övriga världen, under de senaste åren genom ökande konkurrens och mer krävande och medvetna kunder insett det allt större behovet av fokus på kvalitet i det man utvecklar och producerar. För att möta de ökade kraven har man i många företag breddat sina sökningar efter hjälpmedel, dels när det gäller arbetsmetoder eller processer som "requirements management" och robust konstruktion, dels när det gäller verktyg för att stöda dessa som Six Sigma med mera.

Det jag lärt under mina första fyra år på Tetra Pak är att det inte är lätt att komma som nyutexaminerad ingenjör med strikt teoretisk bakgrund ut i den hetsiga industrin. Ofta blir man den person som hamnar i centrum för de teoretiska delarna av arbetet, som till exempel att definiera och leda olika statistiska arbeten och analyser. Vidare är min erfarenhet från Teknisk fysik att de kunskaper man får med sig inom statistik från högskolan på många sätt är otillräckliga. Innehållet i sig är förvisso bra men man skulle behöva mycket mer för att ha en chans att utan extra kompletteringar på egen hand kunna hantera de situationer man hamnar i. Typiska tillämpningar där jag känt att mina högskolekunskaper är bristfälliga är statistisk processtyrning, tillförlitlighetsteori samt allmänkunskapen om våra vanligaste fördelningar, när de uppkommer och vad de visar.

En del i detta är också det att man på universitet och högskola träffar på många olika ämnen som man ska tillskansa sig kunskap inom, där man som student ofta saknar vetskap om vilka områden som är speciellt viktiga för olika typer av industrier eller forskningsområden. Ett sätt att styra detta kan vara att låta ämnenas relevans motsvara den troliga användningsgraden i större utsträckning än vad de gör idag. Istället är denna relation snarare speglad av svårighetsgraden mellan ämne-

na. Under min civilingenjörsutbildning läste jag endast 5 poäng statistik, vilket inte alls står i proportion till hur mycket jag använder statistik i arbetslivet. Nu är jag naturligtvis bara en person med en åsikt och jag är övertygad om att för många andra väger prioriteringen över åt något annat håll. Faktum kvarstår dock att statistik är på väg mot en mer framträdande roll i industrin och det vore då lämpligt att låta detta spegla också kursutbudet, både det obligatoriska och det valfria, på våra universitet och högskolor.

Det kan även tolkas som ett mer generellt problem att kursutbuden speglas mer av hur det varit historiskt än av det behov som finns idag. I industrin är det livsavgörande hur väl man lyckas fånga in det kunderna uttrycker önskan för och detta borde gälla även högre utbildningar.

Hur som helst ses statistik av industrin som något fint, svårt och åtråvärt att kunna. Ibland speglar sig detta i att förväntningarna på vad statistik kan tillföra tar sig orealistiska proportioner. Jag har förstått att det är viktigt att belysa begränsningarna och framhålla statistik endast som ett verktyg för att få hjälp att fatta rätt beslut. Fatta besluten som sådana får man fortfarande själv göra. Att utföra de nödvändiga beräkningarna är inte svårt och det finns hjälpmedel för detta. Svårigheterna ligger i hur man ska angripa problemet, då det finns flera olika metoder som inte sällan ger olika resultat. Dessutom står man inför ett svårt beslut när analysen är genomförd. Ibland är det till och med svårare att ta än vad det skulle ha varit om analysen inte gjorts. Det kan vara lättare att göra en visuell bedömning utifrån två grafer med rådata än att avgöra om ett visst p-värde är tillräckligt litet för att kunna förkasta en nollhypotes samtidigt som man ska bedöma om detta i så fall har någon praktisk signifikans.

Fredrik Mattsson, Tetra Pak

Surveyföreningens styrelse

Föreningen är ett forum för diskussioner, stimulerar till utbyte av yrkeskunskaper och främjar forskning, praktisk metodutveckling och utbildning på området. Vi verkar för en förbättrad användning av statistiska data, surveyundersökningar och metoder inom samhällsplanering, vetenskap, debatt och i samhället i övrigt. Detta sker genom anordnande av föredrag, arbetsseminarier och temadagar, genom kontakter, diskussioner och informationsutbyte på webbplatsen och i den gemensamma medlemstidningen Qvintensen.



Jörgen Svensson, ordförande

Jag är elva "år" (född den 29 februari), bor i Örebro och är ordförande i Surveyföreningen sedan oktober 2006.

Till vardags arbetar jag på Statistiska centralbyrån som processansvarig för ett av fem processområden: Bearbeta och analysera. Tjänsten innebär ett slags övergripande ansvar för bland annat kodning, mikro- och outputgranskning, imputering, skattning, statistisk analys, röjandekontroll och redovisningsprinciper.

Mina största intressen som statistiker har varit inom surveyområdet, främst urvalsfrågor, bortfallsproblematik och kvalitetsaspekter. Jag vill så gott jag kan förankra god metodik i det praktiska undersökningsarbetet och önskar att vi statistiker ska bli bättre på att presentera och kommunicera kring våra metoder och resultat.

Sofia Ekvall, sekreterare

Jag är en ung östgötska som har läst statistikprogrammet med företagsekonomi som inriktning vid Linköpings universitet. Sedan drygt fyra år tillbaka jobbar jag som projektledare på undersökningsföretaget Norstat. Jag håller mestadels på med telefon- och webbpanelundersökningar och arbetar rent praktiskt med metoderna, men även med teoretiska frågor som de olika tillvägagångssätten medför. Det handlar bland annat om representativitet i urvalsramar, viktningar och rena metodjämförelser.

Arbetet i Surveyföreningen känns spännande och jag hoppas kunna bidra med lite inspel från den privata företagsvärlden.

Anton Grafström, kassör

Jag är doktorand i matematisk statistik vid Umeå universitet. Min forskning handlar om olika metoder för att ta ett stickprov då enheterna i populationen har olika stor sannolikhet att tas med. Jag är också intresserad av visualisering som ett verktyg för att förmedla statistik och statistiska metoder till både statistiker och icke-statistiker.

En stor del av fritiden går åt till att jobba på mitt och min frus nyinköpta hus och på garaget som vi bygger själva.

Karin Dahmström

Jag är nybliven pensionär efter en lång tid som universitetslektor i statistik vid Stockholms universitet. Jag disputerade 1980 på en avhandling om binär regression och sedan har jag i min forskning också ägnat mig åt grafteori och något om urvalsmetoder. Sedan början av 1980-talet har jag ägnat mycken tid åt att utveckla olika former av kursmaterial.

Hösten 2006 blev jag ledamot i styrelsen för Surveyföreningen. Här har jag särskilt engagerat mig i arbetet med att utse vinnaren av den bästa uppsatsen inom surveyområdet. Jag har också varit medlem i en bortfallskommitté för utformande av en standard för bortfallsredovisning.

Anders Holmberg

De senaste åtta åren har jag arbetat med många olika delar förknippade med statistikproduktion på Statistiska centralbyråns utvecklingsavdelning i Örebro. Mitt huvudområde är teori och metodik för surveyundersökningar och min doktorsavhandling handlade om flervariabelproblemet. Några intressanta praktiska problem jag arbetat med är undersökningar med blandade insamlingsmetoder, varians estimation för orderstatistikor och uppsnabbning av ekonomisk statistik.

På senare tid har jag även ägnat mig åt metodfrågor i registerstatistiska undersökningar. Jag har varit med i Surveyföreningens styrelse sedan 2005 och sitter även i Statistikfrämjandets styrelse.

David Sundén

Sedan 2005 har jag varit ledamot i Surveyföreningens styrelse. Jag arbetar idag som enhetschef på Statisticon och ansvarar för företagets verksamhet som omfattar samhällsrelaterad statistik. En stor del av arbetet är riktat mot den offentliga sektorn och större företag. Jag är disputerad nationalekonom från Handelshögskolan och har forskat om den offentliga sektorn och pensionssystem.

I Surveyföreningens styrelsen ingår också Niklas Lundvall. □

Surveyföreningen

Föreningens vårprogram är i skrivande stund ännu inte fastlagt, men vi planerar för ett seminarium om uppgiftslämnarfrågor fredagen den 6 februari i Stockholm. De medverkande kommer från Sveriges Kommuner och Landsting, Näringslivets regelnämnd och Statistiska centralbyrån. På vårt årsmöte i mars blir det ett blandat program med bland annat Tore Dalenius-föredrag och utdelning av uppsatspriset.

Inspirera till uppsatser

Kom ihåg att inspirera till och anmäla uppsatser inom surveyområdet till tävlingen om bästa uppsats! Uppsatsen ska vara framlagd mellan den 1 september 2007 och den 25 januari 2009. Det senare datumet är sista dag för nominering. Prissumman är 5 000 kr, och syftet med tävlingen är att stimulera intresset för samhällsstatistisk

undersökningsmetodik bland studenter och handledare i grundutbildningen.

Om du som är medlem inte redan har gjort det, betala in medlemsavgiften à 50 kr för verksamhetsåret 1 september 2007–31 december 2008. Både personer och organisationer är välkomna som medlemmar i Surveyföreningen. Läs om fördelarna med medlemskap på föreningens webbplats www.statistikersamfundet.se/survey/. Där går det också att anmäla intresse för medlemskap via ett webbformulär. Det går även bra att kontakta föreningens sekreterare Sofia Ekvall, sekreteraren.survey@statistikersamfundet.se, 013-489 40 23.

Styrelsen genom Jörgen Svensson, ordförande

Bildande av en förening för forskning och utbildning

Då Svenska statistikfrämjandet bildades som en samslagning av Svenska statistikersamfundet och Statistiska föreningen gick en uppmaning ut till medlemmarna att, vid behov, bilda nya föreningar. Några av oss har tagit fasta på det och börjat diskutera inrättandet av en förening för forskning och utbildning. Föreningen skulle, precis som FMS med flera, fungera som en sektion under Främjandet. Vi tycker att det finns flera anledningar till att en sådan kan behövas.

Av naturliga skäl är det inte lika stort fokus på den universitetsrelaterade delen av vår verksamhet inom Främjandet jämfört med hur det var inom Samfundet. De flesta av Samfundets styrelsemedlemmar var verksamma inom universitetsvärlden, medan endast två av Främjandets styrelsemedlemmar är anställda vid ett universitet. Det kan därför behövas en ny förening under Främjandet som lyfter fram forsknings- och utbildningsfrågor, på samma sätt som Samfundet förut gjorde. Föreningen skulle kunna ansvara för en del av de universitetsrelaterade uppgifter som Samfundet tidigare hade, till exempel att arrangera sommarskolor, delta i arrangemanget av Nordiska konferensen i matematisk statistik (Nordstat) och att vara med och driva *Scandinavian Journal of Statistics*. Detta kunde utökas med doktorandaktiviteter (några slags doktoranddag?) och att hålla i de ämnes-

konferenser i statistik och matematisk statistik som till och från har förekommit och varit mycket uppskattade. Man kan även försöka ha en viss kontroll över arbetsmarknaden för statistiker, till exempel i form av alumniverksamhet och kontaktdagar med arbetsgivare utanför universitetsvärlden. Vidare bör föreningen bevaka utbildningsfrågor, och den kan eventuellt ta rollen av utbildningskommitté.

En arbetsgrupp har börjat titta på frågan med målet att bilda den nya föreningen vid Främjandets årsmöte den 26–27 mars 2009. Medlemmar i arbetsgruppen är förutom de tre initiativtagarna Xavier de Luna, Umeå, Rolf Larsson, Uppsala, och Hans Nyqvist, Stockholm, även Tom Britton, Stockholm, Aila Särkkä, Göteborg, och Lena Zetterqvist från Lund. Detta ger en jämn fördelning mellan statistik och matematisk statistik, och en rimlig geografisk spridning.

Alla som har idéer och synpunkter är varmt välkomna att höra av sig till någon av oss! Till exempel är det inte bestämt vad föreningen ska heta. Ett förslag som har kommit fram i våra diskussioner är Cramérsällskapet.

*Arbetsgruppen för en förening
för forskning och utbildning
genom Rolf Larsson, Uppsala universitet*

Höstkonferens om samverkan mellan statistik och medicin

Årets höstkonferens hade titeln Hälsosam statistik och arrangerades gemensamt av Svenska statistikfrämjandet och Föreningen för Medicinsk Statistik. Konferensen gick av stapeln den 6 november på Citykonferensen i Stockholm och hade lockat ett 40-tal åhörare. Efter den vetenskapliga delen av programmet valde några av deltagarna att avsluta dagen med en gemensam middag. Deltagarna verkade över lag vara mycket nöjda med dagen och uppskattade de intresseväckande presentationerna.

Temat för dagen innehöll några etiska frågeställningar i gränslandet mellan statistik och medicin. Avsikten var att belysa hur statistik och medicin kan samverka för att ge människor ett bra beslutsunderlag i ställningstaganden som påverkar deras liv och hälsa. Programmet var uppbyggt kring tre block, och inom varje block var två föredragshållare inbjudna, med avsikten att belysa ämnet från olika perspektiv. En av föreläsarna gav en medicinsk bakgrund medan den andre gav ett statistiskt och kvantitativt perspektiv.

En sammanfattning av programmets tre teman ges i följande referat. Föreläsarnas presentationer finns också tillgängliga på Främjandets webbplats.

*Bernhard Huitfeldt, Främjandet
Stig Johan Wiklund, FMS*

Prioriteringar och val inom vården

Under konferensens förmiddag fick vi två perspektiv på behovsstyrning inom vården. Lena Lundgren, hälso- och sjukvårdsdirektör vid Östergötlands läns landsting, talade om användning av statistik för att planera och styra vården. Lars Valter, statistiker vid Folkhälsovetenskapligt centrum, Östergötlands läns landsting, talade om arbetet att ta fram statistik som underlag för beslut inom vården. Beslutsfattare å ena sidan, framtagare av underlag å andra sidan.

Att just Östergötlands läns landsting fick representera detta föredrag beror på landstingets offensiva förhållningssätt till problemet med prioriteringar inom vården. Under de senaste åren har landstinget valt att precisera och offentliggöra vilka vårdinsatser som prioriteras respektive inte prioriteras. Dessa så kallade "listor" har varit föremål för debatt och omfattande uppmärksamhet i media.

Utgångspunkten för de landstingsbeslut som fattas inom länet är invånarnas behov av hälso- och sjukvård. Målet är att genom en effektiv verksamhet verka för en god hälsa i länet. Som bakgrund fick vi bland annat lära oss att Sverige tillhör den tredjedel av Europas länder där hälso- och sjukvården fungerar bäst med avseende på patienträttigheter och information, väntetider, medicinska resultat, "generositet" och läkemedelsanvändning. Sverige uppvisar en generellt sett hög effektivitet inom vården vid internationella jämförelser. Östgötarnas upplevda hälsa avtar starkt med ålder och kvinnor förefal-

ler må sämre än män. Det psykiska välbefinnandet hos unga kvinnor är något sämre i Östergötland än i övriga riket, vilket föranledde uppmärksamhet och omfattande insatser. De mest kostnadskrävande sjukdomsgrupperna i Östergötland år 2007 var tumörer (700 Mkr) och psykiska störningar (520 Mkr).

