

Qvintensen

NR 1 2019



Om ålders-
bedömningar
och att lära sig
av misstagen

SIDORNA 9-12

Ranking av
universitet

SIDORNA 14-15

VALBAROMETRAR
Har mätningarna
blivit bättre eller
sämre?

SIDORNA 16-17

Är statistikforskning
industristyrd?

SIDORNA 2-5

A photograph of a forest with sunlight streaming through the trees, creating a dramatic, hazy atmosphere. The sunbeams are visible as bright rays cutting through the dark, misty air. The forest floor is covered in moss and fallen branches. The trees are tall and slender, with some showing signs of decay or being dead. The overall mood is serene and somewhat mysterious.

Skogens påverkan på nutida statistisk forskning är direkt genom den starka statistiska traditionen inom skogsforskningen, nuförtiden vidgad till fjärranalys, markundersökning mm, som sammanställs till uppföljning av nationella miljömål.

Är statistikforskningen industristyrd

– ELLER ÄR DET TVÄRTOM?

Statistikfrämjandet arbetar ju energiskt med att föra ut statistiska metoder och statistiskt tänkande i olika områden, inte minst industri. I denna artikel vill vi vända på steken och ge några exempel där ett industriproblem direkt har påverkat uppkomsten eller inriktningen av en statistisk forskningsgren. Det finns flera exempel på när initiativet till ny forskning har kommit från en industrigren, och där resultatet så småningom blivit till en generell statistisk metod eller teori med tillämpning långt utanför den ursprungliga sfären.

Vi har identifierat fyra industrigrenar, försäkring, skog och skogsindustri, teleindustri, samt bilindustri och annan transportindustri, där specifika industriella problem har lett till ny och ibland oväntad statistisk forskning. Vi har utelämnat den stora läkemedelssektorn, vilken förtjänar en egen behandling. Våra fyra områden har haft mycket stor betydelse både för svensk statistisk forskning och forskarutbildning och för utbildning på grundnivå, inte minst inom tekniska utbildningsprogram.

Vi har också identifierat några gemensamma drag för de fyra exemplen, som är högst aktuella för statistisk forskning idag: pensioner och försäkring av alla slag, klimat- och miljöförändringar, global ekonomi, fokus på säkerhet och riskbegränsning. Fjärranalys av det globala havsklimatet och dess förändringar leder till bättre regler för sjöfartssäkerhet, banker försäkras sig mot finansiella risker och försäkringsbolag anpassar sig till nya klimatrisker, trafikfärdigheten ökar genom bättre modeller för trafik- och vägmiljö och av förarbeteende. Statistisk vetenskap har skarpa verktyg!

Statistik och försäkringsindustrin

1881 anställde Stockholms nya högskola sin förste professor i matematik, Gösta Mittag-Leffler. Den nye professorn var med och skapade livförsäkringsbolaget Victoria, och blev dess chefsaktuarie. År 1904 var han en av grundarna av Svenska Aktuariereningen. Några av Mit-

tag-Lefflers studenter (främst Ivar Fredholm och Edvard Phragmén) var också aktiva i försäkringsindustrin. På grund av Mittag-Lefflers inflytande på svensk matematik (i synnerhet i Uppsala och Stockholm) blev försäkringsmatematik en viktig gren av tillämpat matematiskt arbete.

I Uppsala tog Filip Lundberg år 1898 en fil. lic.-examen med högsta betyg i matematik och astronomi. Han ville skriva en doktorsavhandling för att kunna bli gymnasielärare. För att försörja sig tog han anställning som kamrerare och matematiker i det nygrundade Allmänna pensionsförsäkringsbolaget. Efter några år blev han anställd av folkförsäkringsbolaget De förenade, vars verkställande direktör han blev 1905. Han disputerade 1903 med en avhandling om kollektiv riskteori, där han bl.a. härledde det som senare kallades Kolmogorovs bakåtekvation för en sammansatt Poissonprocess. Avhandlingen ansågs mycket svårsläst, inte minst för att stokastiska modeller inte tidigare hade använts på detta sätt i försäkringsmatematiken.

Harald Cramér, matematikstuderande på Stockholms högskola, disputerade på en avhandling om Dirichlet-serier i maj 1917. Följande år fick han, med hjälp från Ivar Fredholm, en anställning på Försäkringsinspektionen, där han stötte på Lundbergs avhandling, som skulle bli ett viktigt led i hans försäkringsmatematiska forskning. I slutet av 1919 blev han erbjuden en tjänst som chefsaktuarie på Svenska Livförsäkringsaktiebolaget. Han accepterade på villkor att han skulle få syssla med matematisk forskning en dag i veckan.

Cramér (1926) sammanfattade Lundbergs arbete i en (ovanligt lång) recension, där han beklagade att det inte slagit genom i industrin. Kanske berodde detta på Lundbergs sätt att skriva matematisk text, vilket Cramér beskrev som att "gå till sydligare luftstreck, till de tropiska urskogarna, där forskaren endast mödosamt, steg för steg banar sig fram genom snårerna, utan

att kunna se långt vare sig fram eller tillbaka. En expedition i urskogen kräver månader, kanske år, men den lönar sig! Röjer man sig väg fram genom snår och motigheter i den Lundbergska teorien, så tvingas man så småningom – kanske motvilligt – att erkänna befintligheten av system i den skenbara oredan. Man går längre, och utsikten vidgas, och slutligen fattas man av en livlig beundran för kraften och originaliteten i den hjärna, som skapat hela denna värld."

1925 började försäkringsindustrin försöka skapa en professur i försäkringsmatematik och matematisk statistik, något som hade diskuterats länge (inte minst av Mittag-Leffler). Fyra år senare utsågs Cramér till professor vid Stockholms högskola. Cramér slutade då i Svenska Liv, men blev i stället aktuarie i Återförsäkringsbolaget Sverige, de svenska livförsäkringsbolagens gemensamma organ för samarbete. Försäkringsindustrin bekostade inte bara professuren, utan byggde också en institutionsbyggnad, som blev färdig 1940.

Cramérs främsta insats i försäkringsmatematiken var teorin om stora avvikelser, som uppkom från att försäkringsbolag sysslade med summor av stokastiska variabler som inte nödvändigtvis var av samma storleksordning. Detta var också en av orsakerna att han studerade hur fördelningen av sådana summor avviker från normalfördelningen.

Ett viktigt problem i försäkringsmatematiken är att bestämma explicita uttryck för ruineringsproblemet, alltså när bolagets förmögenhet först blir negativ. Cramér (1930) skrev ner en Volterra-ekvation för ruinerings sannolikheten vars Laplace-transform ger den kända Pollaczek-Khintchin-formeln (denna borde alltså kallas Cramér-Pollaczek-Khintchins formel). Flera andra medlemmar i Cramérs institution arbetade på ruinerings sannolikheter, t.ex. Segerdahl, Täcklind och Arfwedsson.

Flera av Cramérs doktorander skrev sina avhandlingar om försäkringsrelaterade problem. Carl-Otto Segerdahl skrev om kollektiv riskteori, Ove Lundberg behandlade stokastiska processer »»»

»»» ÄR STATISTIKFORSKNINGEN INDUSTRISTYRD – ELLER ...?

i sjukdoms- och olycksstatistik, och F. Malmqvist skrev om livförsäkring under minskande räntetryck.

Statistik och skogsindustrin

Svensk skogsindustris påverkan på statistisk forskning personifieras naturligtvis av Bertil Matérn, professor i Skoglig matematisk statistik vid Skogshögskolan från 1963 till sin pensionering 1981. Hans doktorsavhandling från 1960, "Spatial variation", är en standardreferens för korrelation i planet, rummet och högre dimensioner. Men innan vi fördjupar oss i skogens påverkan på Matérn och statistisk forskning måste vi nämna en annan inflytelserik person, jägmästaren (1925), skogsvetenskapsmannen och landshövdingen i Norrbottens län, Manfred Näslund (1899-1988).

Under sina jägmästarstudier vid Skogshögskolan tog Näslund kontakt med Harald Cramér, som då var docent vid Stockholms högskola, och lärde sig med hans hjälp den tidens bästa statistiska metoder, bland annat regressions- och multivariatanalys, kunskaper som han utnyttjade i flera förslag till förbättrad uppskattning av Sveriges skogstillgångar, "Riksskogstaxeringen".

Efter att ha varit sekreterare och "skogsstatematisk expert" för 1937 års riksskogstaxeringsnämnd blev Näslund 1944 chef för Statens skogsförsöksanstalt (från 1945 skogsforskningsinstitut). För riksskogstaxeringen inrättades en statistisk avdelning med Näslund som chef. För att få råd om bemanningen vände sig Näslund till Harald Cramér, som föreslog sin amanuens, Bertil Matérn.

Matérn anställdes vid skogsforskningsinstitutet 1945 och fick av Näslund som sin huvuduppgift att utreda noggrannheten vid linje- och provytetaxering av skogsbeståndet. Ett speciellt problem var hur beroendet mellan mätpunkter skulle modelleras – att mäta träd i en skog utan ett tätt vägnät är tidsö-

dande och dyrt! Matérn var väl förtrogen med teorin för stokastiska processer, korrelationsfunktioner och spektralframställning, ämnen som förekommit flitigt vid seminarierna vid Cramérs institut. I sin licentiatavhandling 1947 använde Matérn en blandning av två exponentialfunktioner för att beskriva korrelationen i den topografiska variationen inom ett geografiskt område. Det var ett tidigt exempel på spatial statistik, som Matérn använde för att hitta lämpliga avstånd mellan taxeringslinjer och mellan provytor och på det sättet få god noggrannhet i skogsuppskattningen.

Manfred Näslund överlämnade 1949 chefskapet för skogsforskningsinstitutets statistiska avdelning till Bertil Matérn, som kunde dela sin tid mellan konkreta skogsstatistiska problem, och att utveckla licentiatavhandlingen till en banbrytande modell för spatial variation. På 1950-talet utvecklades teorin för stationära stokastiska processer i försiktiga steg från en-dimensionella modeller, till modeller för två, tre och högre dimensioner. Den modell Matérn använde sig av är "isotrop", dvs. korrelationen beror bara på avståndet mellan mätpunkterna, inte på riktningen mellan dem.

Varje isotrop korrelationsfunktion är en blandning av s.k. Besselfunktioner, och det Matérn gjorde i doktorsavhandlingen var att välja blandningsfunktionen så att korrelationen fick en enkel generell form, användbar i alla dimensioner. Klassen av Matérn-kovariansfunktioner är numera en rekommenderad standard i många geostatistiska sammanhang. Den går även under namnet Whittle-Matérn-klassen, eftersom Peter Whittle (1954, 1963) visade att den beskriver korrelationsstrukturen för lösningarna till en speciellt enkel stokastisk partiell differentialekvation. Ett plus är att det är lätt att göra modellen icke-stationär, simulera och skatta parametrar.

Skogens påverkan på nutida statistisk forskning är både direkt och indirekt. Direkt, genom den starka statistiska traditionen inom skogsforskningen, nuförtiden vidgad till fjärranalys, markundersökning mm, som sammanställs till uppföljning av nationella miljömål. Den indirekta effekten

ligger i Matérn-kovariansfunktionen, som underlättar modellering av omfattande geostatistiska data, långt utanför skogssektorn.

Statistik och teleteknik

Den amerikanske uppfinnaren Alexander Graham Bell tog patent på telefonen 1876. Patentet gällde inte i Norden, så Lars Magnus Ericsson kunde sälja sin egen variant av "den talande telegraf", som Dagens Nyheter kallade den.

I Köpenhamn hade International Bell Telephone Company ett dotterföretag som 1908 anställde matematikläraren Agner Krarup Erlang. Erlangs första uppgift blev att fundera ut hur många linjer som behövs i en växel. Han härledde att samtalen följer en Poissonprocess (under rimliga antaganden som han jämförde med data från en Köpenhamnsväxel). Metoden var likartad den som försäkringsmatematikern Filip Lundberg hade använt i sin avhandling några år tidigare, men det är inte troligt att Erlang kände till Lundbergs arbete.

I Stockholm hade Ericsson börjat samarbete med en konkurrent till Bells dotterföretag och snart hade de tagit över marknaden. 1918 slogs de två bolagen hop (Bell-företaget hade de redan köpt upp) under namnet Allmänna Telefonaktiebolaget LM Ericsson.

Teleingenjören Conny Palm arbetade på LM Ericsson. Han doktorerade vid KTH, men medverkade också i Harald Cramérs sannolikhetsseminarium på Stockholms universitet på slutet av 30-talet. Cramér beskriver Palm i sina minnen som en originell telefontekniker med unika idéer inom sannolikheteorin. Palm betraktade telefonsamtal i en växel som en punktprocess, ett namn som han själv funnit på (Stockholmsgruppen, som också innehöll Willy Feller, var mycket kreativ när det gällde namn på processer). Hans avhandling, skriven på tyska, handlade om intensitetsvariationer i telefontrafik och publicerades av LM Ericsson 1943. Teorin baserades på något som han kallade ekvilibriumpunkter (numera kallas de regenereringspunkter), som är sådana att processen börjar om, oberoende av vad som skett före den punkten. Detta generaliserar Poissonprocessen, som inte alltid passar särskilt bra till telefonidata.

Ett av Palms viktigaste resultat var att när man överlagrar många punktprocesser på så sätt att summan av deras

intensiteter konvergerar till en konstant, då är gränsprocessen en Poissonprocess.

Palm samlade in data från Östermalms telefonstation i Stockholm. Han hade börjat samla in data 1937, och i sin avhandling använde han 400 000 samtal från två månader under 1941. Ett stort problem när man samlade in telefonidata var att det var svårt att lagra data. Palm byggde apparatur för att beräkna vissa kvantiteter, men om han velat observera något annat så gick det inte. Därför började han fundera på matematikmaskiner, och 1950 ledde han konstruktionen av BARK, Sveriges första dator.

Telefontrafik är inte bara telefoner och telefontrafik utan också telefonsystem, kodning och signalbehandling. Lars Henning Zetterberg (1925-2017) var professor i Teletransmissions-teori, 1964-1965 i Lund och därefter under 25 år vid KTH. Han försåg signalbehandlingen med statistiska verktyg, inspirerade bl a från Bell Laboratories i USA och från Cramérs stockholmsinstitution. Tre namn från Bell Labs, Steve Rice, Claude Shannon och David Slepian, samt Ulf Grenander från Stockholm, är med bland referenserna i Zetterbergs arbeten i början av 1960-talet. Hans elever förde den statistiska traditionen vidare inom telekommunikationsområdet, något som starkt påverkade den statistiska utbildningen och forskningen vid de tekniska högskolorna.

Statistik och transporter till lands, till sjöss och i luften

Utmattning kan drabba människor efter långvarig stress. Bilar, båtar och flygplan kan drabbas av utmattning när de utsätts för långvarig stress i form av varierande belastning. Begreppet "random vibration" blev högaktuellt på 1950-talet, bland annat efter några spektakulära flygkatastrofer. Världens första jet-drivna passagerarflygplan, De Havilland Comet I, sattes i kommersiell trafik 1952. Efter två totalhaverier 1954 konstaterades att utmattningssprickor som startat runt de fyrkantiga fönstren hade lett till att hela flygkroppen bröts sönder.

Fenomenet utmattning var välkänt långt innan Comet-katastroferna. Dåtidens värsta järnvägsolycka ägde rum i Versailles 1842 och orsakades av ett utmattningsbrott i en lokomotivaxel. Andra drabbade konstruktioner har varit stålbroar, trycktankar, fartyg, och på senare år oljerigg. En nyhet under 1950-talet var att

den varierande belastningen beskrevs statistiskt som en stokastisk process med fördelningar, korrelationer och energispektrum.

Den statistiska inspirationen kom genom två artiklar av den svensktättade elektroingenjören Steve O. Rice vid Bell Laboratories: "Mathematical analysis of random noise", 1944, 1945. Artiklarna beskrev brus i telekommunikationssystem, men de statistiska idéerna anammades mycket snabbt av andra teknik- och industriområden. För svensk del gäller det främst lastbilstillverkaren Scania-Vabis i Södertälje och skeppsvärven, bland andra Kockums och Lindholmen, men även flygindustrin tog del av den internationella utvecklingen.

Sverker Sjöström (1920-2005) anställdes 1947 vid Scania-Vabis, som ansvarig för beräkning och provning. Där ledde han mätkampanjer av utmattningsbelastning på lastbilar och bussar och jämförde resultaten med en matematisk-statistisk analys baserad på Gaussiska processer. Han disputerade på KTH 1961 på avhandlingen "On random load analysis", en viktig ingrediens i Scantias framgångar som lastbilstillverkare. Från 1961 var Sjöström teknisk direktör för Scania-Vabis (även Scania och Saab-Scania).

"On random load analysis" kom att få stor betydelse för den matematisk-statistiska forskningen både vid LTH i Lund och vid Chalmers i Göteborg, och hur det blev så är en historia i sig. Opponenten vid disputationen var docenten i hållfasthetslära Bertram Broberg (1925-2005). Han upptäckte ett matematiskt glapp i avhandlingen och han lyckades, efter ett tips från professorn i matematisk statistik, Carl-Gustav Esseen, "nästan" fylla igen glappet. Under ett (obevisat) antagande om ett speciellt oberoende kunde Broberg visa att belastningsväxlingarnas amplitud är Rayleigh-fördelad.

En extra poäng är att Bertram Broberg under 1950-talet arbetade som assistent åt Waloddi Weibull (1887-1979), professor i maskinelement vid KTH sedan 1941. Han hade sedan 1930-talet arbetat med statistiska metoder inom hållfasthet och utmattning. När han 1951 argumenterade för det som nu kallas Weibullfördelningen, gjorde han det med "svagaste länken" argumentet: en kedja brister i sin svagaste länk. Fördelningen möttes med skepsis i matematisk-statistiska kretsar, bl a när han presenterade den för Cramérs stockholmsinstitution (enligt muntlig uppgift från Gunnar Blom, som var närvarande vid tillfället). Waloddi Weibull rönt mycket stor

internationell uppskattning och belönades i Sverige med ingenjörsvetenskapsakademiens stora guldmedalj, bl a för sina vetenskapliga insatser inom hållfasthetsstatistik. Weibull var handledare till Bertram Broberg.

Bertram Broberg blev professor vid LTH 1961 och förde med sig det obevisade oberoendetagandet som ett lämpligt doktorandproblem i matematisk statistik. Georg Lindgren lyckades motbevisa antagandet! Ett negativt resultat, men med oanade konsekvenser: kontakt med Ross Leadbetter i Chapel Hill, extremvärdesteori tillsammans med Holger Rootzén, nya doktorander, till en början Jacques de Maré och Igor Rychlik, och så småningom deras doktorander i matematisk statistik, många i samarbete med både Scania och Volvo. Av de ca 30 doktorer som disputerat arbetar nu närmare 10 som forsknings- och utvecklingsledare inom fordons, marin- och annan industri!

