

MEDLEMSTOMBOLAN/
MÖT SVEN BERG

SIDAN 6

NOMINERA ÅRETS
STATISTIKFRÄMJARE

SIDAN 19



Qvintensen

MEDLEMSTIDNING FÖR SVENSKA STATISTIKFRÄMJANDET NR 4 2010

TEMA/
NATIONELL STRATEGI
SIDAN 8



**Resultatet
har blivit
ett rama-
skri av in-
dignation
och ilska.»**

Carl-Erik Särndal
SIDAN 16

SURVEYFÖRENINGEN/
**Opinions-
undersökningar
– för-och nackdelar
med webbpaneler** SIDAN 24

HAVETS FÖRSURNING

– data, modeller och statistikerns roller

SIDAN 14

PRICKSÄKRA KUNSKAPER OM IMORGON. BARA SÅ DU VET.

DET FINNS TRE SLAGS LÖGNER SA MARK TWAIN EN GÅNG: LÖGN, FÖRBANNAD LÖGN OCH STATISTIK.

Vi håller inte med. I vår värld finns det tre slags sanningar: fakta, avancerad analys och statistik.

Det är med de grunderna som SAS Institute skapar kunskap och beredskap inför framtidens scenarion. Just lösningar för beslutsstöd med avancerad analys är vi världsledande på.

På sas.com/sweden hittar du alltid aktuella seminarier, webb-seminarier och utbildningar, vare sig du är befintlig eller blivande kund.

Välkommen in.



SAS Institute är ledande på lösningar för beslutsstöd med avancerad analys. SAS Institute, även världens största privatägda mjukvaruföretag, omsatte 2009 2,31 miljarder dollar i 118 länder och har sedan 1976 erfarenhet av att utveckla verktyg och metoder som låter stora organisationer lära av sin historia, mäta och kommunicera pågående aktiviteter och inte minst att skapa insikt om framtiden. Världen runt har SAS Institute totalt gjort 45 000 kundinstallationer, bland annat i 92 procent av Fortune 500-företagen. I Sverige startade SAS Institute AB år 1986 och har idag drygt 100 anställda på kontoret i Stockholm. Bland de svenska kunderna finns landets mest betydande företag och organisationer. För mer information besök: www.sas.com/sweden



Test: När föddes han?



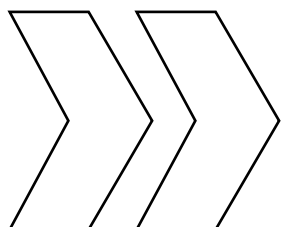
Dagens miljöproblem är mer komplexa och storskaliga.



Carl-Erik Särndal.

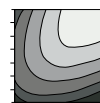


Korskullen i Litauen.



Statistikern behöver inte fastna i traditionella fack»

Anders Grimvall
och Anders Omstedt, sidan 14



Fasta spalter

I HUVUDET PÅ EN REDAKTÖR	4
ORDFÖRANDEN HAR ORDET	5
KALENDARIUM	30

Våra föreningar

SURVEYFÖRENINGEN

Ordföranden har ordet	20
Register-based statistics	20
Registerstatistik, sannolikhetsurval och bortfall	22
Opinionsundersökningar	24

FÖRENINGEN FÖR MEDICINSK STATISTIK

Ordföranden har ordet	25
Vårmöte: Vad är gemensamt?	26

CRAMÉRSÄLLSKAPET

Ordföranden har ordet	27
-----------------------------	----

INDUSTRIELL STATISTIK

Ordföranden har ordet	28
Fortsättning: Fungerar/fungerar inte	29

Ansvarig utgivare

Hans Nyquist

Redaktör

Dan Hedlin, 070-578 4334

Ingegerd Jansson (biträdande)

Redaktion

Hans Ekholm, Främjandet

Alf Fyhrlund, Främjandet

Marie Linder, FMS

Johan Lyhagen, Främjandet

Jonny Olofsson, Surveyföreningen

Lars Söderström, Industriell statistik

Adress

Qvintensen, c/o Dan Hedlin

Runbyvägen 31, 194 44 Upplands Väsby

E-post qvintensen@gmail.com

Produktion

Form och redigering: Mezzo Media AB

Tryckeri: Trydells Tryckeri AB

Annonser

Annonser i Qvintensen bokas med redaktören.

Annonsskick per e-post bokas med Svenska statistikfrämjandets sekreterare. Priset är 5 000 kronor för institutionsmedlem eller motsvarande, och 7 500 kronor för övriga annonsörer. På alla priser tillkommer 25 % moms.



SVENSKA STATISTIKFRÄMJANDETS

STYRELSE

Ordförande

Hans Nyquist, 08-16 11 32,

hans.nyquist@stat.stat.su.se

Vice ordförande

Sune Karlsson, 019-30 12 57

Sekreterare

Chandra Adolffson

Skattmästare

Jonas Selling

Klubbmästare Therese Andersson

Surveyföreningen

Karin Dahmström, 08-36 19 43

FMS Mats Rudholm, 031-703 73 72

Industriell statistik

Ann Lindqvist, 070-297 55 58

Cramérssällskapet

Jan-Eric Englund 040-41 52 36

Övriga ledamöter

Henrik Hult

Andreas Nordvall Lagerås

Adress

Svenska statistikfrämjandet

c/o Chandra Adolffson

Statistiska centralbyrån

701 89 Örebro

E-post sekrfram@gmail.com

Webbplats www.statistikframjandet.se

Fantastiska tillämpningar



Det här numret är en hyllning av de praktiska tillämpningarna. Hur säkerställer man att skruvarna i konstgjorda öron sitter fast? Hur låter man patienter med Parkinsons sjukdom slippa åka in till sjukhuset för att testa sin rörelsekontroll? Hur ser man om vattnet i Östersjön försuras?

”När fakta saknas är alla åsikter lika mycket värda”, säger Liselotte Jansson. Det låter trevligt att alla åsikter ska vara lika mycket värda. Men om vi har åsikten att torsken i Kattegatt är lika stor nu som den alltid har varit mot åsikten att den har blivit avsevärt mindre i storleken under de senaste 40 åren, är det inte bra att det finns statistik med vars hjälp felaktiga, osanna uttalanden kan förkastas? Alla åsikter är inte lika mycket värda.

»Statistiker kan vilja hålla empirin helig eftersom den är deras levebröd.»

Alla som har arbetat med verkliga data har stött på fel i data. Ibland framträder felet först efter tålmodig analys. Karl Wahlin lade märke till överraskande samvariation mellan vattendrag som inte borde ha så mycket med varandra att göra. Läs om hans märkliga upptäckt en dag vid datorn.

En känslig fråga för statistiker är hur mycket man ska tro på vad empiriska data säger och hur mycket man ska lita på de modeller som naturvetare ställer upp. Statistiker kan

vilja hålla empirin helig eftersom den är deras levebröd. I artikeln om havets försurning går Anders Grimvall och Anders Omstedt in på den frågan.

Statistiska centralbyråer i demokratier fungerar i stort sett utan oönskad politisk styrning. Vilka delar av samhället som ska belysas med statistik är självklart en politisk fråga. Hur tung obligatorisk uppgiftslämnarbörd som kan läggas på individer, företag och organisationer är också en fråga för folkvalda. De statistiska metoder med vilka undersökningarna genomförs är det inte. Läs om bråket om Kanadas folkräkning som ledde till att generaldirektören för Kanadas statistiska centralbyrå tog sin mats ur skolan.

Som vanligt innehåller Qvintensen mycket mer! Stort tack till alla – författare, redaktionskommitté, biträdande redaktör, fotografer, layoutare, tryckare – som bidragit till detta nummer!



DAN HEDLIN

FOTO: MONICA HOLMBERG

ORDFÖRANDEN HAR ORDET

PÅ GÅNG I SVENSKA STATISTIKFRÄMJANDET

Först har jag äran och glädjen att meddela att Svenska statistikfrämjandets sekreterare, Chandra Adolfsson, har fött ett gossebarn. Ett stort grattis till Chandra! Detta innebär att Chandra behöver trappa ner lite på sitt sekreteraruppdrag. För den skull har vi utökat styrelsen med Paulina von Rahmel. Paulina kommer nu tills vidare att utföra sekreteraruppdragen.

Jag ser Främjandets viktigaste roll som paraplyorganisation till de olika medlemsföreningarna, dvs. att stötta medlemsföreningarna på olika sätt, bland annat administrativt och ekonomiskt. En viktig del i detta är att kunna erbjuda en aktuell och uppdaterad webbplats. Att åstadkomma detta har visat sig vara svårare än beräknat. Det har definitivt varit svårt att uppdatera Statistikfrämjandets webbplats och hålla den aktuell. För att komma till rätta med problemen har styrelsen därför vidtagit några åtgärder. En åtgärd är att flytta webbplatsen till ett nytt webbhotell som är lättare åtkomligt för oss som ska skriva på den. En annan åtgärd är att lägga upp den med en ny och mer lättanvänd teknik. Förhoppningen är att vi nu lättare ska kunna komma åt webbplatsen och enkelt göra löpande uppdateringar. I ett första steg överfördes Statistikfrämjandets webbplats och därefter flyttades medlemsföreningarnas webbplatser. Jag är övertygad om att det nu finns många idéer och åsikter om hur webbplatsen kan göras bättre. Alla förbättringsförslag är välkomna. När det gäller förslag om Statistikfrämjandets webbplats kan de lämnas till mig eller till sekreteraren. Se sidan 3 för kontaktuppgifter. När det gäller förslag om medlemsföreningarnas webbplatser kan de med fördel lämnas direkt till medlemsföreningarna.



Hans Nyquist,
ordförande.

**»Det bästa
hos en förening
är medlemmarna»**

Det bästa hos en förening är medlemmarna, i alla fall tror jag att det är så i fallet Svenska statistikfrämjandet. Av det bör ju också följa att ju fler medlemmar vi blir, desto bättre blir föreningen. Jag skulle därför vilja uppmana alla att hjälpa till med rekrytering av nya medlemmar. Faktum är att om varje medlem bidrar med en ny medlem blir vi dubbelt så många. Lärare på statistikkurser, berätta för era studenter om vår förening! Ni som jobbar i företag, berätta för era kollegor om Statistikfrämjandet! Kom dessutom ihåg att den bästa tiden att gå med i en förening är i början av ett år, det är ju då man får mest valuta för medlemsavgiften. Det finns en broschyr om Statistikfrämjandet att dela ut och affischer att sätta upp.

Vid det senaste årsmötet beslutades att frågan om eventuell auktorisering av statistiker ska utredas. En sådan kommitté är nu tillsatt och har börjat sitt arbete. Kommittén består av Mats Rudholm (sammankallande), Lilli Japac, Mikael Möller och Magnus Persson.

Statistikfrämjandets nästa årsmöte kommer att äga rum den 24 mars i Örebro. Dagen innan, den 23 mars, är reserverad för medlemsföreningarnas årsmöten och andra aktiviteter. I samband med årsmötet arrangeras vårkonferensen som vi nu håller på att planera. Temat kommer att kretsa kring statistikens historia, speciellt dess utveckling i vårt land.

En ständig punkt på vårkonferensen är att presentera årets statistikfrämjare. Jag vill uppmärksamma alla på att det nu är möjligt att lämna förslag till mottagare av denna hedervärda titel. En "blänkare" om detta finns på sidan 19 i detta nummer av Qvintensen.



■ **Sven, hur kom det sig att du intresserade dig för matematik och statistik?**

– Jag ville gå latinlinjen. Men det fick jag inte för rektorn som sa: ”Berg kan räkna. Han ska gå reallinjen.”

■ **Hur kom det sig att du började med statistik senare på universitetet?**

– Jag läste tre betyg matte och behövde fylla ut med något annat för att få examen. Jag funderade på att satsa på matte helt och hållet, men var tvungen att läsa andra ämnen, och jag hade inte någon väldigt påtaglig mattebegåvning. Mekaniik provade jag på men statistik var roligare. Sen fortsatte jag med nationalekonomi, men då var jag redan anställd som amanuens i statistik.

■ **Varför fortsatte du med forskarutbildning då?**

– Det fanns mycket jobb på institutionen på den tiden och jag behövde inte flytta. Jag blev kvar i Göteborg i tio år efter examen. Jag tänkte bli lärare. Men jag kom aldrig ut ur universitetsvärlden. Ja, jag blev ju lärare ändå, fast på universitetet. Mitt första jobb var en prope-
deutisk kurs i matte för samhällsvetare. Jag tog över den kursen efter Lennart Råde, som var en skicklig pedagog. Det var runt 1960.

– Vi var en stor institution på 60-talet i Göteborg. Sen dess har statistik gått tillbaka som självständigt ämne i Göteborg. Vi hade ett femmannalag i korpfotbollen och vann serien ett år. Det var en rolig tid.

■ **Hur har undervisningen i statistik ändrats sen dess?**

– Den stora förändringen är bättre och bekvämare datahantering. Då hade vi klumpiga mekaniska apparater. Det var så tråkigt, det kunde man inte plåga studenterna med för mycket. Det är stor skillnad också i forskarutbildningen och forskningen också. Man behöver inte alltid ägna sig lika mycket åt matematik, man kan simulera mer idag.

– Det var en stor expansion av statistikämnet på 60-talet. Till exempel införde Ulf af Trolle statistik i undervisningen på Handels i Göteborg när han var rektor där. Han trodde mycket på statistik. Sen lades Handelshögskolan ner 1967 och Ulf af Trolle blev senare mer känd som ”företagsdoktor”.

■ **Vad gjorde du efter 60-talet?**

– Jag var två terminer i Umeå. En termin var jag vikarie för Carl-Erik Särndal. Umeå var för mörkt och kallt för mig som kom söderifrån. 1971-72 var jag på SCB. Där höll jag på med skördeuppskattningarna, alltså den urvalsundersökning där man mätte skörden av olika

»Jag har haft en bra resa»

MEDLEMSTOMBOLAN/*Sven Berg*

Vi drar en person ur medlemsregistret slumpmässigt med lika sannolikhet för alla. Denna gång hamnade slumptalsgeneratoren på Sven Berg.



Hur blir skörden?

grödor. Sen fick jag ett lektorat i Lund. Där blev jag så småningom docent och professor. 1999 gick jag formellt i pension. Jag har arbetat för Linköpings universitet dessutom. Men nu har jag rensat ut tjänsterummet.

■ **Hur var SCB?**

– Det var intressant och roligt. Jag hade skrivit avhandling i urvalsundersökningar så skördeuppskattningarna, med dess urval i flera steg, passade mig. Jag tycker att det är roligt att skriva. SCB hade nästan ingen dokumentation av skördeskattningarna alls, så de blev glada att jag skrev sådan dokumentation.

– Det har varit bra på alla ställen där jag varit. Jag har haft en privilegierad tillvaro som jag är tacksam för.



Sven Berg. Bild från 1999.

■ **Du fick Linköpings universitets förtjänst-medalj.**

– De måste ha tyckt att jag gjort något som var bra. Det var väldigt högtidligt.

■ **Hur kom det sig att du började forska på beslutsmodeller?**

– Statsvetare hade sådana problem. Social choice kallas det området. Jag såg en möjlighet att tillämpa enkla urnmodeller.

■ **Hur kom du på det?**

– Det kommer jag inte ihåg. Urnmodeller är enkla men samtidigt flexibla så de täcker många situationer trots sin enkelhet. Man låtsas att man drar en kula ur en urna, och nästa kula kan bero av vad man fick förra gången. Det är enkla modeller, men mycket kan liknas vid sådana modeller.

■ **En situation du har studerat är en grupp människor, till exempel en jury, som har tydliga personliga preferenser. När de röstar kanske en del av dem röstar i strid mot sina preferenser. Om det finns ett starkt beroende mellan hur de röstar kan jag förstå att resultatet hamnar långt från jurymedlemmarnas preferenser, till exempel om en jurymedlem tar över helt så att alla röstar likadant oavsett vad de egentligen tycker. Men i en artikel kommer du fram till att negativ korrelation kan vara bra. Kan du förklara det?**

– Ja, modellen faller faktiskt ut så. Men utrymmet för negativ korrelation är så litet att i praktiken kan man säga att oberoende mellan beslutsfattarna är bäst.