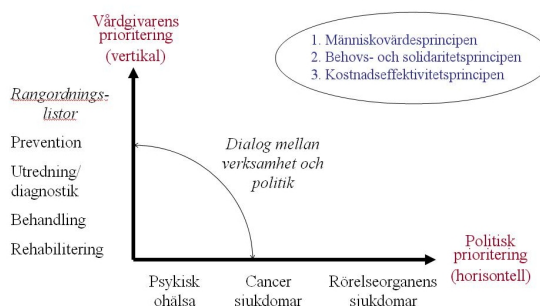
Studier av hälsa och livsstil

Ett centralt instrument som utgör ett viktigt och flitigt använt beslutsunderlag vid behovsanalysen är återkommande befolkningsstudier inom länet. Dessa genomförs medelst enkät med några års mellanrum. Undersökningen baseras på ett stratifierat urval (ålder, kön och kommun) med totalt 13 440 individer. Frågorna kretsar kring hälsa och livsstil. Svarsfrekvensen uppgick vid den senaste undersökningen till 54 procent. Bortfallsjustering med hjälp av kalibrering görs med variablerna födelse-land, civilstånd, utbildningsnivå och sysselsättning. Detta är möjligt eftersom det finns registerinformation om individerna i bortfallet.

Det är inte bekymmersfritt att arbeta med kalibreringsmetoder. För att göra statistik på insamlad data krävs förhållandevis god kunskap om statistisk metodik om kalibreringens metodologiska konsekvenser – något som det råder brist på i de olika grupperingar inom landstinget som behöver arbeta med datamaterialet. Problemet upplevs så stort att kalibreringsmetodiken ifrågasätts. Landstinget försöker istället minska problemet med bortfall genom att höja svarsfrekvensen i undersökningen. Experiment med enkätens utformning (som av många upplevs som alltför omfattande), partiell anonymitet och utskicksmetoder genomförs. Hittills verkar resultaten lovande. Som mest ökade svarsfrekvensen från 54 till 64 procent vid ett försök där enkäten gjordes anonym och enkäten begränsades i omfattning.

Efter ovan redovisade föredragning av genomförda undersökningar som beslutsunderlag, övergick det till att handla om resursfördelning och prioritering.

Horisontell och vertikal prioritering



Figur 1: Modell för prioritering inom Östergötlands läns landsting.

Prioritering görs utifrån en mängd parametrar och intressen. De tre principerna människovärdesprincipen, behovs- och solidaritetsprincipen samt kostnadseffektivitetsprincipen är vägledande. Landstinget tillämpar en

modell som kan illustreras med en vertikal och en horisontell axel.

Som framgår av Figur 1 är den fastlagda prioriteringen ett resultat av en dialog mellan verksamhet och politik. Genomförda undersökningar och behovsanalyser fungerar som underlag för initiativ och diskussioner. Ett underlag som fortfarande är behäftat med stor osäkerhet är kostnader för olika sjukdomsgrupper. Därtill saknas väl anpassade redovisnings- och journalsystem. Gjorda prioriteringar konkretiseras så småningom i resursfördelningsmodeller. Inom landstinget i Östergötland kommer detta bland annat att ske med utgångspunkt från en vårdvalsmodell. Sådana modeller införs runt om i landet och är föremål för många diskussioner och debatter. En av idéerna är att låta medborgare lista sig på en valfri vårdcentral. Därmed förväntar sig landstingen att individernas efterfrågan och behov ska bidra till att effektivisera vårdutbudet.

Det arbetssätt som Östergötlands läns landsting valt förutsätter en inte obetydlig kompetens och förståelse för statistik. Lena Lundgren har till exempel själv forskarbakgrund och är väl bevandrad inom statistikämnet. Vanligtvis har höga beslutsfattare inom landstingen en bakgrund inom vårdsektorns mer administrativa delar.

Mats Forsberg, Statisticon

Screening för prostatacancer

Blocket inleddes av professor Jonas Hugosson, Avdelningen för urologi vid Sahlgrenska Universitetssjukhuset, Göteborg, som gav en medicinsk översikt av området. Screening för prostatacancer med PSA är kontroversiell i Sverige. PSA (Prostate Specific Antigen) är ett enzym som läcker ut från prostatakörteln till blodet vid prostatacancer. Av män som har ett lätt förhöjt värde har cirka 25 procent prostatacancer. Den låga specificiteten är ett problem.

Ett antal viktiga kriterier bör vara uppfyllda om allmän screening för tidig upptäckt av sjukdom ska införas. Bland andra ska ett viktigt hälsoproblem föreligga, en effektiv behandling finnas, en etablerad policy för vem som skall behandlas som patient vara fastställd, tillräckliga resurser för diagnos och behandling finnas, ett igenkännbart tidigt sjukdomsstadium, ett lämpligt diagnostiskt test som är allmänt accepterat i populationen och naturalförloppet av sjukdomen ska vara tillräckligt kartlagt. Många av dessa kriterier är idag tveksamt uppfyllda för PSA och prostatacancer.

Det är stora skillnader beträffande vad som rekommenderas i världen. I USA rekommenderar man allmänt screening sedan 90-talet och där har dödligheten av prostatacancer minskat med omkring 30 procent. Men det har blivit en enorm incidensökning, från tidigare 10 till 20 procent. I Skandinavien är man mer tveksam. Tolkningen av de data som finns är väldigt olika. Men många anser även här att det finns påtagliga effekter av att införa screening.

Problematiken belyses av följande siffror. Av de som upptäcks vid screening botas en av åtta. En av åtta dör ändå och sex av åtta får diagnos men dör inte av sjukdomen. De som botas lever i genomsnitt 5–7 år längre. Det tar i genomsnitt 6–7 år innan en positiv PSA-screening leder till symptom av sjukdomen. Hur många är det värt att överbehandla för att rädda livet på en person? Risken för biverkningar av behandling (strålning, operation, cellgift) är stor, särskilt impotens och inkontinens.

Stor studie snart avslutad

Den långsiktiga nyttan av PSA-screening på diagnostik, behandlingseffekter och överlevnad kan undersökas i prospektiva kohortstudier. Just nu pågår en stor studie i åtta länder (bland annat Sverige) som startades för tio år sen av organisationen *The European Randomised study of Screening for Prostate Cancer* (ERSPC). Den är av hög kvalitet och löper på bra. Sjukdomen utvecklas långsamt, därav den långa uppföljningstiden. Cirka 200 000 personer bjöds in att delta, varav 20 000 randomiserades i Göteborg. Initialt förhöjda värden leder till vidare undersökning och PSA-mätningar vartannat år. Studien beräknas vara avslutad år 2010 och då förhoppningsvis ge det definitiva svaret på nyttan med PSA-screening.

Hans Garmo, statistiker på regionalt onkologiskt centrum i Uppsala, följde upp med att ge den statistiska bakgrunden för PSA-screening.

Vi har sett en påtaglig ökning av prostatacancer i Sverige sedan 1970, osäkert på om den beror på ökande ålder, fler PSA-prov eller något annat. Screening leder till tidigare diagnos och längre överlevnad. Men denna överlevnad kan inte jämföras med den som mäts om man inte screenar, eftersom diagnosen då inträffar i ett senare skede. Däremot kan man påvisa en positiv överlevnadseffekt om död tidpunkten kan förskjutas på grund av screening orsakad av bättre behandlingseffekt.

En frågeställning som funnits är om D-vitamin, som är förhöjd under sommaren på grund av solljuset, påverkar diagnosen. Överlevnaden för de som diagnostiseras på sommaren är nämligen kortare, men färre diagnoser ställs då. Prostatacancer med kliniska symptom upptäcks på sommaren, tidiga former av cancer (på grund av PSA-prov) upptäcks resten av året. Man går inte frivilligt till sjukvården på sommaren utan att ha symptom. Eftersom randomiserade studier inte är möjliga att genomföra på detta område är registerstudier det enda alternativet. Tre patientgrupper är aktuella att studera:

- De som aldrig skulle ha fått diagnos utan screening, det vill säga aldrig hade blivit sjuka.
- De som inte påverkas nämnvärt, till exempel de som ändå dör.
- De som har nytta av PSA-screeningen, det vill säga de som kan behandlas och därmed leva längre (de flesta dör av annan orsak).

En viktig fråga vid PSA-screening är att välja kritiskt PSA-värde för när förhöjd risk för prostatacancer fö-

religger. Ett sätt att utvärdera data är i form av ROC-kurvor, som kan beskriva sambandet mellan sensitivitet och specificitet vid val av olika kritiska värden.

En annan frågeställning är hur rökning ska hanteras vid analysen, då rökning generellt leder till tidigare död. Händelser kan antingen ignoreras, censureras eller betraktas som konkurrerande risk-händelser. Beroende på om man ser på rökning som en konkurrerande risk eller inte får man olika resultat. Det finns också andra omvärldsfaktorer som kan påverka resultatet i teorin, till exempel diet, solljus, sexuell överförbarhet och folsyra.

Sammanfattningsvis fick vi en mycket intressant belysning av ett viktigt medicinsk problem där det visade på hur statistik och medicin kan samverka för att underlätta beslutsfattandet.

*Anna Wiklund
och Inger Persson, AstraZeneca*

Prediktion av sjukdom och död

Dagens tredje tema introducerades genom att visa exempel från Internet och media, i vilka läsarna erbjuds genetiskt kartläggning och test av sjukdomsrisker. Förste föreläsare var Niklas Juth, praktisk filosof på Karolinska Institutet, som talade om "Gen-etik och framtidens skuggor". Genteknikens framsteg har genererat en omfattande diskussion om olika typer av gentest.

Presymptomatiska gentest kan avse diagnos eller latens-tid för uppkomst av sjukdom. Bör man göra ett sådant test? Vilket värde har det och vad tjänar man på det? Vem eller vilka skall få information om resultatet? Utöver testpersonen, kan släktingar, försäkringsbolag och arbetsgivare ha intresse av informationen. Värde kan vara att uppkomsten av de sjukdomar som kan botas skulle kunna förhindras. Å andra sidan finns sjukdomar, som inte går att bota. Man talar här om att värdet ligger i att människors möjlighet att styra sina liv gynnas.

Rätten till information om testresultat kan gälla släktingar, då det innehåller information om till exempel farfar, far och son. Skall släktingar tillfrågas innan ett test utförs? Det kan vara av nytta om risken för behandlingsbara sjukdomar. Informationsintresse kan finnas hos tredje part, till exempel försäkringsbolag. Argument mot att delge (negativ) information brukar anges som orättvisa, då resultatet kan leda till att man inte får någon försäkring. Att inte delge resultat kan å andra sidan utnyttjas av försäkringstagaren genom att antingen teckna eller inte teckna en försäkring.

Kontroversiell fosterdiagnostik

En speciell situation utgörs av genetisk fosterdiagnostik, som går att utföra redan på embryonal nivå. Extrema strategier är att man utför test för att de går att utföra, eller också utför man dem inte alls. En kompromiss skulle vara att göra det för att utesluta allvarliga sjukdomstillstånd, men det är svårt att dra en gräns. Handikapprelsen har framfört invändningar mot den här typen av argumentation. Om folk får välja embryon fritt, utvecklas risk för så kallade "backdoor eugenics", men utan

gräns kan man få barn som lever en kort tid med svåra plågor.

Tillförlitligheten hos genetiska test varierar. Test som berör en enda gen är tillförlitliga i termer av ja och nej, medan test som berör fler gener är otillförlitliga i termer av sannolikhet. Lagstiftning rörande genetiska test, antingen den tillåter eller förbjuder, är svår att tillämpa.

Riskbedömning vid försäkring

Erik Alm, aktuarie, Hannover Life Re, Sweden hade valt titeln "Vem dör när och varför?" Hans föredrag behandlade individuell och kollektiv riskbedömning vid beräkning vid premier av pensions- och livförsäkringar. Premien för en grupp bestäms utifrån dess sammansättning eller skadehistoria. Dödligheten ökar med åldern, varför en gruppgemensam premie sätts så att försäkringspremierna för de yngre med lägre dödlighet balanserar kostnaderna för de äldre med högre dödlighet. Om kunderna utnyttjar denna information till sin fördel så att enbart äldre tecknar försäkring föreligger moturval.

Dödligheten skiljer sig mellan män och kvinnor, vilket talaren hävdade är en biologisk skillnad och inte en social konstruktion. Standardtariffer för livförsäkring följer ålder och kön. Rökning (20 cigaretter per dag) medför dubblad risk att dö i förtid och motsvarar sju års kortare livslängd. Allmänt innebär dålig hälsa ökad risk att dö. En livförsäkring är en försäkring vid dödsfall, men egentligen mot att bli sjuk och sedan dö. Försäkringsbolaget behöver information för sin riskbedömning och agerar utifrån den information de har.

Individuell riskbedömning görs med hjälp av läkarundersökning och följer försäkringsbeloppets storlek och i viss mån försäljningssituationen på grund av risk för moturval. Individuell riskbedömning görs även med hjälp av frågor och test. Sannolikheten för sjukdom, det vill säga ökad skadekostnad vid sjukdom, jämförs för kostnad för testet, vilket är en kostnad för både individen och bolaget. Resultatet av en individuell riskbedömning kan leda till premie enligt standardtariff, premiehöjning, införande av klausul, det vill säga uteslutning av vissa sjukdomar, eller avslag.

Den enskilde har valmöjlighet och väljer den mest fördelaktiga lösningen. Den med stor risk försäkrar sig förmodligen oftare. Försäkringsbolaget bör ha tillgång till samma information som individen, antingen genom att resultatet av redan genomförda gentest delges båda parter, eller att bolaget ställer krav på individen att genomgå gentest för att försäkring skall få tecknas. Genetiska test utgör emellertid ett speciellt problem, då det anses kränkande att kräva gentest. Bolaget får dock fråga om man genomgått gentest, när försäkringsbeloppet är någorlunda stort. Det är emellertid svårt att kontrollera riktigheten av resultaten från genomgångna test. Numera finns det dock regelverk för att utvärdera individens uttalanden om gentestresultat i efterhand. Riskbedömning var förr mera konst än vetenskap. Nu försöker man vara så objektiv som möjligt.

Jan Henriksson

Jubileum för den officiella statistiken

Med anledning av att SCB och den officiella statistiken firar 150 år uppmärksammas detta med ett referat från SCB:s jubileum och en artikel som mer fokuserar på den officiella statistikens historia i stort.

I nästa nummer av medlemstidningen bidrar Ulf Jorner med en artikel baserad på hans nyligen utkomna bok om SCB:s historia, Summa summarum.

Statistiska centralbyrån firade 150 år med inbjudna talare och jubileumsbok

Anders Rickardsson, SCB

SCB firar 150 år i år och den 29 oktober var det jubileumsdag. Det var en dag då all personal samlades på Folkets hus i Stockholm för att lyssna på ett antal föredrag kring SCB:s gångna 150 år. Föredragen speglade den allmänna historiska utvecklingen under perioden, framväxten av några av de viktigare statistikområdena, media och teknikens betydelse. Dagen avslutades med middag och underhållning i Stadshuset. Det blev en intressant och mycket trivsamt dag.

Det kan i sammanhanget noteras att det firades 250-årsjubileum för några år sedan. Tabellverket inrättades år 1749 med ansvar för den löpande befolkningsstatistiken, och var därmed världens första i sitt slag. Tabellverket kan ses som en föregångare till SCB, men benämningen syftar inte på något ämbetsverk utan på själva tabellerna. Denna gång var det alltså myndigheten SCB som firade.

Samhället omkring SCB

SCB sorterar under Finansdepartementet och finansminister Anders Borg var inbjuden som talare. Han beskrev sig själv som en flitig statistikanalytiker genom åren. Viktig statistik för honom är bland annat Konsumentprisindex (KPI), Arbetskraftsundersökningen, lönestatistiken och handelsstatistiken. I sitt tal tog han upp den aktuella finansiella krisen och vad som nationellt och globalt behöver göras för att stärka förtroendet för marknaden.

