

HALTANDE
FASTIGHETSINDEX

SIDAN 6

ATT GÖRA SIG FÖR-
STÅDD PÅ ENGELSKA

SIDAN 17

Qvintensen

MEDLEMSTIDNING FÖR SVENSKA STATISTIKFRÄMJANDET NR 1 2011

FÖRENINGEN FÖR
MEDICINSK STATISTIK/

Ingredienser:
3 tsk framtid och
1 msk dåtid

SIDAN 26



**Det finns
stora veten-
skapliga för-
delar med att
läkemedels-
företagen har
fast anställda
statistiker.»**

Bernhard Huitfeldt,
SIDAN 7

DISKUSSION OM TEST

SIDAN 20

KRÖNIKA/
**GLOBAL UPPVÄRMNING OCH
STATISTISK SIGNIFIKANS**

SIDAN 12

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



Undersökning.



Fastighetsindex.



Uttorkad saltsjö.



Beyza Björkman.

Innehåll

Goda nyheter för alla som äger
en kommersiell fastighet! 6

OBEROENDE ELLER KOMPETENS?

En fråga om trovärdigheten för statistiker
i läkemedelsindustrin 7

KRÖNIKA/

Olle Häggström 12

Att utvärdera frågor i en statistisk
undersökning 14

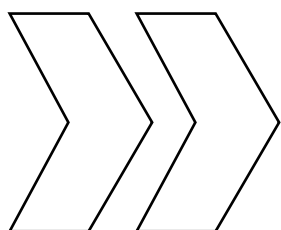
Att göra sig förstådd på engelska 17

DISKUSSION/

Slutreplik i diskussionen om
"Mjuka fakta och osannolik noggrannhet" 20

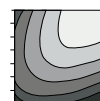
Nollhypoteser och statistisk signifikans 20

STATISTICS INFORMS ENERGY POLICY 31



**Illutformade frågor leder till
dålig datakvalitet i form av mätfel
eller bortfall»**

Andreas Persson
och Fredrik Scheffer, sidan 14



Fasta spalter

I HUVUDET PÅ EN REDAKTÖR 4

ORDFÖRANDEN HAR ORDET 5

KALENDARIUM 30

Våra föreningar

INDUSTRIELL STATISTIK

Osäkerhet 26

FÖRENINGEN FÖR MEDICINSK STATISTIK

Ordföranden har ordet 26

Sol och konkurrerande risker i Brasilien 28

SURVEYFÖRENINGEN

Ordföranden har ordet 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@statstat.su.se

Vice ordförande

Sune Karlsson, 019-30 12 57

Sekreterare

Chandra Adolfsson

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 Adolfsson

Statistiska centralbyrån

701 89 Örebro

E-post sekrfram@gmail.com

Webbplats www.statistikframjandet.se

Förtroende och tillit



Det är när man byter miljö som skillnader i förtroende och tillit framträder. Jag tyckte att det var trivsamt att köra bil i Mogadishu i Somalia. De trafikregler som respekterades begränsade sig i stort sett till att man körde på höger sida. I övrigt kom man överens om hur man skulle göra. Det var för det mesta lugnt och samarbetsvänligt.

I **Stockholm slipper** jag helst att köra bil. I både Mogadishu och Stockholm förutsätter trafiken att alla har förtroende för att medtrafikanterna gör det man förväntar sig. Det förtroendet bygger på trafikregler, polisens uppföljning och en oskriven uppförandekod. Det finns också finansiella villkor som påverkar trafikanterna. Den ekonomiska katastrof för en Mogadishu-bil att smälla bilen föranledde det försiktiga framförandet av fordonen där.

»Det finns inget annat statistiskt område som är så genomreglerat men där förtroendet ändå är så känsligt.»

Förtroende och tillit är mångfaceterade begrepp. Hurvida man litar på "systemet" är också märkligt splittrat. Man ser ofta att samma personer både litar och inte litar på något.

Att Sveriges ekonomiska utveckling just nu är god, tror de flesta. Men hur vet man det? "Statistik" å andra sidan, är ju bara båg.

I **olika statistiska** tillämpningsområden diskuteras förtroendet för statistik. Läs Bernhard Huitfeldts stora artikel i det här

numret av Qvintensen om förtroendet för läkemedelsforskning. Det finns inget annat statistiskt område som är så genomreglerat men där förtroendet ändå är så känsligt.

Det ingår också i tilliten om "grejerna funkar". Fungerar inte saker i rent funktionell mening tynar förtroendet. Statistiker ägnar sin tid till att både säkerställa funktionaliteten och att förklara, dokumentera och prata för varan. Förtroendet för om det språk vi kommunicerar med fungerar är känsligt. Qvintensen har den här gången två artiklar om språket. Den ena handlar om utvärdera frågor i en statistisk undersökning, den andra om användning av engelska i miljöer där ingen är modersmålstalare. I båda fallen krävs det tillit till att språket fungerar.

Och som vanligt finns det mycket mer i Qvintensen, framfört av alla hängivna skribenter. Hur fungerar hypotestest egentligen? Kan vi vara säkra trots att vi alltid är osäkra?

DAN HEDLIN



FOTO: MONICA HOLMBERG

ORDFÖRANDEN HAR ORDET

PÅ GÅNG I SVENSKA STATISTIKFRÄMJANDET

Vi har många intressanta utmaningar: förbättringar för medlemmarna, den etiska koden för statistiker och auktoriseringsfrågan.

Vi börjar lägga det gamla verksamhetsåret bakom oss. För oss i Svenska statistikfrämjandets styrelse har arbetet med vår webbplats varit den stora uppgiften under det gångna året. Det visade sig vara betydligt svårare och mer omfattande än vad vi trodde från början att hitta rätt på alla länkar och filer som behövdes och att lösa upp alla gamla förbindelser. Tack vare en stor arbetsinsats av Silke Burenstam har vi nu till slut kunnat sjösätta den nya webbplatsen. Silke skall ha ett stort tack och en stor eloge för sin insats. När du läser detta kan du gå in på den nya webbplatsen. Efterhand som Främjandets webbplats nu kommer igång, kommer arbetet med medlemsföreningarnas sidor att påbörjas.

Nästa stora problem som styrelsen behöver ta itu med är att arbeta fram ett nytt register som är gemensamt för alla medlemsföreningar och som alla medlemsföreningar kan ha tillgång till. En målsättning är att när en enskild medlem ska betala sin årsavgift är det via en gemensam räkning där alla medlemsföreningarna är representerade. Med ett enkelt kryss markerar man vilken eller vilka föreningar man vill vara medlem i. En annan målsättning är att varje medlemsförening har tillgång till registret. Om en förening vill göra ett utskick till sina medlemmar eller till Främjandets alla medlemmar skall man kunna göra det utan att gå via en central utskicksfunktion. Jag föreställer mig att



Hans Nyquist,
ordförande.

**Hjärtligt välkomna
till en intressant
vårkonferens!**

detta förenklar för både medlemmar och föreningar samt att kontakten mellan föreningar och medlemmar blir bättre och snabbare.

Som en avstamp till det nya verksamhetsåret kommer årsmötet och den tillhörande vårkonferensen den 23-24 mars vid Örebro universitet. Den första dagen är reserverad för medlemsföreningarnas årsmöten och deras aktiviteter. Dagen avslutas med den traditionsenliga middagen. Den andra dagen reserveras för Främjandets aktiviteter och årsmöte. Bland annat kommer vi att få lyssna till intressanta föredrag av Jef Teugels, Torbjörn Tedeén, Hans Stenlund och Tommy Johnsson. Vidare kommer arbetsgruppen med uppgift att arbeta med den etiska koden att rapportera erfarenheter från

kodens första år. På annan plats i det här numret av Qvintensen skriver Bernhard Huitfeldt om statistikers oberoende och integritet i läkemedelsindustrin, som är en bransch där förtroendet för statistikers yrkesetik är särskilt viktig. Även i andra branscher är statistikers trovärdighet helt avgörande för statistikers ställning.

På Främjandets årsmöte kommer även arbetsgruppen som arbetar med frågan om eventuell auktorisering av statistiker att rapportera om hur den har lösts i andra länder.

En höjdpunkt blir då Årets statistikfrämjare presenteras.

Alla medlemmar är hjärtligt välkomna till en intressant vårkonferens och jag hoppas träffa så många som möjligt av er i Örebro den 23-24 mars.



Läs mer

Mer information finns på www.fastighetsindex.se där man kan hämta den så kallade Värderingshandledningen om man vill se mer hur fastighetsindexet är konstruerat (under fliken "Hur värderar man?"). För det internationella indexet kan man läsa mer på www.ipd.com

Goda nyheter

– FÖR ALLA SOM ÄGER EN KOMMERSIELL FASTIGHET!

"Direktägende i fastigheter har med en totalavkastning på 7,5 procent per år slagit både aktie- och obligationsplaceringar på tio års sikt."

(Svenskt Fastighetsindex. Pressmeddelande 2010-02-25)

Men vänta nu – hur kan man veta det? Att aktiernas utveckling bevakas nogga med index som rapporteras till oss dagligen har väl ingen missat, men finns det alltså ett jämförbart index även för fastigheter? Ja, sedan 1997 finns Svenskt Fastighetsindex (SFI) som ger indexserier årligen vilka skulle kunna jämföras mot andra för investerare intressanta index; det må vara guld, tulpanlökar eller aktier. Fastighetsindexet visar hur mycket totalavkastning som investerade pengar har gett under ett år. Formeln är enkel: totalavkastning = direktavkastning + värdeförändring.

Och vad är direktavkastning? Jo, det är hur mycket pengar som tjänats på en fastighets kassaflöde under perioden. Inkomsterna är i regel samma sak som hyran, kostnaderna är drift, underhåll, räntor med mera. Inkomster minus utgifter blir vinsten, och om vinsten delas med marknadsvärdet så får man direktavkastningen.

Vad är då marknadsvärdet? Enklast uttryckt är marknadsvärdet det troligaste priset för en fastighet om den säljs på en öppen marknad, alltså en köpeskillning.

Och slutligen värdeförändringen? Jo, det är relationen mellan marknadsvärdet vid periodens början och slut. För SFI:s del är perioden ett år.

Men det finns en begränsning för hur SFI kan användas som gör att jämförelsen med ak-

tier haltar något. Den har att göra med vilka fastigheter som ingår i indexet. Många av de stora fastighetsbolagen i Sverige – Hufvudstaden, Faberge med flera – är börsnoterade. Dessa bolag är inte med i SFI:s indexserie, eftersom de hellre vill bli värderade för värdet på sin aktie än för marknadsvärdet av sina fastigheter. Börsen tillhandahåller ett särskilt fastighetsaktieindex för de börsnoterade fastighetsbolagen.

Istället är det de så kallade institutionella fastighetsägarna som mäter sitt värde via SFI. Dessa är t ex försäkringsbolagens fastighetsbolag såsom AMF och Folksam och det stora statligt ägda fastighetsbolaget Vasakronan. Gemensamt för dessa fastighetsbolag är att de inte i första hand har som affärsidé att köpa och sälja fastigheter, och därför finns det inga aktuella köpeskillningar att använda för att hitta värdet på fastigheterna. Istället måste man genomföra en konstruerad marknadsvärdering. SFI ger en modell för en sådan marknadsvärdering.

Vad är då poängen med ett index för värdet på fastigheter om inte alla fastighetsbolag deltar?

Ja, poängen beskrivs ofta som själva processen i sig och inte nödvändigtvis resultatet. Fastighetsbolag som deltar i SFI rapporterar in omfattande data om allt från kostnader för trappstädning till intäkter från parkeringsplatser. Uppgifterna kan användas för jämförelser mellan

bolagen och ger i sin tur, om de används klokt, en effektivare fastighetsförvaltning. Ofta är det genom jämförelser med andra som ett bolag kan få syn på sina egna styrkor och svagheter. Inför större beslut om förändring kan det också vara klokt att fråga sig hur andra gör och vad som ger bäst effekt. Genom att använda SFI som tredje part behöver inte företagen riskera att avslöja affärshemligheter. Data som lämnas ut är väl avkodade och aggregerade till en nivå där ingen enskild aktör kan identifieras.

SFI är en ekonomisk förening, grundad av 12 fastighetsbolag, som har uppdraget att samla in data och beräkna fastighetsindex. SFI har sin motsvarighet i andra länder och kallas internationellt för IPD, Investment Property Databank. Andra länder som använder ett fastighetsindex enligt IPD:s modell är bland annat USA, Japan, Kanada, Australien och Sydafrika. Det är frestande att göra jämförelser mellan länder, men även här får man se upp. I Sverige är det t ex kanske självklart att viss inredning ska ingå på ett kontor, så som köksutrymme. Men i många länder tar man med sig spisen när man flyttar.

RAGNHILD FAHLERYD,
LANTMÄTERIET





BERNHARD
HUITFELDT

Oberoende eller kompetens?

**EN FRÅGA
OM TROVÄRDIGHETEN
FÖR STATISTIKER
I LÄKEMEDELSINDUSTRIN**

I Qvintensen nr 2/2010 skrev jag en artikel med rubriken "Statistiskt arbete är i förtroendebranschen" i anslutning till att Svenska statistikfrämjandets förslag till etisk kod presenterades. Jag följer nu upp den med utgångspunkt i de redaktionella principer som Journal of the American Medical Association tillämpar sedan en tid tillbaka. Dessa ställer bl a ökade krav på att industrisponsrade studier ska analyseras av oberoende statistiker för att accepteras för publicering. Inledningsvis ledde detta till att statistiker i läkemedelsindustrin kände sig utpekade som mindre trovärdiga. Men det har också fört med sig en nyttig diskussion om hur oberoende och kompetens ska kunna förenas när man arbetar i en miljö med risk för intressekonflikter.

»»»»



»»» OBEROENDE ELLER KOMPETENS?

Intressekonflikter hotar den vetenskapliga integriteten, i synnerhet om stora ekonomiska intressen är inblandade. Detta anser en av de ledande amerikanska medicintidskrifterna, Journal of the American Medical Association (JAMA), som vill skydda sig mot att resultat från forskning med dolda intressekonflikter blir publicerade i tidskriften. När det gäller kliniska prövningar av läkemedel kan intressekonflikterna uppstå mellan läkare, som utför den kliniska delen av undersökningen, och dem som beställer och betalar för sådana undersökningar (sponsorer) - ofta läkemedelsföretag. Men även statistiker kan beröras av detta, vilket ska visa sig senare i denna artikel.

Sedan länge kräver JAMA full redovisning av författarna när det gäller eventuella egna ekonomiska intressen i den forskning som de vill publicera. Författare tillåts i och för sig att ha intressekonflikter, men de ska redovisas fullt ut så att läsaren kan göra en egen bedömning om dessa kan påverka artikelns trovärdighet. Uppgifterna om intressekonflikter hos författarna skickas inte med till de vetenskapliga granskarna (peer reviewers), så i princip ska sådana inte påverka bedömningen om en artikel är publicerbar eller inte. Även granskarna måste deklarerat eventuella intressekonflikter i relation till en artikel de fått för granskning. Granskarnas deklARATIONER publiceras inte i tidskriften. Finns intressekonflikter som kan påverka en saklig bedömning måste granskarna avstå från uppdraget. Allt detta är gott och väl och har tillämpats sedan länge i olika varianter av de flesta medicinska tidskrifter.

JAMA skärper sina krav på författare

I en ledare för tio år sedan tar JAMA upp frågan om intressekonflikter i kliniska prövningar till förnyad diskussion, se referens (1) i listan på sidan 11. Den leder till en skärpning av JAMA:s policy avseende kraven på författare. Bland annat tillkommer krav på tydligare redovisning av

sponsorns roll i forskningen när det gäller försöksplan, genomförande av studien, insamling, analys och tolkning av data, samt även vilken roll sponsorn har i arbetet med författandet av artikeln. Anställda hos sponsorn, som med sedvanliga kriterier kvalificerar som författare, skall finnas med i författarlistan, och det inkluderar även statistiker. Man vill helt komma ifrån modellen med spökskrivande där sponsorn lägger fram färdiga texter för de tilltänkta författarna som sedan sätter sina namn på manuskriptet utan att ha satt sig in i det underliggande materialet. Och ofta också får bra betalt för detta.