■ **Vilka andra tillämpningar finns det?**

– Göran Ejlerstsson arbetade på en vårdcentral här utanför Lund. De var intresserade av att mäta kontinuiteten, alltså om man fick samma läkare vid varje besök. Man kunde likna det vid en urnmodell där läkaren dras med någon sannolikhet ur en grupp av läkare. Det blir en enkel parameter som ger ett mått på hur god kontinuiteten är. Det var roligt att se att en så enkel modell med en liten parameter kunde bli ett enkelt index som sammanfattade det man var intresserad av i ett enda mått.

■ **Du var med på Högskoleverkets rapport om statistikämnet år 2006.**

– Det fanns en diskussion om statistikämnets profil. Vi såg en tendens att man anpassade sig

mycket efter ekonomiutbildningarna. Fördelen är att det är stora utbildningar så man kan anställa många lärare. Men man kanske tappar andra sidor av statistikämnet som är roliga. Det fanns två personer från Finland i arbetsgruppen. De hade svårt att förstå att vi i Sverige har ett ämne som heter matematisk statistik och ett som heter statistik.

■ **Ni skriver att det är viktigt för universitetslärare att få utrymme för forskning. Varför är det viktigt?**

– Det är viktigt om man ska utvecklas! Om man undervisar på heltid, med tyngdpunkten på A-nivå, som statistikämnet ser ut, så utvecklar man inte sina statistik-kunskaper. Det finns ett alternativ till forskning och det är undervisning på högre nivå. I Borlänge hade de många små kurser i statistik, till exempel för sjuksköterskor. Men de hade också ett väldigt fint masterprogram, som var underfinansierat. Men det var utvecklande för alla inblandade studenter och lärare. Lund har till exempel en kurs i finansiell ekonomi.

■ **Mätningar som görs pekar på att mattekunskaperna har blivit sämre bland studenter.**

– Ja, så är det tyvärr. Man kan inte komma ifrån att det behövs en del matte för statistik.

■ **Vad kan man göra åt det?**

– Det vet jag inte. Jag hjälper min sondotter, som är 17 år, med matten. Jag tycker att läroboken är så tråkig. Det är tal av typen minus 1 gånger minus 4 sida upp och sida ner. Matematik kan vara så roligt, och så kan det göras så vansinnigt tråkigt också.

■ **Vad hade du gjort annorlunda om du hade fått börja om igen?**

– Jag vet inte om jag gjort något annorlunda. Jag ångrar inget. Jag skulle kanske ha tjänat mer på annat håll, men jag hade inte varit lyckligare.

■ **Vad gör du idag?**

– Jag är intresserad av språk, jag läser mycket. Jag är intresserad av romanska språk, franska, italienska och spanska.

■ **Vilka råd skulle du vilja ge till dem som är unga idag och funderar på om de ska satsa på statistik?**

– Det skulle kanske vara att det finns goda jobb- och karriärmöjligheter. Det finns också

relativt få studenter. Om man kombinerar med något, t.ex. nationalekonomi, finns det väldigt goda möjligheter. Även om man tröttnar på statistik kan man ta med sig de kunskaperna till alla andra ämnen. Det finns mycket roligt man kan göra med statistik i statskunskap, eller sociologi, i nästan alla andra discipliner. I Lund är ekonomisk historia framstående, där behöver man statistik. Goda kunskaper i statistik är aldrig bortkastade.

■ **Ska man satsa på någon tillämpning eller på grunderna?**

– Svårt att säga. Det är bra att skaffa sig en viss bredd.

■ **Gäller det också doktorander?**

– Ja, inom ramen för kursdelen kan det vara bra att gå till andra institutioner.

■ **Vilka råd har du gett dina egna doktorander?**

– De har varit så duktiga och självständiga att de inte behövt så många råd. Jag har uppmuntrat dem att publicera sin forskning. Och att komma ut på konferenser. De statistiska institutionerna i Sverige är så små att man träffar för få människor där.

■ **Har du själv haft omfattande samarbete?**

– Javisst. Ove Frank, min företrädare på professuren, väckte mitt intresse för kombinatorik. Också med nationalekonomer och statsvetare, särskilt i Israel, där forskar man mycket på området beslutsfattande i grupp. Det är väldigt roligt att samarbeta internationellt. I Caen har jag varit en del. Staden ligger i Normandie och blev totalt sönderbombad 1944. När man samarbetar internationellt är man inte statistiker. Man bryr sig inte om vilken bakgrund folk har, man har ett gemensamt intresse som förenar. Det är inte lika lätt och självklart att samarbeta i Sverige över institutions- och fakultetsgränser. Det är min erfarenhet. Men det kanske beror på mig, att det är jag som är knepig.

Två eller tre gånger under alla år har jag blivit otrevligt bemött. Det har varit när jag har haft kritiska synpunkter på avhandlingar i andra ämnen, då jag suttit i betygsnämnder. Det tas inte alltid väl emot. Men jag har blivit väldigt väl bemött nästan alltid. Jag har haft en bra resa.

DAN HEDLIN

Vilken serie är slumpmässigt genererad och vilken är det inte?

SERIE 1

```
01101101100011100000101011111110101010010110010110  
011101000110000011011010011101011110111110000010  
11000111110100100011101001010110110011100110110100  
01011111100000010001011011011010111001111011001100
```

SERIE 2

```
10101101001111001010101001100100110110101010100011  
01101010100101101001001011010011000001011010100100  
10110100010101101011011101101010100010101101110101  
11010110101001011001010101100101010101001011000110
```

Bild 1.

BEHOVET AV NATIONELL STRATEGI FÖR STATISTISK KOMPETENSFÖRSÖRJNING

Möte i riks

Mötet organiserades av Anders Grimvall tillsammans med kollegor vid Linköpings universitet. Värd var Johan Löfstrand, riksdagsman för Socialdemokraterna. Anders Grimvall gav exempel på slående statistik: ett av dem var torsken i Kattegatt. Där har de största exemplaren av torsk *halverats* i storlek under de senaste 40 åren enligt Fiskeriverkets statistik.

»När fakta saknas är alla åsikter lika mycket värda«

Bengt Westerbergs försvarstal för statistik, matematik och ett jämlikt samhälle återges kortfattat på sidan 11.

Liselotte Jansson är vd för den svenska delen av SAS Institute. Hon berättade att på den tiden hon arbetade i finansbranschen gick man på magkänslan. Hade hon vetat det hon vet idag om statistik hade hon agerat annorlunda. Det hade även andra gjort: det fanns information om vad som höll på att hända strax före finanskrisen. Nu förstår ledningarna på banker att statistik och analys av statistiska data är en ledningsfråga. »När fakta saknas är alla åsikter lika mycket värda«, påminde Liselotte Jansson om.

Optimering – ett starkt område

I företag idag går man från att framställa rapporter till att istället analysera skeenden. Rapporterna ger i sig ingen insikt om vad framtiden har att komma med. Optimering är ett annat område

som kommer starkt. I detaljhandeln handlar det om hundratals miljoner kronor som kan sparas i lageroptimering.

För att få studenter intresserade av statistik tror Liselotte Jansson att vi måste visa på den konkreta nyttan. Till exempel kan vi visa att man genom att titta på pengatransaktioner kan identifiera mönster som tyder på penningtvätt.

En Estonia i veckan. Det hemska datumet då Estonia sjönk var 28 september 1994. Men hur många vet att samma antal människor som fanns på Estonia avlider i Europa varje vecka pga av felaktig sjukvård? För att komma till rätta med det krävs fakta, dvs statistik, påpekade Liselotte Jansson.

Machine learning

Mattias Villani är statistiker på Sveriges Riksbank och Stockholms universitet. Han gav många exempel på hur statistik idag kan användas till



Bild 2. Kenneth Carlings tillämpningsområden finns bokstavligen utanför fönstret.

dagshuset

sådant som man inte kunde göra för bara några år sedan. Många problem som man stöter på ser inte ut som traditionella statistiska problem. Ibland finns andra sätt att hantera dem, men då förlorar man resonemang om slumpens effekter i bland annat felkällor och samband. Ett exempel är pixelprediktion. I delvis förstörda bilder kan borttappade pixlar predikteras. Eller prediktion av vad en människa säger, om meddelandet uppfattats bara till viss del. Det finns en diskussion om statistiska metoder är användbara i den här typen av problem eller om problemen hanteras bättre med algoritmer där t.ex. en från början löst definierad funktion anpassas till data (se t.ex. "Statistical modeling: two cultures" av Breiman, Statistical Science, 2001). Sådana algoritmer har använts i synnerhet för stora och komplexa datamängder. Det kallas ibland för machine learning. Mattias Villani tyckte att man ska förena machine learning med statistisk stringens, och han förordade en Bayesian ansats i detta arbete. Sverige är väl skickat för framgångsrikt arbete inom det området: det finns spetskompetens i områden som datavetenskap och hantering av stora datamängder.

Vårt opålitliga omdöme

Olle Häggström, Chalmers, visade två serier med 0:or och 1:or. Vilken är slumpmässigt genererad och vilken är det inte? Se Bild 1. Den professionella publiken identifierade omedelbart

serie 2 som icke-slumpmässig. Den alternerar för ofta mellan 0:or och 1:or. En tankeväckande lek är att försöka ge intervall som med 90 % säkerhet täcker det korrekta svaret. Läsaren kan prova sin egen förmåga på tio frågor hämtade från Financial Times, översatta och återgivna på nästa sida. Båda exemplen är statistiska problem och det svaga resultat som vi människor brukar åstadkomma i den här typen av uppgifter visar att vårt mänskliga omdöme är opålitligt. Vetenskap kan ses som ett systematiskt försök att komma runt vårt opålitliga omdöme.

Vi har en explosiv ökning av data, och framtida framgångar hänger rätt mycket på hur vi kan hantera data, menade Olle Häggström. Man kan till exempel tänka på de stora datamängder som säger något om hur klimatet på jorden utvecklas och hur viktigt det är att den analysen görs rätt. De flesta forskargrupper skulle behöva ha kontakt med en statistiker. Men då behöver man mångdubbla antalet statistiker.

Utanför fönstret

Kenneth Carling, Högskolan i Dalarna, visade på olika typer av problem där statistikern kan göra nytta. De tillämpningsområden han talade om finns bokstavligen utanför fönstret (se Bild 2). Det handlade om så spridda problemområden som medicinering av Parkinsonpatienter, leveranseffektivisering vid sågverk, ökad stålproduktion och resultatrankning vid ungdomstävlingar

i motionsidrotter. Patienter med Parkinsons sjukdom har svårt att ta sig till sjukhuset. Istället kan de numera använda ett provinstrument hemma och skicka data till läkaren. I statistiska försök har dessa data utvärderats mot mer traditionell expertbedömning och visat att patienterna inte behöver åka in till sjukhuset lika ofta.

Outokumpu valdar ut tunt stål. Deras fråga var om valsverken kan köras fortare än det tempo kvalitetsmanualen föreskriver. Ett destruktivt försök (stålet förstördes i experimenten) visade att man kunde köra fortare och Outokumpus tekniske direktör fattade beslutet att så också skulle ske. Stål är dyrbart: försöket kostade miljontals kronor i förstört stål och utebliven produktion. Därför ställdes höga krav på effektiv försöksplan.

Henrik Brandes, CMA Research AB, gav en intressant inblick i statistikerns roll i kunskapsföretag. CMA Research startade 1995 som det första dotterbolaget till ett universitet i Sverige och är idag en av Sveriges främsta företag inom marknadsundersökningar. Man anställer gärna folk som är som en själv, sa Henrik Brandes. Men under åren har han märkt hur oumbärliga statistiker är i ett företag som CMA Research. Man måste "ner i grottan", dra urval på ett optimalt sätt, samla in data och analysera dem – och det är bara statistiker som kan det.

Sverige har unika förutsättningar

Juni Palmgren, Stockholms universitet och Ka- »»»

»»» MÖTE I RIKSDAGSHUSET

rolinska Institutet, talade om statistiken som ett verktyg för att kartlägga och förstå sjukdomsspridning. Hon visade exempel på statistiska metoder som dragit nytta av datorernas explosionsartade hastighetsökning. Beräkningar som förr tog år och kostade miljontals kronor kan numera göras på den enskilde forskarens dator över natten. Men det krävs kompetens för att driva storskaliga försök i stora forskargrupper.

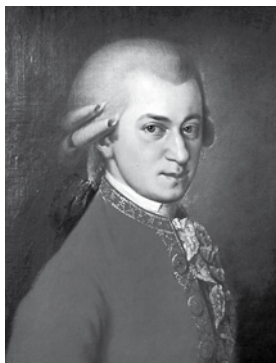
Många utbredda sjukdomar är komplexa: de beror på komplicerade genetiska faktorer och på livsstil. För att studera sådana sjukdomar behövs stora befolkningsstudier med både sjuka och friska. Eftersom man studerar många genetiska varianter är det praktiskt att arbeta tillsammans internationellt. Sverige har unika förutsättningar med befolkningsregister, sjukvårdens kvalitetsregister och personnummer som gör det lätt att matcha individer i olika register. Sverige har också spetskompetens inom genetik, påpekade Juni Palmgren.

En obehaglig tendens

Den avslutande diskussionen handlade bland annat om skolan. Bengt Westerberg sa att de som känner motstånd mot matematik i skolan knappast börjar med statistik på universitetet. Vi vet inte varför både intresse för och kunskaper i matematik minskar i Sverige. Detsamma gäller en del andra ämnen, som kemi. Det är en obehaglig tendens i samhället. Det finns andra länder som lyckas bättre, så det måste finnas något att lära, underströk Bengt Westerberg.

Diskussionen slutade konstruktivt. Ett informationsblad om statistikstudier ska tas fram, med inspiration från det blad som Linköpings universitet har goda erfarenheter av. Tomas Boding på SAS Institute, Lars Wahlgren vid Lunds universitet och Gunbritt Strömwall, SCB, gör det tillsammans med Karl Wahlin, som arbetar på Linköpings universitet. Dessutom ska Kenneth Carling, Anders Grimvall och Liselotte Jansson ta fram ett förslag till "kunskapslyft" för dem som redan arbetar med dataanalys i arbetslivet men som saknar adekvat utbildning.

DAN HEDLIN
KARL WAHLIN



Mozart.



Luna.



Nilen.

Övertro på sin egen förmåga

Här kommer en ovanlig frågesport. Som svar på varje fråga, ge ett intervall som du bedömer täcker det korrekta svaret med 90 % sannolikhet. Googla inte! Frågorna kommer från en artikel i the Financial Times.

1. Hur gammal var Martin Luther King vid sin död?
2. Hur lång är Nilen, i km?
3. Hur många länder ingår i OECD?
4. Hur många böcker består Gamla Testamentet av?
5. Vilken är månens diameter, i km?
6. Hur mycket väger en tom Boeing 747, i kg?
7. Vilket år föddes Mozart?
8. Hur många dagar varar en asiatisk elefants dräktighet i genomsnitt?
9. Hur lång är sträckan London – Tokyo, i km?
10. Hur djupt är det på det djupaste stället i Stilla Havet, i meter?

De rätta svaren finns på sidan 19.

DÄRFÖR VALDE VI ATT BÖRJA

Eftersom många är bekymrade över den svaga rekryteringen till statistikutbildningarna och vad man kan göra åt det, gick frågan till tre Linköpingsstudenter.

Lotta Järvstråt:

– Jag valde att läsa på Statistik- och dataanalysprogrammet eftersom jag är fascinerad av hur man kan ta fram information ur bara en massa siffror. Tidigare läste jag bara matematik och jag ville ha något att tillämpa den på. Jag ville också ha ett tillämpningsområde som det går att få jobb inom.