Författaren Herman Lindqvist talade om samhället omkring SCB under de gångna 150 åren. Han pekade inledningsvis på den långa svenska traditionen med folkräkningar och löpande befolkningsstatistik från Tabellverket. Med Tabellverket följde bland annat en naturvetenskaplig granskning av förhållandena i Sverige. Det fanns på den tiden flera ledande vetenskapsmän med intresse för statistikfrågor, till exempel Pehr W Wargentin, astronom och aktiv i Tabellverket. Den svenska befolkningen var liten och antalet invånare okänt. Tabellverket behövdes för att visa upp och beskriva samhällets olika aktiviteter.

I den andra delen av sitt föredrag beskrev han de snabba samhällsförändringarna under andra delen av 1800-talet. Det fanns flera problem, ny teknik var på väg men man var försiktig att ta in det nya till Sverige. Det fanns till exempel en tidig rädsla för järnvägen och tågtrafiken. Järnvägsutbyggnaden bidrog dock så småningom till ett av dåtidens viktiga standardiseringsprojekt, en för Sverige gemensam normaltid.

Välfärd och kvinnor i statistiken

Eskil Wadensjö, professor i arbetsmarknadspolitik, Institutet för social forskning vid Stockholms universitet, gjorde några nedslag i utvecklingen av den svenska välfärdsstatistiken. Någon sådan fanns inte på Tabellverkets tid. De allra första beskrivningarna sågs som politiskt sprängstoff. Ansvariga myndigheter var i första hand Kommerskollegium och Socialstyrelsen. I slutet av 1960-talet genomfördes en låginkomstutredning. Senare kom levnadsnivåundersökningarna, där bland annat Sten Johansson, senare generaldirektör vid SCB, hade en aktiv roll. Undersökningen om levnadsförhållanden har genomförts sedan 1974.

På temat jämställdhet och gömda kvinnor i statistiken talade den tidigare jämställdhetsutredaren Gertrud Åström. Hon nämnde bland annat Florence Nightingale, som även är känd för sina insatser inom statistiken. I sitt föredrag pekade hon på underskattningen av kvinnornas förvärvsarbete i statistiken. Det gäller till exempel kvinnornas arbete i jordbruket ända till tämligen sen tid. Kvinnorna har dock numera en hög sysselsättningsgrad, men löneläget är fortsatt under diskussion.

Ekonomi, media och teknik

Klas Eklund, SEB, var inbjuden för att tala ekonomi. Han konstaterade att den ekonomiska statistiken har en mycket stor betydelse idag och att en betydande del har sitt ursprung i utvecklingen av nationalräkenskaperna under 1950- och 60-talen. I takt med att ekonomin har blivit allt mer komplex och sammanflätad har den blivit svårare att mäta. Hur mäter vi till exempel kvaliteten i KPI? Ett nytt begrepp i den politiska debatten är utanförskap, som innebär en ny utmaning att beskriva statistiskt. Det blir allt svårare att mäta, men vi måste mäta, därför har SCB och "räknenissarna" flera utmaningar i framtiden, enligt Klas Eklund.

Journalisten Pia Brandelius gav ett antal exempel på journalistens dilemma att tolka och hantera statistik. En undersökning kan lätt få stort genomslag i media, men resultaten kanske inte håller vid en granskning. Räkneexempel med effekter för den vanliga medborgaren kan också bli svårtolkade och ifrågasättas. Arbetskraftsundersökningen och Partisynpatiuundersökningen är alltid viktiga källor för media.

Den tekniska utvecklingen har alltid varit viktig för SCB:s statistikproduktion. Jubileumsdagen avslutades med att Olov Östberg, professor emeritus i informatik,

gjorde några nedslag i teknikhistorien och den spännande framtiden med bäring på SCB:s verksamhet och kommande utveckling. Hans tes var att teknik- och datorutvecklingen kommer att fortsätta ett bra tag till och att nya möjligheter och tillämpningar kommer att bli aktuella.

Jubileumsbok

Inför jubileumsfirandet har SCB presenterat en bok om myndighetens första 150 år med titeln *Summa summa-rum*, skriven av Ulf Jorner. Information om boken finns på SCB:s webbplats, www.scb.se/, men några intressanta händelser återges nedan.

När SCB grundades 1858 producerades redan en hel del statistik, men en statlig kommitté ansåg bland annat att ”statistiken i det hela var mycket planlös, efterbliven och tillbakasatt”. Chef blev Fredrik Teodor Berg. Kommittén föreskrev att ordinarie tjänstemän skulle ha högre vetenskaplig utbildning, allmän erfarenhet som tjänstemän och kunna skriva på flera utländska språk. Under första tiden vilade SCB:s produktion knappast på någon vetenskaplig grund. Det var snarare praxis än teori som användes eftersom det saknades statistisk teori för officiell statistikproduktion. Statistikens kvalitet var dock i fokus.

Den mest framträdande statistikern runt förra sekelskiftet var Gustav Sundbärg, forskare och publicist, innehavare av Sveriges första professur i statistik och under flera år anställd vid SCB, dock inte som chef. Han var även initiativtagare till bildandet av Statistiska förening-

en år 1901. År 1905 utnämndes Ludvig Widell till chef för SCB. Han var docent i statskunskap och statistik i Lund. Det var först vid den här tiden som den första räknemaskinen köptes in, medan andra myndigheter varit före. Den första undersökning som använde hålkortsteknik var 1920 års folkräkning.

Nästan 50 år efter det att SCB grundades bildades en ny statistisk kommitté, vars uppgift bland annat var att se på frågan om ökad centralisering. Året efter första världskrigets slut fick SCB sin andra instruktion, vilken bland annat angav kompetenskraven för anställning (universitetsexamen i till exempel statistik, nationalekonomi, matematik, statskunskap eller moderna språk). I 1948 års statistikutredning diskuterades kompetenskraven och licentiatkravet för första aktuarie. Professor Sten Wahlund, som var utredare, argumenterade med kraft att kravet skulle vara kvar. Ett par år senare inrättades en maskincentral med en IBM-utrustning. Myndigheter som inte hade egna anläggningar skulle vända sig till SCB.

På 50-talet inrättades ett utredningsinstitut vars förste chef blev Tore Dalenius. Han var i mycket en föregångsman med anknytning till SCB. Den första uppgiften blev att bygga upp en rikstäckande intervjuorganisation. Modellen med så kallade basurval etablerades. Tidiga uppdrag var bland annat i anslutning till konsumentprisindex åt Socialstyrelsen. Moderna stickprovsundersökningar började användas omkring 1950. I modern tid har SCB byggt egna register över både individer och företag. Metoden med basurval blev därmed överflödig. ■

1858 – den officiella statistikens nytändning

Olle Sjöström

*En intressant och småningom dynamisk utveckling av svensk officiell statistik kom igång 1858 när ett nytt ämbetsverk etablerades, Statistiska tabellkommissionen, med Fredrik Theodor Berg som planläggare och uppbyggare av flera statistikgrenar. Han upptäckte och blev mentor för Gustav Sundbärg, som blev en internationellt känd statistiker och Sveriges förste professor i statistik 1910. Denna utveckling ledde till etableringen av statistiska byråer och mycket aktiv statistisk profession under de första decennierna av 1900-talet. De ekonomiska och sociala undersökningar som då genomfördes gav viktiga kunskaper, vilka bidrog till den politiska enighet som krävdes för det demokratiska genombrottet i Sverige.**

Vi har under de senaste 50 åren vant oss vid ett ekonomiskt och tekniskt språk om statistik. Statistik är produktion eller uttag av data, som sedan konsumeras. För samhällsstatistikens del kom denna drastiska förändring i det statistiska språkbruket kring 1960 då ”den stora centrala datorn” blev en påtaglig produktionsapparat kring vilken tänkande och språk formades. Detta skedde samtidigt som en stark centralisering av den statliga statistiken till SCB påbörjades.

Därmed avslutades en period från 1858 med decentraliserad statlig statistik som hade stått i öppen förbindelse med folkrörelserna, folkbildningen och den politiska utvecklingen. En statistisk uppbyggnadsperiod under senare delen av 1800-talet följdes av en expansiv utveck-

ling i början av 1900-talet av sociala och ekonomiska undersökningar. De kom att spela en viktig politisk roll fram till genombrottet för den svenska demokratin 1920. Statistikens språk hade kunskapsbildning, genom analys och tolkning av statistik, i fokus och det fanns en medveten anknytning till en empirisk-filosofisk tradition och dess kunskapsteori.

Statistik som väckelse

För den europeiska statistiken har perioden 1830–50 kallats för ”entusiasmens era”. Framväxande liberalism och allmän idealism, närd av allvarliga missförhållanden i olika länders industriregioner, skapade privata statistiska föreningar och privat finansierade och genom-

* Artikeln är baserad på kapitel 4–6 i Sjöström, O. (2002). *Svensk statistikhistoria – En undanskymd kritisk tradition*. Gidlunds förlag.

förda undersökningar av sociala och ekonomiska förhållanden. Det ledde också till att flera stater reformerade sin statistik och inrättade statistiska byråer med motsvarande uppgifter. En viktig pådrivare för en mer systematisk statlig statistik var den belgiske astronomen och statistikern Adolphe Quetelet. Han initierade internationella statistiska kongresser som tog sin början 1853. Kongresserna återkom med cirka tre års intervall och de blev pådrivande för den kommande utvecklingen i Europa. Nu kom också Sverige med i bilden.

Statistiska tabellkommissionen

”Vår officiella statistik saknar drift och liv”, skrev en riksdagsman i en motion till 1853–54 års riksdag. En statistisk kommitté tillsattes med läkaren Fredrik Theodor Berg som ordförande. Berg var medicine doktor, tidigare professor i pediatrik vid Karolinska Institutet och under 1840-talet överläkare vid Allmänna barnhuset. Han hade tidigt tagit starka intryck av den franska medicinalstatistiken och publicerat flera vetenskapliga skrifter.

År 1855 gjorde Berg en studieresa i Europa och knöt inflytelserika kontakter med ledande statistiker. Kommitténs arbete ledde till förslaget att ett nytt ämbetsverk skulle inrättas, *Statistiska tabellkommissionen*. Den skulle bestå av dels ett rådgivande och beställande organ, *Statistiska beredningen*, och dels ett verkställande ämbetsverk, *Statistiska centralbyrån*. Den statliga statistiken skulle i huvudsak, som tidigare, vara decentraliserad till ett 25-tal andra ämbetsverk. Kommitténs förslag blev riksdagens beslut. Statistiska beredningen höll sitt första sammanträde den 30 oktober 1858 under närvaro av kronprins Oskar, svenska och norska statsråd, vetenskapsmän, chefredaktörer och andra bemärkta. Totalt var det ett 70-tal deltagare!

Som ordförande i beredningen höll Berg ett längre anförande om den officiella statistikens uppgifter. Dessa var bland annat

att anordna insamlandet ur rätta källor, att strängt kontrollera beskaflenheten (tillförlitligheten) och att ombesörja det ändamålsenliga sammanställandet av sådana för statsförvaltningen och vetenskapen nödiga och nyttiga statistiska primäriakttagelser, som utan statens medverkan och understöd skulle förbliva otillgängliga.

En grundplan för den officiella statistiken

Ett viktigt resultat av Statistiska beredningens inledande verksamhet blev en samordnad plan för publicering av all officiell statistik. En ny serie publikationer, *Bidrag till Sveriges officiella statistik*, med ett antal bokstaverade littera (ämnesområden) och skrifter med titeln *Statistiska berättelser*, började utges. Till den ambitiösa publiceringsplanen hörde också *Statistisk tidskrift* som startade 1860. Inte minst betydelsefullt var en mer medveten rekrytering av kompetens för statistiskt arbete.

Berg hade under sin långa studieresa 1855 systema-

tiskt undersökt statistikens organisation och de statistiska verksamheterna i flera länder. Han besökte de statistiska byråerna i Berlin, Dresden, Wien, München, Stuttgart, Bryssel, Haag och London. Han samlade statistisk litteratur och även formulären för olika slag av primärdata; formulär som andra länders historiska forskare har saknat men på senare tid återfunnit i Sverige! Han började tidigt forma idéerna till vad som kom att kallas ”en grundplan för den svenska officiella statistiken”. Han lämnade flera utkast för diskussioner i Statistiska beredningen och Berg kunde stärka sina argument med hänvisning till de internationella statistik-kongressernas rekommendationer. Han hade under 1860-talet deltagit i alla kongresserna, i London 1860, Berlin 1863 och Florens 1867, och lät förstå att andra länder hade den svenska utvecklingen under observation. Denna grundplan publicerade Berg till slut 1870 i *Statistisk tidskrift*. Berg skrev där att kartläggningen av olika samhällsförhållanden, som är ”den officiella statistikens stora uppgift”, inte på långt när hade genomförts i Sverige.

Viktigt för framtiden var att Berg framhävde en identitet för professionellt statistiskt arbete i offentlig förvaltning. Bakom låg Bergs teoretiska insikt om att statistik inte var enkel bokföring att utföra på order utan hade ett mer omfattande och viktigt ansvar, nämligen att det statistiska arbetet krävde en medveten metodik och en avsevärd kritisk och konstruktiv förmåga. Han hade stöd i en europeisk teoriutveckling för tillämpad statistik som också började bli känd av flera svenska statistiker.

Den statistiska kompetensens framväxt

En ung gymnasist i Falun, Gustav Sundbärg, sände sitt första befolkningsstatistiska arbete till Fredrik Theodor Berg 1875. Denna debut för en 19-åring publicerades sedan 1876 i *Statistisk tidskrift* och Berg fick en entusiastisk lärjunge som skulle föra hans kunskaper och målmedvetenhet vidare, inte minst efter Bergs död 1883. Sundbärg skulle få sin mest framgångsrika verksamhet efter 1890 och fick ett starkt inflytande på kolleger och rekrytering av statistiker. Av hans mer omfattande och kända arbeten kan nämnas hans revidering av den äldre svenska befolkningsstatistiken och den omfattande emigrationsutredningen. Genom att Sundbärg fick en docentur vid Stockholms högskola 1903 fick statistiken ett första fäste i den högre utbildningen. När Sundbärg sedan utnämndes till professor i statistik vid Uppsala universitet 1910 etablerades statistik för första gången som ett självständigt universitetsämne i Sverige.

Vid denna tid hade det vuxit fram en bred statistisk profession i Sverige. Förutom av Sundbärg befordrades denna framväxt av Pontus Fahlbeck, professor i historia och statsvetenskap i Lund. Han ”gjorde statistiker” av David Bergström, Theophil Andersson och Henning Elmquist genom att de doktorerade på statistiskt orienterade avhandlingar. Från Uppsala kom Gunnar Huss, doktor i historia, och senare Karl Arvid Edin, docent i statistik i Uppsala. De skulle alla bli betydelsefulla i den kommande utvecklingen av den statliga statistiken.

Ett av Gustav Sundbärgs många initiativ var att 1901 starta *Statistiska föreningen* – en förening för statistikens idkare och vänner. Den kom att få stor betydelse som diskussionsklubb och inspiration till olika statistiska initiativ de närmaste åren.

En stark statistisk aktivitet utvecklades i Kommerskollegium i början av 1900-talet. Där fanns ansvaret för näringsstatistiken men också för statistik över arbetsförhållanden i allmänhet och över olycksfall i arbete. Flera undersökningar om arbetsförhållanden i olika yrken genomfördes och en särskild kår av intervjuare anställdes för löpande arbeten. Genom att den statistiska verksamheten kom att omfatta två tredjedelar av kollegiets personal föreslog en utredning att ett nytt ämbetsverk, *Socialstyrelsen*, inrättades. Det skedde också 1912 och det nya verket kallades ”ett statistiskt verk”.