Därför kräver man i ledaren att minst en av sponsorn oberoende författare, vanligen den läkare som ansvarat för den kliniska delen av undersökningen, ska deklarerat att hon eller han haft full tillgång till alla data, tar ansvar för dataintegriteten och för riktigheten av analysen. Ett ganska tufft krav för

någon som inte är statistiker kan man tycka. I denna första version av den skärpta policyn "uppmuntrar" man till att den statistiska analysen av industrisponsrade studier utförs av en universitetsbaserad oberoende statistiker, och inte bara av statistiker anställda hos sponsorn. Denna policy blev sedermera accepterad som standard hos alla större medicinska tidskrifter som framgår av referens (2).

JAMA skärper sina krav ytterligare

Vi ser nu att strålkastarljuset börjar söka sig åt vårt håll, åt det som har med data och statistiken att göra. Och det är nästa steg i denna utveckling som börjar röra om i grytan ordentligt. För drygt fem år sedan skärper nämligen JAMA kraven på tydlig redovisning ytterligare, se referens (3).

Det som väcker diskussion bland våra kollegor är kravet på att en studie sponsrad av ett läkemedelsföretag, som enbart är analyserad av en statistiker anställd av sponsorn, inte kommer att accepteras för publicering i JAMA. För sådana studier måste en ytterligare analys utföras av en oberoende statistiker knuten till en akademisk institution. Denne ska ha tillgång

till alla rådata, protokoll och statistiska analysplaner och självständigt genomföra en statistisk analys. Det framgår också att det är den analysen som ska redovisas i det publicerade manuskriptet.

Av texten kan man få intrycket av att det mest handlar om en verifiering av den gjorda analysen, men om det ska vara någon mening med denna extra analys så måste det förstås uppkomma skillnader ibland. Och då gäller som sagt den oberoende statistikerns analys.

Även här kan det uppstå en intressekonflikt beroende på hur den oberoende statistikern finansieras. Därför måste det också framgå i artikeln vem som betalt för detta arbete. Men det som ger den extra säkerheten för JAMA är den akademiska tillsynen över studien som en universitetsbaserad statistiker kan erbjuda, även om hon eller han får betalt av sponsorn.

Hur blev reaktionen bland statistiker?

Detta träffar en mycket öm tå, och inte bara bland statistiker i läkemedelsindustrin. Troligen fanns det en rädsla för att liknande principer skall sprida sig till andra medicinska tidskrifter, och kanske i slutändan även till andra tillämpningsområden för statistiker. Yrkeskåren känner sig kränkt och hotad och den första instinkten var just att tala om det.

Flera statistiska organisationer skriver nu öppna brev till JAMA, varav två blir publicerade i JAMA 2005, de från Pharmaceutical Research and Manufacturers of America (PhRMA) och International Biometric Society (IBS). Royal Statistical Society (RSS) och European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) skickar var för sig också in brev till JAMA, men dessa blir aldrig accepterade för publicering. Några inlägg publiceras på annat håll, t.ex. från ordföranden i American Statistical Association (ASA) Fritz Scheuren i AmStat News, från Frank Rockhold, chefstatistiker på GlaxoSmithKline, i tidskriften Pharmaceutical Statistics och en ledare i British Medical Journal (BMJ) skriven av Kenneth Rothman från RTI Health Solutions och Steven Evans från London School of Hygiene and Tropical Medicine

(4). Dessa kommentarer kommer alltså från många olika kategorier av statistiker, statistiska organisationer i USA och Europa, statistiker i läkemedelsindustrin och från sådana engagerade i andra medicinska tidskrifter.

Det finns många gemensamma nämnare i dessa rektioner. Den starkaste reaktionen gäller det asymmetriska förhållningssättet till studier sponsrade av läkemedelsindustrin och till annan medicinsk forskning som JAMA:s policy ger uttryck för. Skribenterna vänder sig emot att policyn indirekt pekar ut statistiker i läkemedelsindustrin som mindre trovärdiga när det gäller att sköta sitt arbete med professionell integritet. Man påpekar att det finns en mycket stark statistisk kompetens inom läkemedelsindustrin som har varit pådrivande när det gäller att utveckla goda processer och sund metodik vid kliniska prövningar.

En vanlig kommentar är också att det finns få andra industrier som är så noggrant kontrollerade av myndigheter, etiska kommittéer och oberoende övervakningsgrupper för kliniska prövningar som läkemedelsindustrin. Datainsamling, felsökning i data och inmatning av data är rigoröst övervakade av interna eller externa specialister. Databaserna som skapas för kliniska prövningar är hårt bevakade och alla ändringar registreras efter att dessa har blivit lästa. Det finns sällan någon motsvarande kontrollerad datamiljö vid traditionell akademisk medicinsk forskning. Detta är allmänt omvittnat, inte minst från personer som kommer från akademisk forskning till läkemedelsindustrin.

Skribenterna hävdar med bestämdhet att vetenskapliga oegentligheter förekommer i minst lika stor omfattning även i annan medicinsk forskning. Det kan finnas starka personliga och ekonomiska intressen i resultatet från sådan forskning bland akademiska forskare. Detta motiverar att bedömningen av medicinsk forskning för publicering ska utgå från innehållet i manuskripten och inte varifrån forskningen har sponsrats, menar man. Ett aktuellt exempel är avslöjandet av fusket i det påstådda sambandet mellan trippelvaccinet och uppkomsten av autism. British Medical Journal anser sig nu ha bevis för att en akademisk forskare har mani-



Vetenskapligt grundat, rigoröst kontrollerat.

pulerat data i sina studier för att kunna påvisa ett sådant samband. Han skall ha fått bra betalt för detta av en advokatfirma som driver skadeståndsmål för familjer med autistiska barn som har fått trippelvaccinet.

Skribenterna påpekar att ansvaret för kvaliteten hos publicerade artiklar måste ligga på tidskriften. Det vill säga, om någon tidskrift vill tillämpa särskilda kvalitetskrav får man säkerställa dessa i sin egen granskningsprocess. För att upprätthålla den statistiska kvaliteten kan det innebära att tidskrifterna i större utsträckning bör involvera statistiker i granskningen. I det sammanhanget betonar man att författare och sponsorer måste vara helt öppna med att tillhandahålla allt vad som behövs för detta, till exempel den statistiska analysplanen och data som ligger till grund för de primära analyserna.

Några av inläggen påpekar också svårigheten i att definiera vad som är en universitetsbaserad oberoende statistiker, vilken kompetens som måste krävas av en sådan och hur den eftersträvade statusen som oberoende kan uppnås i praktiken. Och om JAMA:s policy blir allmän regel bland medicinska tidskrifter uppstår ett oöverstigit resurs- och kompetensproblem.

Hur reagerade JAMA på statistikernas kritik?

I ett direkt svar på breven från PhRMA och

»Policyn pekar indirekt ut statistiker i läkemedelsindustrin som mindre trovärdiga när det gäller att sköta sitt arbete med professionell integritet.«

IBS försvarade JAMA sin policy genom att ge ett antal exempel där man menar att läkemedelsföretag undanhållit viktig information om allvarliga biverkningar i publicerade artiklar om sina kliniska prövningar. Dessutom antyder man att statistiker i läkemedelsindustrin kan bli pressade att utesluta vissa data eller inte fullständigt rapportera vissa resultat från sina analyser. Man antyder vidare att anställda statistiker i läkemedelsföretag kan ha individuella ekonomiska motiv som på ett subtilt sätt kan påverka deras analyser och rapporteringen av dessa. Man vidhåller att man inte kommer att ändra sin policy med anledning av dessa kommentarer.

Speciellt illa berörda blev JAMA:s chefredaktörer DeAngelis och Fontanarosa av den ledare av Rothman och Evans (4) som publicerades av konkurrenten British Medical Journal. Detta ledde till en skarp ledare i JAMA (5), där de ger nya exempel på brister i publicerade kliniska prövningar. De hänger också upp sig på vissa ordval, men framför allt tycker de att ett påhopp av inte tillräckligt pålästa kollegor är förvånande.

Vad har hänt sedan 2006?

Efter den inledande samlade reaktionen från statistiker hände inte så mycket. Inga andra medicinska tidskrifter har hoppat på tåget och många har uppfattningen att detta kommer att själv dö. Och det finns rykten att läkemedelsfö-

»»»

»»» OBEROENDE ELLER KOMPETENS?

tag undviker att försöka få sina studier publicerade i JAMA. Detta tycker jag också bekräftas av en granskning av ett antal nummer av JAMA där man numera hittar mycket få industrisponsrade läkemedelsstudier.

Nej, det är mer på det principiella planet som diskussionen fortsätter. Frågan togs upp på International Statistical Institutes stora konferens i Lissabon 2007 i en session som Niels Keiding ledde, den dåvarande presidenten för ISI. Han hänvisar till den diskussionen i (6) och sätter fingret på just det som jag tror har stört oss statistiker mest med en överprövning av slutanalysen av en oberoende statistiker. Han skriver (i min översättning):

”Själva slutanalysen av en låst databas är bara en liten del av den totala statistiska analysen. Det förlöjliga för vår profession att tro att det är en seriös kontroll som utförs av reproducerbarheten om en annans dator kommer fram till samma slutsatser genom att använda samma programkod och samma databas som den ursprungliga statistikern. Vad en statistiker som sitter i ett elfenbenstorn kan tro är en oantastlig databas är i själva verket resultatet av otaliga konkreta bedömningar som görs under studiens gång, främst av vetenskapliga specialister på området, men oftast med insiktsfulla bidrag från den som är ansvarig statistiker för studien. Det är där som en viktig del av statistikerns ansvar ligger och den processen kan inte reproduceras i den policy som JAMA tillämpar. Att dra några slutsatser från en återanalys utan kunskap om denna förhistoria är inte professionellt och trivialiserar biostatistiken.”

DeAngelis och Fontanarosa (som alltså var JAMA:s chefredaktörer under hela den aktuella perioden) besvarar denna artikel i samma nummer av Biostatistics genom att upprepa samma argument som tidigare, dock utan att kommentera Keidings kärnfråga, och i övrigt känna sig allmänt missförstådda.

Nya oegentligheter får JAMA att reagera ännu skarpare

I en ledare förra året blev frågan aktuell igen (7) med anledning av vad som utspelade sig kring diabetesläkemedlet Avandia. Detta läkemedel

var en riktig storsäljare för Glaxo-SmithKline över hela världen. Efter ett tag gjordes en stor meta-analys av tillgängliga studier, där det påstods att Avandia bl a ökar risken för hjärtinfarkt (8). Efter spelet kring denna publikation blir en riktig rysare med stora ekonomiska, vetenskapliga och mänskliga värden på spel. Företaget anklagas för att undanhålla data och för att inte ha meddelat myndigheter och allmänheten om denna allvarliga biverkning långt tidigare. USA:s kongress tillsatte en senatskommitté för att utreda företagets hantering av ärendet. Detta innebar bl a att vartenda mötesprotokoll, brev och e-brev mellan inblandade parter krävdes in för att ta reda på om och när man började misstänka en ökad risk för hjärtinfarkt. Enligt senatsrapporten kände företaget sannolikt till biverkningarna långt innan de informerade om dem. Slutligen drar den europeiska myndigheten European Medicines Agency (EMA) in Avandia från marknaden. Den amerikanska motsvarigheten FDA ger företaget mycket strikta begränsningar för Avandias användning, bl a att inga andra läkemedel ska ha fungerat tillfredställande för att läkemedlet ska få skrivas ut.

I detta fall riktar man primärt uppmärksamheten på läkarna i företaget och på företagsledningen, men återigen är det statistikerna som tas som gisslan och som anses möjliggöra dessa manipulationer. När det gäller kraven på statistiken i industrisponsrade studier höjer JAMA nu tonläget ytterligare något. JAMA inkluderar kommersiella forskningsorganisationer bland dem som är diskvalificerade att utföra primära analyser och likställer dem med sponsorföretaget. Dessutom vill man att den oberoende statistiska analysen ska planeras i förväg och specificeras i studieprotokollet, dvs långt innan sponsorföretaget har beslutat om och var man ska försöka få resultatet publicerat. Man uppmanar också i denna ledare alla andra medicinska tidskrifter att tillämpa samma policy som JAMA.

Förebygg misstroendet istället för att bli kränkt

Det stegrande misstroendet gentemot statistiker i läkemedelsindustrin, som kommit till uttryck i JAMA:s redaktionella policy under de senaste tio åren, kan ses som ett orätfärdigt påhopp på

»Man antyder vidare att anställda statistiker i läkemedelsföretag kan ha individuella ekonomiska motiv som på ett subtilt sätt kan påverka deras analyser och rapportering av dessa.»

en stor grupp av yrkesverksamma statistiker. Bara i Europa finns ca 2 000 statistiker i läkemedelsföretag. Och de inledande reaktionerna speglar också en sådan uppfattning.

Men med lite eftertanke kan man se det som ett erkännande av den stora betydelse statistiken har i medicinsk forskning och

att detta är ett gyllene tillfälle för oss att stödja en utveckling som förebygger att ett misstroende uppstår. Och detta oberoende av vem som är vår uppdragsgivare. Pharmaceutical Statistics, den officiella tidskriften för Statisticians in the Pharmaceutical Industry (PSI) i Storbritannien har tagit fasta på denna inställning och publicerade förra året tre artiklar med olika inriktning.

I den första artikeln (9) betonar författarna den s k studiestatistikerns ansvar vid publiceringen. Oftast är det denne som har den mest detaljerade kunskapen om data i studien och har därför ett särskilt ansvar för att studieresultaten blir fullständigt och korrekt rapporterade och ges en välvägd statistisk tolkning. Studiestatistikern ska också se till att alla författare har tillgång till alla relevanta data och analyser så att dessa kan bilda sig en egen uppfattning om hur resultaten ska tolkas. Dessutom rekommenderas att studiestatistikern är den som är ansvarig för det statistiska innehållet i publikationen och att denne ska vara medförfattare. Idag publiceras synopsis av alla studieprotokoll i förväg på dedicerade webbplatser för detta och studiestatistikern ska se till att alla viktiga statistiska aspekter av studieprotokollet ingår i detta dokument. Denne ska också medverka till att studieresultaten blir publicerade så att de på ett eller annat sätt kan bli allmänt kända.

Den andra artikeln tar sin utgångspunkt i JAMA:s krav på oberoende analys av industrisponsrade studier (10). Författarna vill inte gå i svaromål mot policyn utan vill i stället granska effektiviteten i nuvarande praxis som syftar till att minimera bias i design, analys och rapportering av kliniska prövningar från industrin. Dessutom vill de peka ut områden där statistikern hos sponsorn faktiskt kan göra mer. Alltså helt i andan att förebygga snarare än att känna sig kränkt. De medger att det kan finnas risk

»Efter ett tag gjordes en stor metaanalys av tillgängliga studier, där det påstods att Avandia ökade risken för hjärtinfarkt. Efterspelet kring denna publikation blev en riktig rysare med stora ekonomiska, vetenskapliga och mänskliga värden på spel.»

för ”positiv” bias i industrisponsrade studier beroende på de stora pengar som är involverade vid ett positivt eller negativt resultat för det egna läkemedlet. Det kan röra sig om valet av hypoteser, valet av studiedesign, genomförande av analysen och tolkningen av resultatet, och slutligen valet av vilka resultat som publiceras. Man föreslår tre åtgärder för att minimera dessa risker:

(I) Registrera alla kliniska studier på www.ClinicalTrials.gov eller på någon annan öppen webbplats innan studien startar. Detta är numera lag i USA. Publicera huvudresultaten på en öppen webbplats, t ex www.ClinicalStudyResults.org senast inom ett år efter det att studien avslutats. De flesta större läkemedelsföretag följer redan nu dessa principer, och flera har egna öppna webbplatser där dessa uppgifter finns tillgängliga.

(II) Lämna in studieprotokollet för granskning innan studien startar till den tidskrift man hoppas få resultatet publicerat i. Den engelska tidskriften *Lancet* tillämpar en sådan modell och de personer som granskat protokollet kommer också att granska manuskriptet.

(III) Statistikerna måste engagera sig mer i publiceringen av kliniska studier och ta det fulla ansvaret som medförfattare.