Statistisk allemansrätt

”Statistisk allemansrätt”, sa Bengt Westerberg – han syftade på att alla borde ha tillgång till det statistiska landskapet: förståelsen av statistik och tillgången till kunskaperna.

Ett exempel är de rapporter och mer eller mindre välgrundade påståenden som man ofta ser om skillnaden mellan mäns och kvinnors förmåga till spatial uppfattning och liknande.

Westerberg menade att den korrekta tolkningen av den typen av resultat inte är att det är skillnad mellan kvinnor och män, utan snarare att det *inte* finns skillnader mellan kvinnor och män. Det följer av att resultaten alltid visar att spridningen inom kön är så oerhört mycket större än mellan kön, och att den genomsnittliga skillnaden mellan kön är liten om den alls finns. För att kunna tolka den typen av resultat krävs statistisk förståelse, något som alla har rätt till.

Matematik – ett internationellt språk

Eftersom matematik är ett internationellt språk lämpar sig matematik-kunskaper väl för jämförelser mellan länder. Enligt Westerberg har kunskaperna hos skolungdomar gått tillbaka i Sverige, både absolut sett och i jämförelse med andra länder där Sverige numera ligger under snit-

tet inom såväl OECD som EU. Matte är det skolämne han själv anser sig ha haft mest nytta av, men alla är inte roade av matte. Man kan kokettera med att man inte kan någon matte alls och det säger något om matematikens status.

Statistik finns på ett eller annat vis med i alla vetenskapsområden. Bengt Westerberg gav flera exempel. Thomas Lindh på Institutet för framtidsstudier i Stockholm forskar i hur demografisk och ekonomisk utveckling hänger ihop. I 50-årsåldern når många sin puckel i sparande och i inkomster.

Sparandet går ner i 60-årsåldern. På så sätt kan man få en bild av ett lands kommande ekonomiska utveckling med hjälp av dess demografiska utveckling. Richard Florida är en Torontobaserad forskare som har argumenterat för att det finns ett samband mellan mångfald i samhället och god ekonomisk utveckling genom hög teknologi. Speciellt starkt är sambandet mellan öppenhet mot homoseksuella och hög teknologi. Det har dock inget med människors sexualitet att göra, snarare är det så att öppenhet leder till välkommande



Bengt Westerberg.

FOTO: MAGNUS BERGSTRÖM

av bredd i bakgrund, kunskaper och erfarenhet som i sin tur leder till stark utveckling.

Evidensbaserade beslut i socialtjänsten

Westerberg tog ännu ett exempel på hur statistik berikar och ibland medför kontroverser. Kerstin Wigzell väckte uppeende när hon för ungefär ett decennium sedan propagerade för evidensbaserade beslut i socialtjänsten. Att ens diskutera randomiserade experiment inom socialtjänsten har tills helt nyligen varit kontroversiellt. Fast egentligen pågår det en mängd naturliga experiment hela tiden eftersom nästan alla kommuner gör olika. Det är en utmaning för statistiker att analysera observationsdata från kommunerna.

Westerberg nämnde att sociologen vid Stockholms universitet Robert Eriksson påpekat att Sverige har en skyldighet att underhålla, forska på och tillgängliggöra våra unika befolkningsdata som omfattar hela landet och sträcker sig så långt tillbaka i tiden. Det var Tabellverket som under Pehr Wargentins ledning 1749 genomförde världens första folkräkning. Den väckte större uppmärksamhet utomlands än i Sverige. När Bengt Westerberg var med i Vetenskapsrådet genomfördes bättre möjligheter för forskare att ta del av svenska befolkningsdata genom bland annat SCB:s system för extern åtkomst till mikrodata (MONA). Det finns behov av att anpassa lagstiftningen för balans mellan forskningens möjligheter och människors integritet, menade Bengt Westerberg.

DAN HEDLIN

»För att kunna tolka den typen av resultat krävs statistisk förståelse, något som alla har rätt till.»

LÄSA STATISTIK

Kristoffer Johansson:

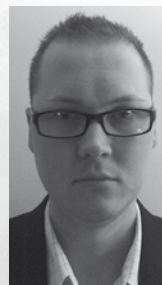
– Jag valde det här programmet för att statistik kan användas inom i princip alla yrkesområden. Blandningen av statistik, matematik och datakurser var också tilltalande. Dessutom kändes det intressant att välja en unik utbildning som inte så många läser. Innan jag sökte hade jag även kollat upp lite om hur arbetsmarknaden såg ut för statistiker, genom arbetsförmedlingens hemsida, och det såg rätt så positivt ut då.

»Jag är fascinerad av hur man kan ta fram information ur bara en massa siffror.»

Tommy Schyman:

– Anledningen till att jag sökte statistikprogrammet var att jag såg statistik som ett problemlösningssverktyg som kan användas för att utreda olika frågeställningar om samhället och andra intressanta områden.

Det är nu ungefär fyrtio år sedan man såg ett behov av att lokalisera och kartlägga förorenade och sårbara miljöer. Vid den tiden kom man också på allvar igång med systematisk provtagning av vår miljö i Sverige. På den tiden var uppföljningen av vidtagna åtgärder förhållandevis enkel, eftersom det framförallt handlade om att klarlägga effekten av att försöka eliminera stora punktutsläpp.



Karl Wahlin.

Bättre datakvalitet med simultan analys

För fyrtio år sedan var uppföljningen av åtgärder förhållandevis enkel. Dagens miljöproblem är mer komplexa och storskaliga. Det finns nu ofantligt mycket mer data, och tolkningen av dessa data kommer allt oftare i centrum av debatten. Miljöövervakningen och uppföljningen av åtgärder kräver statistiska analyser med inslag av informationsvetenskap. Det behövs nya arbetsmetoder och nya prioriteringar.

Vi hade inom ett forskningsprojekt på Linköpings universitet, ENGO (Assessment of Environmental Goal Achievement under Uncertainty, se referenslistan), som utgångspunkt att göra en kartläggning av vattenkvaliteten i de svenska vattendrag som ingår i den nationella miljöövervakningen. Projektets syfte var att på Naturvårdsverkets uppdrag ta fram statistiska metoder för uppföljning av Sveriges 16 miljömål och ingick som en del i mitt avhandlingsarbete.

Svenska vattenkvalitetsdata brukar anses vara av mycket god kvalitet. Baserat på granskning av tidsserieplottar och mängden saknade värden valde vi ut 34 vattendrag från Sveriges sju vattendistrikt samt tre provtagningsplatser i sjöar där data bedömdes ha särskilt god kvalitet. Man tar prover i ytvattnet varje månad; då mäts ett femtiotal kemiska variabler. Så långt det var möjligt arbetade vi med tidsserier mellan 1980 och 2005, en tidsperiod under vilken samma

laboratorium gjort samtliga kemiska analyser.

Till vår hjälp hade vi ett Visual Basic-baserat verktyg som tidigare utvecklats av Anders Grimvall med medarbetare vid Linköpings universitet: Multitrend, en semiparametrisk regressionsmodell vars grafiska output erbjuder simultan överblick av data från många tidsserier. Genom att analysera flera tidsserier samtidigt skapas möjligheter att beskriva trender i insamlade data med hjälp av trendytor. Figur 1 visar en trendyta över normaliserad totalfosforkoncentration i 15 vattendrag som rinner ut i Bottenvikens och Bottenhavets vattendistrikt. Älvarna är sorterade efter medelkoncentration, så att den med lägst koncentration (Indalsälven) ligger främst i figuren. Vattenföring, alltså vattenflödet i ett vattendrag per given tidsenhet, och mängden partikulärt material, mätt som skillnaden i absorptions mellan filtrerat och ofiltrerat prov, har använts som kovariater i modellen.

Vi kan se att synkrona förändringar har skett i alla stationer i detta vattendistrikt: i 1980-talets början, omkring 1983, har det skett en nedgång av totalfosforhalten i samtliga studerade vattendrag. Sedan steg koncentrationerna, men omkring 1996 kom det en ny nedgång, återi-

gen för alla vattendrag samtidigt. Motsvarande mönster återfinns i de vattendrag som flödar ut på västkusten, Västerhavets vattendistrikt. Detta var en mycket intressant och överraskande iakttagelse, eftersom dessa avrinningsområden är mycket olika geografiskt och hydrologiskt, och för att de genomsnittliga koncentrationerna ligger på helt olika nivåer. Vi gick sedan vidare och studerade totalfosforkoncentrationen på tre olika djup i Vänern. I Figur 2 framgår att samma mönster återkommer ännu en gång, trots att mätningarna gjorts ned till 20 meters djup i en sjö med omloppstid på åtta år. Samma mönster återfinns även i Vättern.

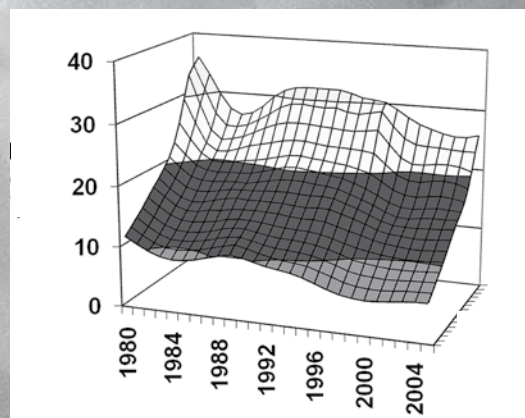
»Dagens miljöproblem är mer komplexa och storskaliga. Det finns nu ofantligt mycket mer data, och tolkningen av dessa data kommer allt oftare i centrum av debatten.»

Vi ställde upp fyra hypoteser för att försöka förklara de synkrona trenderna.

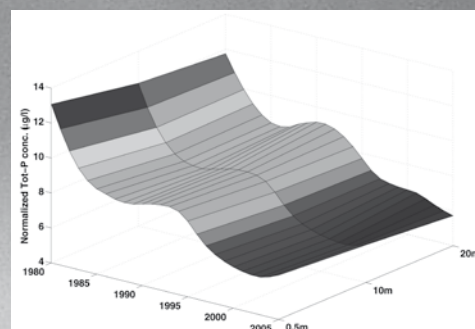
Vår första hypotes var att det var våra analysverktyg som or-

sakade dem. Vi kunde dock snabbt avskriva denna hypotes. Nu när vi visste vad vi letade efter gick det att se trenderna direkt i rådata.

Vi funderade istället över hypotesen att mönstren orsakats av mänskliga aktiviteter, såsom kraftiga utsläpp eller andra föroreningar. Även denna hypotes kunde avskrivas: mönstren återfinns i vattendrag som är starkt geografiskt och hydrologiskt åtskilda. Om man dessutom be-



Figur 1. Diagrammet finns i färg på sista sidan.



Figur 2. Diagrammet finns i färg på sista sidan.

Mer läsning

■ Mer information och en förteckning över publikationer inom ramen för forskningsprojektet ENGO finns på adressen <http://www.ida.liu.se/divisions/stat/research/engo/index.html>

Sveriges 16 miljömål finns beskrivna på: <http://www.miljomal.nu/>

»»» BÄTTRE DATAKVALITET MED SIMULTAN ANALYS

tänker hur långa uppehållstider det handlar om i de stora sjöarna, och att upp- och nedgångarna i halterna var så pass stora, är det osannolikt att människan skulle kunna ha åstadkommit dem.

Nästa hypotes handlade om vädereffekter. Det är välkänt att extrema väderförhållanden kan orsaka mycket snabba förändringar hos vattenkemiska parametrar. Men när vi granskade tidsserier över väder vid de aktuella tidsperioderna såg vi inget utöver det normala. Mönstren fanns kvar, trots att vi justerat för vattenföring i vattendragen och temperatur i sjöarna.

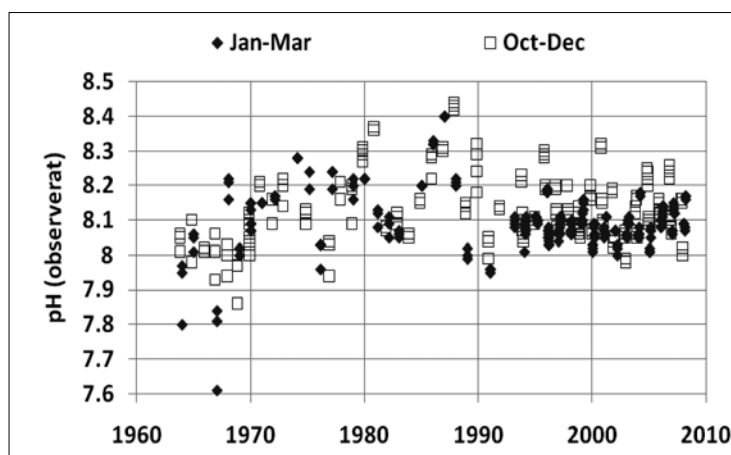
Vår sista hypotes var att mönstren orsakats av medvetna eller omedvetna förändringar i det laboratorium som genomfört de kemiska analyserna. Det visade sig att laboratoriet som utfört analyserna bytt instrument för fosforanalyser under 1996 och att det gamla instrumentet under lång tid haft en systematisk drift uppåt på baslinjen. Vi kunde därför dra slutsatsen att nivåskiftet 1996 berodde på förändringar i laboratoriet. Genom ett liknande resonemang och uteslutningsmetoden kunde vi också avslöja att medvetna eller omedvetna förändringar i laboratoriet låg bakom den synkrona förändringen 1983. Fortsatta studier av vattenföringsdata avslöjade fler kemiska variabler med liknande synkrona trender, trendbrott och andra fel.

Den stora mängden data är det främsta skälet till att data blir väldigt svåra att överblicka och analysera på traditionellt sätt, och därför uppstår denna typ av problem.

Vi har dock kunnat konstatera att våra verktyg för simultan analys av många tidsserier har varit oerhört användbara för att extrahera information som man inte visste fanns i databaserna över nationella miljöövervakningsdata.

Verktygen har avslöjat allvarliga brister i den svenska miljöövervakningen, och vi har visat på ett behov av att göra information om kända kvalitetsproblem mer lättillgänglig, exempelvis genom dokumentation på datavärdens webbplats eller genom flaggningar av osäkra värden i databasen. Det är dock glädjande att vi redan kunnat se att förändringar sker. Till följd av våra upptäckter har Naturvårdsverket tillsatt en oberoende granskningsgrupp bestående av statistiker och miljövetare som ska utveckla kvalitetsarbetet för den svenska miljöövervakningen. Förhoppningsvis kommer detta i förlängningen att leda till en utveckling av miljöövervakningsprogrammets utformning.

KARL WAHLIN, LINKÖPINGS UNIVERSITET



Figur 1. Observerade pH-värden i höst- och vinterprover av ytligt havsvatten från Östra Gotlandsdjupet. Data från SMHI.

Koldioxidhalten i atmosfären har stigit kraftigt under de senaste 100 åren och allt tyder på att den kommer att öka ytterligare. En stark majoritet av klimatforskarna hävdar att detta kommer att leda till ett förändrat klimat. Men den ökade koldioxidhalten kan påverka vår miljö även på andra sätt. När koldioxid löses i vatten så kan vattnets surhetsgrad ändras. Detta har skapat en debatt om att havsmiljön hotas av en global försurning. Vilken roll kan statistiken och statistikern spela i denna debatt?

Vid en första anblick kan det tyckas självklart att högre koldioxidhalt i atmosfären leder till surare havsvatten. Om man ställer ett glas "rent" vatten i ett rum med förhöjd koldioxidhalt så är det ett faktum att vattnet i glaset blir surare. Men havet är självfallet ett mer komplicerat system än ett glas vatten. För det första är havsvattnet motståndskraftigt mot försurning. För det andra påverkar både väder och havsströmmar utbytet av koldioxid mellan atmosfär och hav. För det tredje finns det biologiska processer som starkt kan påverka havsvattnets surhet. Det har t.ex. länge varit känt att vattnets pH kan stiga kraftigt i samband med algbloomningar. För det fjärde påverkar tillrinnande flodvatten med substanser havets kolsystem och dess syra-basbalans. För att man ska belysa försurningstrender i havet krävs därför en noggrann granskning av både teori och empiriska data.