I samband med att Scania 2005 donerade en grundplåt till en Sverker Sjöström-professur i hållfasthetslära vid KTH poängterade Scantias forskningschef vilken stor betydelse Sjöströms avhandling haft för utvecklingen i både hållfasthetslära och matematisk statistik.

Svensk sjöfartsnäring fick en statistisk injektion tidigt på 1950-talet när Chalmers rekryterade Rutger Bennet (1922-2008) att leda arbetet med statistisk behandling av hur fartyg uppträder i havsvågor. Bennet fick många internationella uppdrag inriktade på statistisk analys av vägbelastningar, och flera av hans doktorander blev ansvariga för införande av statistiska metoder inom skeppscertifiering och försäkring.

Sjöfartens påverkan på svensk statistisk forskning är av senare datum och är kopplad till lastanalysen. En mycket viktig faktor är Igor Rychliks analys av de statistiska egenskaperna hos den så kallade Rain Flow metoden att beskriva utmattningscykler. Den väckte intresse även inom marin forskning och ledde med början 1993 till ett samarbete bl a med det franska forskningsinstitutet IFREMER i Brest. Inom ramen för olika europeiska projekt rekryterades internationella doktorander, bl a den kinesiske skeppingenjören Wengang Mao, som efter disputation i matematisk statistik vid Chalmers nu är docent i Marin teknik där. Cirkeln är sluten.

PETER GUTTORP, NORSK REGNESENTRAL
GEORG LINDGREN, LUNDS UNIVERSITET

CERTIFIERING

– på väg mot målet

■ **Frågan om certifiering** av statistiker har aldrig varit mer aktuell eller närmare en lösning! I en värld där antalet påståenden om olika fakta ökar i en skrämmande hastighet, behövs mer kunskap om hur sådana fakta framställs, och hur de ska värderas. Statistik kunskap är här viktigt. Under 2019 kommer ett system för certifiering av statistiker att lanseras av FENStatS – Federation of European Statistical Associations.

FENStatS föreslår en gemensam minsta nivå som kan användas för certifiering – Accredited European Statistician. Certifiering av statistiker innebär att meriter som påvisar kunskap, utbildning och erfarenhet sammanställs och dokumenteras. För att bli certifierad måste den sökande uppfylla krav, som innebär att han/hon kan utföra sitt yrke som statistiker, oberoende och säkert tillsammans med andra statistiker och professioner.

System för certifiering finns redan i drift i USA, Storbritannien, Kanada, Nederländerna och några länder till. Dessa har en gemensam miniminivå med en MSc, 5 års yrkeserfarenhet samt olika dokumenterade exempel på pedagogik, kommunikation, rekommendationer etc.

FENStatS förslag bygger i stora drag på dessa system. De fordringar som föreslås är:

- A – Universitetsexamen, MSc
- B – Fem års yrkeserfarenhet
- C – Pågående aktiv professionell utveckling
- D – Kommunikationsförmåga
- E – Följa etiska standarder
- F – Medlem i en nationell organisation

Målet med en certifiering av statistiker är att skapa en plattform för professionalitet och acceptans inom ett område som blir allt viktigare. Genom en gemensam syn på vad en "statistiker" är ges detta yrke en högre och tydligare status och roll. Certifiering kommer inte att vara obligatorisk – men förhoppningsvis ska värdet av processen överstiga det arbete som krävs.

Systemet som FENStatS föreslår bygger på en gemensam standard som omfattas av alla FENStatS medlemsorganisationer. När en medlemsorganisation har anslutit sig till systemet kan den påbörja certifieringar av de medlemmar som ansöker om detta. Certifieringen blir erkänt likvärdig inom samtliga länder där FENStatS är aktiva. Standardens krav är lagda för att vara i paritet med American Statistical Association och Royal Statistical Society.

Förslaget till nya "by-laws" (förfordningar) kommer att behandlas på FENStatS under våren 2019. Medlemsorganisationerna kommer att kunna ansluta sig redan under 2019.

Arbetet med att förverkliga ett system för certifiering har pågått under många år. Genom gemensamma ansträngningar har FENStatS Accreditation Committee nått fram till en lösning som vi menar kommer att främja statistiken och statistikerna i hela Europa. Målet är nästan nått och systemet är klart att sjösättas under 2019.

MAGNUS PETTERSSON,
FENSTATS ACCREDITATION COMMITTEE



Några få ord om ett tjockt nummer

det här numret har vi för första gången på länge fått en debatt. Det är två artiklar i föregående nummer, som har lockat till kommentarer, vilka därefter bemötts. Det rör sig om olika syn på sannolikhetsurval och bortfallsproblem, huvudsakligen vid partisympatiundersökningar, men även mer allmänt om motsättningar mellan olika surveykulturer. Det har blivit ett ovanligt tjockt nummer.

För att inte i onödan göra numret ännu tjockare väljer jag nu att vara så kortfattad som möjligt. Två iakttagelser av omvärlden har fastnat i mitt minne: (1) I SJ:s gratistidning Kupé ser jag den braskande rubriken "Slumpen är död!". Det är en ekonomiprofessor vid Handelshögskolan i Stockholm som intervjuas. "Finns slumpen över huvud taget?" är intervjuarens första fråga. "Nej!" svarar professorn kort och bestämt. (2) Vid spekulationer om hur det ska gå i en kommande match säger en sportreporter vid Sveriges radio ungefär så här: "Chansen att lag A ska vinna över lag B är osannolikt teoretisk." Fundera över den formuleringen.

JAN WRETMAN

Qvintensen

Ansvarig utgivare

Joakim Malmudin

Redaktör

Jan Wretman, 070-781 78 35

Redaktion

Hans Alberg

Jens Malmros

Rolf Larsson

Marie Linder

Marika Wenemark

E-post wretman.jan@gmail.com

Produktion

Form och redigering: Mezzo Media AB

Tryckeri: Trydells Tryckeri AB

Annonser

Annonser i Qvintensen bokas med redaktören.

Annonssutskick per e-post bokas med Statistikfrämjandets sekreterare. Priset är 6 000 kronor för institutionsmedlem eller motsvarande, och 8 500 kronor för övriga annonsörer. På alla priser tillkommer 25 % moms.



SVENSKA STATISTIKFRÄMJANDETS STYRELSE

Ordförande

Joakim Malmudin 010-479 49 45,
ordforande.framjandet@gmail.com

Vice ordförande Maria Karlsson

Skattmästare Anders Haraldsson

Sekreterare Mattias Strandberg

Hemsidesansvarig Jens Malmros

Klubbmästare Moa Thyni

Surveyföreningen

Åke Wissing

FMS

Fredrik Norström

Industriell statistik

Vakant

Cramérsällskapet

Vakant

Övrig ledamot

Linnea Johansson Kreuger

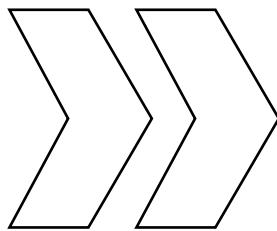
Adress

Svenska statistikfrämjandet

Sekreterare: Mattias Strandberg

E-post sekrfram@gmail.com

Webbplats www.statistikframjandet.se



**Rent allmänt kan sägas att
relativt många svenska uni-
versitet fått hyggliga place-
ringar ... »**

HANS ALBERG, SIDORNA 14-15

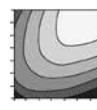
Qvintensen

STATISTIKFRÄMJANDET

NR 1 2019

Innehåll

ÄR STATISTIKFORSKNINGEN INDUSTRISTYRD – ELLER TVÄRTOM.....	2-5
CERTIFIERING – PÅ VÄG MOT MÅLET	6
REDAKTÖRENS RUTA/ Jan Wretman	6
MINNESORD/ Torbjörn Thedéen	8
OM ÅLDERSBEDÖMNINGAR	9-12
FÖREDRAG PÅ ÅRSMÖTE	13
VÄRLDSREKORD	13
RANKING AV UNIVERSITET	14-15
VALBAROMETRAR/ Har mätningarna blivit bättre eller sämre?	16-17
DEBATT/ Fyra debatter kring artiklar	18-33



Våra föreningar

INDUSTRIELL STATISTIK

Ordförande har ordet/

Är Indstats verksamhet

allmännyttig?34-35

SURVEYFÖRENINGEN

Från årsmötet 35

Vinnare av MarC Awards 2018 36

Kvalitetsseminariet 2019 37

Frimis 2018 38-39

ISSN 2000-1819

Författare och andra som utan ersättning bidrar med material för publicering behåller upphovsrätten till sina verk. Svenska statistikfrämjandet äger rätten att publicera dessa verk i valfri form och upplaga samt att behålla alla eventuella intäkter från sådan publicering, om inget annat avtalas. Författare och andra som utan ersättning bidrar med material för publicering garanteras rätten att själva få sprida sitt verk under förutsättning att Qvintensen anges som originalkälla med angivande av nummer och år. Svenska statistikfrämjandet äger upphovsrätten till publicerat material då det är framställt av redaktionsmedlem, styrelsemedlem, någon till Främjandet knuten förenings styrelse eller person som erhållit arvode för sitt bidrag. © Svenska statistikfrämjandet 2019.

MINNESORD ÖVER TORBJÖRN THEDÉEN

Från medarbetare vid SVT:s valvakor och vallokalsundersökningar

Vår vän och arbetskollega Torbjörn Thedéen, professor i statistik vid Kungl. Tekniska Högskolan, svarade under två decennier för valprognoser i Sveriges Televisions valvakor vid allmänna val, EU-val och folkomröstningar. För ett programföretag i allmänhetens tjänst är det av särdeles vikt att dess prognoser över medborgarnas val och preferenser är opartiska och sakliga. Torbjörn Thedéens utomordentliga insatser blev avgörande och han blev med tiden även en välbekant och uppskattad personlighet i de svenska TV-valvakorna.

De fakta som fanns tillgängliga för prognos- och riskbedömning på 1980-talet begränsades till geografisk strukturerings såsom bebyggelse- och arbetsmarknadsstatistik. Trots ett reducerat antal valresultat utvecklade Torbjörn allt säkrare prognoser. Svenska folket behövde därmed inte sitta uppe till långt inpå natten för att invänta det avgörande beskedet om regeringen skulle kvarstå eller avgå.

Varje valvaka brukade utlösa en kamp om tolkningen av valresultatets orsaker och varför väljarna röstat som de gjort. Först året därpå kom forskningsrapporter med vetenskapliga analyser som påvisade att många av antagandena saknade grund. Det stod alltmer klart att väljarna själva borde tillfrågas även om sådant på valdagen.

Inför riksdagsvalet 1991 fick Torbjörn Thedéen och undertecknade därför möjlighet att genomföra Sveriges första vallokalsundersökning, Valu. Så snart vallokalerna hade stängts kunde vi därmed offentliggöra ett omfattande material för opartisk och saklig analys av väljarnas preferenser. Torbjörn Thedéen – alltid med nyfikenhet, gott humör och uppmuntrande tillrop – var en garant för dessa undersökningars solida statistikgrund och att de direkt etablerades som grund för svenska valanalyser. Torbjörn Thedéen förändrade för alltid kunskapsläget inom sin profession.

HANS HERNBORN, JOURNALIST,
VALPROGRAMANSVARIG 1987 – 2007

SÖREN HOLMBERG, PROFESSOR EMERITUS
STATSVETENSKAP

PER NÄSMAN, DOCENT KTH



Från vännen Jan Grandell

»En matematisk statistiker är en person som på fikapausen berättar om en riktigt usel tillämpning av statistik och som hävdar att man borde verkligen göra något åt detta. Efter denna insats går han tillbaka till sitt rum och försöker göra satsen han arbetar med ännu lite generellare.»

Den här "definitionen" lärde jag mig av professor Torbjörn Thedéen som föddes 3 februari 1931 och avled 2 februari 2019. Han blev, så när som på en dag, 88 år.

Torbjörn disputerade 1967 i matematisk statistik på KTH. Han ägnade sig på den tiden åt trafikteori och gav en teoretisk motivering till att gles trafik lämpligen beskrivs av en Poissonprocess.

Sedan slutet av 1960-talet har Torbjörn och jag haft mycket med varandra att göra såväl yrkesmässigt och ämnesmässigt som privat. Torbjörn blev snabbt en av mina bästa vänner. Vid den tiden var han universitetslektor i matematisk statistik på KTH. Senare fick han en docenttjänst i matematisk statistik vid Stockholms Universitet, där jag då var doktorand.

Efter Torbjörns tid vid mat.stat. på SU så arbetade han på Brottsförebyggande rådet och sedermera som professor i statistik på SU. Efter detta återgick Torbjörn till KTH och bildade "Centrum för säkerhetsforskning", där han under många år till långt efter sin pensionering samarbetade främst med Pär Näsman. Utanför statistikerkeretsar är

Torbjörn säkert mest känd för alla valår då han, Sören Holmberg och Pär Näsman deltog i valvakorna på SVT och ansvarade för vallokalsundersökningarna.

Yrkesmässigt följde jag Torbjörn i spåren, och när även jag blev lektor på KTH, så långtidsvikarierade jag i flera år för honom. Vårt gemensamma intresse för punktprocesser gjorde att vi vid flera tillfällen deltog i samma konferenser, mest i Östeuropa. Det nog mest spännande var första Bernoullikongressen i Tasjkent. Vi besökte också Litauen och Östtyskland ett flertal gånger.

Torbjörn och hans hustru Ulla bodde under flera år nära min bostad, och vi umgicks då även familjevis, vilket fördjupade vår vänskap. Förutom ämnet var Torbjörn mycket musikintresserad, främst då det gällde klassisk musik. Torbjörn och Ulla hade ett landställe på Muskö i Stockholms skärgård, vilket förmodligen bidrog till hans seglingsintresse. Vidare var Torbjörn en stor hundvän. Hans kärlek till hundar gör att jag tror att han skulle ha uppskattat att här avbildas med sin äldste sons boxer.

JAN GRANDELL



Magnetrontgen av knä har använts på över 10 000 ensamkommande ungdomar för åldersbedömning.

Vi vill dela med oss av våra gemensamma erfarenheter och funderingar under det senaste året, då vi har jobbat med granskning av myndigheters missbruk av statistik. Vår förhoppning är att vi därmed kan ge goda råd för hur myndigheter ska undvika liknande misstag, hur vi som forskare bör agera och hur vi bör utveckla utbildningen för studenter som läser statistik.

Om åldersbedömningar och att lära sig av misstagen

Vi har granskat Socialstyrelsens rapport *Metoder för radiologisk åldersbedömning - En systematisk översikt* (Socialstyrelsen 2016), och vi med flera har kritiserat myndigheten för att inte följa god vetenskaplig sed¹.

Socialstyrelsens rapport har fått konsekvenser för tusentals ungdomar i Sverige. Man rekommenderar användningen av magnetkamera-analys av knä som metod för att bedöma om asylsökande är över 18 år eller inte. Metoden har sedan tillämpats av Rättsmedicinalverket (RMV) som beslutsstöd åt Migrationsverket. Ett barn som är asylsökande har långtgående rättigheter jämfört med vuxna och åldersbedömningen kan därför ha livsavgörande konsekvenser (Azeri 2017).

Innan vi kommer in på vår granskning av Socialstyrelsen och RMV är det rimligt att fundera över vilket ansvar myndigheter har enligt svensk lag och förordning. I regeringsproposition² från 2009 framgår det hur myndigheter styrs av förvaltningspolitiska mål: "Det övergripande målet för förvaltningspolitiken ska vara en innovativ och samverkande statsförvaltning som är rättssäker och effektiv, har en väl utvecklad kvalitet, service och tillgänglighet och därigenom bidrar till Sveriges utveckling och ett effektivt EU-arbete". Rättssäkerhet innebär att vi ska kunna lita på myndigheters beslut, och välfungerande myndigheter bidrar till att stärka förtroendet hos medborgarna. Samtidigt ska myndigheter uppfylla regeringens mål och göra det så kostnadseffektivt och med så hög kvali-

tet som möjligt. I det följande framgår det att RMV valde en dyr metod som är vetenskapligt dåligt underbyggd. Detta gjordes utifrån Socialstyrelsens inrådan.

Magnetrontgen av knä har använts på över 10 000 ensamkommande ungdomar för åldersbedömning. Leder i kroppen utvecklas med ålder och när lederna är fullt utvecklade kan benen i kroppen inte längre växa. Leden är då mogen och har nått sitt slutstadium i utvecklingen. För de flesta är knät fullt utvecklat vid 16 till 25 års ålder, och i Socialstyrelsens rapport föreslås att personer med moget knä ska klassificeras som vuxna och drar bl.a. följande slutsats vad gäller analysen av knä:

"18-årsgränsen bedöms med en risk att missta ett barn för en vuxen på 3% för pojkar." (sid 16)

»»»

»»» OM ÅLDERSBEDÖMNINGAR OCH ATT LÄRA SIG AV MISSTAGEN

Utifrån detta påstående skulle man kunna tro att metoden är träffsäker för att göra åldersbedömningar. Denna slutsats drog RMV. De rekommenderade användningen av metoden, och då den ansågs träffsäker baserade Migrationsverket sina beslut i huvudsak på RMV:s analyser. Även Barnläkarföreningen var nöjd med detta metodval då de uppfattade den nya metoden som mycket säkrare jämfört med tidigare använda metoder.

Vi har funnit både misstag och rena felaktigheter i det statistiska underlaget som använts i Socialstyrelsens rapport, vilket får konsekvenser på hur RMV:s analyser ska tolkas. Vi är medvetna om att statistik och sannolikhet kan vara svårt. En viktig fråga är hur liknande misstag kan undvikas i framtiden. Grundläggande är att det måste finnas resurser och kompetens för att räkna och tolka resultat korrekt.

För att åskådliggöra de misstag som Socialstyrelsen begick har vi valt att exemplifiera med ovanstående citat ur rapporten. Det är dessutom en av de mest citerade slutsatserna ur rapporten. Tolkningen är att ett barn som testas med metoden löper en risk på 3% att bedömas som vuxen. Nu är det tyvärr inte detta som Socialstyrelsen har beräknat i sin analys. De har beräknat någonting annat. Även ett stort antal andra resultat i deras rapport är felräknade och missvisande.

Vilka fel begicks?