Den tredje artikeln (11) utvecklar frågan om större öppenhet från läkemedelsföretagen när det gäller kliniska prövningar. Författarna konstaterar att det under de sista tio åren skett en markant förändring i och med att information om alla studier finns förr eller senare öppet tillgängliga för allmänheten vare sig de är positiva eller negativa för företaget.

Denna öppenhet gäller tills vidare bara enskilda studier. Men det förekommer många

andra typer av bearbetningar av data från kliniska studier som ligger till grund för olika ställningstagande både från läkemedelsföretag och från myndigheter, främst FDA i USA. Det kan vara metaanalyser av flera studier för att belysa speciella frågeställningar, olika specialanalyser av de samlade data som finns om läkemedlets biverkningar eller olika sammanställningar från hela databasen för att stödja marknadsföringen. Här sätts statistikern många gånger på hårda prov när det gäller att upprätthålla sin professionella integritet.

Tankar om framtiden

Det finns stora vetenskapliga fördelar med att läkemedelsföretagen har fast anställda statistiker. Som sådan blir man involverad i hela läkemedelsprojektet, från de prekliniska försöken, de tidiga studierna på människa och slutligen till de stora bekräftande studierna som ska ge de avgörande svaren på läkemedlets nytta och risker. Och i vissa fall även i de aktiviteter som äger rum i samband med myndighetsgranskningen och när produkten väl är registrerad. I enskilda kliniska studier blir man involverad i hela kedjan från planeringen, datainsamlingen, datakontrollen, analysen till rapporteringen och publiceringen. I samspelet med många olika specialister såsom läkare, apotekare, hälsoekonomer, epidemiologer och dataprogrammerare får man insyn i helheten. Statistikern får därigenom en möjlighet att bidra med sin kompetens under hela resans gång, såväl i läkemedelsprojekt som i enskilda studier. Detta ger statistikerna i läkemedelsindustrin en unik kompetens som en statistiker som enbart utför en avgränsad uppgift, såsom att utföra en ”oberoende” analys av en klinisk prövning inför en publicering, har mycket svårt att mäta sig med.

Den fråga som ställdes i rubriken om ”oberoende eller kompetens” har inget enkelt svar. Jag påstår ändå att det oberoende som JAMA eftersträvar sker på bekostnad av kompetens. Men för att företagen i framtiden ska kunna dra nytta av egna anställda statistiker, och statistikerna av den unika möjlighet som ett läkemedelsföretag ger till ett spännande arbete, måste statistikerna kunna stå för både oberoende och kompetens. Oberoende i betydelsen att alltid sträva efter sanningen, inte låta sig otillbörligt påverkas, stå för etiska värden såsom de uttrycks i Svenska statistikfrämjandets etiska kod och hävda sin professionella integritet. I det långa loppet kommer det att vara till gagn även för företaget och därmed uppfyller statistikern det rimliga kravet att vara lojal mot sin uppdragsgivare.

När det gäller kompetensen räcker det inte med bara teknisk kompetens. Statistiker i läkemedelsindustrin måste vara trygga i sin roll, vänja sig vid att ta ett större ansvar, och utöva ledarskap för att tydligare kunna ta detta större ansvar.

Något som skulle kunna stärka trovärdigheten hos statistiker i läkemedelsindustrin och på andra ställen är en auktorisering av statistikeryrket. Tillsammans med en aktiv tillämpning av den etiska koden skulle det ge en ökad tyngd till vår verksamhet men även ställa ökade krav på utövarnas, ja just det, ”oberoende och kompetens”.

BERNHARD HUITFELDT

Referenser

- 1 DeAngelis CD, Fontanarosa PB, Flanagin A (2001): Reporting financial conflicts of interest and relationships between investigators and research sponsors. *JAMA*, Vol 286, 89-91.
- 2 Editorial (2001): Sponsorship, authorship, and accountability. *Annals of Internal Medicine*, Vol 135, No 6, 463-466.
- 3 DeAngelis CD, Fontanarosa PB, Flanagin A (2005): Reporting conflicts of interest, financial aspects of research, and the role of sponsors in funded studies. *JAMA*, Vol 294, 110-111.
- 4 Rothman, KJ, Evans, S (2005): Extra scrutiny for industry funded trials. *BMJ*, Vol 331, 1350.
- 5 Fontanarosa PB, DeAngelis CD (2006): JAMA's policy on industry sponsored studies. *BMJ*, Vol 332, 177.
- 6 Keiding N (2010): Reproducible research and the substantive context. *Biostatistics*, Vol 11, No 3, 376-378.
- 7 DeAngelis CD, Fontanarosa PB (2010): Ensuring integrity in industry-sponsored research – primum non nocere revisited. *JAMA*, Vol 303, 1196-1198.
- 8 Nissen SE, Wolski K (2007): Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular disease. *New England Journal of Medicine*, Vol 356, 2457-2471.
- 9 Matcham J, et al. (2010): Proposed best practice for statisticians in the reporting and publication of pharmaceutical industry-sponsored clinical trials. <http://onlinelibrary.wiley.com/doi/10.1002/pst.417/abstract>
- 10 Pyke S, et al. (2010): The potential for bias in reporting of industry-sponsored clinical trials. <http://onlinelibrary.wiley.com/doi/10.1002/pst.429/abstract>
- 11 O'Kelly M, et al. (2010): Making available information from studies sponsored by the pharmaceutical industry: some current practices. <http://onlinelibrary.wiley.com/doi/10.1002/pst.430/abstract>



KRÖNIKA/ *Olle Häggström*

Global uppvärmning och statistisk signifikans

Vintern 09/10 var kall här i Sverige, men klimatdebatten hetare än någonsin tidigare. Köpenhamns-mötet hade misslyckats, och det surrades intensivt kring det som kom att kallas Climategate, dvs det påstått skandalösa innehållet i den stora uppsättning e-brev mellan klimatforskare som kommit till allmän kännedom efter ett datorintrång vid Climate Research Unit (CRU) på University of East Anglia i Storbritannien. Climategate kom att bli en stor framgång för dem som önskar sprida förvirring i klimatfrågan¹. Flera av de inblandade klimatforskarna blev mycket hårt ansatta, och trots att den ena utredningen efter den andra slagit fast att materialet inte ger grund för misstanke om forskningsfusk, så finns en kvardröjande känsla hos allmänheten att forskarna skulle ha manipulerat data för att överdriva uppvärmningen.

Hårdast ansatt av alla var CRU:s chef Phil Jones. Och döm om min häpnad då vi måndagen den 15 februari 2010 i Dagens Eko fick höra följande.

"Den kritiska granskningen av FN:s klimatpanel IPCC och dess arbete rullar på och nu är det en av huvudpersonerna som i en intervju i brittiska BBC säger att han inte längre är lika säker i frågan om den globala uppvärmningen. Det är självaste chefen för det brittiska forskningsinstitutet vid University of East Anglia, professor Phil Jones, som säger att det inte finns några statistiskt säkerställda bevis för en global uppvärmning de senaste femton åren."

Påståendet att Phil Jones skulle ha svängt i frågan om den globala uppvärmningen

är mycket märkligt med tanke på vad han faktiskt säger i den BBC-intervju som Dagens Eko hänvisade till:

Q: How confident are you that warming has taken place and that humans are mainly responsible?

A: I'm 100% confident that the climate has warmed. As to the second question, I would go along with IPCC Chapter 9 - there's evidence that most of the warming since the 1950s is due to human activity.

Inslaget i Dagens Eko kom mycket riktigt att fällas av Radionämnden. Varifrån fick då Dagens Eko sin idé om att Jones skulle ha svängt? Svaret på det finns i en annan del av intervjun.

Q: Do you agree that from 1995 to the present there has been no statistically significant global warming?

A: Yes, but only just. I also calculated the trend for the period 1995 to 2009. This trend (0.12C per decade) is positive, but not significant at the 95% significance level. [...] Achieving statistical significance in scientific terms is much more likely for longer periods, and much less likely for shorter periods.

Jones syftar här på den så kallade Hadley-CRU-serien, en av de gängse tidsserierna över global temperaturutveckling². Denna uppvisar stora fluktuationer, till följd av vädrets kaotiska växlingar, ovanpå den trend som vi kallar global uppvärmning. Quintensens läsare är givetvis klara över att det blir svårare att detektera en trend i en brusig tidsserie ju kortare tidsintervall man beaktar, och att trenden på tillräckligt korta tidsskalor drunknar i bruset

»Detta är ett gott skäl till varför matematisk statistik behöver flytta fram sina positioner inom högre utbildning.»

(i det här fallet var frågan dessutom riggad, då året 1995 var valt så tidigt det gick utan att få 95%-ig signifikans). Likaledes klara är såklart Qvintensens läsare över att ett p-värde om 0,07 (som i detta fall) inte är något mått på hur sannolik nollhypotesen är (i detta fall att trenden är noll). Detta dels då vi har lärt oss att inte begå the fallacy of the transposed conditional (att förväxla sannolikheten för data givet nollhypotesen med sannolikheten för nollhypotesen givet data), men också för att en bedömning av hur trolig nollhypotesen är behöver baseras på allt vi känner till i frågan (inklusive vad vi vet om växthuseffekten och annan bakomliggande fysik) och inte bara på (en del av) en enskild dataserie³.

Allt detta har sagts många gånger tidigare, och visar på det föga centrala i bollandet med p-värden i samband med de senaste årens globala uppvärmning. Vad som emellertid inte framförts lika ofta är ett annat problem med denna p-värdesexercis, nämligen de bakomliggande antaganden som p-värdesberäkningen förutsätter. När klimatforskare och andra uppskattar trenden i ett fall som detta använder de som regel linjär regression i något av de vanliga statistiska programpaketen. Den nollhypotes som ligger till grund för p-värdesberäkningen säger att trenden är noll, men den innebär också en del andra (sällan uttalade) antaganden. Mest bekymmersamt av dessa är att tidsserien antas sakna autokorrelation, dvs vi antar att bruset vid olika tidpunkter är okorrelerat. Detta är helt klart inte sant, då vi vet att väderleken uppvisar korrelationer

över olika tidsskalor. Hur grovt felaktigt antagandet är beror på tidsdiskretiseringen: ju kortare tidssteg desto mer orealistiskt är antagandet. Men även om tidsdiskretiseringen är så grov att man bara tittar på årliga data (vilket ofta är fallet) finns korrelationer som knappast är försumbara, orsakade såväl av klimatsystemets inneboende kaoticitet (t.ex. El Nino-oscillationen) som av externa faktorer (t.ex. solfläckscykler).

För att få fram ett seriöst p-värde för existensen av positiv trend behövs en nollhypotes som någorlunda realistiskt fångar tidsseriens autokorrelation. Detta kräver i sin tur att vi har en bra uppskattning av autokorrelationsfunktionen, vilket är ett intrikat problem där det inte räcker med att kika på den aktuella tidsserien. Exempelvis kan vi ju aldrig i den enskilda tidsserien detektera autokorrelationer på längre tidsskalor än tidsseriens längd. En aldrig så tydlig uppgående trend i tidsserien skulle ju mycket väl kunna vara blott en fluktuation till följd av korrelationer på längre tidsskalor. Och sådana korrelationer finns ju faktiskt – se t.ex. den senaste årmiljonens kraftiga växlingar mellan istider och perioderna däremellan. Detta väcker i sin tur frågan om begreppet trend kanske borde definieras relativt någon given tidsskala. Här räcker det knappast med statistiska färdigheter – vad som också måste till är kunskaper om klimatsystemets mekanismer. Som vanligt behöver vi bara skrapa litet på problemets yta för att inse att det är långt mer komplicerat än det först verkar.

En beräkning av p-värden som inte seriöst beaktar autokorrelationen framstår för mig mer som en ritual än som något av verkligt vetenskapligt värde. Vad jag vill ha sagt är inte att klimatforskarna skulle vara omedvetna om autokorrelationer i naturliga fluktuationer – tvärtom finns en hel del litteratur på området. Däremot verkar det vara lite för vanligt, i klimatvetenskapen liksom i snart sagt alla empiriska ämnen, att forskare genomför signifikanstest och p-värdesberäkningar utan att göra tillräckligt klart för sig vad dessa egentligen betyder. Jag tror att för den forskare som saknar gedigna grundkunskaper i statistik är det lätt att förbise vilka modellantaganden som ligger till grund för beräkningarna och som kan ha betydelse. Detta är ett gott skäl (bland många) till varför matematisk statistik behöver flytta fram sina positioner inom högre utbildning.

Fotnoter

[1] Varför vill de skapa förvirring? Som svar på denna fråga är det svårt att erbjuda mer än spekulationer. Se emellertid min text "Om vetenskapens dimridåmakare" i Axess 6/2010, <http://www.math.chalmers.se/~olleh/OreskesConway.pdf>

[2] Den intresserade läsaren rekommenderas uppsöka webbsajten Wood for Trees (<http://www.woodfortrees.org/>) där alla de vanliga globala temperaturserierna finns, jämte enkla verktyg för att experimentera med exempelvis regressionslinjer.

[3] För en förtjänstfull diskussion kring relaterat (miss-) bruk av signifikanstest, se Sven Ove Hanssons artikel Signifikansmissbruket i pseudovetenskap, Qvintensen 3/2010

Hur man formulerar och ställer frågor är viktigt i många statistiska undersökningar. Det är genom frågorna som vi kommunicerar till respondenter vad det är vi vill veta, vilken typ av information vi vill att de ska lämna. Illa utformade frågor leder till dålig datakvalitet i form av mätfel eller bortfall. Därför är det viktigt att utvärdera frågorna innan de används.

Att utvärdera frågor i en statistisk undersökning

Det finns många olika metoder och tekniker att utvärdera frågor med. De skiljer sig åt i olika dimensioner, t.ex. genom att ha olika syften. Ett syfte kan vara att snabbt få en översikt över eventuella problem. Ett annat syfte kan vara att få ett precist statistiskt värde på t.ex. skillnader i svar mellan två olika frågeformuleringar. Metoderna skiljer sig åt även vad gäller ambition och resurser, vissa är betydligt dyrare och tar mer tid än andra. Av dessa skäl är olika metoder naturligt knutna till olika faser i genomförandet av en statistisk undersökning. Vissa metoder används främst före datainsamling, så kallade för-test. Andra metoder används senare i processen och använder data från insamlingen för att undersöka mätinstrumentet. Vi går igenom ett antal metoder som har använts hos oss på SCB.

Tidigt i utvecklingen av enkäten

Fokusgrupper (strukturerad gruppdiskussion, se Morgan 1993 för ytterligare beskrivning) och djupintervjuer (kvalitativ intervju, läs mer i Kvale 1997) är kvalitativa explorativa metoder. I surveysammanhang förekommer de framförallt tidigt i processen när många strategiskt viktiga beslut tas. Syftet kan vara att identifiera känsliga ämnen, upptäcka hur människor förstår termer, lära sig från respondenter vilka aspekter av ett ämne som är relevanta eller för att avgöra vilken information som är rimlig att fråga efter.

Fokusgrupper och djupintervjuer är vanliga metoder när en undersöknings koncept ska definieras och avgränsas. Om man t.ex. ska genomföra en enkät om självupplevd diskriminering är det viktigt att veta hur begreppet självupplevd diskriminering förhåller sig till

angränsande begrepp som trakasserier, mobbing, stereotyper och kränkning. Hur upplever människor diskriminering? I vilka sammanhang och av vilka orsaker? Hur tänker och talar de om diskriminering? Utan den här typen av kunskap är det svårt att gå vidare och formulera enkätfrågor. Därför kan dessa metoder fylla en viktig funktion tidigt i utvecklingen av enkäten, speciellt om man ska undersöka ett relativt nytt område där etablerade teorier och mätinstrument saknas.

Under datainsamlingen

Under datainsamlingen genereras data som kan användas för att utvärdera hur frågorna fungerat. Om undersökningen genomförts som telefonintervju kan intervjuer medlyssnas (Biemer & Lyberg 2003). Medlyssning innebär att en tredje part lyssnar på intervjuerna i fält. Det kan ske i realtid men oftast sker medlyssningen på inspelat material. I samband med medlyssningen genomförs ofta beteendekodning (Fowler 1995) – intervjupersonens och intervjuarens beteenden under intervjun kodas enligt ett schema.