En statistisk byrå hade inrättats 1908 vid Kommerskollegium och kom att följas av många flera. I början av 1920-talet var antalet sådana byråer ett tiotal och antalet anställda statistiker vid dessa närmare 80. Karakteristiskt för dessa byråer var att de hade initiativrätt till nya undersökningar och att deras organisation innebar en reglerad praktiktjänstgöring för nyanställda statistiker. Ett exempel på sådan reglering var att självständigt utförda praktiktjänsterna diskuterades och bedömdes, något som hade att göra med att den statistiska verksamheten betraktades som ett slags forskning inom förvaltningen och att praktiktjänsterna var en vidareutbildning för sådan. De statistiska byråerna kallades ibland för ”samhällets observatorier”.

Den decentraliserade statistiken

Den som tydligast formulerade argumenten för en decentraliserad statistik var den tidigare docenten i statistik vid Lunds universitet, Theophil Andersson, med en karriär som statistiker vid Socialstyrelsen och senare som generaldirektör. Detta skedde 1916 i en festskrift till Pontus Fahlbeck, hans tidigare lärare, handledare och arbetsgivare. Från hans argumentering har följande exempel hämtats:

Liksom man aldrig lär känna ett land endast genom att studera kartan utan först genom att verkligen beresa detsamma, på samma sätt lär man sig aldrig förstå statistikens resultat, om man inte får tillfälle att så att säga dyka ned i dess sifferkolumner och

se hur dessa växa fram såsom sammanfattningen av de individuella fallen, hur varje siffra utgör en avspeglning av ett helt komplex av skiftande företeelser. Summor och medeltal är statistikens slutstenar. Men i dessa summor och tal gå så många detaljer förlorade, förblekna så många skiftningar, som kunna vara av betydelse.

Detta prov på resonemang illustrerar att den statistiska kunskapsbildningens problem stod i centrum och att ordet ”statistikkonsumtion” då skulle ha uppfattats som en rå trivialisering. Anderssons uppsats skrevs när de största framgångarna för den av Berg anlagda grundplanen skördades. En avsevärd breddning av de statistiska verksamheterna hade ägt rum liksom framväxten av en erkänd yrkeskår av statistiker, erkänd för sin förmåga till initiativ och att på ett kompetent sätt genomföra viktiga statistiska undersökningar. Förutom ovan nämnda undersökningar om arbetsförhållanden i olika yrken kan nämnas en ny förmögenhets- och inkomststatistik samt en omfattande bostadsräkning. Resultaten av dessa undersökningar spelade en politisk roll under rösträttsstriderna i början av 1900-talet och för det slutliga genombrottet för svensk demokrati i och med att kvinnor fick rösträtt 1921.

Olika perioder för officiell statistik

Namnet ”Statistiska centralbyrån” fyller 150 år, men det svarar inte för en kontinuerlig utveckling utan rymmer två mycket olika perioder för den officiella statistiken. Först var det namnet på en av avdelningarna i Statistiska tabellkommissionen och blev småningom en av de många statistiska byråerna i den decentraliserade statliga statistikens historia. Efter 1960 står namnet för en helt motsatt utveckling, en stark centralisering som innebar eliminering av de tidigare statistiska byråerna och även med en helt annan grundsyn på statistisk verksamhet. Det blev nu närmast ett slags planekonomi för statistiska ”dataprodukter” som genomfördes, och är en historia som kräver en särskild artikel.

Ett angeläget uppdrag för samhällsvetenskaplig forskning är jämförande studier av dessa mycket olika perioder i svensk statlig statistik, bland annat med hänsyn till förhållandet mellan statistik och politik och de olika slag av kunskapsteori som ligger bakom decentraliserad respektive centraliserad statistik. ■

Webbaserat medlemsregister

Alla medlemmar kan med hjälp av sitt medlemsnummer och sitt tillfälliga lösenord logga in på www2.bokningsplatsen.se/statistiker för att kontrollera eller uppdatera sina uppgifter i medlemsregistret.

Medlemsnumret består av sex siffror och du finner det på sista sidan av tidskriften. Det tillfälliga lösenordet är en sammanslagning av de tre första bokstäverna i ditt efternamn med inledande stor bokstav, och postnumret

till den adress till vilken du fick denna tidskrift. Till exempel And12345 för en person som heter Andersson med postnumret 123 45. För mer utförlig information, följ länken

www.statistikframjandet.se/Portals/28/docs/Instruktion%20webbregister.pdf

för att se artikeln om medlemsregistret som publicerades i Qvintensen 1-1. □

År 1908 publicerade tidskriften *Biometrika* två artiklar av pseudonymen "Student", vilka kom att få stor betydelse för den statistiska vetenskapens utveckling. Vi skall här se litet närmare på vad som hände, på bakgrunden, och på efterverkningarna.

Vem var Student? Författaren hette William Sealy Gosset och var anställd hos Guinness i Dublin. Han var dock inte alls irländare, utan född i Canterbury i England, och namnet Gosset kom från en hugenottsläkt som tvangs lämna Frankrike i slutet av 1600-talet. William föddes år 1876 som första barnet till en överste i The Royal Engineers. Sonen hade nog gärna gått den militära banan, men som tur var för vetenskapen hade han inte tillräckligt bra syn för att duga till det.



William Sealy Gosset, 1908.

I stället blev det akademiska studier då han erhöll ett stipendium för att läsa matematik och kemi vid det gamla New College i Oxford. År 1899 gick ut med toppbetyg i kemi och samma år anställdes han vid Guinness bryggerier i Dublin, vid St James' Gate. Företagets nya ledning hade bestämt sig för att satsa på vetenskapligt baserad tillverkning och att rekrytera ett antal förstklassiga studenter från Oxford och Cambridge. De umgicks säkert mycket sinsemellan. En av dem blev hans sväger, när Gosset 1906 gifte sig med Marjory Phillpotts.

Gosset vid Guinness

Gosset var anställd vid Guinness hela sitt återstående liv, från 1899 till dess att han avled vid 61 års ålder i en hjärtattack den 16 oktober 1937. Han blev ansvarig för forsknings- och utvecklingsavdelningen (Experimental Brewery), och han var en viktig person vid Guinness. De sista åren hade han ett ännu tyngre ansvar som Head Brewer för det nya bryggeri som Guinness startade utanför London.

Företaget Guinness satsade alltså på vetenskaplig kompetens i tillverkningen. Kompetensutveckling var därmed också önskvärd, och Gosset gick en kurs i "Brown beers" vid Birmingham University, vilken faktiskt var betydelsefull för hans fortsatta statistiska utveckling, för då började han fundera på den statistiska osäkerheten i jästcellsräkningar. Den andra studieresan gick till London och Karl Pearson's *Biometric School*. Guinness tillät Gosset att publicera sig vetenskapligt bara han inte avslöjade något som skulle kunna utnyttjas av konkurrenterna. Därför fick Gosset publicera under pseudonym, och han fick inte använda data från Guinness, utan hitta illustrationsexempel från andra håll. Deras restriktioner kvarstod nästan ända fram till andra världskriget, och Gosset var inte ensam om att publicera under pseudonym. Varför det blev just "Student", det vet jag inte.

I sitt arbete vid Guinness skulle Gosset planlägga och utvärdera odlings- och laboratorieförsök samt inflytandet av variation i råvaror och produktionsbetingelser, alltså experimentella data i relativt korta serier. Till en början studerade Gosset några böcker om felteori (till exempel Airy: *On the algebraic and numerical theory of errors of observations and the combination of observations*, 1:a upplagan 1861). Det handlade främst om minsta-kvadrat-metoden och dess användning på mätdata, med traditioner tillbaks till Gauss och Legendre och tillämpningar från astronomi och geodesi. Den enda statistiska fördelningen i sammanhanget var normalfördelningen, *the law of errors*. Statistisk precision uttrycktes med *probable error*, vilket inte är vårt nutida medelfel, utan kvartilernas avstånd från väntevärdet i en normalfördelning. Måtten skiljer sig med en faktor 0,6745.

I slutet av år 1904 lämnade Gosset en rapport till styrelsen för Guinness, *The application of the 'law of error' to the work of the brewery*. Bland rapportens slutsatser var att man borde ta kontakt med en matematiker för att bedöma "the degree of probability to be accepted as proving various propositions". Litet senare lyckades Gosset få kontakt med Karl Pearson och träffade honom en julidag i hans sommarhus i Oxfordstrakten. Det var ett mycket fruktbart besök för Gosset, som senare beskrev det med orden

He was able in about half an hour to put me in the way of learning the practice of nearly all the methods then in use.

Karl Pearson och Biometric School

Vem var Karl Pearson, och vilka metoder handlade det om? Pearson (1857–1936) hade en hög vetenskaplig status. Efter studier i diverse ämnen bland annat i Tyskland (där han tog stavningen Karl, eventuellt inspirerat av Karl Marx), hade Pearson blivit professor i *Applied Mathematics and Mechanics*, senare även i geometri, vid University College, London. Hans bok *The Grammar of*

Science från 1892 inspirerade Albert Einstein, som rekommenderade den till sina studiekamrater. Avgörande för Pearsons fortsatta inriktning mot biometri, genetik och eugeni (rashygien med mera) blev de intryck han fick av Francis Galton (korrelations- och regressionsbegreppens skapare). Pearson grundade Biometric School, med ett Biometric Lab och ett Eugenic Lab, samt tidskriften *Biometrika*. Han var oerhört produktiv, och till det han hade gjort innan Gosset mötte honom hörde definitionen av korrelationskoefficienten såsom vi nu känner den, anpassningstestet χ^2 , och ett system av fördelningsfamiljer (Pearson typ I, II, III, etcetera) för att modellera skeva och på annat sätt icke-normala fördelningar. Teoretisk momenträkning var viktig, inte minst för modellanpassning (momentmetoden). Han gjorde också några principiellt viktiga fel, som senare blev betydelsefulla för Fishers utveckling av statistikteorin (mest känt är välfelet i frihetsgraderna i χ^2 -testet). För att bättre förstå vetenskapens problem vid den tiden bör också nämnas, att man inte var så noga med distinktionen mellan parameter och stickprovstorhet, till exempel inte i mått på medelvärdets osäkerhet. Dessutom hade man vanligen ett bayesianskt synsätt, baserat på en likformig apriorifördelning för parametern av intresse, vilket var något som Gosset hade svårt för.



Karl Pearson, 1924.

Teckning: F.A. de Biden Footner

Gosset i London

De intryck från mötet med Pearson, som Gosset tog med sig till Guinness, ledde till att han fick tillbringa läsåret 1906–1907 i Pearsons skola. Gosset gick på Pearsons föreläsningar, men de gav honom inte så mycket. Större nytta hade han av att få diskutera sina problem och egna tankar med Pearson. Han arbetade då med sina första artiklar. Den allra första kom redan i februari-numret 1907 av *Biometrika*.

Denna första Student-artikel handlar om osäkerheten i en metod för att räkna frekvensen av blodkroppar eller celler, till exempel jästceller i öltillverkning, "On the error of counting with a haemocytometer". Det var något

Gosset tänkt på sedan kursen i Birmingham. Nu härleder han en gränsfördelning av binomial, som han kallar *the exponential series* och anpassar till verkliga data. Som läsaren kanske redan gissat så var det Poissonfördelningen han härledde, och den fördelningsfamiljen var alltså okänd i London vid den tiden, åtminstone för Gosset och Pearson (fram till 1912). Artikeln visar i alla fall Gossets förmåga, och Pearson försöker värva honom till en tjänst vid University College. Det lyckas dock inte alls, för Gosset konstaterar att han har klart högre lön som forskningschef vid Guinness än en professor har vid universitetet. Redan då kunde alltså universiteten ha svårt att konkurrera med näringslivet.

z-statistikan som blev t-statistikan

Gosset jobbade vidare hemma i Dublin, och i mars 1908 kom nästa artikel, "The probable error of the mean". Det är en välstrukturerad och välskriven artikel, som vi skall se på punkt för punkt enligt Gossets innehållsförteckning:

I. Finn fördelningen för stickprovsvariansen s^2 och för s (normalfördelningen för stickprovsmedelvärdet m var välkänd). Gosset klarar inte att direkt härleda fördelningen, men han går via momenten och har tur. Han beräknar de första fyra momenten för s^2 och finner överensstämmelse med Pearson typ III. Han antar fortsättningsvis att det är just den fördelningen. Övergången från s^2 till s är däremot oproblematiske, "all we must do is to compress the base line suitably". Var Gosset därmed först med fördelningen för s^2 ? Nej, den var känd bland tyska geodeter ända sedan 1876 (Helmert), men detta förblev länge än okänt för Pearson, som snarast föraktade felteoristatistiken.

II. Härnäst "There is shown to be no kind of correlation between m and s ". Det oberoende som behövs för fortsättningen kändes väl litet för svårt, och Gosset fortsätter på momenträkningsvägen. Han inser att det inte räcker med att påvisa att m och s är okorrelerade utan visar att även m^2 och s^2 är okorrelerade, för att stärka sin sak.

III. Fördelningen för m/s är nu ett integrationsproblem som Gosset klarar av.

IV–V. Här diskuteras de täthetskurvor som Gosset kommit fram till ovan.

VI. I detta innovativa avsnitt jämför Gosset de teoretiska fördelningarna med verkliga fördelningar i en omfattande simuleringsstudie, eller bootstrap skulle vi kunna säga. Han utnyttjar biometriska data från 3 000 brittiska kriminella. Han väljer två responsvariabler, kroppslängd och långfingerlängd, som båda visar god överensstämmelse med normalfördelningen. Han skriver ned värdena på 3 000 kort som han sedan blandar (det vill säga randomiserar ordningsföljden på). Därefter tar han ut 750 stickprov av storleken $n = 4$ (dragning utan återläggning) och beräknar s^2 och m/s för vart och ett. Fördelningarna stämmer bra! Han har också en diskussion av robusthet mot avvikelser från grundförutsättningen om normalfördelning.

VII–X. Tabeller med förklaringar och tillämpningsexempel, samt slutsatser.

Ännu en kommentar är på sin plats. Gosset ville helst använda $n - 1$ i nämnaren för s^2 , men lyder här Pearson, som förespråkade n . Observera också att Gosset inte dividerar med medelfelet utan med standardavvikelsen. Han har inte heller beteckningen t , utan kallar sin statistika för z . Fisher ändrade senare i samråd med Gosset till den t -statistika vi alla känner.

Fisher in på scenen

I septembernumret av Biometrika samma år kom Students tredje artikel, "Probable error of a correlation coefficient". Nu handlade det om korrelationskoefficientens fördelning, och det är ett svårare problem. Gosset börjar här med att resampla igen från sitt 3 000 enheter stora stickprov, och nu utnyttjar han de två responsvariablerna simultant. De har korrelation 0,66, men genom att förskjuta den ena mot den andra får han dem också okorrelerade. Han räknar på stickprovsstorlekar 4, 8 och 30, och så jämför han de simulerade fördelningarna med fördelningar ur Pearsons fördelningssystem. Han har partiell framgång, ty för okorrelerade data passar nämligen Pearson typ II utmärkt. För korrelerade data måste han dock konstatera "I cannot suggest an equation" och "probably cannot be any of Prof. Pearson's type".