Socialstyrelsen har beräknat sannolikheten för att en person med moget knä faktiskt är ett barn. Det är inte samma som risken att missta ett barn för att vara en vuxen. Alltså, sannolikheten för en händelse *A* givet *B* är inte samma sak som sannolikheten för en händelse *B* givet *A*. Matematiskt skrivs detta $P(A|B) \neq P(B|A)$. Vi är medvetna om att det kan vara svårt att inse skillnaden mellan dessa två "betingade sannolikheter", trots att det är två helt olika saker. Det är till och med ett så pass vanligt fel att det finns ett specifikt begrepp för detta: "Confusion of the inverse". Socialstyrelsen har räknat ut den ena sannolikheten $P(\text{Barn} | \text{Moget knä})$ men skriver i sin tolkning att det är den andra, $P(\text{Moget knä} | \text{Barn})$. För en lekman är det onekligen lätt att blanda ihop de två sannolikheterna, och det kan vara en utmaning att förklara skillnaden. Därför har vi lagt ut ett antal förklarande filmer på www.snackastatistik.se.

Socialstyrelsen har alltså blandat ihop dessa två sannolikheter. Men även om Socialstyrelsen hade undvikit detta misstag, så är siffrorna de

presenterar inte jämförbara. De är framräknade utifrån hur stor åldersspridning var bland studiedeltagarna med mogna knäleder. Men denna åldersspridning påverkas naturligtvis av åldrarna på studiedeltagarna, som i originalstudien var 10 till 30 år. Det är klart att om man testar många 30-åringar så blir sannolikheten att klassas som barn låg. Siffrorna som Socialstyrelsen använder är alltså styrda av *inklusionskriterierna* (valet av vilka som tas med) i studien, inte baserat på åldrarna på de som testas i verkligheten. En viktig insikt är alltså att risken att fatta fel beslut beror på ålderssammansättningen av de personer som testas. Hur många rätt respektive fel det blir beror på vilken grupp av människor som testet appliceras på. Skulle vi bara testa 17- och 18-åringar så skulle vi göra fler felbeslut jämfört om vi testar fler i yngre och äldre åldrar. Eventuella uttalande om risker för felaktiga beslut måste baseras på antaganden om åldersfördelningen bland de personer som faktiskt testas. Så i sin jämförelse mellan olika metoder (knä, tand, handled) blandas olika betingningar samt olika åldersfördelningar, vilket gör jämförelsen felaktig och missvisande.

Vi har hittills talat om logiska fel och fel beräkningar. Nu till ett annat metodproblem. Socialstyrelsen skriver i sin rapport om sannolikheten att missta ett barn för en vuxen. Men barn har olika sannolikheter att klassificeras som vuxna beroende på hur gamla de är. En 10-åring har minimal risk att bli felklassificerad medan en 17-åring har betydligt större risk. Att generellt tala om en risk för gruppen barn är meningslöst. En stor del av debatten om åldersbedömningar har alltså förts på felaktiga premisser. Om vi anpassar metoden så att den blir rättssäker för "gruppen barn" så kan den mycket väl vara icke-rättssäker för en 17-åring. Metoden måste vara rättssäker för 17-åringar då dessa löper störst risk för felklassificering. Sedan måste vi också beakta risken att en 18-åring klassas som barn. Dessa risker måste vägas emot varandra vid framtagande av lämplig metod.

Vi är inte ensamma om att kritisera metoderna för åldersbedömningar. Det har under det senaste året förts en livlig debatt om RMV:s metod för att kombinera resultat från olika tester.

Enligt EU-direktiv 2013/32/EU, artikel 25.5 skall beslutsfattaren, om det finns tvivel om den asylsökandes ålder, utgå från att hen är under 18 år. RMV har i sin metod underlag från un-

dersökning av både tand och knä. Resultaten från dessa två test sammanfattas till ett omdöme som Migrationsverket har som underlag till sitt beslut. Om båda testen visar att personen är över 18 år så används följande formulering. *"Resultatet av den genomförda undersökningen talar för att den undersökte är 18 år eller äldre."* Men även om bara ett av testen indikerar att personen är över 18 år så används exakt samma formulering. På RMV:s hemsida anges att en person med bara en mogen kroppsdel (knä eller tand) har en mindre sannolikhet att vara över 18 år jämfört med en person med två mogna kroppsdelar. Men detta framgår alltså inte i den sannolikhetsskala som de använder. Att inte på något sätt gradera skillnaden mellan de två olika resultaten är missvisande. Risken för felklassificering av barn ökar; det räcker ju med att en av metoderna visar fel för att slutsatsen skall bli fel. RMV måste göra mer nyanserade uttalanden.

Socialstyrelsens och RMV:s statistiska beräkningar och resonemang uppvisar stora brister. Dessa myndigheter är dock inte ensamma om att brottas med hur man skall hantera kvantitativ information. Förvirringen i den medicinska debatten om åldersbedömningar bottnar i genuina svårigheter att tolka siffror på rätt sätt. Nobelpristagaren Daniel Kahneman med flera har i sin forskning visat på människans begränsade förmåga att hantera sannolikheter. Men då statistik och statistiska resonemang är centrala i många viktiga beslut i dagens samhälle så måste vi arbeta för att höja kvaliteten och hitta rutiner för adekvat kvalitetssäkring.

Hur skulle Socialstyrelsen ha kunnat undvika misstaget?

Misstagen i Socialstyrelsen rapport från 2016 hade troligen undvikits om följande grundläggande steg hade vidtagits:

- **En tydlig metodbeskrivning i rapporten.** Om tid hade ägnats åt att skriva en metodbeskrivning så hade rapportförfattarna troligen upptäckt misstaget själva innan den publicerades. Om de inte hade upptäckt felet så hade utomstående granskare upptäckt det på ett tidigt stadium.
- **Att resultaten presenteras i heltal i stället för som sannolikheter.** Gärna som enkla figurer där resultatet framgår enkelt.
- **Ordentlig extern granskning av rapporten.** Det finns två typer av rapporter som myndigheter tar fram: sådana som de själva

skriver och sådana som de beställer av forskare vid högskolor eller forskningsinstitut. Socialstyrelsen rapport från 2016 var av den första typen och dessa är i särskilt behov av extern granskning. Förslagsvis bör de seminariebehandlas vid en extern institution. Framför allt måste den läsas av någon med kritiska ögon, med uppdraget att lyfta fram alla eventuella svagheter med rapporten.

Det här är steg som myndigheter i allmänhet bör beakta då de tar på sig projekt som kan komma att få stora konsekvenser. Det är viktigt att dessa rutiner finns då beslut mer och mer förväntas vara statistiskt underbyggda, samtidigt som det finns en tendens att beslut ska tas allt snabbare. Dessutom finns det flera fall där myndigheter skyller på att de var tvungna att ta snabba beslut, vilket i deras ögon berättigar dåliga underlag (t ex Nya Karolinska, Transportstyrelsen). Samtidigt blir det allt viktigare att vi kan lita på myndigheters underlag, då det finns en trend om att hänvisa till "alternativa fakta", inte minst efter att Donald Trump tillträdde som USAs president.

Förslag på riktlinjer för externt beställda rapporter

Våren 2018 publicerade Socialstyrelsen en rapport utifrån ett uppdrag att fördjupa kunskapen om magnetkameraundersökning (Socialstyrelsen 2018). Denna rapport är av den andra typen, dvs en rapport skriven av externa forskare. Här har vi erfarit att det är viktigt att beställaren inte styr arbetet, att den granskas av externa bedömare (helst utländska) och att beställaren inte heller styr hur resultaten framförs.

En möjlighet som beställaren i princip har är att inte publicera rapporten om resultaten är kontroversiella. Denna möjlighet måste vara tydligt utesluten när arbetet påbörjas och därför bör den som godkänner rapporten också vara en extern person som baserar sitt beslut på de externa bedömarens utlåtanden.

Hur agerade Socialstyrelsen och RMV på vår kritik?

Under våren 2018 fick vi tillfälle att träffa



Hösten 2018 ansåg socialstyrelsen att rapporten *Metoder för radiologisk åldersbestämning* är inaktuell.

och "Myndighetens översyn förändrar däremot inte rapportens huvudsakliga slutsats – att åldersbedömning via magnetkameraundersökning är en metod som utifrån kunskapsläget 2016 hade stöd men borde studeras vidare, vilket också har gjorts." Myndigheten hänvisar till sin rapport från 2018 men även i denna studie har inklusionskriterierna påverkat resultatet⁴ och det går inte att dra långtgående slutsatser om de ensamkommande som faktiskt har åldersbedömts av RMV.

Vi fick också tillfälle att träffa ansvariga personer på RMV och kunde påpeka bristerna i metodvalet. Vid mötet var dock representanterna för RMV ointresserade av att diskutera svagheter i analysen och ansåg inte att de

ansvariga personer på Socialstyrelsen för att diskutera de felaktigheter vi hittade i deras rapport. De var mycket tillmötesgående och en av deras statistiker kunde beskriva de misstag de hade begått vilket visade att de hade förstått problematiken. Efter mötet var vi inte överens vad gäller hantering av rapporten. Vi hävdade att det enda rätta vore att ta bort rapporten helt eller att tydligt påpeka felaktigheterna i den medan Socialstyrelsen valde att skriva att metodbeskrivningen behövde förtydligas. Från och med hösten 2018 framgår det att rapporten anses vara inaktuell⁵

hade någon skyldighet att svara på våra frågor. Mötet blev kort.

När är det rimligt för forskare att gå ut i dagspressen och kritisera en myndighets rapport?

Först och främst gäller yttrandefrihet vilket medför att alla har rätt att kritisera myndighetsrapporter oavsett hur det görs. Det finns dock svårigheter och dilemman som bör beaktas.

Ett problem kan vara om en forskare utnyttjar sin position för att föra fram ett budskap som inte är vetenskapligt grundat. Kritik kan röra sådant som handlar om skillnaden mellan olika vetenskapliga skolor. Om så är fallet behöver det förklaras tydligt. Kritiken får betydligt större tyngd om den är vetenskapligt prövad och publicerad innan den framförs i pressen. Detta är därför något som varje forskare bör sträva efter.

I vårt fall hittade vi ett uppenbart fel i rapporten som vi kunde förklara för en journalist, Negra Efendic, som i sin tur kunde stå för innehållet i artikeln där kritiken framfördes av oss; ett grundläggande fel som inte var öppet för vetenskaplig diskussion. Vi kontaktade inte i första hand berörd myndighet, då det var uppenbart att för mycket prestige var investerad i rapporten för att de skulle vara mottagliga för att föra en diskussion kring felen och hur det skulle kunna rättas till. Dessutom insåg vi att situationen var akut, då metoden hade applicerats på tusentals berörda personer.

RMV är den myndighet som har tillämpat metoden och skickat underlag till Migrationsverket för beslut. Här har Mostad & Tamsen



Ett uppenbart fel i rapporten förklarades för en journalist som i sin tur kunde stå för innehållet i artikeln där kritiken framfördes av oss; ett grundläggande fel som inte var öppet för vetenskaplig diskussion.

»»» OM ÅLDERSBEDÖMNINGAR OCH ATT LÄRA SIG AV MISSTAGEN

(2018) publicerat en vetenskaplig artikel som visar på att den icke-validerade metoden som nu RMV tillämpar inte är rättssäker och att deras egna resultat är motsägelsefulla. En rimlig förklaring till RMVs motsägelsefulla resultat är att metoden för MR-knä innehåller någon typ av systematiskt fel. Genom att basera kritiken på vetenskapligt publicerade resultat får denna kritik stor tyngd. Dock har inte RMV bemött kritiken. En viktig fråga är därför hur politiker säkerställer att myndigheter är öppna för kritik och är beredda att rätta till rena felaktigheter i verksamheten.

Det fel som Socialstyrelsen gjorde i sin rapport går möjligen att förstå då de uppenbarligen arbetade under tidspress och inte hade bra rutiner för hur rapporten skulle granskas. Att man sedan inte rättade till felen, när det uppenbarades utifrån vår kritik, anser vi inte vara försvarbart. För att myndigheter fortlopande ska utveckla verksamheten⁵ behövs en kultur där man lär sig av misstagen.

Vilka lärdomar drar vi om den utbildning vi bör ge studenter i statistik?

Utifrån våra erfarenheter vill vi lyfta fram vad som är viktigt för dagens studenter i statistik. Mycket genomförs redan på de flesta utbildningar i statistik men vår förhoppning är att råden nedan belyser behovet ännu tydligare.

Att lära sig skriva *tydliga metodbeskrivningar* är en naturlig del av statistikstudier, och Socialstyrelsens rapport från 2016 visar på ett tydligt sätt varför det är så centralt. Ett ytterligare steg är att skriva rapporter som är helt *reproducerbara* genom att hänvisa till grunddata och bifoga programkod. På så sätt kan en läsare följa alla steg för hur ett resultat har tagits fram, och om resultaten ifrågasätts kan författarna lätt redovisa hur de har tagits fram. Vår rekommendation är att förmågan att kunna skriva reproducerbara rapporter

bör ingå i statistikutbildningar på magister- eller masternivå.

När en student i statistik har tagit examen bör den känna att *extern granskning* är en naturlig del av den högre utbildningen och förhoppningsvis ta med sig det ut i arbetslivet. Det bästa är om studenterna får träning i att granska varandras arbeten genom seminariebehandling av uppsatser och medrättande av övningsuppgifter. En viktig del i övningen är att upptäcka hur lätt det är att göra fel.

Därmed kommer vi in på kanske den viktigaste förmågan som vi behöver träna våra studenter i, förmågan att inse att *man kan göra fel* och att kunna erkänna misstag. I utbildningen behöver det finnas frihet och öppenhet för att studenter ska våga göra fel, vilket gör att vi lärare rimligtvis bör vara öppna med de fel vi gör. Ett sätt är att visa på andra forskares misstag i publicerade artiklar men också genom att förklara enkla misstag vi själva gjort som lärare.

Myter och felaktigheter rättas inte automatiskt till genom att framföra fakta, och studenter behöver tränas i att *framföra kritik på ett klokt sätt*. Hur detta görs på bästa sätt är en vetenskap i sig (Cook & Lewandowsky 2011, Lewandowsky et al. 2012, Frans 2017), men några enkla grundtips är att visualisera fakta med enkla figurer (s.k. infographics), att inte göra misstaget att förstärka myterna genom upprepningar i texten men däremot förklara varför myten har uppstått.

Vi hoppas att vi härmed har kunnat ge några råd för hur myndigheter ska undvika liknande misstag i framtiden, hur vi som forskare bör agera, och hur vi bör utveckla utbildningen för studenter i statistik.

JOHAN BRING, STATISTICON AB,
ADJUNGERAD PROFESSOR I STATISTIK VID
HÖGSKOLAN I GÄVLE

LARS RÖNNEGÅRD, PROFESSOR I STATISTIK,
HÖGSKOLAN DALARNA

Referenser

- Azeri, N. (2017) "Åldersbedömningar av ensamkommande barn - särskilt om metoderna för medicinsk åldersbedömning". Examensarbete i Offentlig rätt, Juridiska institutionen Stockholms universitet.
- Cook, J., Lewandowsky, S. (2011), *The Debunking Handbook*. St. Lucia, Australia: University of Queensland. ISBN 978-0-646-56812-6.
- Frans, E (2017) *Larmrapporten – att skilja vetenskap från trams*. Volante förlag. ISBN 9789188123930.
- Lewandowsky, S., H., U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131.
- Mostad, P. & Tamsen, F. (2018). Error rates for unvalidated medical age assessment procedures. *International Journal of Legal Medicine*. <http://dx.doi.org/10.1007/s00414-018-1916-3>.
- Socialstyrelsen (2016). Metoder för radiologisk åldersbedömning – En systematisk översikt. (Rapport)
- Socialstyrelsen (2018). Om magnetkamera vid bedömning av ålder – En studie av validiteten i radiologisk undersökning. (Rapport)

Fotnoter

- 1) Efendic, N., "Forskare pekar ut fler fel: 'De bör dra tillbaka rapporten'". Svenska Dagbladet 15 mars 2018.
- 2) Prop. 2009/10:175.
https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/myndighetsforordning-2007515_sfs-2007-515
- 3) <https://www.socialstyrelsen.se/publikationer2016/2016-7-4>
- 4) Efendic, N. "Experter: Varannan 18-åring kan klassas som barn" Svenska Dagbladet 6 juni 2018
- 5) I enlighet med Myndighetsförordningen §6.

Föredrag på årsmöte

Statistikfrämjandets årsmöte 2019 ägde rum den 21 mars i Göteborg. Förutom det formella årsmötet gavs ett antal föredrag av inbjudna föredragshållare.

Magnus Pettersson presenterade det förslag till certifiering av statistiker som utformats av en projektgrupp inom FENStatS (Federation of European National Statistical Societies). Se sid 6 i detta nummer.

Carl-Fredrik Burman diskuterade problem om hur man ska analysera och tolka resultat från en viss typ av läkemedelsstudier. Det handlade om att analysera effekten av två alternativa läkemedel, när man finner att deras överlevnadskurvor korsar varandra.

Anders Holmberg berättade om erfarenheter från arbete vid nationella statistikbyråer i

Sverige, Norge, Nya Zeeland och Kambodja. Skillnader mellan länderna redovisades i termer av (1) hög/låg tillit till myndigheter, (2) hög/låg tillit till officiell statistik, (3) centraliserat/decentraliserat statistiksystem, (4) traditionellt/modernt statistiksystem, (5) stort/litet inflytande från metodstatistiker, (6) stor/liten tonvikt på officiell statistikproduktion i ländernas utbildningssystem.

Till årets statistikfrämjare 2018 hade styrelsen utsett Lars Lyberg. Han har under 2018 bland annat medverkat i två böcker: *Total Survey Error in Practice*, där han varit redaktör, och *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts*, där han författat två av kapitlen. Under 2018 genomfördes dessutom den stora, internationella konferensen "Big Surv 18" i Barcelona som Lyberg varit upphovsman till.

JAN WRETMAN

Statistikfrämjandets nya styrelse

- Ordförande: Joakim Malmelin (omval)
- Vice ordförande: Maria Karlsson (nyval)
- Sekreterare: Mattias Strandberg (omval)
- Skattmästare: Anders Haraldsson (nyval)
- Webbansvarig: Jens Malmros (nyval)
- Klubbmästare: Moa Thyni (omval)
- Ledamot: Linnea Johansson Kreuger (tidigare webbansvarig)
- Repr. FMS: Fredrik Norström (omval)
- Repr. Surveyföreningen: Åke Wissing (omval)
- Repr. Industriell statistik: Ännu ej klart
- Repr. Cramérsällskapet: Ännu ej klart



Har du flyttat?