Beteendekodning bygger på idén att man utifrån intervjuarens och intervjupersonens verbala och kroppsliga uttryck kan få insikt i interaktionen dem emellan, hur intervjupersonen uppfattar frågorna och hur frågorna fungerar i praktiken. Metoden är ett strukturerat sätt att få insyn i vad som händer under intervjun. För att kodningen ska fungera väl ska man ha ett väl utprövat schema som är anpassat till de mål man har, tränade kodare och inspelade intervjuer. Oftast sker kodningen på frågenivå. Exempel på koder som rör intervjuarens beteende kan vara 1) läste precis enligt ordalydelsen, 2) mindre avvikelser och 3) större avvikelser. Exempel på koder som rör uppgiftslämnarens beteende kan vara 1) ber om förtydliganden och 2) ber inte

om förtydliganden. Resultaten kan t.ex. visa hur stor proportion av intervjuarna som läste frågan precis enligt ordalydelsen. Genom närmare studier kan man även upptäcka samband, t.ex. att avbrott i intervjun ofta sker när frågan följs av många exempel.

Fördelarna med beteendekodning är att det är förhållandevis billigt och enkelt att implementera i ett fälttest. Utvärderingarna sker på material insamlat under naturliga förutsättningar, en "riktig" intervju. Resultatet är kvantitativt, vilket kan vara enklare än kvalitativa resultat att förmedla vidare till andra intressenter. Nackdelar är att det kan vara svårt att tolka orsakerna till resultatet och veta hur man ska lösa eventuella problem. En annan nackdel är att beteendekodningen registrerar avvikelser från en optimal sekvens av fråga och svar, men det kan vara svårt att veta hur den avvikelserna påverkar svars-kvaliteten. Vissa typer av problem, som är mer implicita till sin natur (t.ex. att respondenten missförstått frågan utan att ge uttryck för det), framkommer inte med beteendekodning.

Ett exempel på beteendekodning tar vi från Undersökningarna av levnadsförhållanden (ULF) genom frågan "Har du i bostaden tillgång

»Frågan "Hur många barn har du?" kan tyckas vara enkel och självklar..»

till dator?". Beteendekodning av intervjuer visade att vissa intervjuare ställde frågan "Har du dator?" vilket är en läsavvikelse. Denna avvikelse innebär tre omskrivningar av betydelse. Det är skillnad på "tillgång" och "har" då "har" i högre grad anspelar på att äga. Det är även skillnad

på "du" och "bostaden" då det i bostaden kan finnas tillgång till sammanboendes dator. Frånvaron av "i bostaden" i "Har du dator?" innebär att t.ex. tillgång till dator på jobbet kan räknas. Förändringar i frågans betydelse utgör en risk att vi inte mäter det vi avser att mäta.

Om undersökningen genomförs med datorstödda blanketter (t.ex. datorstödd telefonin-

»Om inte sådana oklarheter definieras i frågan kommer uppgiftslämnarna att göra det på egen hand, medvetet eller omedvetet, men troligtvis på lite olika sätt.»

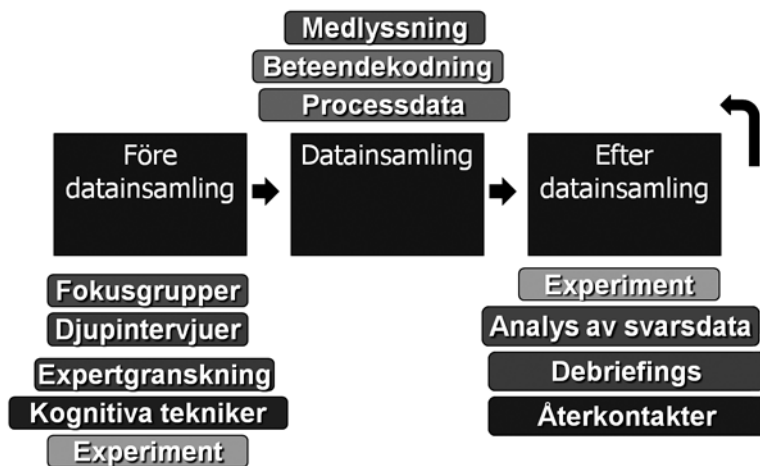
tervju eller webbenkät) kan data som beskriver datainsamlingsprocessen samlas in, i många fall automatiskt. Exempel på data som kan vara intressant ur mättekniskt perspektiv är tidsåtgång för att besvara enskilda frågor, var i blanketten uppgiftslämnarna avbryter sitt deltagande eller noteringar som intervjuaren gör under intervjun. Den typen av data ger viktig och billig feedback på mätinstrumentet och kan därför, framförallt vid återkommande undersökningar, utgöra en värdefull informationskälla inför andra frågetest.

Inbäddade experiment

I datainsamlingen kan man även "bädda in" experiment i en pågående undersökning (Tourangeau 2004). Till exempel kan en slumpmässigt vald mindre del av urvalet få en annan frågeformulering än övriga i urvalet. Experiment i allmänhet innebär att man med riktade insatser undersöker hur någonting påverkar någonting annat. Olika grupper utsätts för olika behandlingar, t.ex. olika frågeformuleringar, samtidigt som andra förhållanden hålls lika. Grupperna jämförs sedan på ett utfallsmått. Om grupperna var i princip lika innan exponeringen av de olika behandlingarna men skiljer sig signifikant åt efteråt är den troligaste förklaringen att skillnaden i behandling orsakat skillnaden i utfallet. I surveysammanhang kan t.ex. frågornas tidsåtgång, svarsandel/svarsmönster eller validitet utvärderas experimentellt. Ett exempel på en annan sorts behandling är att skicka uppgiftslämnaren enbart pappersblankett eller hänvisa henne eller honom enbart till en elektronisk blankett på Internet eller skicka både pappersblankett och en hänvisning i samma brev. Ett sådant experiment beskrevs i Qvintensen nr 4/09 i artikeln "Vad är nästa steg för att öka tillförlitligheten i webbaserade undersökningar?". Syftet var att se om det går att öka andelen som använder webblanketterna



Att utvärdera frågor



och därmed minska det totala bortfallet och kostnaden för datainsamlingen men utan att förlora något i undersökningens kvalitet. Det numret av Qvintensen innehöll också korta beskrivningar av experiment med webblanketers utseende ("Internationell kurs om mätfel").

Fördelar med experiment är bland annat att informationen är kvantitativ med få och relativt väl kontrollerade felkällor. En nackdel är att svaren från en del av urvalet kan bli svåra att använda om behandlingarna ger överraskande stora skillnader. Den risken kan minskas med pilotstudier eller mindre experiment men då tar hela utredningen längre tid.

Två varianter av experimentell design är

inom-individdesign (att samma individer deltar vid två olika mätningar som sedan jämförs) och mellan-individdesign (t.ex. att mätningar av två grupper av individer jämförs). Mellan-individdesignen har fördelen att den minimerar uppgiftslämnarbördan. Inom-individdesignen har fördelen att det är lättare att urskilja effekter, men nackdelen att inlärningseffekter kan förekomma (carry-over effects), det vill säga att svaren vid de två tillfällena inte är oberoende av varandra (Field & Hole 2003). Det finns fall då man vill studera just inlärningseffekterna och då är inom-individdesignen förstås lämplig.

Svarsdata kan även analyseras för att utvärdera mätinstrumentet, oavsett experimentell

Vad är probing?

■ **I artikeln** "Att utvärdera frågor i en statistisk undersökning" förekommer det engelska ordet probe. Qvintensen frågade Terminologiscentrum TNC vad vi ska kalla "probing" och "probes" på svenska. Terminologiscentrum TNC svarade följande:

Det naturliga i ett sådant här fall är att försvenska "probe" till "prob". Det böjs då förstås på svenskt vis: en prob, den där proben, flera prober, de där proberna. Försvenskningen prob finns redan inom många andra fackområden,

t.ex. elektroniken och medicinen. Även om termen där oftast står för något mer fysiskt finns det en gemensam begreppskärna ("något som tränger djupare och djupare in") kvar även inom intervjumetodiken och psykologin.

Verbet proba förekommer redan, och en försvenskning av probing är förstås probning. Man kan också bilda nya sammansättningar med prob-: om man t.ex. vill specificera skulle man kunna använda "probfråga", "probteknik" eller "probmetodik".

»»» ATT UTVÄRDERA FRÅGOR I EN STATISTISK UNDERSÖKNING

design. Svarsdata kan jämföras med en annan datakälla, t.ex. ett register eller en annan undersökning, till vilken man har hög tilltro. Ett annat sätt är att undersöka om variabler relaterar till varandra som förväntat. Man kan även studera om olika individers svar överensstämmer som de borde, t.ex. partners eller andra hushållsmedlemmars svar om gemensamma områden. Andelen partiellt bortfall samt orimliga eller direkt felaktiga värden ger också feedback på hur mätinstrumentet presterat (Tourangeau, Rips & Rasinski 2000).

Samspel med den kognitiva psykologin

Den kognitiva psykologin (läran om hur människor behandlar information, t.ex. tanke- och minnesprocesser) har sedan några decennier tillbaka ett starkt fäste inom litteraturen om frågekonstruktion (Tourangeau, Rips & Rasinski 2000). Enligt en kognitiv modell består svarsprocessen av fyra steg som bör gås igenom av respondenten för att denna ska kunna lämna ett adekvat svar. De fyra stegen är:

- 1) Förståelsen av frågan, vad det är som efterfrågas
- 2) Återhämtning av relevant information, oftast från minnet, för att kunna besvara frågan
- 3) Bedömning av den återhämtade informationen som bör ingå i svaret till frågan
- 4) Anpassning av det individuellt formulerade svaret till de befintliga svarsalternativen i blanketten

I alla dessa steg finns potentiella felkällor och modellen kan därför användas som en utgångspunkt när frågorna utvärderas. Det finns också förslag på en utökad modell av svarsprocessen när det är företag som besvarar frågorna (Willemack & Nichols 2010).

En del testmetoder är också inspirerade av eller lånade från den kognitiva psykologin. En sådan metod är "tänka-högt"-metodiken som ursprungligen utvecklades inom kognitiv forskning om problemlösning (Willis 2005). Metoden innebär, precis som det låter, att testpersoner

återger sitt tankeflöde i realtid medan frågorna besvaras. Genom att analysera testpersonernas tankeflöde kan man få insikt i hur de förstår frågan och vad de innefattar i begrepp och i sitt svar. En annan teknik är probing (ungefär "utredningsfrågor") som ligger närmare kvalitativ intervjumetodik men med en kognitiv utgångspunkt (Willis 2005). Antagandet är att man genom att ställa frågor till testpersoner kan få insikt i deras tänkande. Utredningsfrågorna kan med fördel knytas till modeller av svarsprocessen. Exempel: "Vad innebär 'ansträngande motion' för dig? Kan du ge exempel?" (förståelsen av frågan). "Hur kom du fram till att det var tre gånger förrförra veckan?" (återhämtning från minnet).

Dessa tekniker kan ge värdefull information om hur frågorna i undersökningen fungerar. Frågan "Hur många barn har du?" kan tyckas vara enkel och självklar. Likväl vet vi från våra kognitiva tester att den väcker många frågetecken hos och tolkas olika av uppgiftslämnarna – Räknas sambos barn? Ingår barn som är myndiga? Innefattas barn som har flyttat hemifrån? Om inte sådana oklarheter definieras i frågan kommer uppgiftslämnarna att göra det på egen hand, medvetet eller omedvetet, men troligtvis på lite olika sätt.

Vinjetter

Vinjetter är ett annat kognitivt verktyg. De kan beskrivas som korta berättelser eller scenarier vilka beskriver hypotetiska situationer eller karaktärer, i text eller med hjälp av bilder. Meningen är att respondenten ska reagera på dessa och vinjetten följs normalt av ett antal frågor om hur respondenten uppfattat densamma. Det kan vara raka ja/nej-frågor, attitydfrågor med svarsskalor eller mer öppna följdfrågor. I surveysammanhang används vinjetter även som ett mindre konfronterande sätt att undersöka och ställa frågor om känsliga ämnen (Martin 2004).

Ur mättekniskt perspektiv används vinjetter främst för att identifiera problem med frågor

och dess design (Martin 2004). Vinjetter är ett kostnadseffektivt sätt att studera olika situationer och begrepp eftersom alternativet att hitta verkliga fall av respondenter eller testpersoner som uppfyller de specifika kriterier vinjetterna beskriver kan kräva stora resurser (Beck 2010). Avsikten är oftast att undersöka den kognitiva process som leder till ett val av svarsalternativ för respondenten, m.a.o. att utvärdera respondenters förståelse av frågornas terminologi och avsikt (Willis 2005).

Ett exempel på hur vinjetter kan användas är för att undersöka förståelsen av begreppet arbeta. En vanlig fråga i survey-sammanhang kan vara: Arbetade du förra veckan? Frågan är då vad respondenten lägger för betydelse i begreppet arbeta. Med hjälp av vinjetter kan man måla upp olika situationer där testpersonen får svara på om han eller hon anser att denna person har arbetat eller inte. En vinjett kan då lyda: Lars hjälpte till med bokföringen i sin frus företag 20 timmar förra veckan, han fick dock ingen lön. Följdfrågan: Anser du att Lars arbetade förra veckan?

Man bör dock vara medveten om att vinjetterna utgör just hypotetiska situationer och är en indirekt metod, förknippad med alla där tillhörande mättekniska problem. Se Martin (2004) och Jergeby (1999) för vidare diskussion om vinjetter.

Den här typen av kognitiva tekniker används oftast i tester på ett begränsat urval (5-15 testpersoner) och resultatet ska givetvis inte ses som ett bevis. Det går inte dra slutsatser om utbreddheten eller konsekvenserna av eventuella problem. Däremot utgör resultatet oerhört viktig information i förbättringsarbetet med frågorna. Om man uppmärksammas på och känner till att problem föreligger kan man också åtgärda dessa problem.

ANDREAS PERSSON
FREDRIK SCHEFFER,
STATISTISKA CENTRALBYRÅN

VARFÖR ÄR DET FEL ATT SÄGA:

"I'm good at English so communication is not my problem; it is their problem if they can't follow".

Att göra sig förstådd på engelska

För många statistiker är engelska ett dagligt arbetsredskap. Man talar engelska i internationellt sammansatta projekt, på konferenser och i arbetsmöten hemma och utomlands. Engelska är arbetsspråket på en del företag i Sverige. Undervisning på masternivå i statistik sker ofta på engelska och vanliga seminarier på svenska universitet hålls ofta på engelska.

■ Jag träffade Beyza Björkman som forskar på "English as a lingua franca" (ELF) i högskoleundervisning i Sverige. Beyza Björkman har nu en post-doctjänst på CALPIU forskningscentrum (www.calpiu.dk) i Roskilde. Tjänsten är finansierad av Vetenskapsrådet. Centrumet jobbar just med frågor runt internationalisering inom högre utbildning i Norden och andra länder i världen.

Numer finns det många miljöer inom högskolan där det talas engelska i grupper där inte en enda av deltagarna har engelska som sitt modersmål. Engelskan har sedan andra världskriget blivit globalt helt dominerande inom åtskilliga språkliga domäner.

Vårt samtal skedde på ELF; här följer några axplock i översättning till svenska.

■ Hur kom du in på ditt forskningsområde?

– Jag var lärare i engelska inom en ELF-miljö och började fundera på hur engelska används i situationer där ingen är moders-

målstalare. Hur gör deltagare, studenter och lärare i det här fallet, för att komma runt kommunikationsproblemen? Jag fick chansen att gå in på frågan på djupet.

■ Ja, hur gör folk då?

– De använder det som inom det här området kallas pragmatiska och kommunikativa strategier. Till exempel upprepar man sig ofta. Man använder "back-channelling", alltså att humma och så för att visa att man hänger med. (Se sidan 19 för exempel på vissa pragmatiska strategier). Även modersmålstalare använder pragmatiska strategier men de blir så mycket viktigare i en situation där alla talar engelska som lingua franca.

■ Jag såg ett exempel i en av dina artiklar på en lärare som sa ungefär "this is very, very, very, very important". Det skulle en modersmålstalare aldrig säga.

– Om inte han eller hon med avsikt vill vara teatralisk.

»»»»

Referenser och litteraturtips

Beck, J. (2010). On the Usefulness of Pre-testing Vignettes in Exploratory Research. Statistical Research Division Research Report Series (Survey Methodology #2010-02). U.S. Census Bureau.