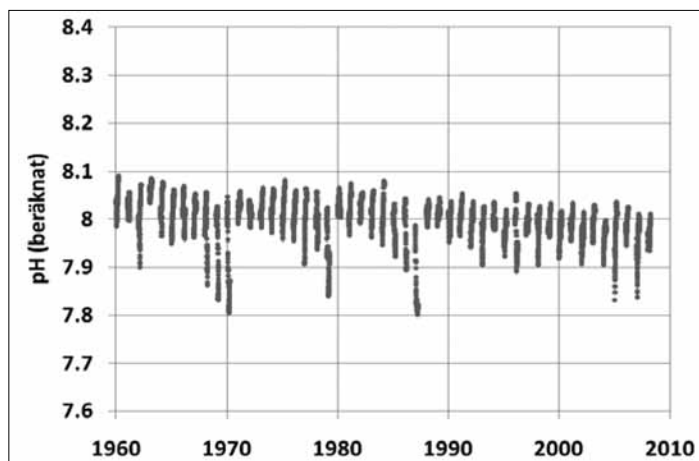
Vad säger observerade data?

Ökningen av koldioxidhalten i atmosfären

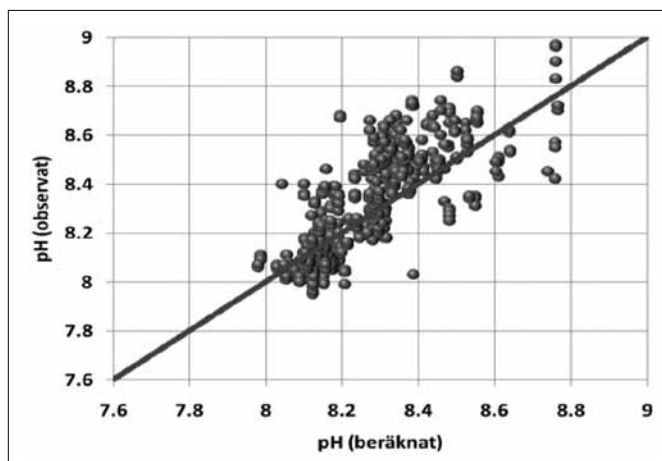
är mycket väl dokumenterad. Bland annat finns en lång observationsserie från Hawaii som tydligt visar både trend och säsongvariation. Mätningar av havsvattnets surhet (pH) och buffringsförmåga (alkalinitet) ger inte lika tydlig information. Enligt teoretiska beräkningar borde pH-värdet i havsvattnet ha sjunkit 0,1 enheter de senaste 100 åren. Variationen i observerade data från Östersjön är mycket större, även om man begränsar sig till vinterdata från ytliga prover (Figur 1). Dessutom kan man med hjälp av animerade diagram över pH och salthalt lätt visa att data insamlade före 1994 har lägre precision. Om man enbart ska lita till observerade data är det därför inte möjligt att dra några bestämda slutsatser om långtidstrender.

Vad säger modellerna?

Östersjön är ett tacksamt studieobjekt om man vill förena kunskap om fysikaliska, kemiska och biologiska processer till en numerisk modell av hur vattnets surhet varierar i tid och rum. För det första är det ett ganska slutet hav som i sin tur kan delas upp i ett antal tämligen



Figur 2. Modellberäknade vintervärden av pH i havsvatten från Östra Gotlandsdjupet. Beräkningar med en processbaserad modell för kopplade havsbassänger (Omstedt).



Figur 3. Observerade och modellberäknade pH värden 1980-2008 i ytligt vatten från havsbassängen öster om Gotland.

STATISTIKERN BEHÖVER INTE FASTNA I TRADITIONELLA FACK

Havets försurning – data, modeller och statistikerns roller

väl åtskilda havsbassänger. För det andra finns gott om data över tillförseln av vatten och substanser från omkringliggande landområden. Enligt en modell utvecklad av Anders Omstedts forskningsgrupp (Göteborgs universitet och Havsmiljöinstitutet) borde pH-värdet i ytvatten öster om Gotland under vintermånaderna ha varierat som i Figur 2. I denna dataserie finns som synes en tydlig nedåtgående trend.

Stämmer överens?

Av flera skäl kan man inte förvänta sig en perfekt överensstämmelse mellan observerade data och modellberäkningar. En modell ger alltid en förenklad bild av verkligheten. I detta fall avser dessutom modellberäkningarna ett medelvärde för en havsbassäng medan mätningarna är gjorda vid en specifik provtagningspunkt. Dessutom kan man sällan bortse från mätfel.

Punktdiagrammet i Figur 3 illustrerar hur väl mät- och modelldata stämmer överens för den havsbassäng i egentliga Östersjön där det finns bäst mätdata. Med tanke på att modellen inte utnyttjar mätdata för att förbättra anpassningen till observerade data får överensstämmelsen

anses tämligen tillfredsställande.

En närmare granskning av data visade att såväl genomsnittsnivån som säsongvariationen i modellens output var tämligen korrekt. Indirekt får därmed också den nedåtgående pH-trenden i modellens resultat en viss trovärdighet. Samordnad analys av mät- och modelldata ger alltså en tydligare bild av havsförsurning än enbart teoretiska beräkningar eller enbart observationsdata.

Vilka roller kan statistikern spela?

Många statistiker har utvecklat en speciell förmåga att se och klarlägga mönster i data som varierar mer eller mindre slumpmässigt. Denna förmåga är viktig för att avslöja kvalitetsbrister i observerade data eller modelloutput. Ett exempel finns här bredvid i Karl Wahllins artikel. Dessutom finns nu animerade (rörliga) grafer som effektivt kan avslöja abrupta förändringar i medelvärde och varians. När sådana verktyg tillämpades på data över pH och alkalinitet i Östersjön så framträdde en tydlig bild. I likhet med de flesta andra långa tidsserier av miljödata innehöll de granskade serierna både större och

mindre fel. Datakvaliteten har blivit bättre med tiden men även under den senaste tioårsperioden finns enstaka starkt avvikande värden som kan ifrågasättas.

Att hitta fel i insamlade data är viktigt men får självfallet inte bli statistikerns enda roll. Processbaserade modeller kommer med största sannolikhet att spela en allt större roll för beslutsfattande i miljöfrågor. I fallet med havsförsurning är det uppenbart att det är svårt att klarlägga trender genom att enbart analysera observerade data. Om man dessutom har ambitionen att klarlägga orsakerna till observerade trender så blir det ännu svårare. Exempelvis kommer det att vara näst intill omöjligt att utan processbaserade modeller avgöra om övergödning motverkar eller förstärker en försurnings-trend i havet. Detta beror på att det finns många olika statistiska sambandsmodeller som ger god anpassning till insamlade data, och vilken modell man då ska välja blir en bedömningsfråga där kunskap om bakomliggande processer och mekanismer blir avgörande. Den statistiker som vill påverka debatten måste därför ge sig in i oceanografernas, klimatforskarnas, marinbio- »»»

Debatten kring Statistics Canadas "long form"



Debatten kring "the long form" för Canada Census 2011 har givit eko i den statistiska världen. Den kretsar kring politikernas och regeringens påverkan och inflytande i framställningen av officiell statistik.

Kanadas regering – konservativ, under premiärminister Stephen Harper – framförde i början på 2010 till Statistics Canada att den detaljerade "long form" som vid tidigare folkräkningar gått ut till vart femte hushåll, med systematiskt urval och med obligatorisk svarsplikt, borde försvinna. Enskilda medborgare hade klagat på att "the long form" efterfrågade högst personlig information, t.ex. om etnisk bakgrund och om sexuell läggning, och på att uteblivet svar kunde bestraffas med böter eller värre påföljd.

En föreslagen ersättare med till stora delar samma variabelinnehåll har skisserats, en National Household Survey. Den skulle gå ut till en tredjedel av Kanadas alla hushåll; medverkan skulle vara frivillig; uppskattningsvis kan man räkna med en svarsandel på cirka 50%. Bortfallet kommer naturligtvis att slå högst olika i olika skikt av samhället. Som presidenten i Statistical Society of Canada påpekar, risken är stor att denna nya survey kommer att visa på ett kanadensiskt samhälle mycket bättre än vad det i själva verket är. Kanadas befolkning kommer att porträtteras som rikare, bättre utbildad och mer konservativ än vad den är, på bekostnad av trovärdig information om små eller utsatta grupper.

Resultatet har blivit ett ramaskri av indignation och ilska från många instanser och sta-

tistikanvändare i det kanadensiska samhället, inte bara specialister i statistik. Man behöver inte vara utbildad statistiker för att ifrågasätta trovärdigheten i resultaten från en datainsamling gjord på sådana okontrollerade grunder.

Hur kunde detta läge uppstå? Det är inte helt klart. Ministern som ansvarar för Statistics Canada – Industry Minister Tony Clement – säger att regeringen underhandlade i början av 2010 med Statistics Canada, och att den nu föreslagna National Household Survey var ett resultat, ett av tre alternativ, som Statistics Canada skulle ha fört fram. Händelseförloppet är delvis höljt i dunkel. Ministern hänvisar till sin ed om sekretess; i vad mån beslutsfattare vid Statistics Canada ska ha haft en nyckelroll framgår inte. Tunga företrädare för statistisk metod inom Statistics Canada tycks i varje fall

»»» HAVETS FÖRSURNING – DATA, MODELLER OCH STATISTIKERNES ROLLER

logernas och kemisternas modellvärld och samtidigt analysera observationsdata och hur olika processer påverkar modellens output. Att utveckla god standard och riktlinjer för mät- och modellberäknade resultat utgör därför en viktig utmaning för dagens statistiker.

Statistikern kan också spela en annan och mycket viktig roll – att överbrygga klyftan mellan modellutvecklare och beslutsfattare. Processbaserade modeller av miljön beskriver fenomen i den tidsskala som är relevant för biologiska, kemiska och fysikaliska processer. En beslutsfattare som ska fatta beslut i miljöfrågor har ofta helt andra tidsperspektiv och vill kanske veta hur en viss åtgärd påverkar

havsmiljön på tio års sikt. Processbaserade modeller har också en tendens att byggas ut med fler och fler processer, somliga viktiga, andra mindre viktiga. Detta gäller såväl inom havsforskningen som inom många andra områden. Då kan statistikern träda in med känslighets- och osäkerhetsanalyser. Vad är det i alla detaljerade indata till modellen som verkligen betyder något för modellens prediktioner i den längre tidsskalan och som intresserar beslutsfattare?

I många fall bör det också vara möjligt att härleda en ganska enkel input-outputmodell som fångar de för beslutsfattaren väsentliga sambanden i en komplex modell med hög

rum- och tidsupplösning. Då öppnas även möjligheter att skapa pedagogiska simuleringsverktyg som kan vara till nytta för både beslutsfattare och modellutvecklare. Statistikern behöver alltså inte fastna i traditionella fack. I fallet med havsmiljön och eventuella försurningstrender så finns uppenbarligen andra uppgifter som kan vara väl så viktiga.

ANDERS GRIMVALL,
LINKÖPINGS UNIVERSITET
OCH HAVSMILJÖINSTITUTET
ANDERS OMSTEDT,
GÖTEBORGS UNIVERSITET
OCH HAVSMILJÖINSTITUTET



Bakgrund

■ De allra flesta länder i världen genomför folkräkningar. Kanada har en folkräkning 2011 som genomförs av Statistics Canada, Kanadas motsvarighet till Statistiska centralbyrån. Kanadas folkräkning har sedan länge bestått av en kort blankett som alla ska fylla i och en lång blankett (long form) som även den har varit obligatorisk, men för en delmängd av befolkningen (20 %). Den obligatoriska långa blanketten ska enligt regeringsbeslut ersättas med en frivillig National Household Survey (NHS). Denna urvalsundersökning designades i all hast. Frågeformuläret är i stort sett detsamma som skulle ha använts i den långa blanketten i folkräkningen. Urvalet till NHS omfattar 4,5 miljoner hushåll. I den ursprungliga planen hade den långa blanketten gått till ca 2,5 miljoner hushåll. Den dåvarande chefen för Statistics Canada, Munir Sheikh, avgick i somras med följande motivering: "The simple reason for my resignation is that I cannot see myself surviving the role of Chief Statistician of an agency whose reputation and integrity come into doubt."

Det finns redan mycket skrivet på Internet i debatten om Kanadas folkräkning, bl.a. på engelskspråkiga Wikipedia. Statistics Canada har dock tagit bort det öppna brev som Munir Sheikh skrev vid sin avgång. Istället har Statistics Canada lagt ut ett annat meddelande på sin webbplats. Det återges här nedan (hämtat från Statistics Canadas webbplats 28 november 2010).

Här är en länk till den korta blanketten och till frågeformuläret för National Household Survey, inklusive kortfattade motiveringar till varför frågorna i NHS ställs: <http://www12.statcan.ca/census-recensement/2011/ref/gazette-eng.cfm>

DAN HEDLIN

inte ha blivit konsulterade; frågan behandlades inte vid mötet i april i år med Statistics Canadas vetenskapliga råd Advisory Committee for Statistical Methods.

Att staten ska kunna ha ett inflytande över innehållet, över det som ska statistiskt belysas i samhället, det är klart. Regeringen ska kunna begära att viktig statistik produceras, att relevanta områden av samhället – det kanadensiska eller något annat – blir belysta.

Den springande punkten,

åtminstone för statistiker som återoppar vetenskapliga grunder för officiell statistikframställning, är att när vi talar om politisk inblandning eller intrång i statistikproduktionen så kommer aspekten statistisk metod i fokus, och inte så sällan kommer den i kläm. Det blir en fråga om skillnaden mellan att "göra statistik" och "göra statistik som ställer höga krav på kvalitet och precision". Den barnsliga föreställningen finns, underligt nog, att "bara vi har många observationer så blir siffrorna bra".

Statistics Canada anför nu att den nya frivilliga undersökningen skulle kunna ge "useful and usable data", men utan de höga krav på kvalitet som fanns på föregångaren. Siffror blir det ju alltid.

Även om han avgick som chef för Statistics Canada 2008 så har Ivan Fellegi blandat sig i debatten. Nyligen har han tillsammans med bland andra en f.d. Governor of the Bank of Canada hemställt till premiärministern om en "amendment of the Statistics Act of Canada" som skulle göra det klart att the Chief Statistician

är den som är ytterst ansvarig för frågor som gäller metoderna.

»Det blir en fråga om skillnaden mellan att "göra statistik" och "göra statistik som ställer höga krav på kvalitet och precision".»

Fellegi hävdar att en "amended Statistics Act" klart bör ange att beslut om metodik ska baseras på vetenskapliga grunder, grunder som Statistics Canadas experter bedömer vara de mest lämpliga, utan inblandning från politiska instanser. "Independence and integrity" är hörnstenar för Statistics Canada; "without it, we might as well not exist", som Fellegi uttrycker det. Det vore olyckligt om det inträffade skulle leda till skada för Statistics

Canadas mycket höga anseende.

"Arm's length autonomy" har tidigare upplevts som en realitet för Statistics Canada, även om den saknat formellt uttryck i lagen. Vad Fellegi och liktänkande syftar till är ett tillägg till lagen i linje med en grundläggande princip formulerad av FN: "To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data."



CARL-ERIK SÄRNDAL

FOTO: MONICA HOLMBERG

Källa: Intervju med Dr. Ivan P. Fellegi rapporterad i *The Hill Times*, Ottawa, 2010-09-27.

Media advisory: 2011 Census
July 16, 2010

OTTAWA — Statistics Canada is not in a position to answer questions on the advice it gave the Minister in relation to recent statements the Minister has made.

Nollhypoteser och statistisk signifikans

Det borgerliga blocket ökade från 45 % till 47 % mellan de två opinionsundersökningarna, men förändringen är inte statistiskt säker.