■ Du kan själv ändra dina uppgifter genom att gå till <https://www.membit.net/m4-member/login> där du loggar in med dina personliga inloggningsuppgifter (finns på din senaste inloggningsavi). Vid frågor kontakta Mattias Strandberg på sekrfram@gmail.com.

Världsrekord!

■ **Det kan kanske** vara av intresse att veta att Statistikfrämjandet bland sina medlemmar har en världsrekordhållare i höjdhopp, nämligen Carl-Erik Särndal. Tidigare statistikprofessor i Umeå, Vancouver och Montreal, och välkänd för alla surveystatistiker. Den 2 mars i år, vid veteran-SM i Malmö, hoppade han 1,40, vilket är världsrekord i åldersklassen 80-84 år.

Som ung var Carl-Erik en av Sveriges bästa höjdhoppare, varfter han gjorde cirka 40 års uppehåll i tävlandet. Som 66-åring återupptog han tävlandet med nyväckt energi, nu inom veteranfriidrottens femåriga åldersklasser. Han har sedan dess tävlat i klasserna M65, M70, M75 och nu i M80 (M 80, till exempel, betyder Män 80-84 år.). Under dessa år har Carl-Erik vunnit flera världsmästerskap i dessa klasser, nu senast i Malmö med världsrekordet 1,40 inomhus i klassen M80. Hans stil är den gamla dykstilen, med sko på vänsterfoten.

Ranking av universitet

Introduktion

Under senare år har intresset för ranking av universitet ökat, vilket på gott och ont kan ses som ytterligare ett tecken på "måtsamhället". Resultaten av PISA-mätningarna har inte varit någon munter läsning för Sveriges del, även om det skedde en viss förbättring senaste gången. Ranking av universitet har däremot sett bättre ut. Rankinglistor som ofta beaktas är:

- Shanghai ranking, även kallad Academic Ranking of World Universities (ARWU)
- World University Ranking från Times Higher Education (THE)
- Quacquarelli Symonds (QS) vilket är ett brittiskt företag specialiserat på utbildning
- Center for World University Rankings (CWUR)

Ranking av brittiska och amerikanska universitet har pågått länge. Det har troligen funnits visst behov av det, som en professor på KTH påpekade våren 1976: "I USA är det en stor skillnad mellan å ena sidan Harvard och MIT, där världsstjärnorna sitter, och ett college i Mellanvästern, där de i princip är gymnasie-lärare". ARWU-rankingen publicerad 2003 var det första försöket till en internationell ranking av universitet och högskolor.

Rankinglistor finns såväl för lärosäten i sin helhet som för en rad ämnen/ämnesområden. Man kan konstatera att såväl metodik som ämnesindelning skiljer sig åt. CWUR redovisar dels en lista på lärosätessnivå, dels de tio främsta i ett stort antal enskilda ämnen.

Vid rankingen beaktas parametrar såsom vunna pris, citeringsindex, akademiskt rykte etc. En introduktion till de olika rankinginstituten kan fås exempelvis via Stockholms universitets webbplats (se referens sist i denna artikel).

En fråga är naturligtvis vilken vikt man skall fästa vid dessa rankinglistor. Jag passade på att gratulera en matematiker som arbetar vid Paris-Sud, som hamnade på andra plats i Shanghai-rankingen i ämnet matematik. Han citerade sin förre rektor som sa att "naturligtvis tror vi inte på rankinglistor, men låt oss ändå glädjas".

Jag upptäckte att QS-rankingen inte bara placerat KTH för "Mechanical, Aeronautical and Manufacturing" på en hedervärd 34:e plats,

utan även gett Stockholms universitet, som ju inte har någon teknisk fakultet, en plats bland de 200 främsta. Detta påtalades för QS som lovade att rätta till det, samt försäkrade att det var ett enstaka misstag och inget systematiskt fel.

Att rangordna olika universitet är således ingen exakt vetenskap, och placeringarna blir olika, men det är ändå ett sätt att få extern utvärdering som kan bilda underlag för en mer djupgående analys. I denna artikel används, om inget annat anges, siffrorna för 2018.

I artikeln används omväxlande begreppen universitet och lärosäte. Lärosäten omfattar såväl universitet som högskolor, där universitet har generell rätt att utfärda forskarexamina. Notera att KTH, CTH och Handelshögskolan räknas som universitet trots att de av historiska skäl har ordet "högskola" i namnet.

Allmän ranking av svenska lärosäten

Enligt senaste Shanghai-rankingen (2018) så har Sverige 11 av världens 500 högst rankade universitet (se Tabell 1). Inget land med färre invånare än Sverige har så pass många lärosäten på den listan. Dock bör påpekas att Schweiz har 8 universitet bland de 500 främsta, varav ETH (Eidgenössische Technische Hochschule) är det enda utanför USA och England bland de 20 främsta. Notera att i vissa listor anges rankingen ibland i intervall, utom för de som ligger främst, se Tabell 1.

Även om placeringarna för de enskilda lärosätena skiljer sig något så kan vi konstatera att såväl Shanghai som Times Higher Education placerar 11 svenska universitet bland de 500 främsta i världen. Det borde ge en tämligen god bild av kvaliteten för svensk högre utbildning

Ranking av svenska lärosäten per ämne/ämnesområde

Nästa fråga är hur det ser ut för de enskilda akademiska disciplinerna. Som framgår av Tabell 1, och nedanstående lista, kan noteras att, trots att SLU intar toppositioner inom fyra ämnesområden enligt CWUR, så hamnar universitetet totalt på en mer modest placering i Tabell 1.

- KI: Allergy 3, Dermatology 9, Endocrinology & Metabolism 5, Hematology 9, Oncology

5, Nursing 1, Pediatrics 10, Biological Psychology 2, Rheumatology 4, Sport Sciences 19, Toxicology 9

- LU: Ornithology
- SLU: Biodiversity conservation 7, Forestry 1, Soil science 5, Veterinary sciences 7
- UU: Evolutionary biology 7, Nursing 10
- SU: Physical geography 6
- GU: Dentistry 4

Man bör nog också beakta att olika ämnen torde ha olika tyngd: således borde GU:s topposition i odontologi väga tyngre än LU:s i ornitologi.

Även enligt ARWU finns svenska lärosäten med ämnen/ämnesområden bland de 10 främsta:

- KI: Pharmacy 3
- SLU: Agricultural science 4
- KI: Public health 5
- SU: Geography 5

Ett ämne där Sverige historiskt har stolta traditioner, och där man kunde förvänta sig bra ranking, är kemi. Ser vi på listan över forskare i världen som upptäckt tre eller fler grundämnen finner vi tre svenskar, nämligen Berzelius, Mosander och Scheele:

- Glenn Seaborg (1912 – 1999), 9 grundämnen
- William Ramsay (1852 – 1916), 5 grundämnen
- Humphry Davy (1778 – 1829), 4 grundämnen
- Jöns Jakob Berzelius (1779 – 1848), 4 grundämnen
- Paul-Emil Lecoq de Boisbaudran (1838 – 1912), 3 grundämnen
- Carl Gustaf Mosander (1797 – 1858), 3 grundämnen
- Morris William Travers (1872 – 1961), 3 grundämnen
- Carl Wilhelm Scheele (1742 – 1786), 3 grundämnen

Ser vi på dagens rankinglistor ser det dock inte fullt så lysande ut. I ämnet kemi finns det inget svenskt universitet bland de 100 främsta

Referenser

1. Stockholms universitet i Rankinglistor: <https://www.su.se/om-oss/fakta/rankning>
2. Wikipedia lista över universitet och högskolor i Sverige: https://sv.wikipedia.org/wiki/Lista_%C3%B6ver_universitet_och_h%C3%B6gskolor_i_Sverige

enligt ARWU-rankingen. SU, KTH och UU finns dock bland de 150 främsta. I kemiteknik ser det något bättre ut: KTH och CTH finns bland de 75 – 100 främsta.

Ranking av svenska lärosäten avseende statistik och närbesläktade ämnen

Det mest intressanta för denna tidskrifts läsare är troligen hur situationen ser ut i statistik och närbesläktade ämnen. Fördelningen av de 25 främsta positionerna i respektive lista framgår av Tabell 2.

Svenska universitets placeringar i de olika rankinglistorna för matematik och statistik framgår av Tabell 3.

Vi kan här se att ämnesindelningen varierar mellan de olika rankinginstituten, exempelvis slår QS ihop statistik och operationsanalys (OA).

Enligt Shanghai-rankingen så intar USA en mycket dominerande position i ämnet statistik. På CWURs "tio i topplista" återfinns endast amerikanska universitet. Sverige har endast två universitet bland de 200 främsta. Med tanke på att mycket tyder på att statistik kommer att få en ökad betydelse i framtiden. ("Mätsamhället", "Industrie 4.0", ökade krav på kvantitativ/probabilistisk riskanalys etc.) så borde detta föranleda en fördjupad analys. En aspekt som redan lyfts fram är att de svenska statistiska institutionerna är relativt fragmentiserade jämfört med de amerikanska.

Slutsatser

Ranking av lärosäten har fått ökad uppmärksamhet under senare år. Avsikten med artikeln har inte varit att göra en fullödig utvärdering av det svenska utbildningsväsendet, utan att förhoppningsvis ge impulser till fortsatta diskussioner. Det gäller såväl metodiken för ranking som det aktuella utfallet. Rent allmänt kan sägas att relativt många svenska universitet fått hyggliga placeringar, även om de verkliga toppositionerna varit färre, förutom KI och SLU. Vad gäller statistikämnet specifikt kan det finnas anledning till en fördjupad analys, inte minst med tanke på att allt tyder på att ämnet får ökad betydelse i framtiden.

HANS ALBERG

	CWUR 2018/19	THE 2018	Shanghai 2018	QS 2018	CWUR ämnen 2017
KI (Karolinska Institutet)	49	40	44	-	11
UU (Uppsala Universitet)	86	87	63	117	2
LU (Lunds Universitet)	95	98	101-150	92	1
SU (Stockholms Universitet)	159	153	77	200	1
KTH (Kungliga Tekniska Högskolan)	156	187	201-300	104	
CTH (Chalmers Tekniska Högskola)	464	201-250	201-300	168	
GU (Göteborgs Universitet)	166	201-250	101-150	269	1
SLU (Sveriges Lantbruks Universitet)	368	201-250	201-300		4
UmU (Umeå Universitet)	255	301-	401-500	338	
LiU (Linköpings Universitet)	357	351-400	301-400	302	
ÖRU (Örebro Universitet)	(730)	401-500			
HHS (Handelshögskolan i Stockholm)	(644)		401-500		

Tabell 1. Kolumn 1-4 visar placering för de svenska universitet som hamnat bland de 500 främsta 2018. Sista kolumnen visar antalet ämnen där ett universitet tillhör de 10 främsta.

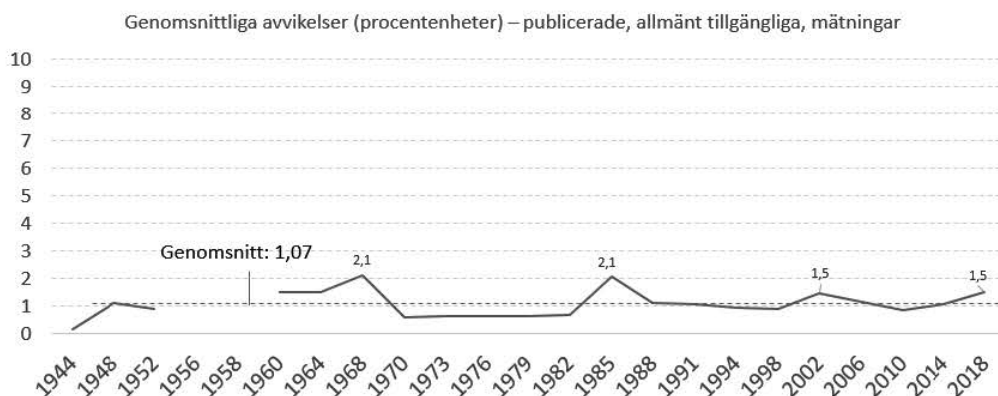
	QS – Stat & OA	Shangha - Stat	QS - Mat	Shanghai - Mat	THE - Mat & Stat
USA	16	21	15	15	18
UK	3	1	3	4	4
Schweiz	1	1	2	1	1
Australien	1		1		
Japan	1		1	1	
Singapore	1	1	1		1
Canada	1	1	1		1
Kina	1				
Frankrike			1	2	
Sverige				1	

Tabell 2. Fördelning på länder för de 25 högst rankade lärosätena i statistik.

	QS Stat & OA	Shanghai Stat	QS Mat	Shanghai Mat	THE Mat & Stat 2019
KTH	51-100		51-100	22	187
CTH	101-150		151-200	101-150	201-250
LU	101-150		101-150		98
SU	101-150	151-200	201-250		153
UU	101-150	151-200	151-200	101-150	87
LiU			251-300	301-400	351-400
GU				101-150	201-250
ÖRU					401-500
UmU					351-400

Tabell 3. Ranking av svenska lärosäten i statistik och närallgande ämnen.

Figur 1. Följande figur visar hur den genomsnittliga absoluta avvikelsen över alla partier för publicerade mätningar har varierat genom åren.



VALBAROMETRAR 1944 – 2018

Har mätningarna blivit

Inledning

Under de senaste åren har artiklar och blogginlägg publicerats som avhandlat hur det gått för opinionsinstituterna i samband med tidigare val. De inlägg i ämnet jag sett har alla varit begränsade till de val som infallit på 2000-talet. Men jag anser att ett 20-årigt tidsfönster är för kort för att man ska kunna uttala sig om mätningarnas träffsäkerhet i ett historiskt perspektiv. Om man på allvar ska försöka besvara frågan om mätningarna var bättre eller sämre förr, måste man se mätningarna över en längre tidshorisont. Det är vad jag nu vill göra. Artikeln är baserad på en presentation som jag höll vid Frimisseminariet 2018.

Sedan den första svenska väljarbarometern publicerades i augusti 1944 (DN, 30 aug.) har det i Sverige hållits 23 val till riksdagen, fem nationella folkomröstningar (1955, 1957, 1980, 1994, 2003) och fem val till Europaparlamentet (1995, 1999, 2004, 2009, 2014).

Vid min sammanställning av mätningar publicerade i samband med dessa val och folkomröstningar har jag använt mig av följande avgränsningar: mätningarna ska vara publicerade före valdagen, mätningarna ska vara genomförda i anslutning till valen (eller folkomröstningen), mätningarna har varit sökbara (eller har en sökbar hänvisning) i KB:s databaser, alternativt att mätningarna är kända (nämnda i böcker, tidningar, tidskrifter etc.), samt att mätningarna ska vara allmänt tillgängliga, d.v.s. publicerade framför betalvägg.

Under mina efterforskningar har jag hittat

	1957	1980	1994	2003	Genomsnitt
Genomsnittlig avvikelse från valresultat (procentenheter)	3,6	1,4	1,2	4,1	2,6

Tabell 1 visar mätvärdenas genomsnittliga avvikelse från det faktiska omröstningsresultatet vid folkomröstningarna 1957, 1980, 1994 och 2003.

	1995	1999	2004	2009	2014	Genomsnitt
Genomsnittlig avvikelse från valresultat (procentenheter)	2,1	1,5	3,8	1,6	1,8	2,1

I tabell 2 redovisas resultat avseende de fem Europaparlamentsvalen på ett sätt som är analogt med redovisningen i föregående avsnitt.

mätningarna i samband med 21 av de 23 riksdagsvalen, för fyra av de fem folkomröstningarna och för alla val till Europaparlamentet. De val och folkomröstningar jag inte lyckats hitta mätningar för är folkomröstningen 1955 (högertrafiken), riksdagsvalet 1956 och extravalet till riksdagen 1958. Antalet institut som publicerat mätningar inför val och folkomröstningar, som jag har lyckats hitta, varierar mellan ett (riksdagsvalen 1944 – 1979) och åtta (riksdagsvalet 2014).

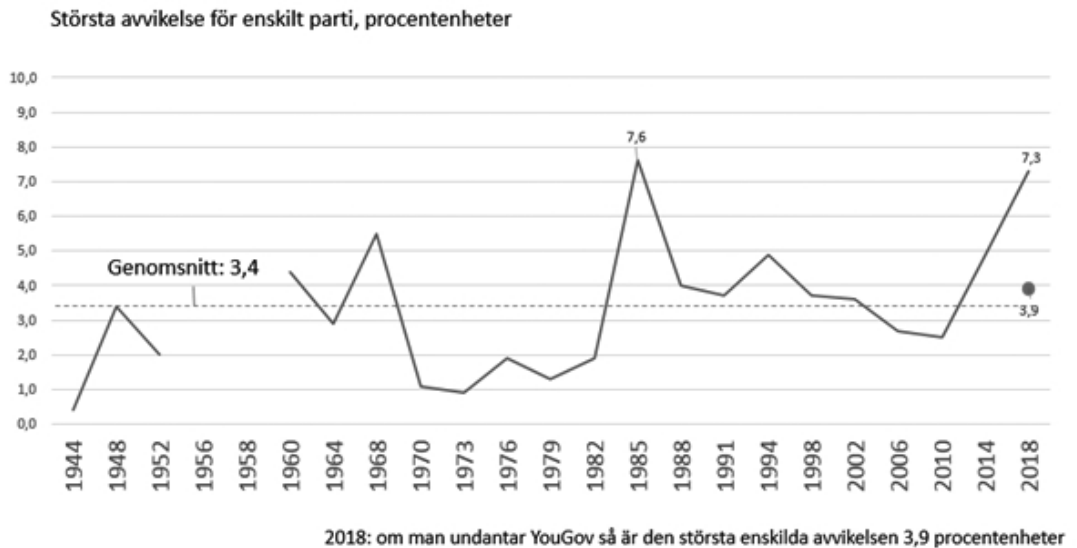
Folkomröstningarna

Tabell 1 visar mätvärdenas genomsnittliga avvikelse från det faktiska omröstningsresultatet vid folkomröstningarna 1957, 1980, 1994 och

2003. Med *mätvärde* menas här genomsnittet över de institut som gjorde mätningar före den aktuella folkomröstningen. Sista kolumnen visar genomsnitt över de fyra aktuella folkomröstningarna.