Biemer, P. & Lyberg, L.E. (2003). Introduction to Survey Quality. New York: Wiley.

Field, A. & Hole, G. (2003). How to Design and Report Experiments. Thousand Oaks, CA: Sage.

Fowler, F.J. (1995). Improving Survey Questions – Design and Evaluation. Thousand Oaks, CA: Sage.

Jergeby, U. (1999). Att bedöma en social situation: Tillämpning av vinjettnetoden. IMS, Socialstyrelsen.

Kvale, S. (1997). Den kvalitativa forskningsintervjun. Lund: Studentlitteratur.

Martin, E. (2004). Vignettes and Respondent Debriefings for Questionnaire Design and Evaluation. In Presser, S., Rothgeb, J.M., Couper, M., Lessler, J.T., Martin, E., Martin, J. & Singer, E. (eds.), Methods for Testing and Evaluating Survey Questionnaires. New York: Wiley, 149-171.

Morgan, D.L. (1993). Successful Focus Groups. Thousand Oaks, CA: Sage.

Tourangeau, R. (2004). Experimental Design Considerations for Testing and Evaluating Questionnaires. In Presser, S., Rothgeb, J.M., Couper, M., Lessler, J.T., Martin, E., Martin, J. & Singer, E. (eds.), Methods for Testing and Evaluating Survey Questionnaires. New York: Wiley, 209-224.

Tourangeau, R., Rips, L.J. & Rasinski, K. (2000). The Psychology of Survey Response. Cambridge: Cambridge University Press.

Willimack, D.K. & Nichols, E. (2010). A Hybrid Response Process Model for Business Surveys. Journal of Official Statistics, 26, 3-24.

Willis, G. (2005). Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA: Sage.



HÄNT OCH HÖRT

■ **LINUS SCHIÖLER**, som är verksam vid Statistiska forskningsenheten, Institutionen för nationalekonomi med statistik vid Göteborgs universitet, disputerade i statistik den 25 november på avhandlingen *Univariate and multivariate surveillance of outbreaks*. Opponent var professor Sven Knoth, Helmut-Schmidt-Universität i Hamburg.

■ Den 4 februari disputerade **MARIA GRÜNEWALD**, Matematisk statistik vid Stockholms universitet, på en avhandling med titeln *Design and analysis of response selective samples in observational studies*. Fakultetsopponent var professor Marie Reilly, Karolinska Institutet.

■ **JENNY HÄGGSTRÖM**, Statistiska institutionen i Umeå, försvarade sin doktorsavhandling i statistik den 4 mars. Avhandlingens titel är *Selection of smoothing parameters with application in causal inference*. Fakultetsopponent var Yanyuan Ma, Texas A&M University.

Har du flyttat?

■ Meddela din nya adress genom att gå in på Svenska statistikfrämjandets webbplats www.statistikframjandet.se, välj Medlemsregister och följ instruktionerna. Ditt användar-id är tryckt ovanför din adress på tidningens baksida.

Bli medlem

■ Svenska statistikfrämjandets syfte är bland annat att främja sund användning av statistik som beslutsunderlag och att väcka och sprida intresse för statistik i samhället. Om du önskar gå med som medlem i Svenska statistikfrämjandet skicka ett e-brev till Chandra Adolfsson på sekrfram@gmail.com. Du får Qvintensen i brevlådan och platsannonser via e-post. Det ställs inga krav för att bli medlem; alla som är intresserade av statistik och vill stödja statistikens roll i samhället är välkomna.

lingua franca

■ [lingwa fraˈŋka] (italienska), ursprungligen namnet på ett språk som användes som gemensamt språk i handel och sjöfart runt Medelhavet under senmedeltiden. Språket var förmodligen ett pidginspråk med ordförråd mest från romanska språk.

Numera används *lingua franca* av språkvetare som teknisk term för alla hjälpspråk som brukas mellan människor med olika språk. T.ex. kan engelska användas som *lingua franca* av en svensk som talar med en ryss. Vissa språk används nästan enbart som *lingua franca*, t.ex. pidginspråk och esperanto. (Ur Nationalencyklopedin.)

»»» GÖRA SIG FÖRSTÅDD PÅ ENGELSKA

Beyza gör en minisketch, och plötsligt känns det som om den teatraliske läraren har man själv haft många gånger.

– En modersmålstalare säger hellre till exempel "extremely important" än upprepar very fyra gånger. Men repetitionen går hem hos studenterna; budskapet att detta är jätteviktigt går fram.

■ **Är inte det dålig engelska att säga "very" fyra gånger istället för att välja något mer nyanserat uttryck?**

Det märks att frågan är gammal och väl genomtröskad hos lingvisten Beyza.

– Jag är mycket försiktig med att säga att något är dåligt språk om syftet är att använda språket som verktyg. Jag är mer intresserad av om språket är funktionellt. Det gäller att kommunicera så deltagarna i kommunikationen förstår varandra. Därför är det inte bra att ha attityden att "jag är så bra på engelska så engelska är inte mitt problem, och om inte du förstår vad jag säger så är det ditt problem". Det är allas ansvar att kommunicera på ett effektivt sätt inom ELF-miljöer.

■ **Men språket är väl mer än bara rå funktionalitet?**

– Visst, man kan uttrycka sig nyanserat och elegant, med *espi* och estetisk skönhet, men i undervisning och i samtal mellan studenter är språket ett kommunikationsmedel. Eller som en av mina informanter formulerade det: "vi är ingenjörer, vi gillar inte kosmetika, vi vill att grejerna ska funka".

– Engelska som *lingua franca* är inte något själlöst språk. Det är ett rikt språk i likhet med andra språk med språkbrukare som kan vara väldigt avancerade.

■ **Är inte det ett trösterikt budskap? Det spelar inte nödvändigtvis så stor roll om inte allt blir helt grammatiskt korrekt.**

– Grammatisk korrekthet ger inte alltid bättre kommunikation. Det är bra med korrekt grammatik men det måste till något mer också. Det finns brukare av engelska som *lingua franca* som struntar i språkelement som inte ger värde. Det finns en massa småsaker i det engelska språket

som är onödiga, till exempel ändelsen "s" i tredje person singular, "he sings" men "they sing". En del bryr sig inte om att använda den ändelsen och det går lika bra ändå.

■ **Kan det till och med vara en fördel att inte vara modersmålstalare?**

– Nej, en fördel är det väl inte. Men det händer ofta att modersmålstalare använder idiomatiska uttryck och fraser även i situationer där de försvårar kommunikationen. Något som är väldigt viktigt i ELF-miljöer är "transparency". Ett idiomatiskt uttryck som "kick the bucket" kan försvåra kommunikationen. Opponenten på min disputation var en framstående forskare och en världskänd professor på området och modersmålstalare av engelska. När han frågade mig om jag tycker att han kommunicerar bra på engelska, svarade jag att det vet jag inte. Alla började skratta. Jag menade att för att kunna

besvara den frågan skulle jag behöva studera hans kommunikation i de situationer vi pratade om. En i publiken sa att vi skulle behöva en manual över pragmatiska knep för modersmålstalare.

■ **Alla skickliga lärare använder väl pragmatiska strategier oavsett språk? Till exempel att markera övergången från ett ämne till ett annat.**

– Absolut, det gör skickliga lärare. Men det blir än mer viktigt att använda dem ofta och tydligt när man talar engelska som *lingua franca*.

■ **Jag upplever att svenskar är överdrivet fokuserade på sitt uttal av engelska. Istället borde de oroa sig för vad som är kulturellt korrekt.**

– Vad roligt att du säger det! Jag forskar nämligen nu på kulturellt specifika pragmatiska strategier, "pragmatic and intercultural ability". Men apropå uttal av engelska så finns det forskning på det också. Bland annat har Jennifer Jenkins vid University of Southampton i England definierat "core features" i uttalet av engelska och "non-core features". Det man ska lära sig uttala och uppfatta är framför allt core features, t.ex. att skilja mellan lång och kort vokal. T.ex. skill-

»Grammatisk korrekthet ger inte alltid bättre kommunikation»



Beyza Björkman.

nad mellan tonande och tonlöst "th" är däremot "non core" eftersom avvikelser från standarduttalet på den punkten inte leder till missförstånd. Att försöka lära sig received pronunciation (Oxford English) är oöverstigit för de allra flesta i vuxen ålder och ganska onödigt.

■ Vilka råd kan du ge lärare som undervisar på engelska?

– Använd stöd som till exempel bilder där nyckelorden framgår tydligt. Strukturera lektionerna nästan överdrivet tydligt och använd fraser som "this brings us to the next point". Repetera ofta. Det är inte tråkigt, repetition gör att studenter känner sig säkra. Använd frågor, även retoriska frågor och frågor som läraren själv besvarar omedelbart. Summera sektioner i undervisningen, och säg ungefär "what I mean by this is...". Man ska tänka på hur man kan ersätta eller komplettera det man inte har i språket.

■ En del är oroade över "domänförluster", alltså att svenska språket förlorar språkliga domäner till förmån för engelska.

– Jag är försiktig med att prata om förluster. Ordet antyder att något fanns där förut. Man kan inte förlora det man aldrig har haft. Jag kan förstå frustrationen över att vi inte kunnat utöka svenskan till att omfatta till exempel IT, men det är inte korrekt att kalla det en domänförlust.

Många i Sverige är ambitiösa med att hitta och använda svenska termer men arbets-språket är de facto engelska för ganska många domäner. Om inte de svenska termerna används aktivt kommer de inte i bruk. Vill man ha ett levande svenskt språk på dessa domäner då måste man skapa miljöer där svenska används, till exempel lokala konferenser. Annars är försöken att införa svenska termer fåfänga.

■ Du använder både kvalitativa och kvantitativa metoder i din forskning. Finns det statistiker på ELF-området?

– Ja bland andra Glen Ole Hellekjær på Universitetet i Oslo. Vi säger att han är "crazy about statistics", säger Beyza och skrattar. Men det är en styrka för forskningsområdet som helhet att ha olika typer av forskningsresultat som vi alla drar nytta av.

■ Vad har du för idéer för din framtida forskning?

– Jag vill jämföra den svenska korpus jag har byggt upp med en motsvarande insamlad i USA för att se om det är skillnad på pragmatiska strategier i svenska situationer med motsvarande i en miljö där engelskan dominerar men där det ändå finns inslag av engelska som lingua franca.

DAN HEDLIN

Faktiska, ordagrant återgivna exempel på pragmatiska strategier

Comment on terms and concepts

"what is steam reforming, this is as i already mentioned a very very commercial way to produce hydrogen"

Comment on details and content of task

"i don't know if we're supposed to know the code during the lab i think we're just supposed to know it so we can model it later"

Comment on discourse structure

"i would like to move on to branch prediction...ok so moving onto branch prediction"

"a few minutes ago i said that"

"you have now seen one way of handling renaming an out-of-order execution"

Comment on discourse content

"so i don't know i i agree with you"

"[yeah yeah i see the (point) but] i am just wondering what could be the the the complicated"

Comment on signaling importance

"it is noteworthy that instructions are taken from the..."

"one thing which is very important from an industrial point of view is that..."

Comment on intent

"yeah i just wanted him to repeat that"

"because i just wanted to know how you motivate continuous equations of motion"

Comment on common ground

"we have to check the distillation process"

"then we need batches for four cubic meter then then it's not (downscaled) any more"

Svar från Björn Jonsson

I min artikel "Mjuka fakta och osannolik noggrannhet" (Qvintensen nr 3/2010) nämnde jag några exempel på missuppfattningar och tveksam statistik användning i debatten om den höga sjukfrånvaron. Huvudpoängen i artikeln var en kritik av Försäkringskassans och FUT-delegationens beräkningar av fusk och felaktiga utbetalningar inom socialförsäkringssystemet. Jag menar att dessa beräkningar inte är stort mer än gissningar. De har dessutom utnyttjats av politiker och organisationer som haft intresse av att sprida myten om svenskarna som ett folk av bidragsfuskare.

I en replik av Jonas Ekstedt, Gunnar Johansson och Peter Juslin (Qvintensen nr 4/2010) avfärdas denna kritik. Dels anför de att både inflöde och sjukfallslängd låg bakom den ökade sjukfrånvaron (vilket jag också nämnde i min artikel). Dels menar de att de expertbedömningar som legat bakom beräkningarna varit välgrundade och redovisats på ett transparent sätt.

När det gäller den första punkten skriver författarna att ökningen av antalet fall kan presenteras på olika sätt, som "bara" 20 procent eller nästan 50 procent. Det är ett underligt påstående, kommande från statistiker. I min artikel talade jag nämligen specifikt om ökningen av antalet sjukpenningdagar 1996–2002 och hur stor del av denna ökning som kan förklaras av ökat inflöde respektive längre sjukskrivningsperioder. Det har genomförts flera oberoende beräkningar av detta, av SCB, Anna Hedborgs socialförsäkringsutredning och Arbetslivsinstitutet. De har alla kommit fram till samma resultat: att mellan

80 och 85 procent av ökningen berodde på de förlängda sjukskrivningsperioderna. Någon "väl dold hemlighet" har detta naturligtvis inte varit, men det har definitivt inte varit känt för någon bredare allmänhet utanför Försäkringskassans analysenhet.

När det gäller beräkningarna av fusk och felaktigheter är det närmast skrattretande att kalla dem "transparenta, välgrundade och objektiva". I FUT-delegationens skrift Vad kostar felen? (2007) redovisas nästan ingenting om hur experterna kommit fram till de siffror som presenteras för de olika socialförsäkringsslagen. När jag läste skriften första gången häpnade jag över hur en så bristfällig rapport kunnat få så stort genomslag.

Min huvudpoäng är att delegationens gissningar rimligen varit starkt påverkade av den offentliga debatten. Mediebilderna av sjukfrånvaron förändrades dramatiskt i början av 2000-talet, vilket jag visar i boken Kampen om sjukfrånvaron (2010). Vid millennieskiftet ansåg nästan alla att sjukfrånvaron främst var ett arbetsmiljöproblem. Sju år senare, då FUT-delegationens beräkningar togs fram, ansågs den höga sjukfrånvaron istället bottna i överutnyttjande och en grasserande "sjukskrivningsepidemi" i befolkningen. Det är min fasta övertygelse att delegationen hade landat i helt andra siffror om man gjort sina beräkningar 2000 istället för 2007. Vad säger det om beräkningarnas trovärdighet? BJÖRN JONSSON, MALMÖ HÖGSKOLA

(Jonas Ekstedt, Gunnar Johansson och Peter Juslin har inbjudits att svara men avböjt.)

Fotnot

FUT-delegationens skrift Vad kostar felen? finns tillgänglig på regeringens webbplats. En pedagogisk analys av inflödets respektive sjukskrivningsperiodernas bidrag till den ökade sjukfrånvaron finns i statistikern Anders Wikmans artikel "Utsorterad från arbetslivet – en fördömd analys" (2006). Den ingår i antologin Väger till en öppnare arbetsmarknad och kan laddas ner från Forskningsrådet för arbetsliv och socialvetenskaps, FAS, webbplats.

Nollhypoteser och statistisk signifikans

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010). I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

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

I tidigare böcker har jag skrivit om statistiska principer och hypoteser i till exempel "Mjuka fakta och osannolik noggrannhet" (2010) och "Mjuka fakta och osannolik noggrannhet" (2010).

Fyra repliker på Nollhypoteser och statistisk signifikans

Fredrik Jonssons debattartikel i Qvintensen nr 4/2010 genererade fyra repliker. Han svarar här på tre av dem. Kurt Pörns inlägg kom för sent för att Fredrik Jonsson skulle hinna formulera ett svar. Alla rubriker är satta av redaktören.

Nollhypotesen är endast en utgångspunkt

Ett tvåsidigt test är ekvivalent med två ensidiga test

”vet ej”, utan hellre skulle säga mer positivt att testutfallet är förenligt med H_0 (eller med både H_0' och H_0''), det är en mindre väsentlig skillnad i uttryckssätt.