I dagsläget finns en vedertagen praxis bakom detta påstående. Statistiker relaterar till hypotesprövning och uppställningen $H_0: p_1 = p_2$ (nollhypotes) och $H_1: p_1 \neq p_2$ (mothypotes) i vilka p_1 och p_2 refererar till populationsandelar i samband med respektive undersökning. I fallet ovan kunde man inte förkasta nollhypotesen till förmån för mothypotesen (ett icke-signifikant resultat), troligen givet en felrisk på 5 %. Jag skulle vilja hävda att denna hypotesuppställning inte är passande för den givna situationen. Är det inte givet på förhand att mothypotesen är den gällande? Vem skulle vilja påstå att populationsandelarna är exakt desamma vid dessa två tillfällen?

Mer intuitivt är att dela upp i fallen $H_1: p_1 > p_2$ och $H_2: p_1 < p_2$ samt tillåta tre svarsalternativ (ökning, minskning och vet ej). Låt mig konkretisera genom ett parallellt exempel med en slumpmässig normalfördelad observation, $X \sim N(\mu, 1)$ med förutsättningen att μ är okänd. Med motsvarande traditionella uppställning förkastar vi nollhypotesen $\mu = 0$ till förmån för

en tvåsidig mothypotes $\mu \neq 0$ i de fall X överstiger 1,96 i absolutbelopp. Med den alternativa uppställningen $H_1: \mu > 0$ och $H_2: \mu < 0$ svarar vi alternativ 1 om X överstiger 1,645, alternativ 2 om X understiger -1,645 och vet ej, dvs. icke-signifikant resultat, i samband med utfallen däremellan. Observera att fler utfall ger signifikant resultat med den andra uppställningen (att överstiga 1,645 istället för 1,96 i absolutbelopp).

Låt oss ta en titt på hur begreppen från den traditionella uppställningen kan överföras till utmanaren. Fel av typ I innebär som tidigare att man avger felaktigt signifikant resultat, medan fel av typ II likställs med icke-signifikant resultat. Felrisk kan man kalla sannolikheten att, för ett givet värde på μ , avge felaktigt signifikant resultat. Testet ovan kännetecknas av att felrisken är högst 5%. Styrka kan man kalla sannolikheten att, för ett givet värde på μ , avge det korrekta svaret. P-värde kan man definiera som den maximala sannolikheten att få ett minst lika avvikande utfall, givet att tecknet för X pekar i fel riktning.

Låt mig nu hoppa till ett annat klassiskt skolexempel: *Läkemedelstillverkaren som avser att påvisa att preparatet har en signifikant positiv effekt*. Låt oss hålla fast vid vårt μ och vårt X . Jag tror läsaren kan se parallellen mellan exemplet och testformuleringen $H_0: \mu = 0$ och $H_1: \mu > 0$. H_0 är en typisk nollhypotes i det att alla parametervärden som den innefattar (bara ett

enda i detta fall) angränsar till parametervärden i konkurrerande hypoteser så att inget utfall med säkerhet kan peka på H_0 . Alla rimliga test väljer därför mellan svaren "vet ej" och " H_1 ". Konsekvensen blir att det bara finns en felrisk, den som motsvarar att svara H_1 när H_0 är sann.

Varför vill jag då inkludera en nollhypotes i samband med läkemedel och placebo, men inte i samband med opinionsundersökningen? I en kontrollerad experimentsituation är idén om att en tillförd substans drunknar i övrig variation relevant. Den svenska väljaropinionen är däremot så pass variabel så att den enda relevanta distinktionen är ökning/minskning. Det är alltså den verklighet vi avser att modellera som avgör.

Jag försöker här väcka intresset för en förändring av teorin för statistisk hypotesprövning. Själva kärnan är att begreppen nollhypotes och signifikant resultat åtskiljs. En hypotesuppställning behöver inte innefatta en nollhypotes. Om en nollhypotes inkluderas i uppställningen så kännetecknas den av att rimliga test avstår från att peka ut detta alternativ (det man vill uppnå är ju enbart att förkasta den). Ett icke-signifikant resultat är detsamma som att testet avstår från att välja bland hypoteserna (= vet ej). Förutom en ökad flexibilitet gentemot olika typer av tillämpningar tror jag att pedagogiska förbättringar kan uppnås.

Låt mig avslutningsvis bemöta ett par invändningar som jag mött rörande opinions-

Replik på "Mjuka fakta och osannolik noggrannhet"

förra numret av Qvintensen hävdar Björn Jonsson att det sätt som statistiken över sjukfrånvaron presenterades och hur den påverkade debatten under början av 2000-talet är ett praktexempel på hur statistik kan missuppfattas och missbrukas (Qvintensen nr 3/10, "Mjuka fakta och osannolik noggrannhet"). Han anför att Sverige upplevde en sjukskrivningsepidemi och att detta berodde på missuppfattningen att fler var sjukskrivna. Vidare skriver han att sjukfrånvaron huvudsakligen orsakades av en ökning av sjukskrivningslängden. Vi menar att såväl inflöde som varaktighet uppenbarligen spelade en viktig roll, men att det på intet sätt var någon väl dold hemlighet. Bilden av ökningen av antalet fall kan

dock presenteras på olika sätt. Ökningen kan redovisas som "bara" 20 procent eller nästan 50 procent. Att sjukfallslängden ökade under perioden, som Björn Jonsson skriver, är dock helt korrekt men som sagt på intet sätt någon väl dold hemlighet.

De bedömningar av omfattningen av felaktigheter inom bidragssystemen som gjorts av bland andra Delegationen mot felaktiga utbetalningar kallar Björn Jonsson rena gissningar. Osäkerhetsintervallen (i storleksordningen 8-24 miljarder kronor), som syftar till att karakterisera osäkerheten kring ett punktestimat, är inte desamma som konfidensintervall – även om det har samma syfte. Det ska ses som en

subjektiv sannolikhetsfördelning baserad på expertbedömningar. Även om dessa omfattningsbedömningar har hög osäkerhet är de mycket mer än rena gissningar och kan användas i brist på data eller statistik över avsiktliga och oavsiktliga fel och vilka belopp det rör sig om. Expertbedömningar ersätter således inte "hårdare" statistiska fakta. Finns sådana tillmäts de i allmänhet avgörande betydelse (även om sammanställning och tolkning kräver omdöme). Policybeslut borde bygga på data om faktorer av central betydelse, dock saknas sådana i många fall. Att inte göra några bedömningar får konsekvenser för de policybeslut som tas. Frågan blir då om det inflytande expertbedömningar får på policybeslut ska vara "implicita" i pro-

»Jag försöker här väcka intresset
för en förändring av teorin för
statistisk hypotesprövning.»

undersökningsexemplet. Den första är att min uppställning är ett exempel på ett betingat test, dvs. att noll- och mothypotes väljs efter det att data samlats in, med konsekvensen att felrisken inte alls är 5% som påstått. Förvisso kan man beskriva proceduren som att noll- och mothypotes väljs efter att data samlats in, men observera att om icke-signifikant resultat därefter erhålls, så rapporterar vi detta som "vet ej", och inte som att nollhypotesen styrkts. Därmed är testet inte betingat, och felrisken är inte 10% som kritikern vill hävda.

En annan invändning är att jag var för hårdhänt i min tolkning av den traditionella hypotesuppställningen, och att nollhypotesen bör tolkas som icke-signifikant skillnad mellan populationsandelarna. Denna tolkning tycker jag snarast sprider konceptuell förvirring. Förvisso bör man kunna få hänvisa till signifikanta förändringar, och då åsyfta de bakomliggande populationerna och inte testresultatet. Värt att komma ihåg är att konfidensintervall för $p_2 - p_1$ är att rekommendera om man vill göra uttalanden om förändringens storlek med hänsyn till felmarginaler. Jag vidhåller att den traditionella uppställningen är opassande för situationen i fråga, vilket i praktiken innebär att testets felrisk är mindre eller lika med 2,5% och inte exakt 5% som påstås.

Synpunkter och kommentarer tas gärna emot, antingen i Qvintensen eller direkt till författaren på fredrik.jonsson@math.uu.se.

FREDRIK JONSSON, UPPSALA UNIVERSITET

cessen, eller transparenta, baserade på bedömningar gjorda av dem som rimligtvis har den bästa tillgängliga kunskapen och genomförs med procedurer som minimerar risken med subjektiva bedömningar.

Den procedur som t ex Försäkringskassan tillämpar syftar till att göra underlagen för beslut så transparenta, välgrundade och objektiva som möjligt. "Expert judgment is pervasive in scientific studies. The question is not whether to use it but whether to use it in an overt manner, documenting its use, or to hide it use".

JOHAN EKSTEDT,
DOKTORAND UPPSALA UNIVERSITET
GUNNAR JOHANSSON,
DIREKTÖR FÖRSÄKRINGSKASSAN
PETER JUSLIN,
PROFESSOR UPPSALA UNIVERSITET



Nominera Årets statistik- främjare

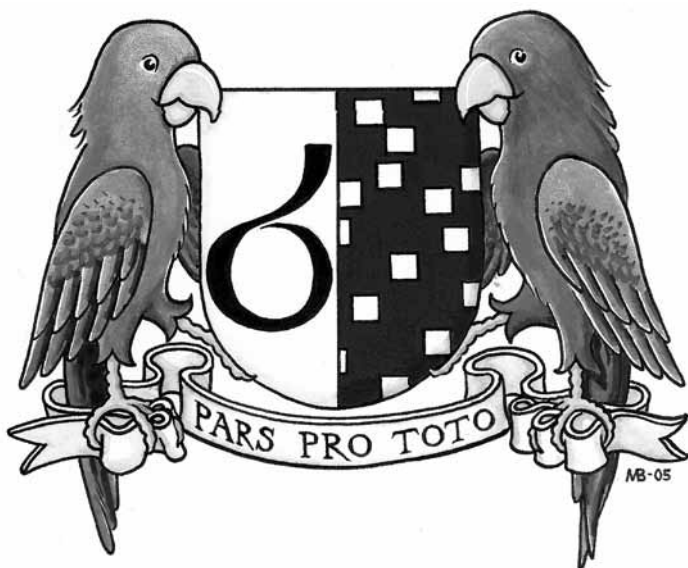
■ **Till Årets statistikfrämjare** utses en eller flera personer bosatta i Sverige som gjort en slagkraftig insats på statistikområdet som under det senaste året givit publik uppmärksamhet. Insatsen kan utgöras av en innovativ metod, en smart analys, en effektiv presentation eller att på ett framgångsrikt sätt ha utmanat en etablerad sanning. 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 alla medlemmar att ge förslag på en eller flera personer som du tycker uppfyller ovanstående kriterier. Skicka förslaget med motivering till hans.nyquist@stat.su.se senast den 30 januari 2010. Utdelningen av priset äger rum vid Svenska statistikfrämjandets årsmöte den 24 mars 2010.

Ulf Jorner fick priset förra året för att ha slutfört arbetet med boken om SCB:s historia – Summa summarum – som beskriver utvecklingen av en samhällsinstitution under 150 år. En imponerande insats med omfattande källforskning som resulterat i livfulla texter illustrerade med många bilder av samtida dokument, porträtt, tabeller och blanketter. I Qvintensen nr 4/09 och 1/10 skriver Ulf om den svenska officiella statistikens historia och i nr 2/10 berättar han om sitt arbete med boken.

Lösning

Rätt svar på frågorna i "Övertro på ens egen förmåga", sidan 10: 1. 39 år; 2. 6700 km; 3. 33 länder; 4. 39 böcker; 5. 3470 km; 6. 180 000 kg; 7. 1756; 8. 645 dagar; 9. 9580 km; 10. 11 000 m. De flesta som gör den här övningen får färre än nio rätt. Det kan tolkas som att man överskattar sin förmåga att gissa rätt med 90% sannolikhet, alltså en övertro på sin förmåga.



Ordföranden har ordet

■ Det är glädjande att kunna berätta att Marika Wenemark har disputerat i Linköping på avhandlingen *The respondent's perspective in health-related surveys. The role of motivation.*

Den handlar om hur man med motivationshöjande åtgärder kan höja svarsandelen. Det är glädjande mot bakgrund av att bortfallet har blivit ett allt större problem för statistikerna. SCB:s alla

individ- och hushållsundersökningar har idag mer än 20 % bortfall

»Behovet av forskning inom bortfallsområdet är alltså mycket stort.»

och ibland en bit över 50 %. I t ex Arbetskraftsundersökningen, där man vill uppskatta andelar runt 5 %, är bortfallet 23,7 % i genomsnitt men betydligt högre i de grupper som har hög arbetslöshet, t ex unga och invandrare. Då är det inte lätt att skatta på ett bra sätt. Det går knappast idag att redovisa resultaten från en urvalsundersökning utan någon form av bortfallskorrigerande åtgärd. Behovet av forskning inom bortfallsområdet är alltså mycket stort. Det kan som i Marikas fall gälla hur man får ner bortfallet men lika viktigt är att studera bortfallets egenskaper och att vidareutveckla metoder för att kompensera för bortfallet.

Den tredje baltisk-nordiska konferensen i Survey Sampling (BaNoCoSS)

arrangeras den 13-17 juni, förmodligen i Norrfällsviken vid Höga Kusten. Jag sitter i programkommittén och kan berätta att Jean-Claude Deville från Frankrike, Steven Thompson från USA och Giovanna Rannalli från Italien blir huvudtalare. Förutom samplingämnen kommer vi att ta upp många andra ämnen som undersökningar via webb, integritet och röjandekontroll, bortfall och granskning. Mer information finns på webbplatsen <https://wiki.helsinki.fi/display/BNU/Home>. De två tidigare konferenserna blev mycket lyckade och jag hoppas att många av er kommer att delta. Jag törs redan nu säga att konferensen kommer att bli mycket intressant. Naturen vid Höga Kusten är underbar.

Vi börjar också planera inför våren. Surveyföreningens grupp för webbpaneler kommer att hålla ett seminarium då man diskuterar det förslag till kvalitetsriktlinjer för webbpaneler som vi planerar att lägga fram. Seminariet blir förmodligen den 3 februari i Stockholm. Vi planerar också inför Surveyföreningens årsmöte den 23 mars vid Örebro universitet, men där är inga ämnen bestämda än utan vi tar gärna emot idéer.

DANIEL THORBURN



FOTO: MARIA SANDQVIST

For some decades now administrative registers have been an important data source for official statistics alongside survey sampling and population census.

Not only do they provide frames and valuable auxiliary information for sample surveys and censuses, systems of inter-linked statistical registers (i.e. registers for statistical use) have been developed on the basis of various available administrative registers to produce a wide range of purely register-based statistics (e.g. Wallgren and Wallgren, 2007).

A summary of the development of some of the key statistical registers in the Nordic countries is given in UNECE (2007). The next census in 2011 will be completely register-based in all Nordic countries. Reduction of response burden, long-term cost efficiency, as well as potential for detailed spatial-demographic and longitudinal statistics are some of the major advantages associated with the statistical use of administrative registers.

The trend towards more extensive use of registers is increasingly being recognized by statisticians around the world. However, also being noticed is a clear lack of statistical theories for assessing the uncertainty of register-based statistics. It is tempting to reflect over the historical development of survey sampling for comparison. The representative method was presented by Kiær at the ISI meeting in 1895. Despite that he was not able to defend the approach on a theoretical ground, the practice kept evolving in the subsequent years. In 1924, the ISI formed a committee to investigate the feasibility of the representative method. The reporter Jensen stated: "When ISI discussed the matter twentytwo years ago, it was the question of the recognition of the method in principle that claimed most interest. Now it is otherwise. I think I may venture to say that nowadays there is hardly one statistician, who in principle will contest the legitimacy of the representative method. Nevertheless, I believe that the representative method is capable of being used to a much greater extent than now is the case". Indeed, breakthroughs did not come until some 30 - 40 years after Kiær's initial presentation. Today, the contributions by Bowley

DEVELOPING STATISTICAL THEORIES FOR REGISTER-BASED STATISTICS

and, in particular, Neyman (1934) are generally taken as the starting point of the theoretical development of the so-called design-based approach to survey sampling.