Vad hände sedan? Gossets artiklar hade inget omedelbart inflytande i den statistiska världen, men vid Guinness använde man Gossets tabeller. År 1912 kom emellertid Fisher in på scenen, och då börjar det hända lite. Ronald A. Fisher hade då nyligen avlagt examen vid Cambridge, och han var ung, ambitiös och begåvad, inte minst matematiskt. Han var visserligen främst eugeniskt intresserad, men då läste han ju förstås Biometrika. Han skickade ett brev till Gosset med en rigorös härledning av fördelningen för m/s , med hjälp av n -dimensionell geometri. Gosset skrev till Pearson:

would you like looking at it for me, . . . , if it's all-right, perhaps you can publish a note. It's so nice and mathematical . . .

men Pearson svarade

I do not follow Mr. Fisher's proof & it is not the kind of proof which appeals to me

så det blev ingen publicering. Det var första gången det skar sig mellan Pearson och Fisher.

Händelsen inledde i alla fall en omfattande brevväxling mellan Fisher och Gosset. Under åren utväxlade de sammantaget omkring 300 brev, där de utbytte idéer om sta-

tistisk metodik, försöksplanering med mera. De var inte alltid överens, men hela tiden vänner. År 1915 hade Fisher löst och publicerat problemet med korrelationskoefficientens fördelning, med samma metoder som för t -fördelningen. Gosset hjälpte 1919 Fisher till jobbet vid Rothamstead Experimental Station, där Fisher fick utlopp för sin kreativitet och utvecklade variansanalysen och statistisk försöksplanering. I ett brev till Fisher 1922 frågade Gosset om fördelningen för regressionskoefficienten. Fisher svarade med t -testen för såväl två-sampelproblemet och för regressionskoefficienten. Dittills hade det alltså varit en teori för ett stickprov, eller för data i par. Den riktigt stora lanseringen av t -testet kom ett par år senare, 1925, med Fishers storsäljande bok *Statistical Methods for Research Workers*. Men då är vi inne i en annan era, som fordrar sin egen historia.

Slutligen ett citat av Egon Pearson, son till Karl Pearson, och ena halvan i paret Neyman-Pearson:

Gosset was a modest, kindly, and tolerant man, who disliked controversy and was absolutely devoid of malice

Hans fördelning för m/s betraktades inte som betydelsefull när den publicerades, men Gosset och hans artiklar kom att bli enormt viktiga som katalysator för Fisher, som gjorde allt på sitt eget sätt och i full generalitet, och gav t -testet en naturlig och stor roll inom ramen för en fullständig teori för statistisk inferens. ■

Litteratur:

Pearson, Egon S. (1939). 'Student' as statistician. *Biometrika*, **30**.

'Student' (1942). *'Student's' Collected Papers* (eds E.S. Pearson and Wishart). Cambridge Univ. Press (nytryckt 1958).

Box, Joan Fisher (1978). *R.A. Fisher. The Life of a Scientist*. Wiley.

Pearson, Egon S. (1990). *'Student', a statistical biography of William Sealy Gosset* (eds R.L. Plackett and G.A. Barnard). Oxford Univ. Press.

Zabell, S.L. (2008). On Student's 1908 article "The probable error of a mean" (w. disc.). *Journal of American Statistical Association*, **103**.

Bildkälla:

Annals of Eugenics (1939). Cambridge Univ. Press.

Debatt

Statistiker behöver utbildning i kommunikation

Mina 15 år som statistiker inom offentlig sektor har lärt mig att det inte är lätt att förmedla resultat av utredningar och undersökningar på ett sätt som förstås av de flesta. I mitt fall har mottagarna bestått av kommunaltjänstemän, landstingstjänstemän och politiker, men också av akademiker som till exempel beteendevetare, psykologer och läkare. Mottagarna för mitt arbete har för det mesta haft en högre utbildning än genomsnittet. Flera av dem har haft ett antal poäng i statistik och när jag har frågat dem om dessa kurser så har svaren varit av typen "usch ja, jag är glad att jag klarade den" eller "den kursen glömde jag raskt".

Statistiken är i många fall en viktig del av deras arbete och de behöver egentligen ha en god förståelse för den särskilt som resultaten av en statistisk analys ofta ligger till grund för viktiga beslut. Om dessa välutbildade människor har svårt med statistiken hur är det då inte för folk i allmänhet? Många har en uppfattning om när det är lämpligt att använda en viss statistisk metod men har däremot ingen aning om de förutsättningar som måste vara uppfyllda. Hur många av era arbetskamrater som inte är statistiker känner till vad residualanalys är för något? Hur ofta redovisas bortfallet och dess konsekvenser för resultatet? Jag är säker på att flera av er av och till har hamnat i diskussioner med era uppdragsgivare om hur den statistiska analysen av ett datamaterial ska genomföras. Många gånger kommer statistikern in i ett senare skede av undersökningen när materialet redan är insamlat och blir ombedd att göra det bästa möjliga av det.

För några år sedan hamnade jag i en roll mellan statistikern och mottagaren när jag för mina dåvarande arbetskamrater skulle förklara ett index som kollegor hade konstruerat. Jag ville peka lite på de problem som kan uppstå när man använder ett index som i det här fallet var underlag för resurstilldelning. Jag tog hjälp av en informatör och bollade med kollegorna för att få hjälp med att formulera presentationen på ett korrekt men ändå begripligt sätt. Jag lade ner mycket jobb på detta men fick hård kritik från kollegorna för att jag hade förenklat för mycket och därigenom delvis förringat betydelsen av deras arbete. Mina arbetskamrater lyssnade uppmärksamt men ställde inte så många frågor. Antingen förstod de allt eller inte så mycket. (Troligtvis det senare.) Jag fick verkligen en tankeställare. Nog hade jag anat att det skulle vara svårt att förklara indexet men inte att det skulle vara så svårt (och ändå hade jag inte visat en enda formel).

Jag har ofta argumenterat för mer grafiska analyser istället för hypotesprövningar. Ett problem är att vetenskapliga resultat nästan aldrig blir publicerade om inte en statistisk analys i form av en hypotesprövning har gjorts.

En grafisk presentation eller sammanställningar i tabeller av data räcker inte. Forskare behöver få in sina resultat i tidskrifterna för att komma vidare i sin akademiska karriär. Jag diskuterade en gång problemet med en bekant med lång erfarenhet inom medicinsk forskning och han menade att många tester görs så att säga "just to please the editors".

Vad kan då vi statistiker göra och hur kan utbildningen förbättras för våra blivande kollegor? En grundläggande förutsättning för bra kommunikation är att förstå mottagaren. Vad vill de veta? Hur förenklad kan en redovisning vara utan att väsentlig information går förlorad? När jag studerade statistik ingick en hel del rapportskrivande och muntliga presentationer men mottagarna var oftast lärare och kurskamrater. Man kanske kan utveckla redovisningsformerna genom att låta studenterna berätta om resultaten från en undersökning för en publik som inte är bevandrad i statistikämnet. Målgruppen kan vara studenter från andra discipliner som till exempel medie- och kommunikationsvetenskap eller journalistik. Även när arbetet är mer matematiskt till sin karaktär kan det vara en utmaning att få lekmän att förstå syftet med det. Någon form av skrivarkurs skulle också kunna erbjudas statistikstudenterna. På motsvarande sätt borde studenterna inom medie- och kommunikationsvetenskap och journalistik ges möjlighet att berätta om sina arbeten där statistik ingår för statistikstudenter. Jag föreslår att statistikinstitutionerna börjar ta kontakt med utbildningarna inom medie- och kommunikationsvetenskap och journalistik och föreslår samarbete. Det är inte omöjligt att en statistikutbildning med mer inriktning på information skulle vara mer lockande för studenterna. Även yrkesverksamma statistiker måste erbjudas vidareutbildning inom informationsområdet. På sikt borde det bidra till att öka förståelsen mellan de som framställer statistik och de som förmedlar den till allmänheten.

*Nils Larsson,
Folkhälsoakademin, Karolinska Institutet*



Medlemstombolan

Från och med förra numret av Qvintensen ber vi registeransvarig att välja en medlem med hjälp av slumpen. På den andra lappen som drogs ur tombolan stod det Anna Blomquist.

Biträdande redaktören

Hej! Du är den andra medlemmen som blivit vald för intervju. Hur känns det?

Hej, det känns kul. Det är en rolig idé. På så vis får man se lite av bredden inom vår förening. Och så får man en chans att läsa om vilka ställen folk hamnar på efter utbildningen.

Vem är du?

Jag heter Anna och är 33 år. Är en Stockholmstjej som sedan ett par år tillbaka arbetar på Storstockholms lokaltrafik (SL). Där jobbar jag med långsiktiga trafikprognoser. På fritiden går jag gärna på konserter och så gillar jag film. Jag löptränar en del också. Vi är ett gäng här på SL som brukar springa ihop. Några halvmaror har det blivit och en helmara. Det var kul, men det gäller att vara väl förberedd.

Berätta lite om ditt jobb. Hur ser en vanlig dag ut?

Jag är ansvarig för en prognosmodell som används när det ska göras konsekvensbeskrivningar av framtida utbyggnader av kollektivtrafiken. Exempelvis så var jag nyligen med och gjorde en förstudie avseende tunnelbana mellan Odenplan och Karolinska. Med hjälp av bland annat befolkningsprognoser som jag hämtar från Regionplane- och trafikkontoret och förväntade åktider som jag hämtar från en annan modell jämför vi åktider för olika färdmedel (bil, kollektivt, cykel och gång). Vi gör jämförelsen dels enligt de förutsättningar som finns idag och dels hur det förväntas se ut år 2030 när den eventuella förändringen av kollektivtrafiken är gjord och med hänsyn till befolkningsprognoserna.

Arbetet innehåller mycket mer än att bara analysera data. Mycket handlar om att stämma av med andra aktörer i länet, till exempel Vägverket, Banverket och kommuner inom området. Den senaste tiden har det varit mycket jobb med den nya regionplanen för Stockholms län. Det är viktigt att vi är med så att det inte planeras exploateringar utan att vi samtidigt utreder kollektivtrafiken.

Så, ja en vanlig dag består oftast av många möten. Vi träffas i arbetsgrupper som rör olika utredningar. Där sitter det med folk med olika intressen, till exempel trafikplanerare. Och är jag inte på möten så sitter jag oftast på mitt rum och jobbar med att "rigga" prognosmodellen med olika scenarion till någon specifik utredning.

Och hur ser en ovanlig dag ut?

Dagarna känns ganska lika. Men en ovanlig dag kanske jag är på ett seminarium på KTH eller SCB. Eller så läser jag in något nytt, en artikel till exempel.

Är du ensam statistiker på din arbetsplats?

Nej, vi är tre stycken här på planeringsavdelningen. Men

vi kallar oss analytiker. Jag känner mig kanske inte riktigt som en statistiker eftersom jag inte direkt arbetar med statistisk metod utan mer med att använda redan tidigare utvecklade metoder.

Vad har du för utbildning?

Jag har läst fristående kurser vid Stockholms universitet och har en magisterexamen i matematisk statistik från 2000. Bland annat så har jag läst demografi. Det var extra kul med lite tillämpningar. Man fick använda teorin lite mer i praktiken då. Sedan gjorde jag mitt examensarbete på en konsultfirma och blev kvar där efter utbildningen.

Vad från din utbildning har du mest användning för nu i arbetslivet?

Det är nog att man tränade sig i problemlösning och logiskt och strukturerat tänkande. Men förstås också att man med hjälp av kunskapen om den bakomliggande teorin kan tolka resultat och har en djupare förståelse för till exempel regressionsanalys.

Finns det något som du gärna läst mer av?

Ja, nu i efterhand har jag insett att det inte hade varit fel om jag hade läst nationalekonomi. Jag har läst lite grann men det hade nog varit nyttigt i största allmänhet med lite mer. Men att ta en kvällskurs nu när man börjat jobba känns inte aktuellt, inte just nu i alla fall.

Har du varit med i någon annan statistikförening före Främjandet?

Ja, jag var med i både Svenska statistikersamfundet och Statistiska föreningen. Och så är jag även med i Demografiska föreningen. De har lite liknande aktiviteter som Främjandet med seminarier och medlemsblad.

Slutligen ska jag ställa en fråga som Ylva, medlemmen som blev intervjuad i förra numret, gav på förslag. Vilka tre egenskaper tycker du utmärker en bra statistiker?

Erfarenhet. Man ska ha sett mycket och jobbat länge för att vara en bra statistiker. Problemen man stöter på i verkliga livet skiljer sig mycket ifrån skolexemplen. Verkligheten är så mycket mer komplicerad. Och för att kunna hantera dessa problem på ett bra sätt så är det till stor nytta om man stött på och löst många olika typer av problem. Sedan ska man förstås också ha en gedigen kunskap i matematik. Att vara programmeringskunnig är också viktigt. Som statistiker hamnar man ofta till slut i situationer där sådana kunskaper är till mycket nytta.

Tack för en trevlig pratstund!

Tack själv! ■

Visualisering på Internet

Visualiseringsverktyg för att studera demografiska data

Mikael Jern, NCVA
Alf Fyhrlund
och Marie Haldorson, SCB

Geografisk informationsvisualisering kombinerad med analytiska verktyg representerar en ny spännande teknik där fokus ligger på interaktivitet och dynamisk utforskning av data. OECD och SCB har observerat ett ökat intresse för den regionala utvecklingen på framförallt europeisk nivå, men även i Sverige. Statistiker behöver nya verktyg för att bättre analysera och förstå de framsteg som görs på regional nivå inom sociala, ekonomiska, utbildnings- och miljörelaterade områden.

Nationellt centrum för visuell analys (NCVA) vid Linköpings universitet har, under ledning av professor Mikael Jern och med ekonomiskt stöd från Stiftelsen för kunskaps- och kompetensutveckling, utvecklat en verktygslåda för analytisk webbvisualisering av geografiska regionala data i nära samarbete med OECD och SCB.

Kommunstatistik på webben

NCVA:s verktygslåda kommer under 2009 att användas i ett stort globalt projekt, *Measuring the Progress of Societies*. I samband med detta projekt släpps programvarans källkod fri. För SCB:s del innebär samarbetet med NCVA nya möjligheter att ge webbesökare ett visualiseringsverktyg för en del av all den kommunstatistik som finns tillgänglig på SCB:s webbplats www.scb.se/.

Figur 1 visar webbprogrammet SCB eXplorers grafiska användargränssnitt med karta, punktdiagram och ett analytiskt verktyg. Diagrammet, kallat parallella koordinater, har utvecklats inom forskningsområdet informationsvisualisering och används också för att jämföra profiler för två eller flera regioner.

I figuren visar punktdiagrammet köpkraft på den horisontella axeln och utbildning på den vertikala. Två kommuner med olika förutsättningar är markerade. Punkternas storlek bestäms av invånarantalet i kommunen och färgen bestäms av köpkraften. Diagrammet parallella koordinater visar samma markerade kommuner som två olika streckade linjer. Värderna för respektive kommun kan läsas i tabellen till höger.

Samarbetet har även lett till utveckling av hur tekniken kan nyttjas för data på låg regional nivå, postnummerområden och kommunernas delområden. Under 2009 ska SCB tillsammans med ett antal kommuner ha ett idéutbyte om hur visualisering kan nyttjas inom den kommunala planeringen, till exempel för analys på bostadsområdesnivå.