Det tycks alltså främst vara för att kunna uttala sig om effektens riktning, som Fredrik Jonsson vill ersätta det tvåsidiga testet med två ensidiga test. Förr förekom nog den ortodoxa uppfattningen, att ett tvåsidigt test inte tillät uttalanden om effektens riktning, men det påstår väl ingen numera? Ett tvåsidigt test är i själva verket ekvivalent med två ensidiga test, det gäller bara att ha koll på det adekvata p -värdet (eller felrisken). Denna tolkning förtjänar att vara mer känd, tycker jag. Ett korrekt p -värde, när man gör dessa två ensidiga test och ser efter ifall det ena eller andra utfaller signifikant, fås genom att dubblera det minsta av de två p -värdena, se t.ex. Cox (2006, Sec. 3.3), dvs vi får det tvåsidiga testets p -värde. Där blir det ett probabilistiskt fel i Fredrik Jonssons resonemang, för han tar inte hänsyn till att han faktiskt gör multipla test (två disjunkta test) och av dem väljer det (mest) signifikanta. Han skulle därför behöva dubblera det minsta p -värdet för att kompensera för selektionen (jfr Bonferroni-gränsen), i stället för att bara ta det minsta p -värdet rakt av.

David Cox inleder för övrigt sitt kapitel Significance tests med att lista sex olika situationer där vi formulerar nollhypoteser av typ $H_0: p_1 - p_2 = 0$ eller $H_0: \mu = 0$. En av dem beskriver Cox så här, vilket jag tror är precis vad Fredrik Jonsson har

i tankarna: “There may be no special reason for thinking that the null hypothesis is true but it is important because it divides the parameter space into two regions with very different interpretations. We are then interested in whether the data establishes reasonably clearly which region is correct” (dvs vilket tecken parametern har). Nollhypotesen behöver inte ens i strikt mening vara möjlig, något som kan utmärka opinionsundersökningarnas ändliga populationer.

Man kan även invända mot några av Fredrik Jonssons begreppsmodifieringar. Speciellt är hans utsaga om felrisken (i hans egen definition) inte korrekt, under just det vi brukar kalla nollhypotesen (H_0 ovan). Han har emellertid av för mig oförståelig anledning valt att exkludera det parametervärdet. Varför denna genomgående aversion mot värdet noll?

Fredrik Jonsson har också ett principexempel från läkemedelsutveckling, där det kan vara motiverat formulera sig i termer av ett enda ensidigt test (i stället för två). Jag håller med om det, och sådan är väl även den gängse uppfattningen. Samtidigt tror och tycker jag dock att läkemedelsforskaren inte bara skulle säga ”vet ej”, om substansens effekt slog oväntat och kraftigt åt ”fel håll”.

ROLF SUNDBERG, STOCKHOLMS UNIVERSITET

Referens: Cox, D.R. (2006). *Principles of Statistical Inference*. Cambridge University Press.

Fredrik Jonsson slår i Qvintensen nr 4/2010 ett slag för en nytolkning av teorin för hypotesprövning. Jag tycker visserligen, som framgår nedan, att han har fel i sitt resonemang. Men eftersom resonemanget är uppfriskande och rör ett centralt område inom statistiken så hans inlägg värt att fundera över.

Som jag ser saken är statistikernas hypotesprövning en form av logikernas Reductio ad absurdum, alltså en bevisföring där man gör ett antagande och visar att dess logiska konsekvenser är orimliga och att antagandet därför måste vara falskt. En liknande ansats använder t.ex. en matematiker som vill visa att roten ur 2 måste vara ett irrationellt tal, alltså ett tal som inte kan uttryckas som en kvot mellan två heltal. Hon antar att roten ur 2 kan uttryckas på formen a/b där a och b är heltal, och visar sedan att i så fall måste ett av talen a eller b samtidigt vara både ett jämt och ett udda tal. Eftersom detta är orimligt, måste antagandet om ett rationellt tal vara fel och alltså är roten ur 2 ett irrationellt tal.

Våra hypotesprövningar utnyttjar i grunden samma maskineri: vi börjar med ett antagande, nollhypotesen, och konfronterar den sedan med verkligheten, våra data. Skillnaden är att vi inte rör oss i logikens och matematikens ideala värld, utan i en värld med slumpvariation. Våra slutsatser blir därför inte av typen ”omöjligt” utan av typen ”otroligt”, eller med vårt fackuttryck ”signifikant”.

Det är i behandlingen av nollhypotesen som jag anser att Fredrik Jonssons principiella fel ligger. Nollhypotesen är inte något man tror på utan endast en utgångspunkt för ett slags motbevis. Om man så vill en slags teoretisk krok som man hänger upp sitt försök till motbevis

»»»»

Nollhypotesen är inte något man tror på utan endast en utgångspunkt för ett slags motbevis

»»» pa. Man kan ju aldrig, som vi predikar för våra studenter, bevisa en nollhypotes. Det saknar alltså betydelse om nollhypotesen är rimlig eller inte, på samma sätt som det för det matematiska beviset är irrelevant om matematikern tror att roten ur 2 är ett rationellt tal eller inte. Och därmed är det ingen principiell skillnad mellan de två fallen som Fredrik diskuterar.

Men det finns mer att fundera över. För det första har Fredrik Jonsson helt rätt i att hypotesprövning är olämpligt i många situationer. Det är oftast bättre att försöka besvara den positiva frågan "Hur stor är skillnaden?" med ett konfidensintervall än den negativa "Hur stor är skillnaden inte?" med ett signifikanstest. Men det finns ändå en pedagogisk poäng med begreppet signifikans, inte minst i den situation som Fredrik Jonsson diskuterar, nämligen opinionsmätningar. Det är nämligen ett av de få områden där journalister (nästan) förstår en statistisk term.

De skillnader som i samband med opinionsmätningar kallas signifikanta skulle med enklare ord kunna beskrivas som t.ex. troliga eller som skillnader som knappast kan bero på urvalsfel. Alla icke-signifikanta skulle i sin tur kunna beskrivas som möjligen orsakade av slumpen. Och, med en liten förökning, är det just dessa icke-signifikanta resultat som sammanfattas i motsvarande konfidensintervall. Ett 95 %-igt konfidensintervall täcker alltså de icke-signifikanta resultaten på 5%-nivån. Och här ser vi åter att Fredrik Jonssons resonemang inte stämmer, eftersom hans icke-signifikanta resultat skulle motsvara ett 90%-igt konfidensintervall.

Fredrik Jonsson bemöter den tänkta kritiken mot att hans test skulle vara betingat utifrån ett resonemang om ifall nollhypotesen styrks eller inte. Som jag ser saken går det dock inte att komma ifrån att Fredrik använder samma data två gånger, en gång för att välja mothypotes och en gång för att studera avvikelens storlek.

Det skulle kanske t.o.m. gå att göra ett Reductio ad absurdum utifrån Fredrik Jonssons påstående. Antag att han har rätt. Och antag

vidare att vi vill göra något så ovanligt som en 50%-ig hypotesprövning. Enligt min lärobok får jag då ett signifikant resultat om jag hamnar utanför gränserna $\pm 0,67$ (om vi kan använda standardnormalfördelningen). Fredrik Jonsson skulle däremot först välja sida och sedan lägga alla de 50 procenten där. Han skulle alltså alltid få signifikans vid test på 50%-nivån. Det förefaller orimligt, och alltså är utgångspunkten fel.

Tja, detta var ju i stort sett bara invändningar. Men låt mig avsluta med att sticka ut huvudet själv när det gäller frågan om enkel- eller dubbelsidiga test. Jag har nämligen länge funderat på om det egentligen någonsin i praktiken är motiverat att göra enkelsidiga test!

När jag för länge sedan sysslade med medicinska konsultationer brukade jag använda en kontrollfråga i situationer när klienten ville göra ett enkelsidigt test: "Vad skulle du göra om vi fick en stark signifikans åt 'fel' håll?" De flesta svarade hederligt att det skulle vara mycket oväntat, men att det måste följas upp: Var det något fel i försöksuppläggningsen, hade man missat något eller var det helt enkelt så att man var inne på fel spår? Och i så fall skulle ju ett dubbelsidigt test vara mer motiverat.

I Fredrik Jonssons exempel med läkemedel och placebo kan man ju fundera på vad man skulle dra för slutsats i ett klassiskt experiment, med mothypotes (enkelsidig med bäst resultat för läkemedlet) och antal observationer bestämda i förväg, om resultatet skulle bli klart åt "fel" håll – säg $z = -2,3$. Skulle det inte vara tjänstefel att bara rycka på axlarna och att inte börja fundera på varför det troligen blev sämre resultat än med placebo? Det är lätt att göra en lista med liknande situationer, t.ex. i de miljöstatistiska situationer som tas upp i Qvintensen 4/2010.

Som jag ser saken så är det oerhört svårt att hitta praktiska situationer där man inte skulle dra slutsatser av ett signifikant resultat åt "fel" håll, och därmed skulle också utrymmet för enkelsidiga test vara mycket begränsat, om det ens skulle finnas.

ULF JORNER

"Att använda samma data två gånger"

[1] Med ovanstående ekvivalensformulering blir detta väldigt tydligt; man utnyttjar i det andra steget information (med avseende på tecknet på teststatistikan) som inte fanns tillgänglig från början. Detta givet, så betingar man rent faktiskt på information från själva testutfallet.

[2] I rättvisans namn skall man kanske påpeka att denna procedur ej uppfyller standardkriterier som att signifikanta utfall motsvarar diton 'varandes lika med eller extremare än rejektionsgränsen'. Jag tar mig här dock friheten att driva denna analogi då Fredrik Jonsson rent faktiskt gör just detta, dock något mer implicit, med sin procedur då man i steg två i detta fall i praktiken plockar bort hälften av de initialt intressanta extremare (i denna mening) värdena motsvarande, så att säga, fel tecken.

[3] Situationen blir ju annorlunda om man antar att det endast finns ett litet antal möjliga sanna värden, men då jämför man väl lämpligen snarare motsvarande hypoteser direkt mot varandra än delar upp dessa i noll- och mothypoteser (baserat exempelvis på likelihoodberäkningar). Dessa fall torde dock vara ganska ovanliga i praktiken.

[4] För att vara lite petig så verkar den introducerade termen felrisk även lite onödig då detta exakt motsvarar signifikansnivån – under varje rimlig tolkning av vad som beskrivs – och fel av typ-II borde nog beskrivas som utfall varandes falskt icke-signifikanta och ej bara icke-signifikanta. Definitionen av felrisken kan även eventuellt anses opassande då, som författaren även beskriver, detta här enbart motsvarar en av de två möjliga feltyperna (typ-I; typ-II).

Referenser

Goodman, S. (2008) A Dirty Dozen: Twelve P-Value Misconceptions. *Seminars in Hematology*, Vol 45, 135-140.

Rothman, K.J., Greenland, S. & Lash, T.L. (2008) Precision and Statistics in Epidemiological studies. I *Modern Epidemiology*, 3rd ed. Red Rothman, Greenland och Lash. Lippincott Williams & Wilkins

Att använda samma data två gånger

Bakgrund: Fredrik Jonsson beskriver i sin artikel i Qvintensen nr 4/2010 ett, enligt egen utsago, nytt angreppssätt för statistisk hypotestestning. Kort uttryckt, om jag har förstått det hela korrekt, så består detta av att ersätta standardmetoden, där man sätter upp en nollhypotes och en tillhörande mothypotes, med en metod där man istället kan dela upp den senare i flera olika fall. Som ett exempel förs fram testning kopplad till typexemplet en normalfördelning $N(\mu, 1)$ där hypotesuppsättningen då sätts till nollhypotes $H_0: \mu=0$; mothypotespar $H_1: \mu>0, H_2: \mu<0$. Svaren från ett sådant test beskrivs då enligt Fredrik Jonsson som varandes signifikant med avseende på H_1 ("μ större än 0"), signifikant med avseende på H_2 ("μ mindre än 0"), samt att man ej funnit signifikans ("vet ej").

Sedan beskriver Fredrik Jonsson hur denna modifiering leder till att man kan motivera mindre stränga rejektionskriterier utan att fördenskill som konsekvens tvingas höja signifikansnivån. Detta i sig själv, och argumentationen bakom, anser jag vara något diskutabel. Detta avses här därför att diskuteras med viss grundlighet nedan.

Huvudsaklig kritik

Grundproblemet här är att man som resultat ändrar den faktiska signifikansnivån om man ändrar tröskelvärdena för rejektion (från, absolutbeloppsmässigt, 1,96 till 1,64 i aktuellt exempel vid nyttjande av signifikansnivån 0,05). Fredrik Jonsson avfärdar sådan tidigare mött kritik, om att detta då kan ses som att motsvara ett betingat test, med ett något vagt argument givet som 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..."

Jag har svårt att riktigt följa logiken i detta. Rent formellt så blir en sådan procedur helt ekvivalent med att först ta fram värdet på teststatistikan, kika på vilket tecken denna har (plus eller minus) och sedan utföra ett ensidigt test (i

traditionell mening) med avseende på den mest lämpliga mothypotesen (i avseende att så ofta som möjligt finna signifikans). Om man gör detta, som antytts ovan, med avseende på signifikansnivån så blir den totala signifikansnivån – i klassiskt förstånd – helt klart 10 %. Och ja, metoden är klart och tydligt (på grund av ekvivalensen ovan) ett exempel på ett betingat test¹.

Ett annat sätt att formulera situationen är att man initialt är intresserad av skillnader oavsett, så att säga, vilket håll dessa pekar åt. Därför utför man två ensidiga test med avseende på en viss signifikansnivå. Då man konsulterar två test – två olika utfall – så befinner man sig i vad som skulle kunna kallas för en multipeltestningssituation (med det modesta antalet av två test). För att då kunna hålla sig inom en justerad total (global) signifikansnivå så måste varje ensidigt test därför utföras på en strängare nivå. Bonferronitestning ger direkt att man klarar sig om man halverar signifikansnivån i det enskilda fallet. Det speciella beroendet mellan dessa två test gör här dessutom att detta vanligtvis gravt konservativa justeringssätt, under antagandena, blir exakt.

För att dra det hela till sin spets så kan man även konstruera ett extremt exempel: Antag normalfördelningssituationen ovan. Exkluderades punkten 0, dela nu in utfallsintervallet i m delar, $m>20$, där varje del motsvarar sannolikhetsmassa $1/m<0,05$ under nollhypotesen. Låt nu varje sådant intervall motsvara en egen mothypotes, det vill säga ett eget rejektionsområde. Via ett resonemang analogt med det som Fredrik Jonsson ger så kan man sedan, helt oberoende av utfall, hävda att mothypotesen inneslutande det faktiska utfallet alltid är signifikant på signifikansnivån 0,05².

Diskussion

Författaren varnar för konceptuell förvirring gällande tolkningen av nollhypotesen och dess relation till icke-signifikanta utfall, men verkar själv något otydlig på dessa punkter. Proceduren med att införa kategorin 'vet ej' som motsvarandes icke-signifikanta utfall har jag svårt att

förstå som något annat än en ren omformulering av exakt samma tolkning som redan tillskrivs sådana utfall med standardformulerad testning? Den generella, och naturliga, tolkningen av icke-signifikanta utfall är ju just att de icke bekräftar den testade mothypotesen snarare än att de ger evidens åt nollhypotesen (som ju normalt endast fungerar som ett slags basantagande)³. För viss relaterad tolkningsdiskussion inom det statistiska signifikansområdet, se exempelvis Sven Ove Hanssons artikel 'Signifikansmissbruket i pseudovetenskap' i Qvintensen 3/2010, Goodman (2008) och – mer allmänt – Rothman et al. (2008, kapitel 10).

Författaren för även fram kritik utgående ifrån att testade värden – som sannolikheter, väntevärden och dylikt – i princip aldrig är exakt lika med varandra (mellan grupper) eller med ett specifikt värde (för en enskild grupp), vilket då leder till att nollhypotesen i samma mening aldrig är helt sann. Denna kritik, vid användandet av den naturliga testtolkningen, blir dock då direkt verkningslös eftersom detta aldrig blir det som man påstår gällande grundantagandet. Vidare så motsvarar den nya p-värdesdefinitionen som förs fram,

"...den maximala sannolikheten att få ett minst lika avvikande utfall, givet att tecknet för X pekar i fel riktning..."

givet varje rimlig tolkning, exakt den traditionella implementeringen givet ett ensidigt test. I denna situation kan ju nollhypotesen ses som en sammansatt hypotes (precis som mothypotesen) och för att säkerställa upprätthållandet av given signifikansnivå under den sammansatta nollhypotesen så maximerar man denna sannolikhet, vilket direkt innebär att man utför relaterade beräkningar (av rejektionsgränser etc.) under grundantagandet att, exempelvis, $\mu=0$. Huvudfrågan blir ju istället, här som ovan, om man initialt antar ett ensidigt eller tvåsidigt förfarande⁴.