Register-based statistics, when viewed in a mirror of history, appear currently pretty much at a pre-Neyman stage in their theoretical development. We believe that the key issue here, from a statistical methodological point of view, is the conceptualization and measurement of the statistical accuracy in register data which will enable us to apply rigorous statistical concepts such as bias, variance, efficiency and consistency, as one is able to do in survey sampling.

Administrative registers certainly do not provide perfect statistical data. In order to identify the various potential error sources, we thought it may be helpful to chart the life cycle of statistical microdata in a way similar to what Groves et al. (2004, Figure 2.5) have done for sample survey data. The resulting two-phase diagram is shown in Figure 1.

The entire life cycle of statistical microdata is generally divided into two phases. The first phase concerns data from a single source, the second phase concerns the possible integration of data from different sources. The life cycle of sample survey data as charted by Groves et al. falls in the first phase. But their concepts have been extended to accommodate data from the administrative sources as well. Let me elaborate along the line of representation. In survey sampling the target set would contain the units of the target statistical population, whereas the accessible set would correspond to the sampling frame. The difference between the two is known as frame error. Next, the accessed set would correspond to the gross sample, whereas the observed set would only contain the respondents. The selection error corresponds then to the sampling error, and so on. But the distinctions are equally applicable to an administrative data source, in which case the difference between the target and accessible sets of objects are often due to feasibility reasons. For instance, a job register may be intended to provide the basis

for the administration of sickness and child-care benefits. The target set should ideally contain all jobs that are 'substantial' enough to qualify for the benefits. To facilitate the reporting in practice, however, only jobs of certain regularity are required to be registered, hence accessible at all. Next, the accessed set would contain all the jobs that are actually registered, and the va-

lidated set the ones that are deemed admissible. Inevitably, some selection errors are associated with the reporting/registration process, while some of the reported jobs may be inadmissible due to confusion surrounding the administrative routines or regulations.

The second phase of integrating data from different sources is an essential feature in register-

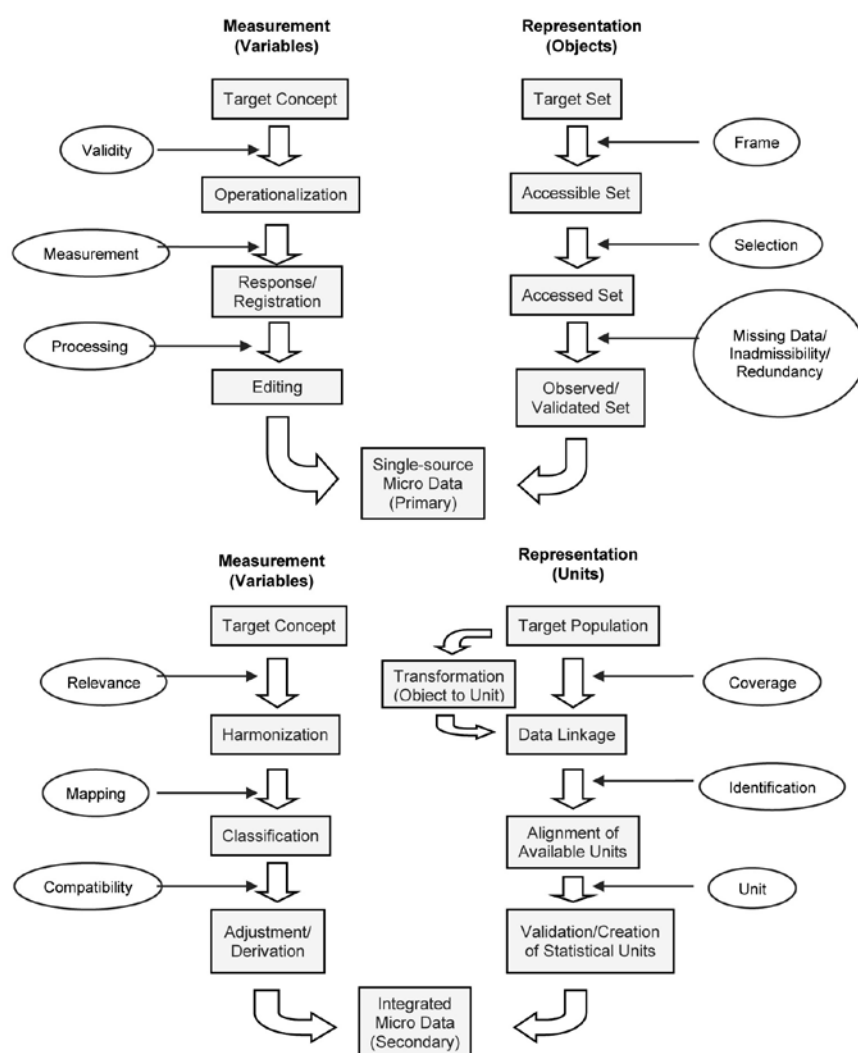


Figure 1. Two-phase (primary and secondary) life cycle of statistical data from a quality perspective. Stage of data production (square frame) and type of potential errors (circle frame). »»»»

»»» DEVELOPING STATISTICAL THEORIES...

based statistical production. The idea is to make secondary use of data that have been collected for a different purpose primarily. Above all, integration with the so-called population base registers, such as the Central Population Register and the Business Register, is necessary in order to incorporate the register data into the statistical system at all, because the target statistical population and variables for integration are almost always different from those of the primary phases. Hence the necessary distinction between objects at the first phase and units at the second phase. In the example of the job register above, what are primarily registered are job events such as hiring and dismissing, which are the objects at the first phase. To turn this information into, say, employment status of persons, i.e. the units for a secondary statistical purpose, conversion from the primary objects must take place. It is important to emphasize that, from the same perspective, also sample survey data should be integrated into the statistical system in order to achieve greater secondary usability, by which the life cycle of sample data will be extended beyond the first phase.

A shared understanding of the potential error sources can help us to collocate and coordinate the different research efforts. Recently, several research initiatives have been undertaken at Statistics Norway. For instance, the uncertainty of the register-based census employment status is being studied. Comparisons are made to the Labor Force Survey data. Since a main cause of difference is the employment definition in each source, which is referred to as relevance error (2nd phase, Figure 1), a theory for valid and equivalent statistical micro data is being developed in connection. Another example is unit errors (2nd phase, Figure 1) in the household register. The statistical

unit 'household' does not exist in any administrative source; it must be constructed from the relevant primary units, such as person, family and dwelling. Errors are unavoidable. Again, the problem is rooted in the secondary use of data. A statistical theory for household unit errors has been developed (Zhang, 2010), which allows one to assess not only the uncertainty in the household statistics, but also the statistics that are based on household units, such as household income and expenditure. Other topics that have been studied include a modeling approach to registration delays and mistakes (i.e. measurement error, 1st phase, Figure 1) for producing statistics at detailed levels (Zhang and Fosen, 2010), and an imputation-based simultaneous prediction approach (Zhang, 2009) for statistical registers (i.e. selection and missing data error, 1st phase, Figure 1). All these topics are highly relevant for the forthcoming register-based census.

To summarize, all national statistical institutes face currently the challenge of finding a way between budgetary constraints and ever increasing demand on statistical information. Efficient use of all data available is naturally an option that must be explored. The 20th century has witnessed the birth and maturing of sample surveys. The 21st century will be the age of data integration, and administrative register data are a major part of it. As statisticians, it is our duty to make it come out right.



LI-CHUN ZHANG,
STATISTISK SENTRALBYRÅ,
NORGE

Referenser

Groves, R.M., Fowler Jr., F.J., Couper, M., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2007). *Survey Methodology*. New York: Wiley.

Neyman, J. (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society*, nr 97, 558-606.

UNECE (2007). *Register-based Statistics in the Nordic Countries: Review of Best Practices with Focus on Population and Social Statistics*. United Nations Publication, ISBN 978-92-1-116963-8.

Wallgren, A. and Wallgren, B. (2007). *Register-based Statistics - Administrative Data for Statistical Purposes*. John Wiley & Sons, Ltd.

Zhang, L.-C. (2009). A Triple-goal Imputation Method for Statistical Registers. Paper presented at the UNECE Work Session on Statistical Data Editing, Neuchatel.

Zhang, L.-C. (2010). A Unit-error Theory for Register-based Household Statistics. To appear in *Journal of Official Statistics*.

Zhang, L.-C. and Fosen, J. (2010). Assessment of Uncertainty in Register-based Small Area Means of a Binary Variable. Submitted.



Baltic-Nordic-Ukrainian network (BNU) är ett nätverk som sedan 1997 anordnat ett flertal konferenser och sommarskolor på temat urvalsundersökningar (survey sampling). I augusti i år genomfördes årets arrangemang i form av en workshop i Vilnius.

51 deltagare från nio olika länder fanns på plats, varav sju särskilt inbjudna talare. Deltagarna bestod av såväl yrkesverksamma surveystatistiker som representanter från den akademiska världen. En stor grupp utgjordes av studenter på master- och doktorandnivå. Under en intensiv vecka fick vi lyssna på föredrag, besöka sevärdheter och prova diverse inhemska specialiteter såsom Skilandis (rökt fläskkorv) och Zeppelinare (en sorts litauiska dumplings).

Huvudtalare på workshopen var Li-Chun Zhang (Statistisk sentralbyrå i Norge) och Carl-Erik Särndal (emeritus på Université de Montréal och Statistiska centralbyrån). Li-Chun Zhangs föreläsningar kretsade kring registerbaserad statistik, medan Carl-Erik Särndal fokuserade på bortfallsfrågor. Övriga inbjudna gäster talade om exempelvis Audit sampling (urval ur räkenskaper) eller om estimationsproblem i undersökningar med roterande urval. Alla deltagare fick dessutom själva hålla kortare presentationer. En eftermiddag vigdes åt rundabordsdiskussioner kring ett antal surveyrelaterade ämnen, och en annan åt datorlaborationer i programvaran Domest (ett användarvänligt Javaprogram med



Korskullen.

FOTO: GUNNAR KULLDORFF

Registerstatistik, sannolikhetsurval och bortfall

vilket modell- och designbaserade skattningar för mindre redovisningsgrupper kan tas fram, konstruerat av Risto Lehtonen och Ari Veijanen).

Li-Chun Zhang talade om behovet av ett teoretiskt ramverk anpassat för registerbaserad statistik och dess särskilda egenskaper. Han menade att en teoretisk grund är nödvändig för att statistikens kvalitet ska kunna utvärderas med hjälp av statistiska metoder. (Se även hans artikel på sidan 20.) Li-Chun Zhang gick igenom de felkällor som är särskilt viktiga inom registerbaserad statistik och teorier för dessa. Ett exempel är så kallade unit errors vilka utgörs av felaktigt avgränsade objekt, t.ex. hushåll som innehåller för få eller för många individer. I den registerbaserade statistiken används ju administrativt datamaterial vilket samlats in för helt andra syften och är ofta baserat på andra objektsdefinitioner. Statistikern måste därför själv skapa de önskade objekten i efterhand. Li-Chun Zhang presenterade en teoretisk modell för att kvantifiera unit errors med avseende på dess påverkan på statistiken.

Carl-Erik Särndal diskuterade bortfallsfrågor och berörde bortfalllets konsekvenser både för

datainsamling och för estimation. Problemet med stigande bortfallsfrekvenser diskuterades särskilt. Särndal frågade sig med anledning av detta om det nu är dags att se över den praxis som råder för sannolikhetsurval. Kanske måste praxis anpassas till rådande omständigheter? Han liknade det hela vid en liten pojke som vuxit ur sin kostym. (Det är inget fel på kostymen, den passade perfekt för ett par år sedan. Men pojken har helt enkelt vuxit ur den.) Särndal redogjorde även för sitt begrepp balanserad svarsmängd (balanced response set). En balanserad mängd svaranden karaktäriseras av att den representerar hela stickprovet på ett bra sätt, då stickprovsmedelvärdena för samtliga hjälpvariabler förblir desamma med eller utan bortfallet. Han introducerade ett kvantitativt mått på hur pass obalanserad en svarsmängd är (lack of balance), samt diskuterade hur detta mått kan användas som underlag för beslut vid bortfallsuppföljningar. Ytterligare indikatorer, kallade bias indicators, introducerades slutligen och Carl-Erik Särndal beskrev hur de kan vara till nytta när olika hjälpvariabler ska utvärderas mot varandra. Indikatorerna är något olika, men kortfattat kan man säga att de tar hänsyn både

till hur pass väl hjälpvariablerna samvarierar med undersökningsvariabeln och till vad de säger om lack of balance.

I övrigt presenterade Markus Gintas Sova, från Office for National Statistics i Wales, en fallstudie där förändringsskattningar baserade på överlappande urval undersöktes. Dan Hedlin (Statistiska centralbyrån och Stockholms universitet) diskuterade hur man kan korrigera för den typ av undertäckning som följer av att det ofta tar lång tid innan nya objekt registreras i en urvalsram. Onsdagen vigdes emellertid åt en längre utflykt till ett antal sevärdheter i närheten av Vilnius. Höjdpunkten utgjordes av Korskullen (Hill of Crosses), ett välkänt resmål för katolska pilgrimer från hela världen. På en liten kulle strax utanför staden Siauliai står tiotusentals kors av olika storlekar uppställda. Känslan som infann sig vid besöket var närmast spöklik och väl värd resan dit. Workshopen avrundades med sedvanlig avskedsmiddag för samtliga deltagare.

OLIVIA HELLBERG,
STOCKHOLMS UNIVERSITET

OPINIONS- UNDERSÖKNINGAR

Två veckor efter valet ordnade Surveyföreningen ett seminarium om opinionsundersökningar.

Opinionsinstitutet gjorde bra ifrån sig i årets val. I alla fall om man ska tro Arne Modig från Novus opinion. Enligt Arne var årets val ett "lätt" val; alla opinionsundersökningar låg nära det slutliga utfallet, även om man numera inte gör anspråk på att göra prognoser om valutgången utan håller sig till det rådande opinionsläget vid mättidpunkten. En orsak till att mätresultaten låg nära det slutliga utfallet kan vara att det inte förekom några dramatiska svängningar i opinionsläget under månaderna närmast före valet. De dramatiska förändringarna inträffade i stället under mellanvalsperioden.

Historiskt sett har opinionsinstitutet klarat sig bra under de senaste tjugo åren, undantagen är valet 2002 när man kraftigt underskattade Socialdemokraterna, samt valet till Europaparlamentet 2004. Generellt är det svårare att göra

en bra mätning i val där valdeltagandet sedan visar sig vara lågt, som det har varit i valen till Europaparlamentet.

Det finns sju opinionsinstitut (oräknat SCB) i Sverige i dag som alla ställer samma fråga:

hur skulle du rösta om det var val idag? Några ställer dessutom ett par kompletterande frågor för att fånga valbeteendet. Den dominerande metoden för datainsamlingen är telefonintervju. På sextioalet fanns det bara ett opinionsinstitut, Sifo. SCB började med partisympatiundersökningen 1972. Temo började med mätningar 1979 och därefter har flera företag tillkommit. Allra flest publicerade mätningar gjordes inför valet 2002, hela 75 stycken. Därefter har det varit färre, i årets val gjordes sammanlagt 45 stycken publicerade mätningar av de olika företagen i branschen.

Novus håller en databas över resultaten från de olika opinionsinstitutens alla mätningar de senaste tio åren. Tidsserierna visar att institutet har sett ungefär samma utveckling över åren. Vissa skillnader finns dock och kan till viss del förklaras av att man använder något olika metoder. Databasen finns fritt tillgänglig i form av en Excelfil på <http://www.novusgroup.se/> under Publicerade undersökningar.