Mätningarnas förmåga att förutsäga resultaten i kommande folkomröstning har varierat kraftigt genom åren. Störst var den genomsnittliga avvikelsen i samband med folkomröstningen om Euron 2003, lägst vid folkomröstningen om EU-medlemskapet 1994. När det gäller mätningarna 2003, så bör det nämnas att där förekom flera institut som normalt inte (vare sig före eller efter) har ägnat sig åt publika mätningar avseende väljarnas röstningsintentioner. Dessa bidrar i

Figur 2.



bättre eller sämre?

väsentlig grad till den stora genomsnittliga avvikelsen. Genomsnittliga avvikelse över folkomröstningarna 1957–2003 är 2,6 procentenheter.

Europaparlamentsvalen

I tabell 2 redovisas resultat avseende de fem Europaparlamentsvalen på ett sätt som är analogt med redovisningen i föregående avsnitt.

Europaparlamentsvalen har ända sedan det första tillfället 1995 uppvisat större avvikelser än vad vi varit vana vid i samband med riksdagsvalen. Det låga valdeltagandet är en svårighet som instituten har att brottas med. Den genomsnittliga avvikelsen per parti är en procentenhet högre än det historiska snittet i samband med riksdagsval. Bäst lyckades instituten vid Europaparlamentsvalet 1999 då den genomsnittliga avvikelsen hamnade på 1,5 procentenheter. Att den genomsnittliga avvikelsen var störst 2004 förklaras av att Junilistan gjorde en mycket stark spurt, som i huvudsak inträffade efter att instituten avslutat sina intervjuer inför valet. Det historiska snittet har landat på 2,1 procentenheter i avvikelse per parti.

Riksdagsvalen

Figur 1 visar hur den genomsnittliga absoluta avvikelsen över alla partier för publicerade mätningar har varierat genom åren.

I genomsnitt har avvikelsen i förhållande till valresultaten 1944 – 2018 varit 1,1 procentenheter (streckad linje i figuren). Mätningen med minst avvikelse i förhållande till valresultatet publicerades av Gallup inför 1944 års val. Mätningen hade där en genomsnittlig avvikelse per parti på 0,2 procentenheter. Störst är avvikelsen

i samband med valen 1968 och 1985 då den genomsnittliga avvikelsen var 2,1 procentenheter. Sedan millennieskiftet har det hållits fem riksdagsval. Vid tre av dessa val har den genomsnittliga avvikelsen legat nära eller under det historiska snittet. Vid två val, 2002 och 2018, har den genomsnittliga avvikelsen landat på 1,5 procentenheter. Sett över de dryga 70 år då det publicerats väljarbarometrar är det svårt att hävda att mätningarna över lag har blivit sämre och uppvisar allt större avvikelser i förhållande till valresultaten. Med undantag av Sifo:s mätningar 1970 – 1982, som uppvisade en betydligt mindre genomsnittlig avvikelse i förhållande till valresultaten, så visar historien exempel på enskilda val då undersökningarna uppvisat större avvikelser än genomsnittet. Men att det är svårt att utifrån det längre perspektivet hävda att instituten *generellt* har blivit sämre på att mäta väljarnas röstningsintentioner i samband med riksdagsvalen.

I likhet med alla genomsnittsjämförelser så löper man risken att med genomsnittet dölja större enskilda avvikelser. Jag har därför valt att utöver att titta på genomsnittet även sammanställa största enskilda avvikelse per institut och enskilt parti; se figur 2.

Här är, av naturliga skäl, variationen betydligt större än när vi tittar på genomsnittet. Den största avvikelsen för ett enskilt parti återfinns i samband med 1985 års val då ett av instituten underskattade Folkpartiets väljarstöd med 7,6

procentenheter. Den näst största avvikelsen hittar vi i samband med 2018 års val, då ett av instituten överskattade Sverigedemokraternas stöd med över 7 procentenheter. Undantar man det institutet från sammanställningen landar den största enskilda avvikelsen på 3,9 procentenheter för ett enskilt parti

i samband med 2018 års riksdagsval. Den genomsnittliga största avvikelsen för enskilt parti 3,4 procentenheter.

När det gäller den största enskilda avvikelsen så kan man dra två

olika slutsatser. Väljer man att betrakta valen 1970–1982 som undantag är det svårt att hävda att mätningarna blivit sämre när det gäller att skatta stödet för enskilda partier. Historien visar flera exempel på stora avvikelser för enskilda partier (exempelvis 1968, 1985 och 1994). Väljer man att ha 70-talsvalen som utgångspunkt kan man med visst fog hävda att instituten har blivit sämre på att skatta enskilda partiers stöd i samband med riksdagsvalen.

Givet de allt större utmaningarna, som tilltagande täckningsbrister, fallande svarsfrekvenser, en allt rörligare väljarkår och ett ökat antal partier innebär, så lyckas instituten alltså väl i samband med valen vid en historisk jämförelse.

NICKLAS KÄLLEBRING, IPSOS

Debatt kring artiklar i Qvintensen

■ I förra numret av Qvintensen (2018:3) fanns bland annat följande två artiklar: (1) På väg mot goda skattningar av Åke Wissing, och (2) Bortfallet ökar i undersökningar av Dan Hedlin. Det har nu kommit synpunkter på innehållet i dessa artiklar från två läsare var för sig, Carl-Erik Särndal och Sixten Lundström. Även SMIF har kommit med ett inlägg. (SMIF är en branschorganisation för undersökningsföretag i Sverige.) Debatten Särndal-Lundström-SMIF-Wissing återspeglar en skillnad i synsätt mellan två olika surveykulturer. Å ena sidan den kultur som kännetecknar officiell statistikproduktion, och å andra sidan den kultur som kännetecknar kommersiella opinionsinstitut.

Debattinläggen återges i det följande. För att läsaren lättare ska kunna orientera sig har framställningen strukturerats på följande sätt, i fyra separata sektioner.

1. Särndals kommentar till Wissings artikel → Wissings svar.
2. Lundströms kommentar till (huvudsakligen) Wissings artikel → Wissings svar.
3. SMIF:s kommentar till Lundström → Lundströms svar.
4. Lundströms kommentar till Hedlins artikel → Hedlins svar → Lundströms kommentar → Hedlins svar.

Här följer nu debattinläggen.

SÄRNDAL:

Att testa opinions- under- sökarnas mätningar

■ ■ ■ På väg mot goda skattningar är rubriken på en artikel i Qvintensen nr. 3 2018 där surveyföreningens ordförande Åke Wissing (ÅW) genomför ett test ämnat att bedöma noggrannheten i opinionsundersökarnas mätning av partisympatierna, deras sista inför septembervalet 2018.

ÅW säger ”Trots felkällor som så gott som omöjliggör vetenskapligt förankrade skattningar visar opinionsundersökningarna att de idag ändå klarar av att ge en acceptabelt god bild av nuläget bland allmänheten.”

Stödd på sina testresultat säger ÅW: ”Vi kan konstatera att flertalet av undersökningsföretagen har resultat som ligger inom eller mycket nära den teoretiska felmarginalen – trots att vi varken har en ideal urvalsram, ett korrekt sannolikhetsurval eller ett lågt bortfall.”

För tillförlitliga siffror om en stor population fordras avancerad metodik. Cirka hundra år av vetenskaplig och praktisk utveckling har lärt oss ganska bra hur man gör det. Göra gällande att man kan ta fram fakta om populationen utan vetenskapligt förankrad metod, det är en attityd som lämnar fältet öppet att publicera siffror som är rena gissningar, fake news. Opinionsundersökarna – eliten bland dem i alla fall – förtjänar att framställas i fördelaktigare dager än så.

Att ta fram tillförlitliga slutsatser utan vetenskaplig metod skulle ge utsikter till snabb bra statistik för lågkostnadspris, om det vore sant. I detta inlägg granskar jag ÅW:s test. Testet är intressant, men med sättet han utför det blir resultaten intetsägende. Jag skisserar ett mera rättvisande utförande av idén i hans test, ett som kan tjäna som verktyg för att skilja de bra undersökarna från de mindre bra.

”Vetenskapligt förankrade” kan surveyskattningar vara på olika sätt. Oavsett hur, så är sådan förankring en viss garanti för vederhäftighet och tilltro. Vetenskaplighet är inte förbehållet undersökningar med riktigt sannolikhetsurval och frånvaro av bortfall. Det finns rik vetenskaplig litteratur som behandlar estimation under icke-sannolikhetsurval och/eller närvaro av bortfall. Det är fritt för



På väg mot goda skattningar
– ur Quintensen nr 3 2018.

undersökare, inklusive de privata instituten, att studera den litteraturen och använda den i sin verksamhet.

Wissings test

Tanken bakom ÅW:s test är enkel och klar; han frågar sig: Om en tänkt undersökning hade gjorts under de ideala betingelserna med perfekt urvalsram, riktigt sannolikhetsurval ifrån den, och frånvaro av bortfall (100% svar), skulle då undersökarnas faktiskt erhållna mätningar (gjorda utan dessa ideala betingelser) bestå provet, inte skilja sig (signifikant) ifrån vad man förväntar sig i vad som han kallar ”en helt perfekt och ideal undersökning”?

ÅW gör ett tankeexperiment: Han tänker sig en undersökning gjord med sannolikhetsurval av n personer (från registret över den röstberättigade svenska befolkningen), och med 100% svar, med syfte att skatta proportionen väljare för ett givet parti. Mer specifikt, han antar sannolikhetsurval enligt simple random sampling (obundet slumpmässigt urval) av n personer, som alla svarar, och att skattningen av ett partis väljarandel görs väntevärdesriktigt med det enkla medeltalet (proportionen)

$$\sum_{i=1}^n y_i/n \quad (1)$$

där $y_i = 1$ eller 0 beroende på om personen i röstar eller inte på partiet i fråga. Under dessa betingelser kan partiets andel av de röstande i valet 2018 tillskrivas ett prediktionsintervall $P \pm 1.96\sqrt{PQ/n}$, centrerat på partiets faktiska valresultat $P = 1 - Q$, med PQ/n som en välmotiverad skattning av variansen, och med en prediktionsgrad mycket nära 95%.

ÅW kallar $\pm 1.96\sqrt{PQ/n}$ för den (teoretiska) felmarginalen. Som han nämner, för ett parti med $P = 20\%$, och när $n = 1000$, är felmarginalen $\pm 2,5\%$; om $P = 4\%$ och $n = 1000$ är den $\pm 1,2\%$

Med P lika med valresultatet för ett parti i september 2018, och med $n = 1000$, de privata institutens schablon, beräknar ÅW prediktionsintervallet. I en grafik för varje parti – åtta stycken i Figur 1 – avsätter ÅW valresultatet

P på vertikalaxeln och förser det med ett horisontellt osäkerhetsband, $P \pm 1.96\sqrt{PQ/n}$, för att se om sista-mätningarna (nio staplar på horisontella axeln, de åtta kommersiella institutens samt SCB:s, gjord redan i maj 2018, med större urval än 1000) ligger inom bandet eller ej. De flesta ligger inom.

När nu testet utfaller fördelaktigt (mätning inom intervallet) för en stor majoritet av alla fallen så anser ÅW – man får tolka honom så – att betingelserna ideal urvalsram, korrekt sannolikhetsurval och lågt bortfall är inte nödvändiga för tillförlitlig statistik, det går bra ändå. Och när skaran av undersökare nu klarar hans test bra, så gratulerar ÅW hela den nationella branschen: ”Opinionsundersökarna i Sverige har visat att de klarar av de stora praktiska problem som har uppstått” genom stora bortfall och oförmåga att realisera riktiga sannolikhetsurval.

Ett intetsägende test

Det är näst intill självklart att majoriteten av institutens mätningar ska komma att falla inom det breda intervallet $P \pm 1.96\sqrt{PQ/n}$. Vi behöver inget test för detta; det är att slå in öppna dörrar. Många av värdena i intervallet är inte mycket bättre än ren gissning, utifrån rudimentär kunskap om det politiska landskapet.

Intervallet är byggt på de triviala omständigheterna simple random sampling och skattning med det oviktade medeltalet (1); därmed är $\pm 1.96\sqrt{PQ/n}$ inte en rättvisande felmarginal.

ÅW:s ambition är att se ”vad man skulle kunna förvänta sig vad gäller statistisk säkerhet om man hade kunnat genomföra en teoretiskt helt perfekt och ideal undersökning, helt i linje med de vetenskapliga krav som den sannolikhetsbaserade urvalsteorin baseras på”.

Simple random sampling och skattning med enkelt medeltal ger inte uttryck för ”en teoretiskt helt perfekt och ideal undersökning” under villkoren som gäller här. Dessa tillåter mycket effektivare sätt att realisera ”de vetenskapliga krav som den sannolikhetsbaserade urvalsteorin baseras på”. Testet ska göras mer informativt. Då ska man finna att många av institutens mätningar faller utanför ett korrekt konstruerat prediktionsintervall. Det

»»» blir att skilja agnarna från vetet, de bra undersökarna från de mindre bra, även om det ändå inte helt eliminerar tillfälligheternas spel, nämligen att en viss undersökare kan "lyckas bra" just i en enskild mätning.

Testet kontra verkligheten

ÅW jämför undersökarnas mätningar med de sämsta möjliga villkoren för sannolikhetsurval och väntevärdesriktig skattning, simple random sampling och det primitiva oviktade medeltalet (1).

Men i verkligheten är institutens teknik är långt mer raffinerad, i flera avseenden, det vet vi, även om vi inte får reda på detaljerna i hur de gör. De har dels mer kontrollerad datainsamling, dels mer avancerad skattning. Båda är uttryck för vetenskaplig metod, låt vara med landvinningar gjorda för länge sedan. Av allt att döma gör instituten så här:

1. Data på surveyvariabeln y (med värde 0 eller 1) hämtas in från en mängd svarande (av storlek $n = 1000$) som har tillrättalagts, kanske genom kvoturval för att respektera bestämda proportioner i grupper, en långt mer kontrollerad och styrd mängd svarande än ett mycket oförutsägbart utfall av simple random sampling.

2. Estimationen utförs med viktad skattning av typ

$$\sum_{i=1}^n w_i y_i / \sum_{i=1}^n w_i, \quad (2)$$

där vikterna w_i har beräknats på hjälpinformation. Viktningen korregerar till god del för slumpmässig variabilitet, för avsaknad av sannolikhetsurval, och för snedvridande bortfall. Skattningen är därigenom mycket säkrare, "närmare sanningen", än det oviktade medeltalet.

3. Mängden svarande, schablonmässigt fixerad till 1000 svar, är ett resultat av att bortfall uppstått i ett betydligt större urval, kanske från en panel. Till exempel, ett verkställt urval av 4000 kan ha gett, med 25% svar (75% bortfall), en svarmängd av storlek precis 1000. Urvalsmängden och svarandemängden är två skilda begrepp, grundläggande för teorin. Prediktionsintervallet ska beräknas inte med $n = 1000$ utan ta hänsyn till storleken på det mycket större urval som gjordes innan bortfallet inträffade.

Härtill kommer att skattningarnas inneboende variabilitet kanske reducerats genom att undersökaren använt sig av förfarandet känt som herding.

Den verklighet som 1 till 3 uttrycker gör att det triviala prediktionsintervallet $P \pm 1.96\sqrt{PQ/n}$ blir alldeles för brett, så nästan alla mätningarna faller där inom. Testet gjort på det sättet utgör inget bevis (som ÅW verkar säga) för att vi klarar oss utan vetenskapligt förankrade metoder från den sannolikhetsurvalsbaseade teorin.

Att göra testet rättvisande

Jag skisserar ett mera rättvisande test grundat på ÅW:s tankeexperiment. Utifrån ett sannolikhetsurval av storlek n och med 100% svar bygger prediktionsintervallet då på en (nästan) unbiased skattning av samma viktade form och med samma hjälpinformation som undersökaren lägger in i sina vikter. Välkända och allmänt använda sådana är GREG (generalized regression estimator) och kalibreringsskattning. Skattningen är av form (2), där $w_i = d_i g_i$ är produkten av d_i = inverterade värdet av inklusionssannolikheten för person i (i ett sannolikhetsurval som helst ska likna hur en undersökare gjort; svårt att få reda på i efterskott, givetvis) och vikt faktorn g_i , som för person i kalibreras fram på precis samma hjälpvariabelinformation som undersökaren förklarar sig ha lagt in i sina vikter w_i .

Informationen kan bestå av kända storlekar på populationsgrupper, kön $\times$ åldersgrupp $\times$ diverse andra indelningar.

Prediktionsintervallet för ett parti, på nivå approximativt 95%, är nu $P \pm 1.96\sqrt{\hat{V}}$, där P är valresultatet för partiet och $\hat{V}$ är variansestimaten för skattningen, under sannolikhetsurval av n personer och utan bortfall. Här beräknas $\hat{V}$ med de välkända resultaten om skattning under hjälpinformation, GREG eller kalibreringsskattning. Olika undersökare får olika värden på $\hat{V}$, och därmed olika intervall, eftersom de använder olika mycket hjälpinformation.

Testet består då i att se om undersökarens sista-mätning för partiet i fråga ligger inom $P \pm 1.96\sqrt{\hat{V}}$. Den springande punkten är att intervallets bredd $\pm 1.96\sqrt{\hat{V}}$ kan vara betydligt mindre än den triviala och missvisande $\pm 1.96\sqrt{PQ/n}$. Hjälpinformation reducerar skattningens varians. Intervallet blir kortare. Många av institutens skattningar ska komma att falla utanför.

Om någon nu invänder att vi kan inte beräkna intervallet i tankeexperiment på det där sättet – om till exempel en undersökare inte vill lämna ut informationen i fråga – ja, då får vi anse ett meningsfullt utförande av ÅW:s test ogenomförbart. Det var en bra tanke i alla fall. Men vi ska inte använda triviala förenklingar av den goda tanken.

Om skillnaders storleksordning

En skillnad, för ett politiskt parti, på en eller två procent från ett konstaterat valresultat, eller som en förändring under en period mellan val, vad betyder den? ÅW inbjuder oss med sitt test att reflektera över vad "komma nära" innebär, över storleksordningen på ganska små procentuella skillnader eller förändringar. Är en eller två procentenheter "något att bry sig om"? Är resultatet signifikant? Det är två väsentligt skilda frågor.

Ord som "statistiskt säkerställt" och "inom felmarginalen" kan förekomma i kommentarerna i media. De är ofta meningslösa. Komplexiteten i datainsamling och estimationsförfarande gör att riktiga mått på dessa statistikteoretiska begrepp måste baseras på teoretisk insikt djupare än den elementära som ligger bakom den naiva formeln $\sqrt{PQ/n}$.