LARS ÄNGQUIST, INSTITUT FÖR SYGDOMSFOREBYGGELSE, KÖPENHAMN

Ett bayesianskt synsätt

Qvintensen, nr 4/2010, försöker Fredrik Jonsson, Uppsala universitet, väcka intresse för en diskussion om förändring av teorin för statistisk hypotesprövning. Jag välkomnar en sådan diskussion eftersom den traditionella uppställningen av nollhypoteser, mycket riktigt, inte alltid har en formulering som passar det givna problemet.

Detta illustreras så tydligt av exemplet med två på varandra följande opinionsundersökningar där man skattar populationsandelarna p_1 resp. p_2 och där ökningen $p_2 - p_1$ inte är statistiskt säkerställd. Hypotesprövningen är baserad på nollhypotesen $H_0: p_1 = p_2$ och mothypotesen $H_1: p_1 \neq p_2$. Som Jonsson säger: "Är det inte givet på förhand att mothypotesen är den gällande? Vem skulle vilja påstå att populationsandelarna är exakt desamma vid dessa två tillfällen". Han föreslår istället att man delar upp problemet i fallen $H_1: p_1 > p_2$ och $H_2: p_1 < p_2$ och tillåter tre svarsalternativ (ökning, minskning och vet ej). Jag håller helt med om att denna alternativa hypotesformulering är mycket mera realistisk samtidigt som den leder till statistiskt säkrare svar.

Jag skulle med mitt inlägg vilja utvidga diskussionen till att omfatta även det bayesianska synsättet beträffande hypotesprövning. I ovan nämnda exempel kan man först konstatera svårigheten att tilldela den enkla nollhypotesen en realistisk a priorifördelning vid sidan om en motsvarande tilldelning för den sammansatta mothypotesen. Definitionen av en a priorifördelning är betydligt enklare i den alternativa uppställningen där fallen H_1 och H_2 till en början kan tilldelas lika stor sannolikhet via en likformig fördelning för parametern p . Vid varje ny opinionsundersökning väljer man som a priorifördelning föregående undersöknings posteriorifördelning. Den nya posteriorifördelningen

sammanfattar all den information vi har om p_2 och fördelningens medelvärde brukar användas som punktskattning av parametern. Sedan kan sannolikheten (P) att $p_1 > p_2$ och sannolikheten ($Q=1-P$) att $p_1 < p_2$ beräknas ur kombinationen av dessa parametrars fördelningar.

Om $Q > P$ har vi, statistiskt sett, haft en ökning av populationsandelen p . Finns det i det bayesianska synsättet någon plats för begreppet "statistiskt säker"? Såvitt jag kan se är svaret nej. Istället för att säga att det nya estimatet av p är statistiskt säkerställt använder vi uttrycket: andelen p har ökat med sannolikheten Q ($Q > P$) eller minskat med sannolikheten P ($P > Q$). Sannolikheten P (eller Q) är det enda måttet på styrkan i erhållet analysresultat. Om $P \approx Q$ har vi en situation där förändringen i p endast är marginell eller där datamaterialet är så svagt att det inte orkat förändra a priorifördelningen.

Fredrik Jonsson nämner tre svarsalternativ i hypotesuppställningen med H_1 och H_2 : ökning, minskning och vet ej, där det sistnämnda alternativet kan vara ganska frekvent. Enligt det bayesianska angreppssättet vet vi alltid något i termer av sannolikhet. Vi tangerar dock svarsalternativet "vet ej" när $P \approx Q$.

Ibland kan det ändå vara svårt att undvika den traditionella uppställningen med nollhypotes H_0 och en mothypotes H_1 . Jag har under många år arbetat med statistisk skattning av felintensiteter för grupper av tekniskt likartade säkerhetskomponenter i nordiska kärnkraftverk. För mig är det naturligt att anta, trots komponenternas tekniska likhet, att varje komponent i gruppen har sin individuella felintensitet men ändå står i en nära relation till andra komponenter i gruppen; detta är en hypotes om inhomogenitet, som vi betecknar H_1 . Här vill jag återigen använda Fredrik Jonssons formulering och säga: vem skulle vilja påstå att felintensi-

teterna är exakt desamma för komponenterna i gruppen (homogenitet, H_0)? Med tanke på olikheter i miljö, underhåll, konditionstester och användning är det väl givet på förhand att mothypotesen (H_1) är den gällande.

Nollhypotesen här är en sammansatt hypotes i och med att felintensiteterna inte har ett givet värde utan endast antas vara sinsemellan lika (homogena), medan H_1 står för motsatsen (inhomogena). Observerade feldata D i form av antalet inträffade fel och drifttid/standbytid för varje komponent i gruppen är den information vi har för att testa hypotesen. Enligt Jeffreys' procedur för bayesiansk hypotesprövning tilldelas hypoteserna H_0 och H_1 a priorisannolikheter $P(H_0)$ och $P(H_1)$. Via den grundläggande likelihoodmodellen, $P(D|H_i)$, beräknas posterorisannolikheterna $P(H_0|D)$ och $P(H_1|D)$, av vilka den större avgör vilken hypotes som favoriseras. Denna testprocedur har tillämpats för några grupper av komponenter. De grupper som utfallit till H_1 :s favör har karakteriserats av relativt många felhändelser, medan grupper med mycket begränsat antal felhändelser ofta resulterat i $P(H_0|D) \approx P(H_1|D)$, d.v.s. ingen övertygande favorit utav de två modellerna. Många komponentgrupper är av det senare slaget, vilket följaktligen innebär svårigheter att avgöra homogenitetsförhållande inom varje grupp med avseende på komponenternas felintensitet.

Om säkerhetskomponenterna kunde antas vara homogena i varje grupp skulle det innebära stora förenklingar i den fortsatta hanteringen av felintensiteter inom kärnkraftverkens probabilistiska säkerhetsanalyser. Därför finns det stora intresset för frågan bland säkerhetsanalytikerna.

KURT PÖRN, PÖRN CONSULTING

SVAR FRÅN FREDRIK JONSSON

Tänk er situationen att ni lyssnar på nyheter på radio, uppfattar en rapport om resultatet av två väljarsympatiundersökningar som utfördes med en månads mellanrum, samt att en viss ökning var statistiskt säkerställd. En rimlig tolkning av det senare vore väl: "det är alltför osannolikt att urvalsosäkerheten orsakat hela denna ökning"?

Vad menas då med att "urvalsosäkerheten orsakat hela denna ökning"? Det naturliga är att utgå från situationen att det i verkligheten i princip inte skett någon förändring, eller ännu mer extremt, från att det i verkligheten skedde en minskning. Från detta kan man uppskatta sannolikheten att urvalsosäkerheten orsakar en minst lika extrem ökning.

Efter att man enats dels om en kritisk nivå (säg 5%), dels om att agera spegelvänt vid minskningar, så har man en automatiserad procedur. Proceduren säkerställer uttalanden om minskningar och öknings. Sannolikheten att säkerställda uttalanden är felaktiga begränsas av den kritiska nivån 5%.

Nu kan man fråga sig vad Rolf Sundberg, Ulf Jorner och Lars Ängquist har för argument för att förkasta ovanstående resonemang. Jo, de utgår från den så kallade nollhypotesen om att det i verkligheten inte skett någon förändring och påpekar, vilket jag håller med om, att sannolikheten då är 10% att avge ett uttalande om en statistiskt säkerställd förändring.

Mitt motargument är, och har hela tiden varit, att när det gäller uttalanden om populationen så finns det i detta fall ingen anledning att skilja på något annat än "ökning" och "minskning". Det döljer sig inget logiskt fel, och ingen orimlighet, i att utgå från att något av svaren alltid är korrekt. Gör man detta begränsas sannolikheten för ett felaktigt, statistiskt säkerställt uttalande av nivån 5%. Om man inte accepterar utgångspunkten att det antingen skett en ökning eller minskning får man en krångligare procedur att erbjuda till allmänheten.

Vilken relevans har då ett signifikanstest i denna situation? Jag skulle säga, först och främst som en varningsklocka som signalerar ifall informationen inte är övertygande i frågan om det skett en ökning eller en minskning. Ur trendkänsliga radiolyssnarens perspektiv, och därmed kvalitetsmedvetna produktansvarigas perspektiv, torde detta vara relevant. Samtidigt torde

det vara mindre relevant med vilken signifikansnivå man kan förkasta det man kallar för "att inget har hänt".

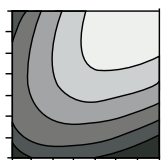
Lars Ängquist är inne på rätt spår då han diskuterar multipla test. Metoden jag föreslår kan nämligen identifieras som ett dubbelt signifikanstest, för ökning mot minskning, samt minskning mot ökning. Vad som skiljer oss åt är att Lars Ängquist (och Rolf Sundberg) menar att hypoteserna måste överlappa varandra i ett så kallat oförändrat tillstånd.

Möjligen kan man vara missnöjd över att kopplingen mellan signifikanser och konfidensintervall bryts i och med mitt förslag (vilket Ulf Jorner är inne på). En eventuell tröst är att felmarginal (konfidensgräns) istället blir det centrala begreppet. Vill man säkerställa en ökning förefaller det naturligt att subtrahera en felmarginal från den skattade ökningen och se efter om värdet fortfarande är positivt. Vid ett nekande svar brukar man säga att "ökningen ligger inom felmarginalen". Med det vanliga synsättet baseras felmarginalen på idén om en 2,5%-kvantil, eftersom det finns två sätt på vilket man kan svara fel under nollhypotesen. Jag föreslår däremot 5%-kvantilen, eftersom det i mitt synsätt bara finns ett sätt att svara fel.

Avslutningsvis, måste inte ett signifikanstest som helt avstår från icke-signifikanta svar ha signifikansnivå 100%? Ulf Jorner tycker att det nekande svaret är orimligt. Jag skulle säga att om man i frågan ökning/minskning arbetar helt utan felmarginaler så är felrisken 50%. Med andra ord, om man gör ett uttalande om en ökning så fort skattningen visar en aldrig så liten ökning så är den maximala felrisken 50%. Om den egentliga skillnaden är väldigt liten, så är det nämligen hugget som stucket i vilket av stickproven som andelen kommer att vara störst. Om skillnaden däremot är större, så är det även större sannolikhet att informationen leder en till rätt svar i frågan. Mitt svar på frågan blir alltså nej (med en mer allmän definition), och att det inte ligger något orimligt i detta.

Allra sist vill jag tacka Ulf, Rolf, Lars och allra övriga som tagit sig tid att diskutera detta!

FREDRIK JONSSON, UPPSALA UNIVERSITET



Osäkerhet

Kan man bli mer osäker än när man får uppdraget att skriva en introduktion till Föreningen industriell statistiks kommande seminarium om just osäkerhet. För osäkerhet är väl något som alla känner emellanåt, kanske främst inför okända händelser och storheter. Framför allt om dessa storheter är personer, men kanske inte lika ofta om de är matematiska storheter. För då hamnar vi ju på säker grund, vi har ju faktiskt mätt storheten, med fyra decimaler. Så då måste den ju vara säker, speciellt om vi avrundar bort några decimaler!

Men är det så enkelt, även för mätningar av storheter gäller detsamma som i verkliga livet. Vissa personer är osäkra inför samtliga händelser och möten, medan andra inte är det minsta osäkra inför dessa situationer. På samma sätt finns det de som hela tiden ser en osäkerhet i sina mätningar och andra som aldrig ser någon osäkerhet, utan mätt är mätt och detta har vi fått, punkt, slut. Det fascinerande här är att ibland verkar detta stå i proportion till hur tekniskt avancerad utrustning man använder, ju mer tekniskt avancerad, desto mer säkra resultat. Jag har faktiskt stött på personer som, efter att ha kämpat och bokat med mjuka variabler på låg skalnivå och sedan äntligen hittat en kvantitativ variabel mätt med någon avancerad apparat, på något sätt lyckas förtränga att även denna variabel har ett mätfel och en osäkerhet. Kanske är detta mänskligt, äntligen hittar vi något konkret som dessutom går att räkna på utan att vi behöver göra alltför vidlyftiga antaganden och där vi kan tillämpa relativt vanliga effektiva metoder.

En annan förklaring kan vara att vi många gånger har nog problem med variationen i de enheter vi mäter och därmed glömmer bort, eller bortser från, mätmetoden som genererat värdena. Men tänk om variationen i mätmetoden är större än variationen mellan enheter vi mäter? Eller om den mätmetod vi använder inte mäter lika från gång till gång?

Fast, om den inte mäter lika från gång till gång kan man ju alltid kalibrera. Jämföra med något som är sant och sedan skruva på någon skruv, så löser det sig nog. Sen får man ju bara hoppas att det man använt som sanning också verkligen är sant och inte bara sant här och nu. För som en kollega en gång uttryckte det: "Sanna värden skall vara sannolika, samt helst troliga och i alla fall möjliga". Det gäller att inte vara osäker!

Så hur hanterar man då osäkerhet och hur kan man kalibrera?

Lite hjälp på traven kan man kanske få genom att delta i Industriell statistiks vårseminarium, Mätosäkerhet och kalibrering, i Örebro den 23 mars. Viss osäkerhet finns dock alltid.

LARS SÖDERSTRÖM FÖR INDUSTRIELL STATISTIK



»Sanna värden skall vara sannolika, samt helst troliga och i alla fall möjliga»



Snart är det dags för årsmöte och det innebär att jag har varit ordförande i ett år. Det har varit en väldigt positiv upplevelse och en nyttig erfarenhet. Allra roligast är att skapa kontakter mellan statistiker från olika arbetsplatser och inriktningar (inom medicin men också lite bredare). Detta gör vi bland annat genom att ordna aktiviteter i form av möten och konferenser till en relativt liten kostnad som lockar ut oss från våra respektive kontor.

Vi är en ganska liten yrkeskår och det finns goda förutsättningar för en medlem i FMS kan känna till majoriteten av de övriga medlemmarna. Målet är också att vi ska främja utvecklingen inom medicinsk statistik och förhoppningsvis väcks funderingar och föds en del idéer under våra möten. I samband med FMS årsmöte den 23/3 på universitetet i Örebro kommer vi att ha ett vårmöte där två av 2010 års stipendiater framträder. Vi har också bjudit in några lokala föreläsare som arbetar med vitt skilda områden, samtliga av intresse för statistiker som jobbar inom medicin och hälsa.

Vi kan också se fram emot en spännande aktivitet i maj då vi arrangerar ett möte med ett synnerligen aktuellt tema, nämligen "Model Based Drug Development". Mötet kommer hållas på KTH i Stockholm och SSL är medarrangör. SSL står för "Statistikot Suomen Lääketeollisuudessa r.y." och är alltså, för er som behöver hjälp med översättningen, en förening för statistiker i den finska läkemedelsindustrin. Även representanter från KTH ingår i programkommittén. Håll utkik efter inbjudan och mötesprogram via e-post.

Höstkonferensen på AstraZeneca i Södertälje den 29/11 2010 lockade drygt 60



ORDFÖRANDEN HAR ORDET

3 tsk framtid och 1 msk dåtid

deltagare från läkemedelsföretag, konsultföretag, Läkemedelsverket, forskningscentrum, universitet och högskolor. Under rubriken "Multiplicity, why, when, where and how?" presenterades och diskuterades olika angreppssätt från skilda tillämpningsområden. FMS arrangerade konferensen tillsammans med Cramér-sällskapet. Huvudtalare var Yudi Pawitan som är professor vid Institutionen för medicinsk epidemiologi och biostatistik på Karolinska Institutet. Yudis föredrag var mycket uppskattat och innehöll både bredd och djup.

Yudi påminde oss om att för forskare inom olika vetenskaper betyder fler variabler bättre beskrivning, mer information och större förutsättningar för upptäckter. Ju mer desto bättre. För en statistiker innebär fler variabler mer brus och extra utmaningar. Framförallt är det tydligt vid hypotesprövning (flera tester = ökad möjlighet för felaktig förkastelse) men också för skattningar (fler skattningar = större sannolikhet för extrema resultat) och regressionsmodeller där det gäller att välja vilka av alla variabler som ska ingå.