Inför årets val har Novus även gjort metaanalyser på väljarbarometrar, dvs. vägt ihop resultat från flera undersökningar. Det har gjorts i USA ett bra tag, men någon sådan analys har inte tidigare publicerats i Sverige. Metaanalyserna fungerar bra tack vare att man i alla valundersökningar ställer samma fråga. Resultaten i publicerade mätningar vägs samman efter antalet

Den dominerande metoden för datainsamling är telefonintervju.

intervjuer i respektive undersökning. I de fall intervjuperioden går över en månadsgräns delas resultatet utifrån ett antagande om att intervjuerna är jämnt fördelade under den redovisade intervjuperioden.

Ett nytt inslag inför årets val var att opinions-

undersökningar gjordes på nätet. Carl Melin från United Minds, som är ett av två opinionsinstitut som endast använder sig av webben vid datainsamlingen, berättade om både för- och nackdelar med icke-

»Allra flest publicerade mätningar gjordes inför valet 2002, hela 75 stycken. Därefter har det varit färre»

slumpmässiga urval i form av webbpaneler. Det blir allt ovanligare med fast telefon och det är inte heller särskilt enkelt att nå mobiltelefoner. Antalet som avböjer att delta på telefon ökar. Det gör det svårt att realisera ett slumpmässigt urval. Människor arbetar och lever i nya cykler vilket opinionsinstitutet behöver anpassa sig till, till exempel genom att göra det valfritt när man kan svara på frågor. Till nackdelarna hör att webbpaneler medför självselektering och inbjuder till "proffstyckande". Man vet att man har dålig representation bland äldre, eftersom det i den gruppen finns många som inte använder Internet. Man vet också att personer inom vissa partitillhörigheter är mer aktiva på nätet, som miljöpartister och sverigedemokrater.

Urvalet till en undersökning hos United Minds görs ur en panel som omfattar ungefär fem procent av Sveriges befolkning. Även om man försöker få urvalet representativt genom kvotering av urvalet och vägning av svaren så är det inte likvärdigt med ett slumpmässigt urval. Detta är man medveten om, men Carl Melin ansåg att eftersom det inte går att göra slumpmässiga urval i en webbundersökning så får man vara nöjd med att försöka göra det näst bästa så bra som möjligt. United Minds har kontrollerat sin panel och anser att den är representativ, med





ORDFÖRANDEN HAR ORDET

undantag för personer som är äldre än 70-75 år.

Hur fungerade webbundersökningarna? United Minds överskattade antalet sverigedemokrater och antalet miljöpartister. Man överskattade också flödet från Alliansen till SD. Antalet moderata anhängare underskattades, liksom antalet socialdemokrater. Man har också insett att det inte finns något tydligt samband mellan att gissa rätt på parti, block respektive valutgång. Metoden med webbundersökningar är fortfarande relativt ny och det finns mycket som måste utvecklas vidare.

För att få lite perspektiv på användningen av webben som insamlingsmetod kan man fundera på Arne Modigs kommentar om hur man på sjuttioalet såg på övergången från besöksintervju till telefon i valundersökningar. I debatten fanns stor tveksamhet huruvida sådana viktiga frågor verkligen kan ställas på ett så lättsinnigt sätt som på telefon.

En fråga från publiken handlade om det faktum att tidningar står för en stor del av finansieringen av de flesta opinionsundersökningar. Förekommer det att tidningar styr eller försöker styra resultaten? Både Arne Modig och Carl Melin menade att så inte är fallet även om det förekommer konspirationsteorier.

INGEGERD JANSSON

Gott nytt år. Tack alla medlemmar som bidragit till våra möten 2010 med sina presentationer, sitt deltagande och engagemang.

Framtiden blir snart nutid och sedan väldigt snabbt dåtid vilket den sedan är för evigt. Visst är det märkligt att vi förundras över att tiden går? Borde vi inte vant oss vid det här laget? Ett nytt år inleds. Ett år som vi vill ska bli särskilt gott, glatt och friskt. Nyårslöfte från mig? Nej, inte i år heller.

Det är även dags att skriva verksamhetsberättelsen eller den egna årskrönikan. Detta kanske inte bara är en viktig uppgift för föreningar, företagsledningar och redaktioner utan även för var och en. Även om tiden aldrig står stilla utan är ett kontinuum, där händelserna sker i sekvens, försöker vi skapa oss en slags paus/andhämtning/eftertänke så här vid årsskiftet. 2010 har blivit dåtid. Det känns naturligt att sammanfatta det som varit och börja om på nytt. För Statistikfrämjandet och dess sektioner sker visserligen nystarten på allvar först efter årsmötet i mars. FMS-medlemmar, vi ses väl på FMS årsmöte onsdagen den 23 mars i Örebro? Boka in detta i era fina nya almanackor redan nu.

När ni läser denna spalt har vår höstkonferens i Södertälje med titeln "Multiplicity, why, when, where and how?" ägt rum. Det mötet arrangerade vi tillsammans med Cramér-sällskapet. Det är givande att korsbefrukta olika sektioner och få höra om olika angreppssätt från skilda tillämpningsområden. Det märktes också tydligt på den mycket berikande Statistikerträffen i Göteborg i september. Läs mer om den i artikeln på nästa sida. Styrelsen vill ännu en gång gratulera och lyckönska våra tre stipendiater år 2010:

Sharon Köhlmann Berenzon, Sandra Eloranta och Anna Johansson. Jag hoppas att era resor till IBS-konferens respektive workshop var både trevliga och professionellt utvecklande.

»Det är givande att korsbefrukta olika sektioner och få höra om olika angreppssätt från skilda tillämpningsområden.»

Vi i styrelsen arbetar förstås med nya framtida aktiviteter. Planer finns på att ordna ett EFSPi-möte i Sverige med tema "Model Based Drug Development". Förkortningen EFSPi står för "European Federations of Statisticians in the Pharmaceutical Industry" där FMS och likasinnade föreningar i Europa är medlemmar. Ett annat ämne som vi hoppas få utforska tillsammans med icke-statistiker är adaptiv design i praktiken. Styrelsen har även fått in förslag på möjliga ämnen för workshopar, t.ex. hantering av saknade värden (missing values). Saknade data är väldigt störande och utmanande för oss statistiker. Vi antar gärna att data är "missing at random" men hur trovärdigt känns det? Nja, inte så värst.

Apropå saknade data: du som är medlem i FMS och vill vara "non-missing": har du flyttat eller bytt mejladress? Skicka in ändringar till fmsstyrelse@gmail.com. Har du ännu inte blivit medlem men önskar det? Du är välkommen att sända en ansökan till oss.



ANNA TORRÅNG

VAD ÄR GEMENSAMT?

Nacksmärta, hållbara investeringar, träd, DNA-sekvensiering, robust konstruktion

Den 22 september samlades 54 förväntansfulla statistiker till Statistikerträff i Göteborg. Programmet var mer späckat än tidigare och dessutom mer varierat. Som vanligt spände det över ett antal olika tillämpningar och nivåer av teori och praktik.

Mest hett var förmodligen det sista föredraget som handlade om internetbaserade panelundersökningar av väljarsympatier.

Valet 2010 kännetecknades av att vara mycket jämnt mellan regeringsalternativen och av "jokern" Sverigedemokraterna. Intresset för väljarundersökningar måste ha slagit rekord! Statsvetaren Stefan Dahlberg redogjorde för hur man i deras internetpanel, E-panelen, testat metoder för att justera för systematiska avvikelser. E-panelen drivs av gruppen för Multidisciplinär forskning om opinion och demokrati vid Göteborgs universitet. Åhörarna hade många förslag och synpunkter kring detta. Eftersom internetpaneler är en metod för attitydundersökningar som lär bli mer vanlig är det angeläget att utveckla metoder för korrigering. Detta arbete pågår som sagt och diskuteras på många platser.

Miljö- och klimatfrågan diskuterades även den i valet. Man kan angripa detta på flera sätt, ett är att inom ramen för det marknadsekonomiska systemet skapa incitament som gör att gröna investeringar även är lönsamma ekonomiskt. För att nå dit måste de som agerar på finansmarknaden ges rätt incitament att välja

»DNA-sekvensiering utmärks av att man använder nästan absurd stora data-material.«



Nästa träff 2011

Nästa Statistikerträff kommer att arrangeras i september 2011 i Göteborg. Datum är ännu ej bestämt.

Opinionsundersökning på Internet.

det som på sikt ger förbättringar antingen för miljön/klimatet eller för mänskliga rättigheter eller liknande. Mattias Sundén bjöd på några av de resultat som forskargruppen han deltar i kommit fram till kring sådana incitamentprogram. Dessa bygger ofta på bonusvillkor där det lönar sig för förvaltare att investera i "gröna" företag om dessa utvecklas bättre än jämförbara konventionella företag.

Statistiskt arbete brukar kännetecknas av att antingen röra minsta möjliga datamängd eller en okontrollerbart stor sådan. DNA-sekvensiering utmärks av att man använder nästan absurd

stora datamaterial. Erik Kristiansson illustrerade detta genom att dra paralleller till datorernas utveckling. Dessa har enligt Moores lag fördubblat sin prestanda var 24:e månad, men DNA-sekvensieringen gör motsvarande ökning på åtta månader. Dessa två processer går alltså för tillfället i otakt, då antalet registrerade baspar ökar fortare än datorerna hinner analysera.

DNA handlar om klara och tydliga data – om än oerhört omfattande. Andra utmaningar möter Anna Ekman som inom kort kommer att disputera på en avhandling om modellbaserade analyser för nacksmärta. Smärta är i sig ett vagt begrepp, vilket ökar variationsmöjligheterna i data. Samtidigt är denna typ av åkomma mer och mer prevalent. Förutom att vara ett stort

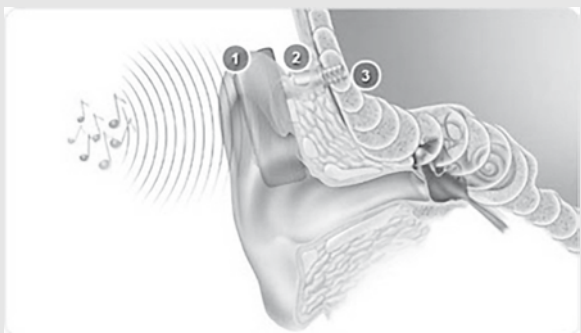


Bild från Cochlears webbplats.

Baha-system

1. En hörapparat fångar upp ljudvibrationerna.
2. En distans är fäst i hörapparaten och implantatet. Distansen överför ljudvibrationerna till implantatet.
3. Ett litet titanimplantat är placerat i benet bakom örat, där det sammansmälter med det levande benet. Denna process kallas osseointegration*. Implantatet överför ljudvibrationerna via skallbenet till det fungerande inreröret.

* Grunden för Baha är Brånemark system®. I början av 1960-talet upptäckte professor Brånemark att titan accepteras av människokroppen och växer ihop med det omgivande benet till en permanent struktur. Han benämnde denna process "osseointegration".

hinder för individen kan vi även konstatera att samhällets kostnader för arbetsrelaterad smärta är omfattande. Alla försök att på ett korrekt sätt mäta och modellera hur detta uppkommer och kan förhindras borde vara prioriterade.

Även Aila Särkkää talade om modeller, vilket i hennes föredrag handlade om hur träd växer i en skog. Tillväxten illustreras av cirklar, vilka motsvarar de anspråk som trädet gör på sin omgivning. Tillväxten kommer att styras utav flera miljöfaktorer samt, vilket är Ailas speciella intresse, den samvariation som finns med andra träd i grannskapet. Modellen kan tillämpas på flera områden där tillväxten begränsas eller gynnas av "grannar".

Robust konstruktion är ett växande område i kölvattnet av sex sigma (Six Sigma på engelska). Martin Arvidsson presenterade ett exempel från företaget Cochlear på hur experiment genomförts på skruvförband där målet är att säkerställa att egenskaperna hos levererade produkter endast varierar inom acceptabel tolerans. Det fortfarande höga intresset för

statistik inom sex sigma innebär att statistiker har goda möjligheter att medverka i industriella utvecklingsprocesser!

Statistikerträffen, som hölls för fjärde året i rad, har utvecklats till en spännande och livaktig mötespunkt. Arrangörerna - Astra Zeneca, FMS, Göteborgs universitet och Statistikkonsulterna - har som ambition att träffarna ska belysa många olika verksamhetsfält där statistiker kan vara inblandade, vad man arbetar med och hur man samverkar med kollegor med annan kompetens. En närliggande tanke är att man ska kunna lära av varandra och kanske upptäcka att man brottas med likartade statistiska problemställningar. Hur ofta förekommer det inte att sådan intressegemenskap döljs av att olika verksamhetsområden är ganska isolerade från varandra och att man har olika språk och begrepp?

MAGNUS PETTERSSON OCH
JONNY OLOFSSON,
STATISTIKKONSULTERNA

ORDFÖRANDEN HAR ORDET

På gång i Cramérsällskapet

Som ordförande i Cramérsällskapet känns det angeläget för mig att utnyttja privilegiet av en egen spalt i Qvintensen. För egen del har jag tillbringat en del av hösten på Nuffield College i Oxford eftersom jag har sabbatstermin från min tjänst på Uppsala universitet. Det är också ett värdefullt privilegium. Här i Sverige är det vad jag vet bara vi som har den "gamla sortens" professurer som åtnjuter det. I många andra länder, t.ex. i England men också i Norge och Danmark, har även lektorer sådana möjligheter. Dessutom har de, liksom professorerna, goda möjligheter till forskning inom tjänsten (ofta 50 %). Detta står återigen i kontrast mot Sverige där ju lektorerna som regel hänvisas till att söka externa medel för forskning. Nu ska jag inte orda mer om detta här, även om dessa frågor är väl värda att ventileras oftare än vad som görs. Men det ska kanske ske även i ett annat forum än Qvintensen? Istället tänkte jag ta tillfället i akt att påminna lite om Cramérsällskapets verksamhet.

Vi har haft vårt mycket lyckade höstmöte som arrangerades i samarbete med FMS. Platsen var AstraZeneca och ämnet var multiplicitet av test. Kopplingarna till genetisk statistik är starka. Där har man ibland till och med fler parametrar än observationer. (Detta fick jag lära mig en del om på den förträffliga Nordstat-konferensen i Voss i somras.)

I vår, den 23-24 mars, är det dags för nästa årsmöte, i Örebro. Förra året diskuterade vi masterutbildningarna i statistik och matematisk statistik, och i år känns det som ett naturligt steg att gå över till forskarutbildningarna. Fungerar de som de ska, och om inte, hur ska vi göra för att förbättra dem? Vad kan man förvänta sig av en doktor i (matematisk) statistik? Kom och bidra till diskussionen med dina synpunkter!

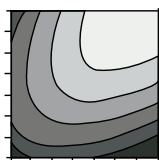
I sommar är det vår ambition att arrangera vår första sommarskola. Vi tar då upp den avsmnade traditionen från Statistikersamfundet med en konferens på några dagar som i huvudsak riktar sig till doktorander (även om alla andra också är hjärtligt välkomna) och med ett par huvudföredragshållare inom ett angeläget ämne. I skrivande stund finns inga specifika planer, men vi återkommer i ämnet när saker och ting har klarnat.

Till sist, om du tycker att vår verksamhet låter intressant men ännu ej har blivit medlem, tveka då inte utan gå in på vår hemsida och bli det!

ROLF LARSSON



»Vad kan man förvänta sig av en doktor i (matematisk) statistik?»



Ordföranden har ordet

Nu när dagarna är korta och mörkret dominerar, åtminstone hos oss som bor i den norra landsändan, känns det bra att kunna tänka tillbaka på en av höstens ljuspunkter när det gäller industriell statistik, nämligen ENBIS10-konferensen.