Att en opinionsundersökare kan "komma hyggligt nära sanningen", resultatet på valdagen, med bara en eller två procents skillnad, det är ju, ur ett visst perspektiv, ett bra resultat. Vi ser det ofta, och det kan imponera på många iakttagare. Men sådan närhet är alls inte förvånande, givet den driven teknik – baserad på längesen etablerad vetenskap – som undersökarna använder: väl representativ mängd svarande, driven teknik i de viktade estimaten. Det är ett hantverk, enkelt för nya aktörer på marknaden att tillägna sig, och kräver ingen djupare kunskap om statistisk metodik.

Å andra sidan är en sådan skillnad ett mycket dåligt resultat, nämligen om vi gör vetenskaplig prövning åt hållet som ÅW vill se: Vi utsätter den där skillnaden på såg en procent för ett formellt hypotesprövningstest och finner den högst signifikant skild från noll. Skattningen får en formell felstämpel.

Allmänheten, eller till och med erfarna samhällsvetare, noterar att vissa undersökare "har lyckats", har kommit nära valresultatet. Vad "nära" egentligen innebär, och vad det är, i en viss undersökares speciella metodkunskande, som ger överlägsna resultat, dessa frågor med intellektuellt innehåll bryr man sig inte om.

Politiska opinionsundersökningar gäller nu inte "en fråga om liv och död", som kan vara fallet i samband med kliniska experiment, ny kontroversiell kirurgi eller revolutionerande läkemedel.

Och i statistik om samhället, en förändring som kan få konsekvenser på hög politisk nivå är om arbetslösheten, som den mäts i SCB:s AKU (Arbetskraftsundersökningen), skulle konstateras gå upp två procent från ett tillfälle till ett annat; närmast "ett jordskred". Ansvarvilar tungt

WISSING:

Svar till Särndal

på SCB att gå i god för metoderna bakom siffran och tillförlitligheten i den.

Officiell statistik och opinionsundersökningar är två skilda världar. Undersökningsvariabeln "Arbetslös" är en objektiv, av ILO definierad, variabel, inte en "opinionsvariabel".

SCB:s AKU görs med sannolikhetsurval, men utan att komma nära de ideala 100% svar i urvalet. Icke dess mindre har SCB metoder – och de tvekar nog inte att kalla dem vetenskapligt förankrade – för hantering av den bortfallssituationen, med känd teori och internationellt fastställda normer för arbetskraftsundersökningar.

Sist i sin artikel framställer sig ÅW som en opinionsundersökarnas beskyddare: "De belackare som ondgör sig över kvaliteten i opinionsundersökningar i Sverige borde söka sakunderlag för sin kritik innan de visar upp sin okunskap om en bransch som de egentligen inte kan uttala sig om." Bland oidentifierade belackare anar läsaren akademiker som forskar på estimation under sannolikhetsurval, eller användare, SCB och andra, som praktiserar sannolikhetsurval för landets officiella statistik. Eller andra som bedöms "okunniga".

I Sverige finns en kuriös klyfta "opinionsundersökarna kontra de andra". Den skärps genom en artikel som ÅW:s. Tonen i den svenska surveyvärlden förvånar mig.

Min karriär var mestadels i länder där jag upplevde att sådan schism saknades. För mig stod det klart: De som framställer landets officiella statistik var en grupp, opinionsundersökarna en annan, den senare perifer för mig, men jag såg inte någon sorts prestigemättad tävlan eller konflikt mellan de två. Min uppgift låg hos de nationella statistikbyråerna, i att vidareutveckla teorin bakom framställning av officiell statistik. Jag och mina medarbetare i olika länder bidrog till den vetenskapliga kultur som idag präglar den miljön.

CARL-ERIK SÄRNDAL

■■■ Tack Carl-Erik. Jag är ytterst tacksam för att du har fördjupat dig i min artikel och engagerat dig för att göra dina tillrättalägganden. Jag måste ha uttryckt mig mycket slarvigt eller felaktigt om du har uppfattat att min intention har varit att genomföra någon form av test. Därför är jag naturligtvis tacksam för ditt inlägg när du skriver *"Jag skisserar ett mera rättvisande utförande av idén i hans test, ett som kan tjäna som verktyg för att skilja de bra undersökarna från de mindre bra"*. Jag har personligen ingen ambition att betygsätta undersökningsföretag som bra eller mindre bra. Jag skriver redan i ingressen att opinionsundersökarna, trots de stora problem som föreligger idag, *"ändå klarar av att ge en acceptabelt god bild av nuläget"*. Och i rubriken använder jag orden *"På väg mot ..."*.

När jag har läst din artikel inser jag att du reagerar över att jag åskådliggjort min illustration med ett grått fält som jag kallar felmarginal. Dels borde det ur ett statistiskt perspektiv hellre kallas prediktionsintervall, dels kunde det ha gjorts smalare.

Jag uppfattar ordet felmarginal som något mer lättförståeligt än konfidensintervall och prediktionsintervall. För beräkning av intervallet har jag använt samma enkla formel som är mest vanlig bland opinionsundersökarna och som du vill beteckna som naiv. Av de fyra undersökarna i min artikel som använder sig av telefonintervjuer är det ett företag som har tagit det strategiska beslutet att inte publicera några mått på felmarginaler (eller intervall). De har en saklig och välgrundad motivering till detta. De övriga tre använder sig av samma enkla formel som jag har gjort. Till detta kommer att dessa tre undersökningsföretag har gjort fler än de 1 000 intervjuer jag har valt i mitt förenklade exempel (1 673, 1 847 respektive 3 480 respondenter). Jag är helt medveten om att jag hade kunnat prestera smalare grå fält men det är inte den typen av sifferexercis som jag varit intresserad av. Jag är positivt överraskad av att de opinionsföretag som redovisat sina resultat från mätningarna före valet har kommit så nära det slutgiltiga valresultatet. På 43 av de 64 mätpunkter som redovisas ligger resultaten mindre än 2 % från valresultaten (dvs. 2 av 3 mätpunkter). 38 mätpunkter (59 % av punkterna) skiljer sig med 1 % eller mindre från valresultaten!

Jag är övertygad om att vi har en gemensam syn när det gäller att bortfallsproblemet idag är ett verkligt problem för oss surveystatistiker. Jag är också rörande överens med dig när du refererar till Tore Dalenius om att vi "silar mygg och sväljer kameler" eftersom det lilla är samplingfelet och det stora är alla de andra felen (din presentation i Örebro 2012). Exempelvis har vi vid de 9 redovisade partisympatimätningarna (inkl. SCB) 7 olika frågekonstruktioner samt en hel del andra, mer eller mindre viktiga, skillnader. Vilka fel dessa olikheter innebär vet vi inte men det understryker din målande beskrivning när du jämför surveymetodik med ett lapptäcke av "practices backed by some theory" här och där – eller Robert Groves (1987): "Survey research is a methodology without a unifying theory".

Vi är också båda medvetna om att det har förelegat stora olikheter i förutsättningarna för partisympatiundersökarna 2018 jämfört med deras mätningar inför valet 2014. Exempelvis har de flesta undersökarna som 2014 använde telefonintervjuer nu kompletterat med webbenkäter från olika typer av paneler.

Jag tror också att vi är överens om att vi står inför eller befinner oss i ett paradigmskifte. Och då kanske vi behöver tänka nytt.

Jag har ingen lösning på hur vi ska ta oss vidare eller bedöma vad som är rätt eller fel, men som surveystatistiker blir jag mycket glad över att se hur nära valresultaten flera av våra opinionsundersökningsföretag har kommit de slutgiltiga valresultaten. Kanske kan vi vara: "På väg mot goda skattningar"?

Avslutningsvis vill jag förtydliga vad jag syftar på när jag har kommenterat "belackare som ondgör sig över kvaliteten i opinionsundersökningar". Det handlar INTE på något sätt om

»Jag är positivt överraskad av att de opinionsföretag som redovisat sina resultat från mätningarna före valet har kommit så nära det slutgiltiga valresultatet.«

»»»

»»» akademiker/forskare, SCB eller andra som praktiserar sannolikhetsurval. Det handlar om dem som utan sakkunskap kritiserar undersökningar. Det förekommer exempelvis att man skyller på fel i undersökningen när man får obekväma resultat. Ännu vanligare är att medier gärna slår upp stora rubriker om de ser resultat som kan locka läsare. Som exempel kan jag här nämna parlamentsvalet i England 2015, Brexit-valet 2016 och presidentvalet i USA 2016. För Brexit-valet var skillnaden i valresultat 3,8 % (51,9 % mot 48,1 %) och vid Presidentvalet i USA var skillnaden 2,1 % (48,2 % mot 46,1 %). I båda fallen var således utgången mycket svårbedömd men det skapade mycket stora rubriker om osäkerheten i de föregående opinionsmätningarna – inte bara i England respektive USA. Däremot skrevs det inte mycket om undersökningsresultaten efter Parlamentsvalet i Nederländerna 2017, Presidentvalet i Frankrike 2017 eller Stortingsvalet i Norge 2017. Den för mig mest störande belackaren är Daniel Nordström – chefredaktör och ansvarig utgivare för VLT – när han för ett år sedan (februari 2018) förbjöd alla VLTs mediekanalet att skriva om opinionsundersökningar inför valet. Hans motivering var att de inte går att lita på: ”Vi publicerar fakta inte spekulationer”. Hans generalisering inkluderar således även SCBs majmätning 2018. Personligen undrar jag hur han ser på alla spekulationer som publicerades från valdagen fram till regeringsbildningen, eller artiklar som informerar om kommande evenemang o.dyl.

Jag har under åren haft relativt mycket kontakter med SCB liksom med olika opinionsundersökare i Sverige. Jag har inte vid något tillfälle märkt någon tendens till klyfta ”opinionsundersökare kontra de andra”. Inte heller någon tendens till prestigemättad tävlan eller konflikt mellan de två. Jag hoppas för båda ”sidornas” skull att jag har rätt i min uppfattning.

ÅKE WISSING

[Wissings svar är baserat på en tidigare version av Särndals kommentar. Därför finns det i Wissings svar några citat från Särndal, som inte stämmer helt med den nuvarande versionen av Särndals kommentar. Jag beklagar att jag inte hunnit rätta till detta. JW]

DEBATT 2: LUNDSTRÖM – WISSING

LUNDSTRÖM:

Kommentarer till 2018:3

■■■ När jag läser Qvintensen påminns jag om ett sammanträde 1970 på SCB med två riksdagsmän. De var utskickade av riksdagen för att diskutera en eventuell partisympatiundersökning. Riksdagen var bekymrad över att det hade dykt upp institut som kunde drivas mer av lönsamhet än av vetenskaplig hederlighet. De befarade att dessas partisympatiundersökningar kunde påverka de allmänna valen och därmed skada demokratin. Riksdagen gav efter en tid SCB i uppdrag att regelbundet genomföra en partisympatiundersökning (PSU).

Vad har hänt under de närmare 50 år som gått? Starten av PSU ledde inte till att institutens mätningar ersattes, utan i stället dök det upp fler institut och mer frekventa mätningar. Jag tror att de två riksdagsmännen skulle vara bedrövade över utvecklingen, särskilt när de inser att metoderna saknar vetenskapligt stöd.

Jag är också bedrövad. Under alla mina år på SCB har jag haft välutbildade surveysstatistiker i min omgivning och vi har alltid använt metoder baserade på vetenskapen. Vi har också dokumenterat metoderna utförligt. Jag ser inget av detta hos instituten.

Jag har också skrivit en doktorsavhandling om hur man i estimationen reducerar urvals- och bortfallsfelen i ett sannolikhetsurval. Förutom närheten till teorin har jag också erfarenhet av bortfallshantering i en mängd undersökningar och efterföljande dokumentation. Nu har instituten t.o.m. intagit Statistikfrämjandets finaste arena, Qvintensen, och då kan jag inte vara tyst.

Jag har arbetat som metodstatistiker i drygt 40 år och tillhör skolan av surveystatistiker som baserar skattningar på sannolikhetsurval. Registret över totalbefolkningen (RTB) utgör vanligtvis urvalsram, och tack vare den löpande uppdateringen av registret är täckningsfelet små. Bortfallsfelet har däremot blivit det stora problemet. Över åren har bortfallet ökat på ett oroande sätt; exempelvis i Arbetskraftsundersökningen från ca 2 % till över 50 %. Men under mina år utvecklades nya estimatorer som kunde utnyttja den ökande mängden hjälpinformation till att reducera bortfallsfelet (och urvalsfelet). För individundersökningar kunde man välja bland mer än 200 registervariabler att användas i GREG-estimatorn eller kalibreringsestimatorn.

När jag nu följer utvecklingen inom ”surveyvärlden” blir jag förfärad över att man okritiskt och lättvindigt lämnar all teori som utvecklats under drygt hundra år med hänvisning till det stora bortfallet. I det senaste numret av Qvintensen publicerar Statistikfrämjandet flera artiklar som hävdar att opinionsinstitutens metoder fungerar väl. Mest handlar det om årets val, men presumtiva beställare av undersökningar tror säkerligen att metoderna fungerar även för andra ämnen. Det bakomliggande resonemanget är att bortfallet nu är så stort att sannolikhetsurval har spelat ut sin roll och att det nu är fritt fram att hitta på vad som helst. Nu kan man t.o.m. basera skattningarna på självrekryterade webbpaneler!

Jag hävdar fortfarande att sannolikhetsurval och utnyttjande av effektiv hjälpinformation i estimatorn är den bästa metoden. Det är en metod som har stöd av samplinglitteraturen och som inte är rädd för dagsljuset. Opinionsinstitutet däremot mörkar sina förfaranden med hänvisning till ”affärshemligheter”. Deras dokumentation är undermålig och därför är det helt omöjligt att värdera deras metoder. Ett institut kan exempelvis hitta på vilka vikter som helst utan att någon utomstående har möjlighet att se om de har vetenskapligt stöd.

surveyartiklar i Qvintensen

I avsaknad av metodbeskrivningar utvärderar artikelförfattarna i Qvintensen institutens förfaranden genom att jämföra skattningarna av partisympatier med valresultatet och finner bl.a. att "... *ingen köpare eller användare av surveyundersökningar torde kunna ha högre krav på statistisk säkerhet än de som erbjuds av flera av våra opinionsundersökare.*" (Å. Wissing i artikeln *På väg mot goda skattningar*; ÅW i fortsättningen.)

Jag vet att en metod inte kan utvärderas av ett enda utfall, men här verkar artikelförfattarna mena att det är möjligt. Dessutom begränsar man sig inte bara till partisympatier utan menar också att metoderna fungerar för vilket ämne som helst. Det här sättet att utvärdera statistiska metoder överensstämmer väl med journalisters och andra metodstatistiskt okunniga personers sätt. Därför kan resultaten på ett förödande vis övertyga presumtiva statistikbeställare. Vore detta inom politik skulle man kunna beskylla Qvintensen för "populism".

Låt oss titta på ett exempel på vad som kan hända när ett enda utfall används för att utvärdera metoden.

Exempel 1. Sentio lyckades tydligen bäst med att mäta partisympatierna för SD i valet 2014. En utvärdering enligt ÅW:s metod efter det valet skapar då hos läsarna en stor tillit till Sentios förmåga att skatta andelen SD-sympatisörer. Vid valet 2018 visade sista Sentiömätningen före valet att SD skulle få 24,0 procent. I verkligheten blev resultatet endast 17,5 procent! Sentios metod var tydligen inget att lita på.

Redaktören skriver att "... *ett allmänt intryck är att samtliga lyckades rätt bra, så bra man nu kan begära på ett så här svajigt område.*" De synbara goda resultaten (för alla utom YouGov och Sentio) tror jag beror på "herding" och att partisympatierna förändrades bara lite under perioden från maj till valdagen.

Första gången jag mötte begreppet herding var i en film av Novus. Där beskrivs de problem som instituten hade vid senaste presidentvalet i USA och valet i UK som ledde till Brexit. Ett av skälen till misslyckandet säger Novus var att "en blind ledde en blind". Inget av länderna hade någon överlägset bra undersökning att luta sig mot. Alla institut lutade sig mot varandra, vilket gav en skenbar säkerhet i skattningarna, men i själva verket gav det dåliga skattningar och samma riktning på felet. Underförstått i filmen är att i Sverige är herding effektivare. Vilket facit använder Novus och de andra instituten? Jag tror att instituten använder SCB:s partiundersökning i maj för att bilda vikter som ger skattningar som väl överensstämmer med SCB:s.

I tabellen i Redaktörens ruta ser man att SCBs skattningar i maj överensstämmer väl med valresultatet, vilket tyder på att partisympatierna förändrats bara lite över tiden. Det innebär att även ett institut med en perfekt metod – sannolikhetsurval och inget bortfall – skulle behöva

ett stort urval för att finna signifikanta förändringar. Men instituten har hittat signifikanta förändringar, hävdar någon. Tyvärr är dessa uppgifter inte att lita på. Beräkningen av signifikanser har gjorts under antagandet att svarmängden är ett obundet slumpmässigt urval från populationen, vilket aldrig varit fallet.

Kan det vara så att partisympatierna i populationen knappt har förändrats alls över tiden och att variationen i skattningarna beror på urvalsfel, skeva urval, bortfallsfel, dåliga estimatorer, mätfel, etc., och att det har tolkats av journalister och statsvetare som verkliga förändringar?

En del kollegor menar att det är väl inte så farligt att instituten håller på med detta. Någon kallade det för en "lekstuga för ledarskribenter". Framträdande statsvetare menar dock att mätningarna är viktiga för demokratin.

Samhällsnyttan med mätningarna av partisympatier kan man ha olika uppfattningar om. Vad mig beträffar ser jag den stora faran i att de metoder, som vinner tillit i bl.a. här diskuterade artiklar, kommer att användas i andra undersökningar som utgör underlag för politiska beslut. Den låga kvaliteten kan också leda till en allmän misstro till alla urvalsundersökningar och därmed även dra med sig officiell statistik i fallet.

De vikter man har bestämt med herding passar inte i andra sammanhang. Ett exempel kan illustrera detta:

Exempel 2. År 2016 och 2017 gjorde Inizio (åt Aftonbladet) och Ipsos (åt DN) en skattning av andelen vuxna som tycker att Sverige bör gå med i Nato. Med ledning av skattningarna skriver DN att andelen som vill att vi går med i Nato har minskat med 4 procentenheter, och Aftonbladet att det har ökat med 6 procentenheter. Båda tidningarna har naturligtvis blivit övertygade om att de institut de anlitar producerar tillförlitlig statistik.