Yudi visade flera exempel och förhållningssätt. Inom agronomi används ANOVA-modeller och om ett F-test ger signifikant resultat görs efterföljande analys med jämförelse av gruppmedelvärden (med procedurer som Scheffe, Tukey, Newman-Keul, Duncan, Dunnett etc.). Inom bland annat läkemedelsindustrin finns ett annat förhållningssätt. Där används metoder som garanterar att planerat typ-I fel (alfa-nivå) inte överskrider, t.ex. genom att använda Bonferroni-metoden eller nyare förbättrade varianter. Med andra ord vill man hålla kontroll på "familywise error rate" (FWER). Inom genetik används istället uppskattning av andelen falska förkastelser ("false discovery rate"). Inom epidemiologi

görs inga justeringar för multipla jämförelser. Detta är förvånande nog vedertaget och försvaras med hänvisning till en publikation av den kände epidemiologen Kenneth Rothman (1990).

Yudi tog också upp nya idéer som inte är så strikta som FWER. Han refererade till en helt färsk publikation av Goeman och Solari (2010).

Övriga föreläsare under dagen var Naveed R. Butt från Lunds universitet (som talade om signalmodellering och detektering av landminor, narkotika och förfälskade mediciner), Olivier Guilbaud (AstraZeneca Södertälje) och Carl-Fredrik Burman (AstraZeneca Mölndal). Olivier visade oss först några basala procedurer för multipel testning av en familj av hypoteser såsom fix sekvens utan och med säkerhetslina ("fallback") samt Holms test. Återvinning är ju modernt i flera sammanhang och så även återvinning av typ-I felet ("alpha recycling"). Olivier visade grafiska scheman med trädstruktur som kan användas för att konstruera och illustrera procedurer för alfaåtervinning på ett enkelt sätt. För den som vill fördjupa sig i detta rekommenderas publikationen av Burman, Sonesson och Guilbaud (2009). Carl-Fredrik hade filosofiska funderingar om varför alfa oftast sätts till 5 % och hur detta ska spenderas per experiment/studie.

Vad utgör egentligen ett experiment? Tänk om varje statistiker bara skulle ha dessa 5 % att spendera under hela sin yrkesverksamma tid. Det kan också vara roligt att känna till att det ursprungliga namnet på återvinningsmetoden var kasinometoden men detta ratades av en "referee". Dagen avslutades med Carl-Fredriks "Casino Multiplicity". Alla del-

tagarna fick 50 marker vardera att spendera på ett antal hypoteser. Valfri procedur fick tillämpas. Fina priser delades ut till segrarna.

Presentationerna från konferensen finns på FMS webbplats.

Till sist något om nästkommande år då vi borde hitta på något extra festligt. Eller hur? Varför undrar ni kanske? Det är viktigt att ta alla chanser att fira och visst är ett 25-årsjubileum något som borde uppmärksammas med ett möte med tema historia och framtid eller liknande. FMS bildades 1987 och mycket har hänt sedan starten. I FMS arkiv finns det mesta sparad, bland annat från tider där korrespondensen inleddes med "Bäste

broder" eller "Hej bröder", men om ni medlemmar har några kuriösa fakta eller spännande skrönor så hör av er. En historisk händelse som är värd att notera: under åren 1994-1996 pågick utredningsarbete inom FMS om auktorisering av medicinska statistiker och därefter följde en omröstning där auktorisering

blev nedröstad. Auktorisering av statistiker är något som blivit aktuellt igen eftersom Främjandet nyligen tillsatt en utredningsgrupp i ämnet.

Mer från FMS på nästa sida – en reseberättelse skriven av Sandra Eloranta, en av våra FMS-stipendiater 2010. Vi ses på årsmötet!

»Till sist något om nästkommande år då vi borde hitta på något extra festligt.«



ANNA TORRÅNG

SOL, SAKNADE VÄRDEN OCH KONKURRERANDE RISKER I BRASILIEN

I likhet med ca 800 andra deltagare från olika delar av världen för jag i mörkaste december till Florianopolis i Brasilien, där det 25:e mötet med International Biometric Conference (IBC) gick av stapeln.

IBC arrangeras vartannat år. Den syftar till att uppmuntra utvecklingen av statistisk samt matematisk teori inom biovetenskaperna. Som arrangör står The International Biometric Society (www.tibs.org). Florianopolis ligger i södra Brasilien och har ca en halv miljon invånare. I staden kan man, förutom att sola och bada, studera vid det federala universitetet Santa Catarina och det var i universitetets lokaler som vi alla hade samlats för 2010 års IBC. Söndagen innan mötets öppnande gavs fyra endagskurser om bland annat modellering av konkurrerande risker, helgenom-associationsstudier, multipel imputering och dess tillämpningar samt statistik för biologiska nätverk.

Jag valde den förstnämnda kursen som presenterades av Ronald Geskus och Hein Putter från Nederländerna. Programmet var fullspäckt men kursmaterialet var oerhört välplanerat och pedagogiskt framlagt. Sammantaget var den första dagen mycket lyckad. Under konferensen hölls fem sessioner parallellt varav en eller två var sessioner där 3-4 personer gemensamt satt ihop en serie längre föredrag i syfte att bättre kunna sammanfatta utvecklingen inom ett specifikt område. Som exempel avhandlades genombrott i longitudinella studier, statistiska problem i vaccinationsstudier, multivariabel modellering av kontinuerliga variabler med hjälp av olika typer av splines, saknade värden i kliniska försök samt utmaningar inom multipla genbaserade analyser

i helgenom-associationsstudier. Jag ägnade den största delen av tiden åt de sessioner som var relaterade till överlevnadsanalys, som är det område som jag själv är verksam inom. Mest intressant med konferenser av den här typen tycker jag dock är att man påminns om hur oerhört brett statistik är som ämne. Som doktorand är det lätt hänt att man efter en tid "glömmer bort" att det finns en värld utanför det egna fältet. Den världen uppdagades dock allteftersom konferensen fortskred och med mig hem har jag således en hel lista på metoder, terminologi mm som jag känner att det är dags att studera lite närmare vid tillfälle.

De övriga sessionerna bestod av något mer traditionsenliga kortare inslag inom en lång rad olika områden, alltifrån traditionella statistiska teman som tidsserieanalys, bayesianska metoder, stokastiska processer, överlevnadsanalys och kausal inferens till något mindre välkända områden (åtminstone för någon som kommer från den epidemiologiska världen) såsom djurvetenskap, tillämpningar inom jordbruksvetenskap och fiskeriforskning. Bakgrunden till mitt eget deltagande i konferensen är att jag hade fått möjligheten att presentera en av mina egna studier som jag arbetar med inom mitt doktorandprojekt. Den studien syftar till att vidareutveckla statistiska metoder som traditionellt används inom registerbaserade epidemiologiska studier. Med de metoderna kan man trots att det föreligger konkurrerande risker modellera och skatta behandlingsrelaterad hjärt- och kärlsjukdomsdöd t.ex. efter en genomgången bröstcancerbehandling. Som bidrag hade jag under 2010 dessutom mottagit FMS stipendium som syftar till att uppmuntra unga (i rent akademisk mening!) statistiker att utvecklas professionellt.

Efter två dagars intensivt springande mellan olika föredrag och postrar bjöds deltagarna på en ledig onsdag. Dagen till ära hade en hel uppsjö av aktiviteter förberetts för dem som kände sig manade att delta i vandringsturer i den vackra naturen eller att åka båt i den brasilianska skärgården.

Under veckan firades även IBC:s 25-årsjubileum då vi fick lyssna till en trevlig sammanställning av anekdoter från tidigare möten. Bland annat hade arrangörerna bett dem som var med på 60- och 70-talen att skicka in sina starkaste, och ofta underhållande, minnen från de tidigare mötena. Vi fick bland annat höra om kontroverser som uppstått som en följd av föredragsdiskussioner och statistiker som tagit emot IBS hederspris och som under tacktalet vägrat att prata om annat än sin passion, hierarkiska modeller.

Den bästa anekdoten måste ändå ha varit den om de statistiker som på väg hem från ett IBC-möte i Sydamerika i slutet av 70-talet hade råkat hamna på ett överbokat flygplan. Istället för att vänta på nästa flyg erbjöds en av personerna att sitta framme med piloten eftersom andrepiloten aldrig dök upp. Det är bara att inse att något sådant troligtvis aldrig kommer att hända igen. Sammanfattningsvis kan jag konstatera att de lokala arrangörerna i Brasilien lyckades med att bjuda oss på en trevlig konferensvecka med många intressanta och roliga möten, diskussioner kring statistik och dess tillämpningar och framförallt...SOL.



SANDRA ELORANTA,
KAROLINSKA INSTITUTET



Ordföranden har ordet

Jag skriver detta under julhelgen och har just hört julevangeliet: "Vid den tiden utfärdade kejsar Augustus en förordning om att hela världen skulle skattskriva sig" (Luk 2:1).

Skattskrivningen var en tidig form av folkräkning om än inte den äldsta. Enligt den romerske historikern Josefus ägde den första rum i Palestina år 6 e.Kr. Så har det fortsatt. Nu är det FN som bestämmer att hela världen skall räknas vart tionde år. Länderna kan välja att göra 2010 års räkning under 2009, 2010 eller 2011. I hela EU görs folkräkningarna under 2011.

I Sverige bygger den för första gången helt på registerinformation och planeras till slutet av 2011 med namnet Hushålls- och Bostadsräkning, HoB. De flesta andra länder har redan genomfört sin räkning och i många länder sker det genom att alla hushåll besöks och varje person räknas. I Kina t ex skulle alla hålla sig hemma en dag i höstas och 6 miljoner personer gick runt och räknade, dvs fler än hela Sveriges befolkning i arbetsför ålder.

Vid de senaste svenska folkräkningarna (t ex FoB 85 och 90) var debatten minst sagt livlig och räkningen till och med obstruerades. Men det var inte heller något nytt. Enligt Josefus förekom svåra oroligheter i samband med skattskrivningen vid Jesu födelse.

Till och med så svåra att Josef tog med sig sin familj och flydde till Egypten "ty Herodes

kommer att söka efter barnet för att döda det" (Matt 2:13). Vi får hoppas att årets räkning blir lugnare.

Ett bra befolkningsregister är en förutsättning för rättvisande statistik. Speciellt viktigt är det att ha en god ram för urvalsundersökningar med bra bakgrundsvariabler, t ex för att kunna belysa bortfallsfelet. En viktig variabel är hushållens sammansättning. För att räkningen skall gå att genomföra med register krävs att alla inom samma hushåll är folkbokförda på en och samma adress och i samma lägenhet. Mot bakgrund av deras betydelse kommer vi att ägna årets årsmöte åt folkräkningar och deras användningar. Bland annat hoppas vi ta upp hur folkräkningar gjordes förr i tiden, hur de kan användas i forskning och en utblick mot framtiden, gärna med tonvikt på hur de kan användas inom surveyundersökningar.

Enligt julevangeliet blev alltså Josef, Maria och Jesus politiska flyktingar i Egypten. Flyktingar och alla andra som flyttar över gränserna innebär i dag en stor utmaning vid surveyundersökningar och leder till många olika problem. Många behärskar inte språket i sitt nya land vilket gör att de lätt blir bortfall t ex vid RDD-undersökningar och även om undersökaren i en postenkät har försökt översätta blanketten är det inte alltid lätt att få rätt valörer på orden t ex i en attitydfråga. Det ligger också nära till hands att t ex svenskar

utomlands eller utlänningar i Sverige inte ser sig som representativa och därför avstår från att svara.

Folk som flyttar över gränserna utgör en ganska heterogen grupp från politiska och ekonomiska flyktingar till studerande och tjänstemän i multinationella koncerner, men det kan i många fall vara speciellt intressant hur snabbt de etablerar sig i sitt nya land. Ett problem i det sammanhanget är att det tar tid att komma in i folkbokföringen och därmed i urvalsramarna. Vid en longitudinell undersökning som arbetskraftsundersökningen kommer några i urvalet från en ram som är nästan tre år gammal. En viktig anledning till folkräkningar ansågs förr vara att uppdatera och kontrollera gamla uppgifter. Kinesernas metod med att begära att alla håller sig hemma och sedan gå från hus till hus och räkna är ett sätt att hitta alla men i en registerräkning måste kvaliteten kontrolleras på annat sätt.

Slutligen vill jag påminna alla om sommarens surveykonferens i Norrfällsviken vid Höga Kusten den 13-17 juni (se kalendariet på nästa sida). Jag tycker själv att programmet ser mycket intressant ut. Där kommer några av världens stora inom samplingområdet att tala, men ämnena blir inte bara ren sampling utan även mycket annat inom surveyområdet.

DANIEL THORBURN



FOTO: MARIA SANDOVIK



New York.



Örebro.



Lissabon.

KALENDARARIUM

DETTA HÄNDER 2011

	KONFERENS	PLATS	WEBBPLATS
23-24 mars	Årsmöteskonferenser, Svenska statistikfrämjandet	Örebro, Sverige	www.statistikframjandet.se
5 maj	SCB:s forskardag 2011	Stockholm, Sverige	http://www.scb.se/Pages/Form____307708.aspx
6-9 juni	3rd Nordic-Baltic Biometric Conference	Åbo, Finland	www.nbbc11.utu.fi
13-17 juni	The 3rd Baltic-Nordic Conference on Survey Statistics (BaNoCoSS)	Norrfällsviken, Sverige	www.mathstat.helsinki.fi/msm/banocoss/2011/
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/
15-17 aug	SCB:s och Örebro universitets sommarskola, Small Area Estimation	Örebro, Sverige	www.oru.se/hh/summerschool2011
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	European Network for Business and Industrial Statistics, 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.

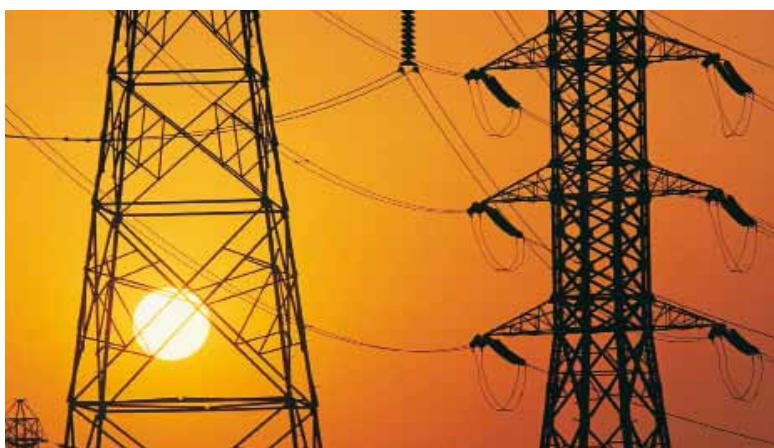
Energy Policy

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.

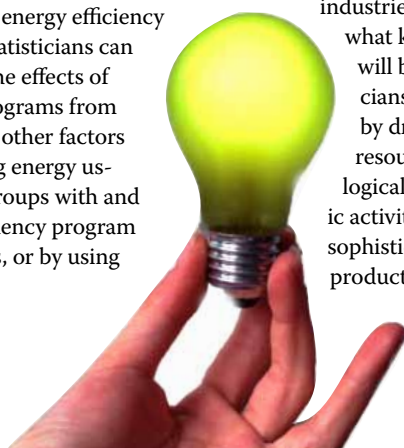
PROJECTING FUTURE ENERGY SUPPLY AND DEMAND:

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.



NYTT UTSEENDE OCH UTÖKAT INNEHÅLL

Välkommen till Statistikfrämjandet på nätet

Svenska Statistikfrämjandets webbplats – statistikframjandet.se – har fått ett modernare utseende och ett utökat innehåll.

Här hittar du aktuell information om kommande evenemang, föreningsnyheter, den etiska koden och mycket annat.

Via Statistikfrämjandets webbplats når du också våra medlemsföreningar: Industriell statistik, Surveyföreningen, Föreningen för medicinsk statistik och Cramérsällskapet.

Välkommen!