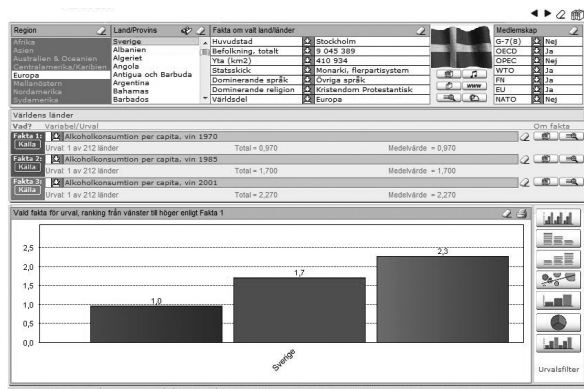
Fler applikationer för visualisering finns redovisade på webbplatsen ncva.itn.liu.se.



Figur 1: Demografiska data för Sveriges kommuner.

Utan att vara statistiker gav vi oss i kast med att utveckla en statistiktjänst för Internet, www.factlab.se/. En tjänst som skulle göra det enkelt och roligt att titta på och arbeta med statistik.

Vad är det då vi utvecklat? Vi gjorde något som vi själva vill ha och som är lätt att begripa. Det är enkelt att klicka med musen och se hur tusentals statistiska uppgifter från hela världen förvandlas till jämförbara parametrar. Tidigare var statistiska data i stort sett omöjliga att jämföra och analysera. Alla databaser är uppbyggda på olika sätt och alla insamlare gör på sitt eget sätt för att gagna sina syften.



Statistiktjänsten är gratis för användaren och finns på webbadressen www.factlab.se.

Det blir inte bara äpplen och päron, utan en hel fruktkorg när man lägger ihop statistik från flera olika insamlare. En metod utvecklades för att kunna förvandla uppgifterna till jämförbara och analyserbara parametrar. Vi kan till exempel snabbt begära att få titta på vilka länder som har störst andel av befolkningen över 65 år och sedan välja att se resultatet som en tabell eller grafisk presentation. För att se om det finns en möjlig relation mellan dessa fakta och hur många läkare det finns per 1 000 invånare, behövs bara en knapptryckning så visas en trendlinje som tyder på det. Det är enkelt gjort att byta ut en eller två variabler.

Olika indelningsgrund

Det finns tusentals variabler att laborera med och de uppdateras och kompletteras fortlöpande med nya. En global tjänst har världens länder som minsta statistiska indelningsgrund och en nationell tjänst har landets kommuner eller motsvarande som minsta indelningsgrund. I den nationella tjänsten för USA används de 3 200 länsområdena (*counties*) som indelningsgrund.

I den nationella tjänsten för Sverige ges en stor mängd information om alla kommuner som kan användas på många sätt. Det går till exempel inte att se att höga betyg i årskurs 9 relaterar till antal lärare per 100 elever, däremot tycks det finnas ett samband mellan höga betyg och om en stor andel av lärarna har pedagogisk utbild-

ning. Det går att söka på vilka kommuner som har flest bränsleslukande bilar och om det finns en koppling mellan dessa och kommuninvånarnas medelinkomst, eller hur stor del av bilarna som ägs av kvinnor. Alla variabler har källhänvisning, och alla definitioner samt eventuella fotnoter finns med liksom en länk direkt till källan.

Statistik i undervisningen

För ett år sedan togs kontakt med Myndigheten för skolutveckling i Göteborg som startat projektet Praktisk IT- och mediekompetens (PIM). I projektet erbjuds alla lärare i kommunerna att gå en utbildning för att bättre lära sig att utnyttja IT i förberedelserna eller undervisningen.

Inledningsvis var det osäkert hur många kommuner som skulle delta i projektet då PIM är förenat med kostnader för utbildning av examinatorer och handledare, men idag är över 55 000 lärare och pedagoger anmälda.

Statistiktjänsten kan användas som en portal där lärare planerar och förbereder lektioner som man delar med sig av till kollegerna, och där svaren finns att tillgå via tjänsten. Eleverna kan enkelt redovisa sina resultat genom att använda ett ordbehandlingsprogram och sedan ett presentationsprogram. Det går också att välja att se materialet i tabellform och använda ett kalkylprogram för att göra ytterligare beräkningar.

Det är viktigt att vi har ett bra samarbete med de stora statistikinsamlarna, både när det gäller den globala och den nationella tjänsten. I Sverige är SCB en mycket viktig samarbetspartner.

Dagens version av statistiktjänsten är den andra betaversionen. Inom kort kommer en ny för PC och Mac med stöd för alla webbläsare. Idag erbjuds full funktionalitet endast för dem som använder sig av PC och webbläsaren Explorer.

Bokrabatt för Främjandets medlemmar

Svenska statistikfrämjandet har i och med samgåendet mellan Statistiska föreningen och Svenska statistikersamfundet tagit över rollen som en av fyra utgivare av *Scandinavian Journal of Statistics*, en roll som Samfundet tidigare hade. Bokförlaget Wiley-Blackwell erbjuder alla medlemmar i Främjandet 20 procents rabatt på statistikböcker. Ange koden SSS om du beställer via Internet, telefon eller post.

Rapporter från konferenser och seminarier

Europeiska industristatistikens årliga konferens

Ann Lindqvist, Scania

Ett nätverk av europeiska industristatistiker möts en gång om året i konferensform. Denna gång samlades vi på University of Economics and Business i Aten. Nätverket har namnet ENBIS (*European Network for Business and Industrial Statistics*) och detta var dess åttonde konferens. Från Sverige är Föreningen industriell statistik medlem i nätverket. I år var vi väl representerade med sex deltagare från föreningens styrelse. Totalt var det drygt hundra statistiker från flera europeiska länder som samlades i Aten. Konferensen pågick i två och en halv dag. I anslutning till konferensen gavs kurser som förlängde evenemanget med två dagar.

Upplägget är ungefär detsamma varje år. Före och efter konferensen ges kurser ideellt, vanligen av konferensens deltagare eller föredragshållare. Under konferensen ges ibland kortare kurser. Till konferensen kommer minst tre inbjudna föreläsare. Fem priser delas ut och dessa renderar i tre föreläsningar. Det är omkring åtta parallella sessioner där man vanligen kan välja mellan tre områden. Varje område har tre föredrag som godkänts av programkommittén. Exempel på områden är försöksplanering, statistisk processtyrning, kvalitetsförbättring (bland annat Six Sigma), ekonomi och marknadsföring, mätsäkerhet och data mining.

Från konsult till effektiv ledare

Min upplevelse är att kurserna i anslutning till konferenserna är mycket givande. Jag och elva andra deltagare satsade därför på en kurs dagen före konferensens öppnande. Det visade sig vara helt rätt. Kursen hade titeln *From Consultant to Effective Leader*. Lärare var Roger Hoerl, General Electric, USA. Han uppmuntrade oss till att inte sitta och vänta på uppdrag utan aktivt propagera för vad vi kan göra med vår kunskap och att ta en ledande roll i projekt. Detta gjorde han utifrån sina egna erfarenheter som chef för en grupp statistiker som han har utvecklat från konsulter till ledande analytiker och chefer. Vi fick i uppgift att beskriva våra personliga mål på fem och tio års sikt och planera en aktivitet att utföra i anslutning till hemkomsten som innebar marknadsföring av våra kunskaper.

Christian Ritter, Ritter and Danielson Consulting, Bryssel, Belgien, har vid en tidigare konferens hållit en kurs om vad man skall tänka på vid presentationer av diagram och tabeller. Nu gav han en kurs vid namn *Advanced Issues in Statistical Graphics* och det var en mycket uppskattad återkomst. Denna kurs var lika bra som den förra, även om en del var upprepning. Ett mycket uppskattat tips berörde en programvara, RExcel, som stöttar grafer gjorda i programspråket R men som möjliggör att graferna visas och justeras i Excel.

Jeroen de Mast, Institute for Business and Industrial Statistics, Amsterdam, höll som inbjuden föreläsare ett mycket uppskattat föredrag under rubriken *Exploratory*

data analysis in quality improvement projects där han jämförde explorativ dataanalys (EDA) med konfirmativ dataanalys, som han använde som ett samlingsnamn på hypotesprövning, skattning och modellering. Det senare området är rikt beskrivet i litteraturen medan det nästan helt saknas referenser till EDA. Han beskrev en process för hur en explorativ dataanalys går till och exemplifierade den med historien om hur John Snow på 1850-talet identifierade källan till ett kolerautbrott i London.

Priser till föredragshållare

Douglas C. Montgomery, Arizona State University, tilldelades "George Box Medal" och höll ett mycket underhållande föredrag. Han berättade hur en grupp som inte hade kunskap om vinframställning köpte en vinodling och med hjälp av försöksplanering lyckades så bra att de vann priser för sina viner. Hans bok *Design and analysis of experiments* bygger till viss del på dessa erfarenheter och använder exempel därifrån.

Priset för bästa föredrag gick till Malin Albing, Luleå tekniska universitet. Malins föredrag handlade om hennes arbete med att utveckla processkapabilitetsindex för Weibullfördelade mätvärden. Ett kapabilitetsindex, som även kallas duglighetsindex, är ett mått på hur pass väl en produktionsprocess lyckas producera enheter som ligger innanför toleransgränserna. Detta genom att, på ett finurligt sätt, jämföra toleransintervallets bredd med storleken på standardavvikelsen. Sådana index för Weibullfördelade data karaktäriseras av att målvärdet är noll och att det endast finns en (övre) toleransgräns. Arbetet har skett i samarbete med Volvo CE i Eskilstuna. Hon har bland annat gjort tillämpningar på kugghjul.



Malin Albing

Detta var höjdpunkterna, men under konferensen fanns naturligtvis många bra föredrag under de parallella sessionerna och trevliga och givande pratstunder under luncher och kaffepauser. Programmet förgylldes också av kvällsaktiviteterna. Efter en intensiv kursdag hann jag och några av kursdeltagarna besöka Akropolis. Konferensmiddagen hölls på restaurangen Dionysos (vinets gud) där fick vi en underbart god middag med utsikt mot ett belyst Akropolis. Maten under konferensen – härliga grekiska rätter – förtjänar beröm.

Nu ser vi fram mot nästa års konferens som hålls på Chalmers i Göteborg. Mer information om ENBIS och den kommande konferensen finns på nätverkets webbplats www.enbis.org/. ■

Surveyföreningens arbetsseminarium om statistik, juridik och etik

David Sundén,
Åsa Greijer,
och Jenny Lagerqvist, Statisticon

Den 11 september 2008 samlades omkring 30 personer i Statisticons lokaler i Uppsala för att delta i Surveyföreningens arbetsseminarium *Statistik – Juridik – Etik*. Inbjudna talare var Sören Öman från Stockholms universitet som talade om personuppgiftslagen, Eva Nilsson från SCB som talade om sekretesslagstiftningen samt Bernhard Huitfeldt från BH Statistik-konsult som talade om etiska riktlinjer för statistiker.



Föredragshållarna på Surveyföreningens arbetsseminarium, Eva Nilsson, Sören Öman och Bernhard Huitfeldt. Foto: David Sundén

Sören Öman inledde arbetsseminariet med att definiera vad en personuppgift är för något: all information som (direkt eller indirekt) kan hänföras till en fysisk person som är i livet. Personuppgiftslagen omfattar all automatiserad och manuell hantering av personuppgifter, det vill säga både i dator och på papper.

Ändamålet viktigt

Grundläggande krav för att få hantera personuppgifter är att ändamålet med hanteringen är preciserat. En konsekvens av det är att man inte vid ett senare tillfälle får återanvända uppgifterna för något annat ändamål än det ursprungliga. Uppgifterna ska vara adekvata, relevanta och inte sparas längre än nödvändigt. Det är tillåtet att hantera personuppgifter om man inhämtat samtycke från de personer som berörs, eller för att utföra en arbetsuppgift som är av allmänt intresse eller för framställning av statistik. Att besvara och lämna in en enkät räknas som samtycke, men om man avser att hämta uppgifter från andra källor, till exempel register, måste uppgiftslämnarna informeras om detta. I princip betyder det att bortfallsanalys inte är tillåten, eftersom samtycke inte finns, men det går att hänföra till allmänt intresse och till att det rör sig om framställning av statistik. Om uppgifterna är avidentifierade är det också tillåtet att hantera dem.

När det gäller känsliga uppgifter (till exempel hälsa, etniskt ursprung, politiska åsikter) gäller särskilda regler. Undantag från dessa är till exempel statistikframställning om allmänintresset klart överstiger integritetsriskerna, om man inhämtat uttryckligt samtycke, eller för forskningsändamål.

Personuppgifter som rör lagöverträdelse får endast behandlas i ett fåtal fall, och då av myndigheter som har tillstånd av regeringen eller Datainspektionen. Även för forskningsändamål kan det vara möjligt att behandla sådana personuppgifter, men då krävs ett etikgodkännande. När det gäller personuppgifter rörande lagöverträdelse räcker det inte med samtycke.

Registrerade personer måste informeras om vem som är personuppgiftsansvarig och om behandlingens ändamål. Man har även rätt att en gång per år få ett registerutdrag med information om vilka uppgifter som behandlats.

Eva Nilsson inledde *Statistik och sekretesslagstiftningen* med att beskriva vilken aktuell lagstiftning som gäller för statistikproduktion. Det finns bland annat en lag som reglerar den officiella statistiken. Normalt gäller offentlighet av statistiska uppgifter, vilket regleras av rätten att ta del av allmänna handlingar. Mot detta ställs förbudet att röja sekretesskyddade uppgifter enligt sekretesslagen. Sekretess innebär förbud mot att röja en uppgift. Det kan vara muntligt, genom en myndighets utlämnande eller via publicering. Viktigt att förstå är att sekretessen inte följer med uppgifterna när de lämnas ut av en myndighet, om detta inte är särskilt föreskrivet i sekretesslagen.

För att en myndighet ska kunna hålla en allmän handling hemlig måste det finnas en bestämmelse i sekretesslagen som det går att hänvisa till. Finns inte det är en handling en offentlig uppgift. Det betyder att hänsyn måste tas till vad uppgifterna handlar om innan det går att avgöra om de är hemliga eller inte. Sekretess gäller uppgifter som avser enskilds personliga eller ekonomiska förhållanden som kan hänföras till den enskilde. Det finns dock undantag – uppgifter som behövs för forskning och statistikändamål, alternativt som inte är direkt hänförliga till den enskilde, får lämnas ut om det står klart att röjandet inte medför skada eller men.

Yrkesetisk kod

Bernhard Huitfeldt fortsatte med föredraget *Etiska riktlinjer för statistiker och några erfarenheter från läkemedelsindustrin*. I föredraget togs de förslag till yrkesetisk kod för statistiker i Sverige upp som utarbetats av en arbetsgrupp i Svenska statistikersamfundet. Yrkesetisk kod för statistiker finns i flera länder, bland andra USA och Storbritannien. Även om dessa koder i många delar

är allmängiltiga och tillämpliga för svenska förhållande valde man att ta fram en yrkesetisk kod på svenska vilken är förankrad hos svenska statistiker.

Förslaget till yrkesetisk kod är uppdelat på fem områden:

Statistikerns ansvar för samhället och allmänheten karaktäriseras av statistikerns roll att ha tillräcklig insikt i för arbetet tillämpliga lagar och regler, främja kompetent användning av statistisk metod, se till att undersökningen inte styrs av att nå ett förutbestämt resultat samt ansvara för att statistiska resultat presenteras på sätt som visar vilka slutsatser som kan dras.