Månader före valet hade jag många kontakter med DN-journalister för att visa att de bröt mot sina skrivna riktlinjer i upphandlingen av statistik från Ipsos partisympatiundersökning. Riktlinjerna kräver att medicinska resultat måste vara publicerade i välrenommerade tidskrifter, men däremot ställs inga sådana krav på opinionsundersökningar. Vissa villkor måste dock vara uppfyllda: T.ex. "*Det spelar ingen roll om antalet svarande ... är väldigt stort – om svarsfrekvensen är väldigt låg.*" I studien presenterad i ovanstående exempel har DN fått information om att svarsfrekvensen är 70-80 procent (gäller andra stegets kvoturval), men en korrekt uppgift är 1 procent. Liknande fel kan man hitta i många instituts information. Detta visar hur svårt det kan vara även för en seriös aktör att följa egna riktlinjer.

Även om jag visade på många allvarliga brister i Ipsos metoder fortsatte DN att publicera deras skattningar. Det verkar som om journalister redan är så beroende av det löpande flödet av "siffror" att det krävs starka krafter

»»»

»»» för komma till rätta med spridandet av desinformation. Dessutom ger statsvetare (omedvetet) stöd till institutens förfarande.

Idag, när fakta är viktigare än någonsin att försvara, är det sorgligt att se att Statistikfrämjandet låter sig utnyttjas på detta sätt. Massmedia är redan övertygade om att opinionsinstitutet har bra metoder, och nu blir de ändå mer säkra. Även SCB är tyst. SCB torde även fortsättningsvis använda de förfaranden jag försvarat, men verket borde känna ett vidare ansvar för samhällsinformationen och inte tyst åse den ökande mängden av undermålig statistik.

Vad kan vi vänta oss av framtiden? Mycket hänger på hur framtidens surveystatistiker är utbildade. Dan Hedlin, professor vid Stockholms universitet, tror *”att synen på slumpmässigt urval kommer att ändra sig under de närmaste årtiondena”*. Vilka kurser kommer universitetet att hålla? Kommer kurser om sannolikhetsurval att upphöra? Jag tror, som sagt, att sannolikhetsurval med effektiv hjälpinformation är den bästa metoden även vid stora bortfall. Innan universitetet överger metoder som ufoforskats under ett århundrande, bör man kunna hänvisa till artiklar publicerade i välenomrade tidskrifter som visar att det finns bättre metoder.

Jag har här skrivit om opinionsinstitutet, som om de utgör en homogen grupp, men naturligtvis finns det kvalitetsskillnader. Därför kan någon känna sig orättvist kritiserad. (Ipsos har blivit utpekad p.g.a. mina kontakter med DN, som anlitar Ipsos; inte av något annat skäl.). Till mitt försvar hänvisar jag till den bristfälliga dokumentationen av metoderna och att artiklarna i Qvintensen inte gör någon åtskillnad utan menar att svenska opinionsinstitut, utan utpekade undantag, har bra metoder.

Jag skulle kunna ge en mängd exempel på hur illa beställt det är med den metodstatistiska kompetensen i opinionsinstitutet, men tills vidare räcker det nog med det jag visat. Institutet lär inte hålla med mig utan komma med repliker, och då kan jag få tillfälle att ytterligare förklara skälen till min hårda kritik. Institutet kan gärna fundera på varför jag utser begreppet ”representativt urval” till det mest missbrukade i deras dokumentation.

Låt oss titta på ett exempel som berör detta begrepp.

Exempel 3. YouGov använder ett urval som är ”representativt för den svenska befolkningen vad gäller kön, ålder, region och röstfördelning vid riksdagsvalet i september 2014”. Ett sådant uttalande ger en överdriven tilltro till metoden; statistik användaren tror att metoden ger bra skattningar av partisympatier. I verkligheten var deras skattningar av sympatierna inför 2018 års val för M och SD mycket avvikande från valresultatet - underskattning med 3,3 procentenheter resp. överskattning med 6,3 procentenheter.

Slutligen vill jag kommentera ÅW:s kritik av belackarna:

”De belackare som ondgör sig över kvaliteten i opinionsundersökningar i Sverige borde söka sakkunskap för sin kritik innan de visar upp sin okunskap om en bransch som de egentligen inte kan uttala sig om.”

Som framgår av det skrivna är jag en som ”ondgör” mig. Jag hade gärna skaffat mig kunskap om institutens metoder, men dokumentationen är så bristfällig att det inte är görligt. Jag har via mejl och telefon försökt få mer information men misslyckats. Att som ÅW använda *ett enda* utfall som bevis för att metoderna fungerar i ”det långa loppet” saknar stöd i samplinglitteraturen. Dessutom är jag inte heller imponerad av utfallet.

Slutligen vill jag ge en rekommendation till massmedia, statsvetare och andra statistik användare: Publicera inte skattningar från institutet i tiden mellan PSU i maj och valet. Detta är två skäl. Det första är att metoderna saknar stöd av vetenskapen. Det andra skälet är att texter baserade på journalistik ger läsaren en riktigare värdering av säkerheten i uttalandena. Nu låter man ofta institutet uttala sig om t.ex. skälen till förändringar i partisympatier, vilket aldrig torde ha stöd i data. Läsaren får en felaktig känsla av att ”experten” vet mer än journalisten. Troligen skulle då mätningarna upphöra med den goda följden att uppgiftslämnarbördan för allmänheten minskar och att bortfallet kanske minskar för seriösa statistikproducenter.

WISSING:

Svar till Lundström

■■■ Jag är tacksam och positiv till att du, liksom jag, värnar om vetenskaplighet och kvalitet när det gäller surveyundersökningar. Det är något som jag har brunnit för ända sedan jag kom in i undersökningsbranschen 1970. Då fanns det endast ett företag i Sverige som publicerade partisynpatiundersökningar, nämligen SIFO (Svenska Institutet för Opinionsundersökningar). Forskningschef, och senare VD, var då Hans Zetterberg som bland annat under 50-talet var assistant professor vid Columbia University och 1977 utsågs till hedersdoktor vid Umeå universitet. Hans var, liksom du och jag, också mycket mån om vetenskaplighet och kvalitet. SIFO var "först på plan" att genomföra partisynpatiundersökningar i Sverige och den första publicerades 1964. Fortfarande 1970 var SIFO ensam om att publicera partisynpatiundersökningar i Sverige. SIFO:s opinionsundersökningar genomfördes på den tiden med personliga intervjuer i hemmen med hjälp av en mycket stor stab av erfarna intervjuare. Urvalet var ett så kallat basurval – ett renodlat sannolikhetsurval i två steg där man först lottade ut orter där sannolikheten att bli utvald var proportionell mot respektive ortsstorlek och sedan drog man ett sannolikhetsurval av individer inom respektive ort. Vid intervjutillfället hade intervjuaren med sig en hel uppsättning valsedlar till varje enskild respondent. Respondenten fick välja en valsedel, lägga den i ett kuvert, försluta det och sedan slänga de övriga valsedlarna så att intervjuaren inte skulle veta vilken valsedel respondenten valt. Vid varje måttillfälle (varannan vecka) genomfördes ca 1 000 intervjuer under en tvåveckorsperiod och med hög svarsfrekvens.

I en informationsskrift skrev SCB 1972 följande: "1967 började man diskutera de metoder som användes vid partisynpatiundersökningar. Denna debatt tog extra fart efter valet 1968. Mest oroade man sig för att publicering av resultat från undersökningar som inte utfördes på ett vetenskapligt sätt skulle påverka väljarkåren". Vidare skrev SCB 1972 följande: "Mot bakgrund av denna debatt erhöll SCB anslag under åren 1969 – 1972 för att bedriva forskning för att förbättra metoderna för partisynpatiundersökningar. Detta förberedande arbete är nu avslutat och SCB har fått i uppdrag av riksdagen att genomföra partisynpatiundersökningar".

Jag vill inte på något sätt ifrågasätta eller förringa ditt sammanträde 1970 på SCB med två riksdagsmän för att diskutera partisynpatiundersökningar. Det var många som redan då var engagerade i utvecklingen av partisynpatiundersökningar i Sverige. SCB hade fått anslag för att bedriva forskning kring partisynpatimätningar under perioden 1969 – 1972. I november 1970 genomfördes provmätning för att kunna starta i full skala 1972.

Jag är inte lika övertygad som du när du skriver att ni under dina år på SCB "har alltid använt metoder baserade på vetenskap" och att du förfäras över att inom surveyvärlden "man okritiskt och lättvindigt lämnar all teori som utvecklats under drygt hundra år". Min uppfattning är att man även på SCB tvingas att ge avkall på och prioritera bort vissa vetenskapliga krav i samband med att budgetar skärs ned.

Samtidigt vet jag att det finns undersökningsföretag som gör allt de kan för att följa den teori som har utvecklats under drygt hundra år. Det finns även institut som har anlitat extern vetenskaplig kompetens för att bygga upp sina mättningsmodeller och i flera fall har det varit experter med hög akademisk kompetens.

Som exempel kan nämnas Inizio som är ett av de tre företagen som avstår från att använda telefonintervjuer utan "enbart" arbetar med en så kallad webbpanel för sina partisynpatimätningar. De hade den näst bästa skattningen av valresultaten. På deras hemsida framgår att de arbetar med ett så kallat Advisory board bestående av fyra internationellt kända och välmeriterade experter. Dessutom har de till sitt team knutit Lars Lyberg som senior rådgivare. Han torde vara välkänd för de flesta läsare av Qvintensen.

Om någon har uppfattat min artikel som att "bortfallet nu är så stort att sannolikhetsurval har spelat ut sin roll och nu är det fritt fram att hitta på vad som helst" så har jag missuppfattats gruvligt. Jag är övertygad om att sannolikhetsurval INTE "har spelat ut sin roll". Jag är lika övertygad om att det INTE är "fritt fram att hitta på vad som helst".

Vi är helt överens när det gäller uppfattningen att SCB, ur ett vetenskapligt och metodologiskt perspektiv, är helt outstanding, jämfört med de privata/kommersiella undersökningsföretagen. Jag håller helt och fullt med dig om att SCB också är outstanding när det gäller att dokumentera vad och hur de har gjort. Däremot vill jag inte påstå att undersökningsföretagen "mörkar förfarandena" och inte heller klassificera deras dokumentation som "undermålig". Varken att det skulle gälla alla opinionsinstitut eller att de skulle vara så värdelösa som du vill framhäva. På flera av institutens hemsidor framgår ganska väl hur de har gjort och om man vill ha djupare information så är åtminstone de flesta öppna för att informera och svara på frågor. Jag har personligen mycket god erfarenhet av det. Jag vet att exempelvis Dan Hedlin (professor i statistik) också har god erfarenhet av att få önskad information från undersökningsföretagen. Jag kan bara beklaga att de inte visar samma öppenhet mot dig.

Ditt "Exempel 1" där du tar ett mätvärde från ett institut och jämför med vad jag skriver om i min artikel antyder varken en seriös eller vetenskaplig ansats. Att du dessutom kallar det för: "En utvärdering enligt ÅWs metod ..." tolkar jag som förtal. Jag har studerat 8 olika partisynpatimätningar med resultat från vardera av 8 olika politiska partier. Det ger mig 64 mätpunkter som kan jämföras med valresultaten, vilket tydligt framgår av min artikel. Jag skriver också att "Flertalet undersökningsföretag visade en förvånansvärt god överensstämmelse – både gentemot varandra och mot de slutgiltiga valresultaten". Jag noterar också att de aktuella undersökningsföretagens mätningar inte uppvisar några nämnvärda oförklarliga hopp i sin mätserier över tiden.

Du tar upp frågan om herding – att undersökningsföretagen bevakar varandra och gör egna viktningar styrda av vad andra undersökningsföretag har redovisat – och jämför det med ordspråket "en blind leder en blind" (Matteus 15:14). Du tror att "De synbara goda resultaten" (för alla

»»»

»»» utom två företag) beror på herding. Samtidigt tror du att instituten använder SCBs majmätning för att bilda vikter. Gäller din tro alla instituten eller bara de 6 som kom närmast valresultaten i sina skattningar? Den föreslagna liknelsen skulle innebära att man betraktar SCB som blind, vilket jag inte vill instämma i. Om vi är överens om att partisympatimätningarna från SCB är överlägsna de övriga företagens mätningar anser jag att liknelsen snarare borde vara att "en sakkunnig leder en blind". Ur statistisk synvinkel vore ju i så fall en vägning, där hänsyn tas till SCBs majmätning, snarare en fördel än en nackdel. Jag kan mycket väl acceptera epitetet "sakkunnig" för SCB men tar klart avstånd från att betrakta alla andra undersökningsföretagen som blinda. Jag tog mig friheten att tillfråga vart och ett av de aktuella undersökningsföretagen om herding generellt och om de före valet 2018 justerat/vägt sina resultat genom att ta hänsyn till SCBs resultat i deras mätning av partisympatier i maj 2018. Jag fick svar från 7 av de 8 företagen (svarsfrekvens 87,5 %). Samtliga svarande garanterade att de inte tagit någon hänsyn till SCBs majmätning när de redovisade resultat från sina respektive mätningar före valet 2018. Som surveyundersökare vill jag gärna tro på mina respondenter, speciellt när de garanterats anonymitet. Därför tolkar jag det som sannolikt att de synbara goda resultaten inte huvudsakligen förklaras av herding, åtminstone inte för alla undersökningsföretagen.

Jag har förståelse för att du värnar om sannolikhetsurval, där bortfall vid estimation hanteras genom hjälpinformation på det sätt som exempelvis GREG använder och, om jag har förstått det rätt, du har varit med att utveckla. Du skriver att du tror att "*sannolikhetsurval med effektiv hjälpinformation är den bästa metoden även vid stora bortfall*". Jag håller helt och hållet med dig när bortfallet ligger i närheten av den storleksordning där Tore Dalenius började förespråka att resultaten borde förkastas – dvs. vid ca 10 %. Och jag högaktar dina insatser som hjälper oss att förbättra skattningar genom att använda hjälpinformation och därigenom förbättra estimat när underlaget drabbats av stora bortfall. MEN. Hur stora bortfall kan vi acceptera? Hur långt håller teorierna kring att använda sig av hjälpinformation? Under gynnsamma förhållanden är SCB nu nere i ca 50 % i svarsfrekvens. I vissa fall lägre. Kan vi nöja oss med 30 % i svarsfrekvens? Varför satsar vi pengar för att hålla uppe en så hög svarsfrekvens som möjligt? Samhällsutvecklingen går emot oss och vi kan inte ändra på att bortfallet fortsätter att öka.

Mot den bakgrunden blir jag glad – nästan lyrisk – när jag kan notera att flera av våra opinionsinstitut ser ut att ha klarat av att ge relativt goda skattningar av partisympatier inför valet 2018. Därav rubriken: "*På väg mot goda skattningar*". Jag säger inte att vi har de slutgiltiga lösningarna. Men mot bakgrund av de, för oss surveystatistiker, minst sagt deprimerande rapporterna om ökat bortfall ser det ut att finnas "ett i ljus i tunneln", vilket bådär gott för framtiden. Frågan är hur vi ska ta oss dit. Det kanske är så att vi är på väg?

ÅKE WISSING

SMIF:

Svar på Sixten

■ ■ ■ Först och främst vill vi rikta ett stort tack till Sixten Lundström för att han lyfter detta viktiga ämne om opinionsmätningarnas utveckling och potentiella framtid. Mycket har hänt sedan sammanträdet på SCB 1970, men våra åsikter kring vad detta innebär skiljer sig dock åt.

I hela den öppna världen har en konkurrensutsatt undersökningsmarknad varit ett fundament för demokratin. Att flera aktörer har kunnat etablera sig och genomföra mätningar av landets medborgares tankar och åsikter kring politik- och samhällsfrågor har såväl drivit på utveckling som sett till att inte staten äger det enda mätinstrumentet. Den typen av samhällsordning där endast en aktör är tillåten finner man företrädesvis i fundamentalistiska och diktatoriska stater, inte i ett öppet och demokratiskt samhälle. Mer än ett ögonbryn i världen har höjts över att Sverige och SCB genomför en partisympatiundersökning finansierad av allmänna medel.

När vi nu slagit fast att en mångfald opinionsaktörer är bra för demokratin, är det värt att uppmärksamma det goda arbete som SCB gör. Deras partisympatiundersökning (PSU) vilar på många års erfarenhet och vetenskaplig grund, precis som anstår en statistikmyndighet. Däremot är det direkt felaktigt att påstå att alla metoder förutom de som SCB använder sig av saknar vetenskapligt stöd. Att hävda att inget opinionsinstitut dokumenterar sina metoder är ytterligare ett osant uttalande. Med bara några klick går det att få reda på hur flera av instituten genomför sina undersökningar. Om det fortfarande finns frågor har samtliga institut listat sina kontaktuppgifter, det är bara att höra av sig. Samtliga medlemmar i SMIF följer dessutom ESOMAR:s etiska regler och ett flertal har dessutom skrivit under den kvalitetscertifiering som SMIF tagit fram i syfte att öka transparens och tydlighet. Båda dokumenten finns på SMIF:s hemsida.

Precis som när undersökare fick lämna dörrknackningen, då portkoder började införas utanför hyres- och bostadsrättshusen, och det därför inte gick att komma in till respondenterna längre, har dagens teknikutveckling gjort det lika svårt, om inte ännu svårare, att nå ett representativt urval av befolkningen genom telefon. Oregistrerade kontantkortsnr, nummerpresentatörer och en allmän ovilja att spendera sin tid på att svara på frågor via telefon gör att det idag är i princip omöjligt att genomföra undersökningar på det sätt som Lundström förespråkar. Att en metod har ett

Lundströms artikel i Qvintensen

hundraårigt förflutet är ju på intet sätt en garanti för att metoden fortfarande fungerar idag. Även SCB känner av problemet. För inte så länge sedan tvingades myndigheten att ställa in en undersökning om Hushållens utgifter eftersom svarsfrekvensen med nöd klarade 20%-strecket. Däremot är det givetvis viktigt att inte ”kasta ut barnen med badvattnet”, vilket är anledningen till att de flesta institut idag har ett flertal välenommerade statistiker knutna till sig för säkra en så hög kvalitet i undersökningarna som möjligt. Att då hävda att de ”okritiskt och lättvindigt lämnar all teori som utvecklats under drygt hundra år” är ett riktigt oförsämligt uttalande.