Föreningen Industriell statistik är ansluten till det europeiska nätverket European Network for Business and Industrial Statistics (ENBIS). Fjolårets ENBIS-konferens var vi med och organiserade på Chalmers i Göteborg. Årets ENBIS-konferens, som dessutom var tioårsjubileum, hölls i september i Antwerpen i Belgien. Som vanligt bjöds det på ett varierande program med en blandning av teori, tillämpning och praktik och med föredragshållare från ett stort antal olika länder. Kvällsaktiviteterna hölls i fascinerande miljöer i denna gamla europeiska stad.

Sverige var väl representerat med två huvudtalare, Marianne Frisén, Göteborgs universitet och Boris Lorenc, SCB, och många andra föredragshållare. Marianne Frisén gav en intressant och givande överblick över området statistisk övervakning i föredraget "Methods and evaluations for surveillance in industry, business, finance, and public health". Det handlar om att kunna upptäcka viktiga förändringar i en process så snart som möjligt efter en ändring har skett. Statistiska teorier för detta har utvecklats inom olika områden, bl a inom den industriella, finansiella och medicinska sektorn. Inom industriell statistik kallas det för SPC (statistisk processtyrning). I grunden är den statistiska

teorin densamma oavsett område och resultat inom ett område kan stödja utvecklingen inom ett annat, vilket är värdefullt att känna till. För den som vill läsa mer om detta rekommenderas webbplatsen www.statistics.gu.se/surveillance.

»ENBIS-konferenserna rekommenderas varmt.»

Boris Lorenc föredrag hade titeln "Modern official establishment statistics: trends and prospects". För oss som inte är så bevandrade i den officiella statistiken gav det en bra överblick. Övriga huvudtalare var Henry Wynn, en känd profil inom industriell statistik från London School of Economics, som talade om "Statistics: the number one discipline for business and industry" samt Christophe Perruchet, UTAC, Frankrike, vars föredrag handlade om statistik och standardisering. Ett underhållande föredrag, med titeln "Three laws of communication" gavs under konferensmiddagen av Jean-luc Doumont, Principiæ, Belgien. Där visade han tydligt och med humor hur dessa "lagar" kan förtydliga ett budskap i en muntlig presentation. Vi var nog många som insåg att här finns mycket att lära. Han har även skrivit en bok i ämnet, "Trees, Maps and Theorems", se <http://www.treesmapsandtheorems.com>.

Mycket intressant att lyssna på var sessionen "The Madison legacy and its implications for the past, present, and future of business and industrial statistics" som var dedicerad till en av ENBIS grundare, Søren Bisgaard, som så sorgligt gick bort för knappt ett år sedan. Søren Bisgaard var under en längre tid verksam vid University of Wisconsin-Madison, USA, där han också disputerade. Han lämnade under sin tid många

och väsentliga bidrag till industriell statistik. Han var också en person som lämnade avtryck hos alla som fick förmånen att lära känna honom. Under sessionen berättade många kända personligheter inom området industriell statistik, som på ett eller annat sätt varit knutna till University of Wisconsin-Madison, om sin tid där. Förutom Søren har kända statistiker som George Box och Bill Hunter varit verksamma där.

För dem som intresserar sig för undervisning fanns också en session där bl a Douglas Stirling, Massey University i Nya Zeeland, berättade om många goda idéer i föredraget "CAST: A Collection of e-books and other resources using dynamic diagrams to help teach statistics", se <http://cast.massey.ac.nz>.

Detta och mycket annat fanns att lyssna på för den som är intresserad av "business and industrial statistics" i vid mening. Programmet och presentationerna är lätt åtkomliga på ENBIS webbplats <http://www.enbis.org>. Nästa års konferens går av stapeln i Coimbra i Portugal, 4-8 september. ENBIS-konferenserna rekommenderas varmt. Det är inte bara för alla intressanta föredrag, posters, workshopar och givande endagskurser, som ges före och efter konferenserna, utan också för att få möjlighet att träffa alla stimulerande och härliga personer som deltar och knyta nya kontakter. Det är ju det som idag kallas för att nätverka.

KERSTIN
VÄNNMAN



Föreningen Industriell statistik är ansluten till det europeiska nätverket European Network for Business and Industrial Statistics (ENBIS).

»Ett binärt resultat kan tyckas ganska enkelt och okomplicerat men rymmer en hel del teoretiska och praktiska svårigheter.»



Binära resultat.

”Fungerar/ fungerar inte” – en fortsättning

Rubriken – Fungerar/fungerar inte – använde vi under seminariet i samband med årsmötet. Vi återanvände den för vårt tvådagars höstseminarium i oktober på Chalmers i Göteborg.

En stor mängd av alla industriella mätdata är nämligen av binär natur. Ett binärt resultat kan tyckas ganska enkelt och okomplicerat men rymmer en hel del teoretiska och praktiska svårigheter. Seminariets första dag dominerades av en introduktion till programspråket R.

Med detta som en grund rullades ärmarna upp för dag två som helt ägnades åt analys av binära data och händelsedata med utgångspunkt från logistisk regression och Poissonregression.

Vi hade engagerat två duktiga och erfarna föreläsare för de två dagarna. Mats Franzén har lång erfarenhet från universitet, industri och offentlig sektor – och en glupande aptit på och oslagbara kunskaper om datorer, datorprogram och källor på Internet. Han lärde oss mycket om R och delade med sig av sina erfarenheter av långt användande av R.

Leif Nilsson, lektor i matematisk statistik i Umeå och med stor erfarenhet av programvaror, uppdragsgivare och knepiga situationer, lotsade oss genom det enkla, och det som vi felaktigt trodde var enkelt, men som Leif förklarade för oss.

Det var en intresserad grupp av deltagare med många frågor och egna erfarenheter. Var och en fick många nya kunskaper och referenser från både Mats och Leif.

Avslutningsvis vill vi tacka Chalmers som låtit oss utnyttja lokaler på vad som tidigare var Vasa sjukhus i Göteborg.

INGEMAR SJÖSTRÖM



New York.



Örebro.



Cancún.

KALENDARARIUM

DETTA HÄNDER 2011

	KONFERENS	PLATS	WEBBPLATS
28 feb – 3 mars	The Second International Symposium on Biopharmaceutical Statistics	Berlin, Tyskland	www.isBioStat.org
6-10 mars	The 42nd Winter Conference in Statistics (Incomplete data: semi parametric and Bayesian methods)	Hemavan, Sverige	snovit.math.umu.se/aktuellt/vinterkonf/v-konf-11/index.asp
23-24 mars	Årsmöteskonferenser, Svenska statistikfrämjandet	Örebro, Sverige	www.statistikframjandet.se
21-24 apr	International Conference on Probability, Statistics and Data Analysis (ICPSDA-2011)	Raleigh, USA	www.iisaconference.info/
10-13 maj	International Conference on Design of Experiments (ICODOE-2011)	Memphis, USA	www.mscl.memphis.edu/ICODOE2011
6-9 juni	3rd Nordic-Baltic Biometric Conference	Åbo, Finland	www.nbbc11.utu.fi
19-21 juni	2nd International Workshop on Integer Valued Time Series	Protaras, Cypern	http://www2.ucy.ac.cy/~wints2011/
19-24 juni	The 35th Conference on Stochastic Processes and their Applications (SPA 2011)	Oaxaca, Mexiko	abalontico.matem.unam.mx/SPA/index.php
26-29 juni	ICSA 2011 Applied Statistics Symposium	New York, USA	www.icsa.org/meetings/symposia/
4-8 juli	PROBASTAT 2011	Smolenice Castle, Slovakien	www.um.sav.sk/en/probastat2011.html
6-8 juli	The 2011 International Conference of Computational Statistics and Data Engineering	London, England	www.iaeng.org/WCE2011/ICCSDE2011.html
30 juli – 4 aug	2011 Joint Statistical Meetings	Miami Beach, USA	www.amstat.org/meetings/
21-26 aug	International Statistical Institute, 58th ISI World Statistics Congress	Dublin, Irland	www.isi2011.ie/
26-29 aug	65th European meeting of the Econometric Society	Oslo, Norge	eea-esem2011oslo.org/
4-8 sep	ENBIS-11	Coimbra, Portugal	www.enbis.org
5-9 sep	17th European Young Statisticians Meeting	Lissabon, Portugal	www.fct.unl.pt/17eysm/
5-7 okt	9th International Conference on Health Policy Statistics	Cleveland, USA	www.amstat.org/meetings/ichps/2011/index.cfm

...OCH LITE LÄNGRE FRAM

10-14 juni 2012	Nordic Conference on Mathematical Statistics	Umeå
-----------------	--	------

STATISTICAL SIGNIFICANCE

Statisticians provide policymakers with a range of information that covers energy production, consumption, imports, exports, inventories, prices, and environmental effects. They also provide projections for future activities and analyses of the possible effects of changes in technology, regulations, and tax policies.

Statistical analyses based on high-quality data promote sound policymaking, efficient markets, and public understanding of energy and its place in the economy and environment.

Energy Policy

TRACKING ENERGY PRODUCTION AND CONSUMPTION:

One of the most basic activities of statisticians in the energy area is tracking energy production and consumption on a global, national and regional basis. This poses many challenges. For example, because it is impractical to measure end-use consumption across millions of homes and businesses, consumption must be estimated from measures such as withdrawals from inventories at fuel storage points. Estimates of production and consumption have implications for the price of energy as well as broader issues such as our national exposure to disruptions in energy supplies.

DETERMINING ENERGY EFFICIENCY:

Statisticians design surveys of end users in the residential, manufacturing, and commercial building sectors to learn about differences in energy consumption. These data can be used to identify the best opportunities for energy savings and measure the effects of energy efficiency programs. Statisticians can distinguish the effects of efficiency programs from the effects of other factors by comparing energy usage among groups with and without efficiency program interventions, or by using



analyses that statistically control for other differences among groups.

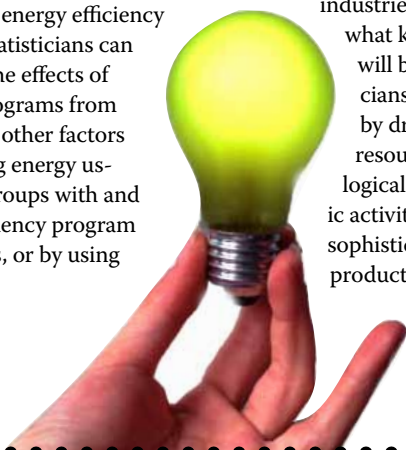
PROJECTING FUTURE ENERGY SUPPLY AND DEMAND:

How much energy might we need to satisfy U.S. and global demand in 10, 20, or 30 years? What energy supplies might be available to meet this demand? The answers to these questions have profound implications, ranging from the expected costs of energy to greenhouse gas mitigation to what kinds of industries will be profitable and what kinds of transportation will be economical. Statisticians answer these questions by drawing on projections of resource availability, technological capabilities, and economic activity, all pulled together with sophisticated models of energy production and consumption.

MODELING EFFECTS OF POLICY INTERVENTIONS:

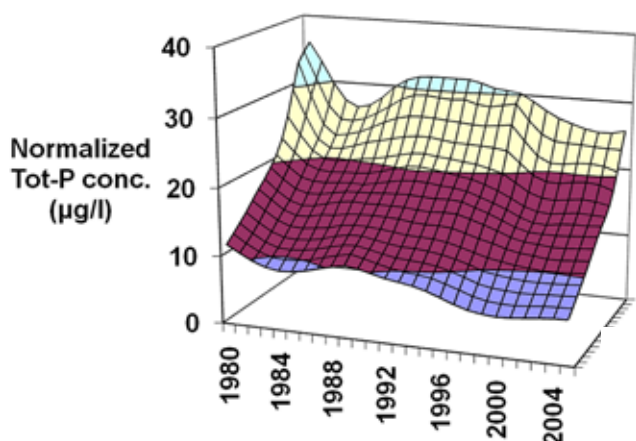
The complex models used to project energy supply and demand can be used to help determine the potential effects of policy initiatives. Policymakers can ask “what if” questions: What if we change automotive fuel efficiency requirements? What if we mandate that electric utilities use some percentage of renewable energy? What if we invest in developing clean fuel technologies? Statisticians can project the possible effects on variables such as the cost of producing energy (which affects economic activity and household energy costs) and energy-related emissions of greenhouse gases.

“Statistics Informs Energy Policy” is part of Statistical Significance, a series from the American Statistical Association documenting the contributions of statistics to our country and society. For more in this series, visit www.amstat.org/outreach/statsig.cfm. The American Statistical Association is the foremost professional society of statisticians, representing 18,000 scientists in industry, government, and academia: www.amstat.org. This *Statistical Significance* was produced under the supervision of the ASA Committee on Energy Statistics.

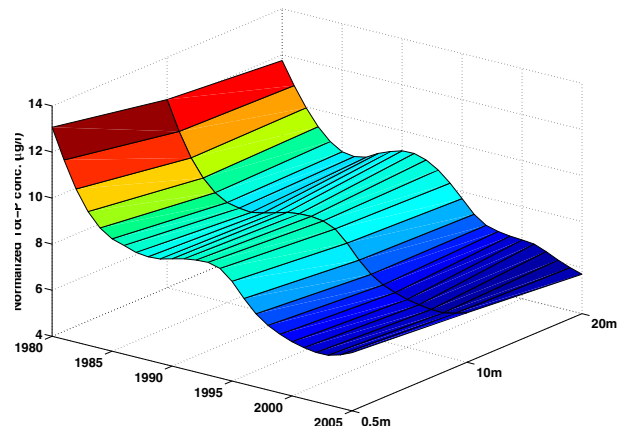


Bättre datakvalitet med simultan analys

Läs mer på sidan 12.



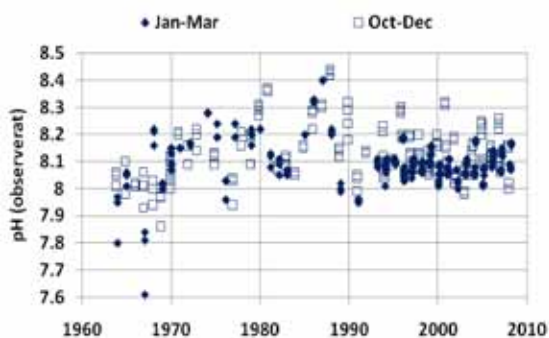
Figur 1. Totalfosforkoncentration för femton vattendrag över åren 1980-2005. Indalsälven med lägst koncentration längst fram, Råne älv med högst koncentration längst bak i figuren.



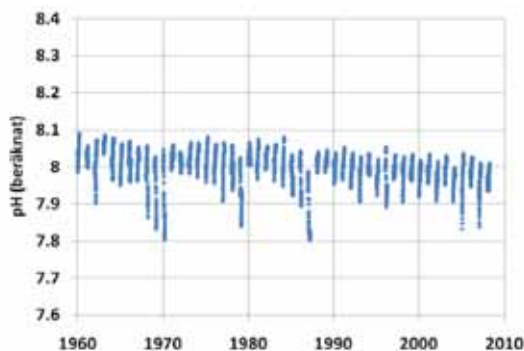
Figur 2. Totalfosforkoncentration på tre djup i Vänern (0,5 m, 10 m och 20 m, år 1980-2005).

Havets försurning – data, modeller och statistikernas roller

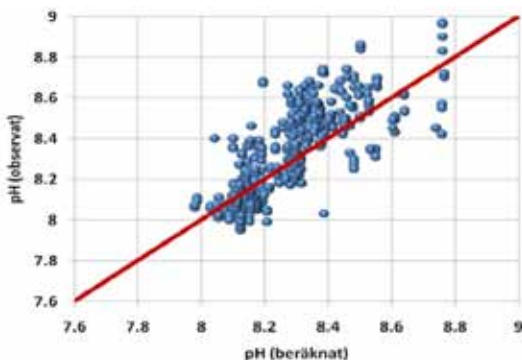
Läs mer på sidan 14.



Figur 1. Uppmätta pH-värden i Östra Gotlandsdjupet (SMHI).



Figur 2. Beräknade vintervärden av pH i Östra Gotlandsdjupet.



Figur 3. Jämförelse av uppmätta och beräknade pH-värden.