Statistikerns ansvar gentemot uppdragsgivare omfattar statistikerns roll att tillvarata uppdragsgivarens intressen genom att utföra uppdraget omsorgsfullt och vetenskapligt, att avråda från undersökningar där det uppenbart att resultatet inte kommer att bli användbart, att skydda uppdragsgivarens data samt att klarlägga sitt ansvar för genomförande och kvalitet.

Statistikerns ansvar gentemot uppgiftslämnare karaktäriseras av statistikerns roll att skydda uppgiftslämnarnas identitet, undvika onödigt stora urval av uppgiftslämnare, undvika att ta för mycket av uppgiftslämnarens tid i anspråk samt att ge all relevant information om undersökningen till uppgiftslämnaren.

Statistikerns ansvar som chef för andra statistiker omfattar att skapa en arbetsplats som präglas av etiskt och kvalitetsmedvetet arbetssätt, se till att värna den professionella integriteten samt understödja att ny statistisk kunskap som är framtagen av medarbetare offentliggörs.

Statistikerns ansvar för sig själv, sina kollegor och för professionen karaktäriseras av att statistikern strävar efter att utveckla sin kompetens, dokumenterar sitt arbete så att data kan återanvändas samt att statistikern stöder och respekterar sina kollegor.

Föredragen hittar ni i sin helhet på Surveyförningens webbplats www.statistikersamfundet.se/survey under Arkiv/Workshopar. ■

Statistikerträff i Göteborg

Mats Rudholm,
Statistikkonsulterna Jostat & Mr Sample

Onsdagen den 17 september genomfördes den andra upplagan av Statistikerträffen. Även denna gång i Matematisk statistik lokaler på Chalmers. En nyhet var att FMS stod som samarrangör tillsammans med AstraZeneca, Matematisk statistik och Statistikkonsulterna Jostat & Mr Sample. Talarlistan hade utökats och antalet anmälningar var fler denna gång. Ambitionen från förra året – att belysa olika intressanta verksamhetsfält för statistiker – behölls, men flera föredrag hade koppling till statistik och epidemiologi.

Eva Andersson, Arbets- och miljömedicin Sahlgrenska (AMM) inledde med att tala runt ämnet *Mellan exponeringsbedömning och symtomrapportering*. Vid AMM arbetar Eva tillsammans med bland annat yrkeshygieniker, läkare, ergonomer och miljöfysiker. Som statistiker kan man bland annat syssla med att modellera exponering för att avgöra hur stor del av variationen som ligger mellan respektive inom individer och att skatta risker för överexponering.

Åsa Hellqvist, AstraZeneca, hade temat *Mycket snack å lite statistik*. Hon lyfte fram ett antal kunskapsområden som man som statistiker vid ett läkemedelsföretag bör behärska. Viktiga icke-statistiska färdigheter är SAS-programmering, engelska, presentationsteknik samt tekniker för kommunikation och förhandlingar.

Eric Järpe, Högskolan i Halmstad, talade om *Modellering av epidemisk smittospridning*. Erik tog på ett inspirerande sätt upp olika problemställningar inom området, till exempel "Hur snabbt sprids smittan?" och "När kommer toppen?". Han nämnde även andra tillämpningsområden där liknande modeller bör kunna tillämpas, till

exempel vid spridning av datavirus och för att analysera köp- och försäljningsinflenser vid aktiebörser.

Max Petzold, Nordiska högskolan för folkhälsovetenskap, höll ett föredrag med titeln *Epidemiologiska utmaningar i södra Afrika och Asien – och om vilken fördelning flugor har*. Max gav intressanta exempel på metodutmaningar man kan stöta på när det gäller studier inom hälsovetenskaperna i Afrika och Asien. Bland annat fick vi bekanta oss med studier kring kopplingen mellan fågelinfluensa och exponering av smittade fåglar.

Petter Mostad, Matematisk statistik, Chalmers, invigde oss i området *Statistik och rättsgenetik*. Petter berättade om statistiska problemställningar som uppkommer när man via DNA-analys vill uttala sig i faderskapsmål. Andra intressanta och brännbara områden som berördes var identifiering av personer från DNA-spår och problematik kring databaser för DNA-identifikation.

Tommy Johnsson, Institutionen för vårdvetenskap och hälsa, Göteborgs universitet, avslutade med temat *Kunskap och/eller statistik*. Tommy anlade ett filosofiskt synsätt och satte oss på det hala genom att diskutera i vad mån man kan ha någon kunskap och vad som i så fall karaktäriserar kunskap. Tre grundläggande sätt att förvärva kunskap belystes, nämligen deduktion, induktion och abduktion.

Dagen avslutades under lättsamma och trevliga former på Club Avanze där en läcker buffé med passande dryck avnjöts. Vi arrangerar har bestämt att fortsätta med Statistikerträffen. Nästa gång blir i september 2009. Om du känner för att vara med och kanske är intresserad av att hålla ett föredrag, så kontakta oss. ■

Summer School on Stochastic Differential Equation

James Blevin, Uppsala universitet

The 2008 BioMedMath Summer School and Workshop were held in Middlefart, Denmark, August 3–12. Focusing on endocrine systems and neurons, an international team of researchers gave lectures and led workshops to more than 50 young researchers.

An introduction to pharmaco-kinetics and pharmacodynamics in endocrinology was given by Professor Andrea de Gaetano, Biomathematics Laboratory, Rome. He and the other instructors announced a pharmacokinetic challenge that would require mathematical modeling and statistical computing.



Students were given a pharmacokinetic-modelling challenge by Professor Andrea De Gaetano. Photo: James Blevin

Solutions were submitted by four teams of participants. Besides collaborating on a winning solution, doctoral student Edmund Watson, AstraZeneca and the University of Warwick, submitted an individual model using the iconic Danish plaything, Lego!

Motivation for stochastic models

In biology, scientists use ordinary differential equations for population-mean models. In recent decades, more attention has been focused on populations using random effects models: typically a random effect perturbs the population-mean.

In some applications, random effects models have difficulty in accounting for the day-to-day variability within individuals. Not only do measurements vary, but also the rate of change. Such fluctuations motivate stochastic differential equations. In an ordinary differential equation, the coefficients are deterministic; in a stochastic differential equation (SDE), the coefficients are random. By allowing the “rate” parameters to vary, SDE-models account for the fluctuations observed in the repeated measurements on individual subjects. Such fluctuations were apparent in the weekly measurements on pig-weights that were presented with SDE-analysis by Helle Sørensen and doctoral student Anders Bjerring Strathe, Royal Veterinary and Agricultural University, Denmark.

Modelling SDEs requires mathematical statistics. Bernt Øksendal gave introductory lectures on the probability-theory of SDEs and volunteered to give a two hour “crash course” on the stochastic calculus of variations

(“Malliavin calculus”), after students expressed interest over dinner. Michael Sørensen and Martin Jacobsen delivered lectures on estimating-functions, which allow inference for discrete time observations using continuous time models.

SDE-modeling has become much more accessible following the publication of quality software. Seema Nanda and Umberto Piccini lectured on the theory and practice of SDE-modeling and then led several laboratory sessions using Matlab and R. Laura Sacerdote, Gilles Wainrib, and Susanne Ditlevesen explained neuronal models and SDE-inference.

The Summer School was organized by Professors Susanne Ditlevesen and Mikael Sørensen, University of Copenhagen. Ditlevesen maintains a website with lectures at www.math.ku.dk/~susanne/SummerSchool2008.

Talks on research

Many young researchers presented talks on their research. At the Department of Automatic Control, Lund University, doctoral students Fredrik Ståhl and Marzia Cescon are working on the EU-funded project, DIAdvisor – a “personal glucose predictive diabetes advisor” [1]. Cesar Palerm, Medtronic Diabetes, discussed another project, which aims to enable diabetics to sleep without fear of falling into an insulin coma, which can be fatal. Several researchers encouraged me to contact Professor Mats O. Karlsson of the Bio Pharmaceutical Faculty at Uppsala, because of his group’s research in diabetes modeling and experimental design.

In the middle of the conference, there was a bus ride to Odense to see the museum for Hans-Christian Andersen. This excursion also stopped at the Ekeskov Castle in Kvaerndrup, which hosts two creations of Piet Hein, a bamboo maze and Hein’s famous “super-ellipsoid”, the latter form having inspired Sergels Torg in Stockholm.

Statisticians should consider applying to the next Summer School. The 2009 BioMedMath Summer School will be held in Sicily during September and will focus on “Parameter estimation in physiological models”. The associated Workshop will have the theme *Mathematical Modelling of the Glucose/Insulin System and Estimation of Insulin Sensitivity*. Full local support and a contribution to travel expenses will be granted to up to 50 EC-supported participants. The BioMedMath programs 2007–10 are supported by the Marie Curie Program of the European Union. ■

Referenser

- [1] Cescon, M., Ståhl, F., Landin-Olsson, M. (2008). Subspace-based Model Identification of Diabetic Blood Glucose Dynamics. Department of Automatic Control, Lunds tekniska högskola and Lund universitet.

Konferens om administrativa datakällor i Shanghai

Anders Holmberg, SCB

The International Association of Official Statistics (IAOS) är en av sektionerna inom ISI som vartannat år håller en större konferens (de år det inte är ISI-konferens). I år stod världsstaden Shanghai och National Bureau of Statistics of China som värd för konferensen dagarna 14–16 oktober. Tillsammans med tre kollegor från SCB, Folke Carlsson, Åke Bruhn och Mats Bergdahl, samt ytterligare cirka 300 deltagare, deltog jag på en mycket välarrangerad sammankomst. Konferensen hade undertiteln *Smart data, Innovative uses – Reshaping official statistics* och innebörden bakom den var starkt fokuserad på ökad användning av administrativa datakällor och register i officiell statistikproduktion.

Det finns två huvudsakliga drivkrafter till att nationella statistikorganisationer världen över intresserar sig alltmer för datakällor vars ändamål ursprungligen är något annat än att göra statistik. Det första och största är ekonomiskt och det andra är ett större tryck på att ta hänsyn till och minska uppgiftslämnandet.

Stort intresse för registeranvändning

Administrativa källor som underlag för officiell statistik betraktades länge med skepsis och misstro och det var främst i de nordiska länderna och Nederländerna som de användes regelbundet. Numera driver FN frågan aktivt bland sina medlemsländer och FN:s statistikkontor var mycket delaktiga i organiserandet av den här konferensen. På FN:s initiativ hölls parallellt med konferensen ett tre dagar långt seminarium med titeln *The Seventh Management Seminar for the Heads of the National Statistical Offices in Asia and the Pacific* (SIAP). Huvudanledningen till mitt och Folkes deltagande var en inbjudan att hålla en presentation på detta seminarium om SCB:s och Sveriges erfarenheter av användning av administrativa datakällor i statistikproduktion. Vårt intryck efter seminariet är att intresset för detta och för

de registerbaserade ansatser vi har i Norden är mycket stort. Ett kort referat från SIAP-seminariet finns att läsa på www.unsiap.or.jp/article/2008/ms7.html.



Åke Bruhn, avdelningschef på SCB:s processavdelning, var inbjuden till konferensen för att tala på temat *Processoriented Statistical Production*. I bakgrunden syns Börje Dernulf, tidigare SCB och numera konsult, som även deltog på konferensen. Foto: Maria Bergdahl

Konferensen var bättre och intressantare än jag på förhand hade förväntat mig. Presentationerna höll god kvalitet, även om det ibland var svårt att hänga med när det simultantolkades från kinesiska till engelska. Organisationen var felfri, lokalerna bra, maten mycket bra och vädret utmärkt. Luften var behaglig och det var varken varmt eller fuktigt. Inte ens trafiken eller trängseln upplevde jag som särskilt besvärande och ska jag nämna något mindre bra så blir det den långa flygresan via Helsingfors och de hundratals gatuförsäljarna. Program och presentationer från konferensen finns om ni klickar er fram på konferenslänken på IAOS webbplats isi.cbs.nl/iaos/. ■

Hänt och hört

MALIN ALBING, Institutionen för matematik vid Luleå tekniska universitet försvarade den 8 december sin doktorsavhandling med titeln "Contributions to Process Capability Indices and Plots". Opponent var professor Geoffrey Vining från Virginia Tech, Blacksburg, USA. Handledare var Kerstin Vännman och biträdande handledare Bjarne Bergqvist.

JOHAN LYHAGEN, Institutionen för informationsvetenskap, enheten för statistik, Uppsala universitet, utnämndes den 15 november till professor i statistik med inriktning mot ekonometri och tidsserieanalys.

Förstår vi världen?

ONSDAGEN DEN 5 NOVEMBER deltog bland andra Allan Gut, professor i matematisk statistik, Uppsala universitet och Hans Rosling, professor i Internationell hälsa vid institutionen för Folkhälsovetenskap, Karolinska Institutet i programmet *Café Einstein* på kanal Axess TV. Under programrubriken *Förstår vi världen?* diskuterades den föråldrade uppdelningen av vår värld som i- och u-länder. I programtablån utannonserades "Matte handlar inte bara om att räkna utan också att förstå. Så hur vore det om alla svenska medborgare fick en grundkurs i statistik?" För den som vill se programmet finns en länk på webbplatsen www.axess.se. □

Ingeborg Waernbaum disputerade den 12 september 2008 på sin avhandling Covariate Selection and Propensity Score Specification in Causal Inference vid Statistiska institutionen, Umeå universitet. Fakultetsopponent var professor Elena Stanghellini, Perugia universitet. Handledare var Xavier de Luna och biträdande handledare var Göran Broström.

Val av kontrollvariabler i observationsstudier

Ingeborg Waernbaum, Umeå universitet

En central frågeställning inom empirisk forskning är skattning av effekten av en behandling på en responsvariabel, exempelvis effekten av rehabilitering på antalet sjukdagar eller effekten av en arbetsmarknadsutbildning på återgång i arbete. Data av denna sort finns ofta tillgänglig i stora register men problemet i detta sammanhang är att de behandlingar som vi vill undersöka inte är slumpmässigt tilldelade. När man skattar behandlingseffekter i observationsstudier kan man inte särskilja effekten av själva behandlingen från effekten av eventuella skillnader i andra variabler. Skillnader i andra variabler som kan ställa till det för oss är variabler som påverkar både behandlingsvariabeln och responsvariabeln. Om vi kan ta hänsyn till, och kontrollera för skillnader i sådana variabler, kan behandlingseffekten identifieras.

Det finns en etablerad teoretisk modell för så kallad kausal inferens, Rubinmodellen [1, 2]. Den definierar en kausal effekt för en individ som skillnaden mellan två potentiella utfall, $Y_1 - Y_0$, där Y_1 är det potentiella utfallet under behandling och Y_0 är det potentiella utfallet under icke-behandling. Om vi till exempel vill studera den kausala effekten av rehabilitering på antalet sjukdagar för en individ så skulle den enligt Rubins definition vara antal dagar individen skulle vara sjuk om denne genomgått rehabiliteringen minus antalet dagar individen skulle vara sjuk om denne inte genomgått rehabiliteringen. Det fundamentala problemet med att mäta en sådan effekt är att av de två potentiella utfallen som varje individ har så kan endast den ena observeras.

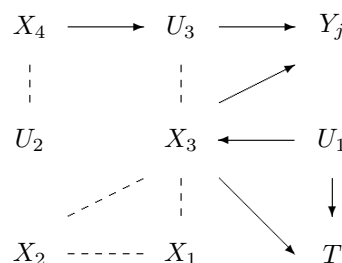
Skattning av effekten

Låt säga att vi vill skatta medelbehandlingseffekten i en population, $E(Y_1 - Y_0)$. Låt T beteckna en binär behandling och $\mathbf{X}$ beteckna en mängd bakgrundsvariabler. Vi kan endast observera $E(Y_1|T=1)$ och $E(Y_0|T=0)$. I en randomiserad studie är behandlingen oberoende av de potentiella utfallen och behandlingseffekten $E(Y_1 - Y_0) = E(Y_1|T=1) - E(Y_0|T=0)$ kan skattas till exempel med differensen mellan medelvärdet i det behandlade stickprovet och medelvärdet bland kontrollerna. För att identifiera behandlingseffekten i en observationsstudie krävs att behandlingen är oberoende av de potentiella utfallen betingat på en mängd bakgrundsvariabler, det vill säga vi måste kontrollera för alla variabler där det finns ett beroende med både behandlingsvariabeln och det potentiella utfallen. På detta sätt kan medelbehandlingseffekten skattas genom $E(Y_1 - Y_0) = E(E(Y_1|T=1, \mathbf{X}) - E(Y_0|T=0, \mathbf{X}))$. Inom det aktuella teoriområdet, kausal inferens i observationsstudier, finns väl studerade stickprovsfunktioner

som används vid skattning av kausala effekter, exempelvis matchningsestimatorer, icke-parametriska regressionssestimatorer och viktningsestimatorer.

Avhandlingen bidrar med nya resultat inom kausal inferens som avser val och användning av bakgrundsvariabler för att skatta en behandlingseffekt. I en utvärderingsstudie är det den aktuella ämnesteorin som styr vilka bakgrundsvariabler vi bör använda, men tidigare forskning har visat att det för vissa stickprovsfunktioner är viktigt att använda så få variabler som möjligt. Denna kunskap tillsammans med tillgången till stora register med många bakgrundsvariabler har lagt tonvikt på behovet av att även kunna använda data för att hjälpa oss att välja variabler när vi inte har tillräcklig ämnesteoretisk kännedom om dessa.

Avhandlingen består av fyra rapporter. I tre av rapporterna behandlas minimala mängder av bakgrundsvariabler så att behandlingseffekten kan identifieras. I de två första rapporterna studeras egenskaper hos minimala mängder av kontrollvariabler i en grafisk representation av Rubinmodellen samt hur dessa kan identifieras. Den grafiska modellen är en riktad acyklisk graf (DAG) som specificeras partiellt. Med hjälp av teori för DAG definieras identifierbara mängder som är minimala för betingat oberoende mellan behandlingsvariabeln och de potentiella utfallen. I den första rapporten [3] antas variablerna $(\mathbf{X}, Y_j, T)$, $j = 0, 1$ i Rubinmodellen kunna representeras av en DAG där en möjlig kant mellan $\mathbf{X}$ och Y_j eller T är riktad från den förra till de senare. Begränsningen till denna riktning representerar att mängden $\mathbf{X}$ är variabler som bestäms före behandlingen. I den andra rapporten [4] generaliseras resultaten till att även inkludera förekomsten av icke-observerade variabler U .



Figur 1: Partiellt specificerad graf över variablerna $(\mathbf{X}, U, T, Y_j)$ i Rubinmodellen där U är variabler som inte observerats.

Den tredje rapporten [5] föreslår en modellfri implementering av algoritmer för variabelselektion som föreslås i de två första rapporterna. Den modellfria metoden som utvecklats av Li med flera [6] har sitt ursprung i dimensionsreducerande metoder inom regressionsana-

lysen. Inom kausal inferens används framför allt icke-parametriska eller semi-parametriska stickprovsfunktioner av behandlingseffekten. En modellfri variabelselektion är därför nödvändig för att behålla den icke-parametriska karaktären på skattningsproceduren.

En funktion av bakgrundsvariablerna som används i många stickprovsfunktioner av behandlingseffekten är *propensity score*, $e(\mathbf{x}) = P(T = 1|\mathbf{X})$, sannolikheten att bli behandlad givet bakgrundsvariablerna. Om behandlingen är oberoende av de potentiella utfallen givet bakgrundsvariablerna så gäller även att behandlingen är oberoende av de potentiella utfallen givet propensity score [7]. För en matchningsestimator betyder detta att istället för att matcha varje individ mot en kontroll med avseende på alla bakgrundsvariabler kan vi matcha individen mot en kontroll med samma värde på propensity score.

I den fjärde rapporten [8] studeras stickprovsfunktioner av behandlingseffekten när en parametrisk modell för propensity score har felspecificerats. Här visas att vissa typer av stickprovsfunktioner är väntevärdesriktiga trots att propensity score har felspecificerats. Modellspecificeringens betydelse för stickprovsfunktionernas varians undersöks också. Här visas att det inte nödvändigtvis är så att den sanna modellen ger den minsta variansen för en stickprovsfunktion av behandlingseffekten. ■

Referenser

[1] Rubin, D.B. (1974). Estimating causal effects of tre-

atments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66**, 688–701.

- [2] Holland, P.W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, **81**, 945–960.
- [3] de Luna, X., Waernbaum, I. (2005). *Covariate selection for non-parametric estimation of treatment effects*. IFAU Working paper, 2005:4, Institute for Labour Market Policy Evaluation, Uppsala.
- [4] de Luna, X., Richardson, T.S., Waernbaum, I. (2007). *Identification of minimal sets of covariates for the non-parametric estimation of an average treatment effects*. Research Report 2007, Department of Statistics, Umeå University, Umeå.
- [5] Waernbaum, I. (2008). *Model-free variable selection in causal inference*. Research Report 2008, Department of Statistics, Umeå University, Umeå.
- [6] Li, L., Cook, R.D., Nachtsheim, C.J. (2005). Model-free variable selection. *Journal of the Royal Statistical Society B*, **67**, 285–299.
- [7] Rosenbaum, P.R., Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55.
- [8] Waernbaum, I. (2008). Propensity score model specification for estimation of average treatment effects. Research Report 2008, Department of Statistics, Umeå University, Umeå.

Benjamin Yip försvarade den 12 september 2008 sin avhandling Statistical Genetics Analysis of Family data vid Karolinska Institutet. Opponent var professor Pak Sham från Hong Kong universitet. Handledare var professor Yudi Pawitan och biträdande handledare professor Paul Lichtenstein, båda från Institutionen för medicinsk epidemiologi och biostatistik vid Karolinska Institutet.

Statistical Genetics Analysis of Family data

Benjamin Yip, Karolinska Institutet

Familial clustering is a common characteristic of genetic related phenotypes, providing vital insights into the etiology of diseases by establishing the relative contribution of genetic and environmental factors. In general, family data analyses involve large datasets with binary outcomes; computation is a stumbling block in the analysis of realistic models.

The overall aim of the thesis is to develop efficient and flexible statistical approaches for population-based family data analysis. The thesis demonstrates that the newly developed methodologies for the analysis of family data are highly flexible and allow straightforward handling of covariates.

In the first paper, hierarchical likelihood is used as an al-

ternative approach to avoid the high-dimensional integration of the random effects. With this approach, joint analysis of the preeclampsia with suggested risk factors and with the random genetic and environmental effects is performed. The results show that being Nordic, older or diabetic are significantly associated with a higher risk of preeclampsia. The results of the joint analysis also show that the genetic factors may only be partly mediated by some known risk factors, such as diabetes and smoking, such that there must be other possibly mechanism involved. These results can not be found in an earlier study of preeclampsia, due to the high-dimensional integration.

Some diseases, like schizophrenia and bipolar disorder,

have similarities in epidemiological features, risk factor and intermediate phenotypes. These similarities strongly indicate that there might be a common etiology. Therefore, in the second paper, the analysis is extended from single binary traits to multivariate binary traits. The model has been applied to study the common etiology of schizophrenia and bipolar disorder. It is shown that the genetic effects explained 65.5 percent of the variability of comorbidity of schizophrenia and bipolar disorder.

Computation more efficient

In paper three and paper four, ascertainment approaches are developed to increase the computation efficiency.

The results from simulation and real data analysis show that the ascertainment approaches provide accurate estimates. Thus, joint analysis of general covariates with genetic and environmental effects is more feasible and efficient by using the ascertainment approaches.

Age-at-onset traits, such as cancer occurrence, were often of interest in genetic studies. Traditionally, frailty models have been used for correlated survival data. However, for family data, the likelihood and estimation for frailty models became very complicated. Instead, in the fourth paper, it is shown that the accelerated failure models are more straightforward to use for family survival data analysis. ■

Chuen Seng Tan försvarade den 5 september 2008 sin avhandling Statistical Methods for Biomarker Discovery in Proteomics vid Karolinska Institutet. Opponent var professor Tomasz Burzykowski, Hasselt Universitet. Handledare var Yudi Pawitan och biträdande handledare var Alexander Ploner, båda från Institutionen för medicinsk epidemiologi och biostatistik vid Karolinska Institutet. Biträdande handledare var även Kee Seng Chia, National University of Singapore.

Statistical Methods for Biomarker Discovery in Proteomics

Chuen Seng Tan, Karolinska Institutet

Surface-enhanced laser desorption and ionization (SELDI) is a promising proteomic technique for discovering biomarkers. There are, however, still problems with the early steps in the analysis of SELDI data, particularly in the detection of peaks in the spectra. With the availability of proteomic and transcriptomic data from the same samples, the search for biomarkers could be enhanced with data integration. But the current data integration approach results in the loss of large amounts of information.

In this thesis, we made improvements to the peak detection step in SELDI by developing the *annotated regions of significance* (ARS) method. It uses a multi-spectral signal detection method, *region of significance* (RS), to identify regions with potential biomarkers. This method had better operating characteristics than existing methods in identifying peaks. Using lung cell line data, at 80 percent sensitivity, the *false discovery rate* (FDR) of existing methods were around 25 to 50 percent, compared to around 8 percent for RS. ARS extracts a peak template from all spectra in the peak region via principal component analysis and fits the template to the spectra. A refinement was made to the estimation of the amplitude via a mixture model. Using patient samples from a clinical study, we showed that ARS detected more peaks and gave more accurate peak quantifications than the stan-

dard method. We implemented ARS as an R package, ProSpect, and also developed a graphical user interface, ProSpectGUI.

Better specificity

Motivated by the performance of ARS in SELDI, we extended ARS to *matrix-assisted laser desorption and ionization* (MALDI) data with isotopic resolution. The extended ARS utilizes the isotopic pattern to filter out peaks which do not adhere to the expected isotopic pattern. Using the spike-in data, we validated the use of the log-transformed intensities for ARS in MALDI. Compared to the standard method, extended ARS generally had better specificity and was better in quantifying the peaks. At low FDR, extended ARS had higher sensitivity than the standard method.

We also investigated the use of *maximum covariance analysis* (MCA) which utilizes all proteomic and transcriptomic information for data integration. Using simulated and real data, we showed that the estimates of the gene and protein pattern-pairs from MCA were consistent and biologically congruent. Therefore the pattern-pairs from MCA could potentially enhance our understanding of the interplay between genes and proteins. ■

Industriell statistik i Göteborg nästa höst

I samband med en konferens som anordnades i Linköping 1999 av kommittén för industriell statistik inom *International Statistical Institute* fick Søren Bisgaard, Isenberg School of Management vid University of Massachusetts Amherst, i uppgift att leda ett arbetsseminarium om hur man skulle kunna skapa ett Europeiskt nätverk för personer med intresse av att utnyttja statistik inom industrin. Søren grep sig verket an med stor energi och *European Network for Business and Industrial Statistics* (ENBIS) skapades.

Nätverket fick en bra start genom det EU-finansierade projektet *pro-ENBIS* som avslutades för ett par år sedan, se Coleman: *Statistical Practice in Business and Industry* (Wiley, 2008). En stabil verksamhet har kommit igång med en årlig konferens och ett stort antal intressegrupper som omfattar allt från undervisning inom industriell statistik till stokastisk modellering, tillförlitlighet och kvalitetsförbättring i industrin. Årets konferens gick av stapeln i Aten och besöktes av något hundratal personer. Från Sverige var vi sex stycken.

Nästa konferens, ENBIS9, kommer att hållas på Chal-

mers i Göteborg den 21–23 september 2009. Den organiseras av den svenska grenen av ENBIS, tillika Föreningen industriell statistik. Kerstin Vännman, Luleå tekniska universitet, är programkommitténs ordförande.

Spännande mötesplats

Vid Chalmers är det främst tillförlitlighetsgruppen inom Gothenburg Mathematical Modelling Centre och Centre for Healthcare Improvement som kommer att vara värdar. Thomas Svensson, SP Bygg- och Mekanik i Borås, är ordförande i den lokala organisationskommittén. Vår ambition är att få till en spännande mötesplats för praktiker, teoretiker och alla andra som är intresserade av att med hjälp av kvantitativa metoder lösa tillämpade problem och finna nya möjligheter inom såväl industri som hälso- och sjukvård.

Mer information om ENBIS9 kommer att finnas på hemsidan för Föreningen industriell statistik och på ENBIS webbplats www.enbis.org/.

Bo Bergman, Chalmers tekniska högskola



Statistik konsulterna
JOSTAT & MR SAMPLE AB

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

Vi arbetar inom områdena:

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

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

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

Konferenskalendarium

Datum 2009	Konferens	Plats	Webbplats
27 mar	Industriell statistiks vårseminarium: Konsten att sätta specifikationer och krav	Uppsala	www.indstat.se/
16–18 apr	Twelfth International Conference on Artificial Intelligence and Statistics	Clearwater, USA	www.ics.uci.edu/~aistats/
19–21 apr	21st Annual Kansas State University Conference on Applied Statistics in Agriculture	Manhattan, USA	www.ksu.edu/stats/agstat.conference
30–2 apr/maj	SDM09: The Ninth SIAM International Conference on Data Mining	Reno, USA	www.siam.org/meetings/sdm09
18–20 jun	Econometrics, Time Series Analysis and Systems Theory: A Conference in Honor of Manfred Deistler	Wien, Österrike	www.ihs.ac.at/etsast
18–20 jun	BISP6 - Bayesian Inference in Stochastic Processes	Bressanone och Brixen, Italien	www.mi.imati.cnr.it/conferences/bisp6.html
21–24 jun	ICSA 2009 Applied Statistics Symposium	San Francisco, USA	icsa2.org/2009
21–24 jun	International Symposium on Forecasting	Hong Kong, Kina	www.forecasters.org/isf/index.html
23–27 jun	18th International Workshop on Matrices and Statistics (IWMS)	Smolenice Castle, Slovakien	www.um.sav.sk/en/iwms2009.html
25–27 jun	United States Conference on Teaching Statistics (USCOTS) '09	Columbus, USA	www.causeweb.org/uscots
30–3 jun/jul	Applied Stochastic Models and Data Analysis (ASMDA)	Vilnius, Litauen	www.asmda.net/asmda2009
1–3 jul	International Conference of Computational Statistics and Data Engineering 2009	London, England	www.iaeng.org/WCE2009/ICCSDE2009.html
20–24 jul	27th European Meeting of Statisticians	Toulouse, Frankrike	http://bs-erc.stat.unipd.it
27–31 jul	The 33rd Conference on Stochastic Processes and their Applications	Berlin, Tyskland	www.okbuero.de/spa/
1–6 aug	2009 Joint Statistical Meetings	Washington, USA	www.amstat.org/meetings/
16–22 aug	International Statistical Institute 57th Biennial Session	Durban, Sydafrika	www.cbs.nl/isi/
23–27 aug	30th Annual Conference of the International Society for Clinical Biostatistics	Prag, Tjeckien	www.iscb2009.info/