Snarare handlar det om att undersökningsbranschen, liksom alla branscher, måste utvecklas för att fortsätta vara relevant. Det innebär att nya metoder måste utvecklas för att gamla metoder visat sig vara otillräckliga eller rent undermåliga. Därför är det inte förvånande att vi ser att OSU-baserade undersökningar minskar, medan nya metoder tar plats som innebär andra möjligheter och utmaningar. Många av Lundströms kolleger inom officiell statistik här och utomlands är nu med om att utveckla metoder för kombinationer av sannolikhetsurval och icke-sannolikhetsurval och att kombinera surveys med andra datakällor som register och big data. Inledningstalaren vid konferensen om big data och survey science i Barcelona i oktober 2018, Tom Smith, chef för Storbritanniens enhet för big data vid Office for National Statistics, gick så långt som att säga att när det gäller att skaffa information om ett koncept så är surveys en sista utväg. Andra datakällor är många gånger mycket effektivare. Attityden är likartad på många andra byråer i världen, som USA, Holland och Eurostat. Användare önskar snabba resultat, och man överlämnar datakvaliteten till utföraren. Nya metoder innebär andra sätt att redovisa resultat. Man måste förstå hur undersökningar görs idag, för att förstå hur man ska läsa den bifogade dokumentationen. Att redovisningen ser annorlunda ut betyder inte att det saknas transparens eller vilja att berätta om vad man gjort. Det betyder däremot att den som vill värdera metoden behöver ha andra utgångspunkter än tidigare.

Så vitt vi vet förekommer inte herding bland svenska institut. Att dessutom utifrån en Novus-film dra slutsatsen att skattningarna inför det amerikanska presidentvalet och Brexit-omröstningen misslyckades på grund av detta fenomen tyder på en osedvanlig ovilja att läsa och förstå vad som faktiskt hände. Flera oberoende utredningar har genomförts för att försöka förklara skillnaderna mellan prognoser och utfall vid valurnorna. De undersökningar som gjordes inför såväl det amerikanska presidentvalet som den brittiska EU-omröstningen visade att det var mycket små skillnader mellan Trump-Clinton respektive Ja-Nej, skillnader som var inom felmarginalen, som dessutom oftast är underskattad. När media då gör sin journalistiska tolkning (vilket är en metod artikelförfattaren till sist förespråkar) och korar ena sidan till vinnare baserat på

punktskattningarna, i strid med institutens rekommendationer, är det inget som kan anses vara ett misslyckande från institutens sida. Sannolikt hade det i båda fallen varit bättre om experterna fått uttala sig, snarare än journalisterna själva.

Vad gäller den utbildning som ges i statistik finns det en hel del diskussioner kring detta i exempelvis tidskrifterna ”Journal of Teaching Statistics”, ”International Statistical Review” och till och med SCB:s Journal of Official Statistics samt på branschorganisationernas hemsidor (t.ex. AAPOR) för den som bemödar sig att leta. Just nu befinner vi oss i en mellanperiod, när de gamla metoderna visar sig inte fungera tillfredsställande, även om hjälpinformation används, och nya metoder håller på att växa fram, men ännu inte har hunnit få fullt genomslag i vetenskapliga tidskrifter på det sätt som artikelförfattaren efterfrågar, även om vi dock talar om hundratals artiklar som handlar om varianter på ett förändrat surveylandskap. Det hela lämnar oss därför med ett traditionellt dilemma: Ska vi klamra oss fast vid metoder som vi ser har stora brister, och som kommer fungera allt sämre i takt med att utvecklingen går framåt och omvärlden förändras, eller ska vi söka oss till nya metoder där mycket utveckling fortfarande sker, men framtiden ser ljusare ut? För många av de institut som artikelförfattaren ondgör sig över är valet enkelt, den som inte utvecklas kommer obönhörligen att avvecklas.

Avslutningsvis: Att nära en förhoppning om att uppgiftslämnarbördan skulle minska om vi upphörde med opinionsmätningar är återigen ett tecken på den mycket begränsade förståelse för undersökningsbranschen som artikelförfattaren har. Opinionsmätningar utgör en mycket liten del av de många undersökningar som genomförs, och att upphöra med dessa skulle inte bara ha en ytterst marginell effekt på uppgiftslämnarbördan, utan skulle även innebära en mycket olycklig monopolställning för SCB. Det kan för vissa intressenter givetvis ses som ett attraktivt alternativ, men erfarenheten visar att monopol missgynnar snarare än stimulerar forskning och utveckling. Det är dessutom, som fastställdes i det inledande stycket, ett förfarande som inte gynnar det demokratiska samhället. En livskraftig undersökningsbransch är beroende av konkurrens för att se till att vårt öppna och demokratiska samhälle lever vidare och att tillförlitliga undersökningar går att genomföra även i framtiden.

SMIF:S STYRELSE:
KARIN NELSSON, INIZIO – ORDFÖRANDE
PER HÖRNSTEN, DEMOSKOP – LEDAMOT
IDA MYKLAND, YUGOV – LEDAMOT
RICHARD STRÖM, CINT – LEDAMOT
MATTIAS STRANDBERG – GENERALSEKRETERARE

LUNDSTRÖM:

Svar till SMIF

■■■ SMIF:s diskussion om demokrati handlar mest om näringsfrihet och en fungerande marknadsekonomi. Det senare kräver att kunderna har möjlighet att ställa kostnad mot kvalitet, vilket inte är möjligt inom den bransch vi här diskuterar. Institutens undersökningar är billiga, men den information kunderna får om kvaliteten är bristfällig och av reklamtyp. Det finns nog ingen bransch där det är så lätt att lura kunderna som denna, eftersom statistik framställs inom områden där fakta saknas och där resultaten sällan kan ställas mot facit. Men just inom området partisympatiundersökningar får man i samband med riksdagsvalet en möjlighet att "utvärdera" metoderna. Institutet menar att utvärderingen visar att metoderna fungerar bra inom alla områden.

Urvalen görs på en mängd olika sätt med mer eller mindre slumpmässiga inslag och utan möjlighet att beräkna inklusionssannolikheterna. Därefter vidtar viktning som ska rädda bristerna i urvalsförfarandena. Här genomför instituten ett arbete som borde kunna dokumenteras, men det enda man får reda på är vilka variabler som används i viktningen. Jag skulle kunna ställa en mängd frågor kring viktningen, men inser att det inte för diskussionen vidare. Jan Wretman, redaktör för Qvintensen, ställer åtta frågor om undersökningsmetoderna (i Qvintensen 2018:3, sid. 6), som han inte finner några svar på i dokumentationen. Jag håller helt med Jan och dessutom har jag kontaktat instituten, men utan framgång. (Varför låter ni inte "de välrenommerade statistikerna" svara på våra frågor?). Trots detta säger SMIF: "Med bara några klick går det att få reda på hur flera av instituten genomför sina undersökningar". I ESOMAR:s etiska regler står det under rubriken "Responsibilities to the general public":

When publishing research findings, researchers must ensure that the public has access to sufficient basic information to assess the quality of the data used and the validity of the conclusions.

Av en del uttalanden kan man sluta sig till hur delar av viktningprocessen går till. Karin Nilsson och Anders Lithner, båda från Inizio, kommer med en "hemlighet" (DN Debatt 2015-01-27):

"Siffrorna ni läser i media är nästan alltid knådade, länge och väl, innan de hamnar där. [...] Det kallas för att väga

datan och är en process som alla undersökningsföretag arbetar med och som kan göras väldigt avancerad. Utöver det jämför man ofta med annan data, eller med tidigare undersökningar av samma sak, och justerar sina siffror."

Man använder alltså "herding", där *flocken rättar sig efter* en källa med god kvalitet. Efter valet skriver Sentio: "Trots den dramatiska skillnaden mellan valresultat och vår sista mätning har vi utan problem kunnat identifiera vad det beror på och även kalibrerat in mätningarna efter utfallet i riksdagsvalet." Sentio använder alltså herding.

Märkligt nog skriver SMIF:s styrelse att "såvitt vi vet förekommer inte herding bland svenska institut", trots att medlemmar i styrelsen använder herding. Dessutom menar man att dokumentationen är tillräcklig, men man vet inte om instituten använder herding. Huruvida ett institut använder herding är viktigt att veta, eftersom herding ger vilka amatörer som helst möjlighet att fixa bra skattningar.

SMIF säger att det är "i princip omöjligt att genomföra undersökningar på det sätt som Lundström förespråkar". Ett märkligt uttalande, eftersom seriösa aktörer gör det. Man säger också "att nya metoder måste utvecklas för att gamla metoder visat sig vara otillräckliga eller rent undermåliga." Var finns de undersökningar som i den svenska kontexten med bra urvalsramar och en mängd hjälpinformation har varit *otillräckliga eller undermåliga*? Jag väntar på en publicerad artikel som visar detta. Ska vi förkasta dessa "gamla" metoder måste det finnas *nya metoder som är bättre*. Skulle det vara institutens gör-hur-som-helst-metoder?

Instituten är certifierade och säger sig använda vetenskapliga metoder och presenterar resultaten som fakta. Dessutom uttalar sig institutens representanter också om förhållanden som inte har stöd av data (vilket strider mot ESOMAR:s etiska regler.)

Exempel 2 i mitt inlägg visar hur det kan gå. När allmänheten får syn på ytterligare fall kommer den att tappa tilltron till fakta och vetenskap och vägra delta i undersökningar. Då tar demokratin allvarlig skada.

SIXTEN LUNDSTRÖM

LUNDSTRÖM:

Kommentarer till Hedlins artikel i Qvintensen 2018:3



Bortfallet ökar i undersökningar – ur Qvintensen nr 3 2018.

■■■ På annat ställe i detta nummer kritiserar jag opinionsinstitutens metoder (under rubriken *Kommentarer till surveyartiklar i Qvintensen 2018:3*). I princip menar instituten att man kan göra hur som helst och ändå få bra resultat. Jag vill här ytterligare kommentera en artikel i det nämnda numret av Qvintensen. Det är artikeln med rubriken *Bortfallet ökar i undersökningar*, av Dan Hedlin (DH), professor i statistik vid Stockholms universitet. Han har en av de viktigaste positionerna i samhället när det gäller att forma surveyteknikens framtid i Sverige, och därför blir jag bekymrad över hans inlägg. T.ex. konstaterar han att ”sedan 1950-talet är slumpmässigt urval ’guldstandard’. Andra sätt att dra urval betraktar många statistiker, med rätt eller orätt, som sämre alternativ.” Han fortsätter med att uttrycka sin känsla att ”synen på slumpmässigt urval kommer att ändra sig under de närmaste årtiondena.”

Grunden för surveyteorin är sannolikhetsurval, dvs. förfaranden där inklusionssannolikheter för alla individer i populationen är kända. DH bryr sig inte om detta och därmed flyter allt ut i allmänt tyckande. Det blir även förvirrande när flera av de texter han rekommenderar kräver kännedom om inklusionssannolikheter, t.ex. Särndal och Lundström (2005) och Särndal och Lundquist (2014). Institutet har ingen kontroll på inklusionssannolikheter och därför blir referenserna meningslösa.

DH gör även ett märkligt och farligt utlåtande om svarsandelen och hur den påverkar skattningarna: ”Själv tror jag att det går en mjuk gräns vid ungefär 25-30 % svar. Över den gränsen är chansen ganska god att bortfallet inte snedvrider alltför mycket...”

Jag har analyserat bortfallet i en mängd enkätundersökningar och genomgående sett att svarsbenägenheten är låg hos vissa grupper, t.ex. män, unga, boende i storstad, lågutbildade, personer med utländskt ursprung, ensamstående, låginkomsttagare och egna företagare. Svarsandelarna låg högt över den mjuka gräns DH nämner. Men hur kan DH veta att varierande svarsbenägenhet mellan olika grupper inte snedvrider alltför mycket?

Jag har mött en mängd kunder med olika utbildningar och tjänster. I regel har de varit okunniga om metoder för urval och estimation, men däremot oroade över stort bortfall. När de nu läser ovanstående uttalande från professorn i statistik suger de säkert tacksamt åt sig innehållet.

DH fortsätter med ”...under gränsen blir det snabbt mycket sämre, om inte korrelationen mellan svarssannolikhet och undersökningsvariabeln är väldigt liten”. Det är en oanvändbar kommentar eftersom svarssannolikheten aldrig är känd, och även om den vore känd så ingår det många undersökningsvariabler i en undersökning, vilket gör det svårt att säga något generellt.

Jag hävdar att en statistisk undersökning ska basera sig på ett sannolikhetsurval och estimationen utföras med GREG- eller kalibreringsestimatorn med hjälpvariabler valda så att bortfallsfelen reduceras för alla undersökningsvariabler. Tyvärr verkar DH inte hålla med.

SIXTEN LUNDSTRÖM,
F.D. CHEFSSTATISTIKER VID SCB

HEDLIN:

Svar till Lundström

■ ■ ■ Sixten Lundströms kritik av min artikel i Qvintensen 2018:3 inspirerade mig att tänka ett varv till på surveyteorin. Här nedan ska jag först försöka skissa Lundströms resonemang, som jag har uppfattat det, och sedan kommentera det. Lundström tillskriver mig också åsikter och ställningstaganden som möjligen ger någon läsare intryck av att jag är en människa med konstiga idéer. Jag lämnar dem därhän.

Lundström tycks utgå från endera av följande premisser:

1. Den enda vetenskapliga grunden för surveyteorin är sannolikhetsurval
2. Den överlägset bästa (kanske i betydelsen mest pålitliga), om än inte den enda, vetenskapliga grunden för surveyteorin är sannolikhetsurval

Han anser vidare att de statistiska slutledningarna bör grundas på vad vi inom surveyteorin kallar designbaserad inferens, där sannolikheterna att komma med i urvalet (inklusionssannolikheter) är väsentliga för den statistiska slutledningen. Den ansatsen är den överlägset mest vanliga på statistiska centralbyråer världen över. Lundström anser även att estimationen bör utföras med den generaliserade regressionsskattningen eller med kalibrering, där Lundström själv har gett viktiga bidrag till forskningen och som jag har lärt mig mycket av. De som inte känner till kalibrering kan tänka på det som ett slags generaliserad vägning av data. Vanlig poststratifiering är ett specialfall. Den typen av estimation är okontroversiell på statistiska centralbyråer. I vissa fall, bland annat när urvalen är små, använder man dock andra metoder.

Så, slumpmässiga urval, designbaserad inferens, kalibrering – inget fel på det, rentav svårt att överträffa, förutsatt att urvalen är tillräckligt stora. Men är den ansatsen den enda eller åtminstone den överlägset bästa, som står på vetenskaplig grund? Jag hävdar att det inte är så.

Här kan jag bara ge referenser till en del av den omfattande, moderna vetenskapliga litteraturen som diskuterar bortfall, icke-slumpmässiga urval och designbaserad inferens och dess alternativ. Två bra ställen att börja lära sig mer om icke-slumpmässiga urval är Elliot och Valliant (2017) och Baker et al. (2013). Det senare är en tjock rapport från en arbetsgrupp tillsatt av AAPOR. Det finns en sammanfattning med efterföljande diskussion i *Journal of Survey Statistics and Methodology* från samma år. Arbetsgruppen fann att de typer av icke-slumpmässiga urval som grundas på explicita statistiska modeller är de mest lovande. Teoretiska grunder för skillnaden mellan slumpmässiga urval utan bortfall och andra typer av urval och situationer ges av Little (1982) och Smith (1983b). Deville (1991) analyserar kvoturval teoretiskt. Kvoturval med webbenkät används idag av bland andra Ipsos.

Ett annat spår som diskuteras flitigt i litteraturen är metoder att kombinera undersökningar och data av olika slag för att få mer

tillförlitliga skattningar. En översikt ges i Lohr och Raghunathan (2017).

Det är också bra att hålla sig informerad om vilka utvecklingsprojekt andra statistiska centralbyråer arbetar med, vilket jag vet att SCB:s metodstatistiker gör, eller försöker göra om de ges tid till det. Weinberg et al. (2018) ger en bred översikt över vad USA:s Census Bureau gör nuförtiden. Där finns mycket att hämta, bland annat resonemang om balans mellan kostnad och nytta, som inte fångas upp i den ansats som jag etikettmässigt benämnde ”slumpmässiga urval, designbaserad inferens, kalibrering”. Även Weinberg et al. tar upp kombination av undersökningar av olika slag och av data från olika källor.

En återkommande workshop, som jag själv tyvärr inte haft möjlighet att delta i, kallas International Total Survey Error Workshop. Den har bland annat lett till antologin Biemer et al. (2017). Totalt surveyfel är något som den ansats som Sixten Lundström förespråkar behöver kompletteras med.

Danny Pfeffermann är en produktiv forskare på surveyområdet, och dessutom chef för Israels statistiska centralbyrå. Ett av de många teman han tog upp i sin Morris-Hansenföreläsning om problem och utmaningar i officiell statistik är icke-slumpmässiga urval och snedvridande bortfall (Pfeffermann 2015).

Man kan också titta i SCB:s egna skrifter. I *Urval: från teori till praktik* (Statistiska centralbyrån, 2008, sid. 90) ges rådet att icke-slumpmässiga urval bör undvikas när alla tre av följande föruttningar gäller: 1) urvalet är huvudurvalet i undersökningen, 2) ett slumpmässigt urval är möjligt att genomföra och 3) ett icke-slumpmässigt urval skulle leda till subjektivt godtycke. Där finns också en varning: ”Man bör dock notera att det finns fall då sannolikhetsurval leder till dålig kvalitet. Det kan bli så t.ex. då ramen är otillräcklig.” Samma skrift ger exempel där SCB drar (eller drog, vid tidpunkten då exemplen skrevs) icke-slumpmässiga urval (sid. 93).

SCB:s vetenskapliga tidskrift *Journal of Official Statistics* är internationellt populär och bidrar kraftigt till att upprätthålla den bild man har utomlands av SCB som en av världens mest framstående statistiska centralbyråer. Tidskriften publicerar artiklar som ligger i framkant, författade både av forskare och statistiker på myndigheter. Jag kan nämna, som bara ett exempel på nya sätt att tackla problem, Li-Chun Zhangs (2015) artikel om användning av proxyvariabler för att åstadkomma skattningar när det saknas tillräcklig information.

Den designbaserade ansatsen har varit kontroversiell sedan 1970, trots att den som sagt är den vanligaste på statistiska centralbyråer. Kritiken går bland annat ut på att designbaserad inferens skiljer sig utan tydlig anledning från det sätt att dra slutsatser som är det vanliga i ”vanlig” frekventistisk statistik (alltså det man lär sig i en universitetsutbildning i statistik). I surveysammanhang kallas detta för modellberoende inferens. Sviten Smith (1983a,

1994, 1997, 2001) belyser en del av utvecklingen av den här diskussionen, och även Fred Smiths personliga utveckling. Jag tror liksom Smith (2001, sid. 180) att det inte går att styrka att designbaserad inferens är av principiella skäl bättre eller sämre än modellberoende inferens, som ofta har utmålats som huvudalternativet.

De referenser som jag nu gett kanske kan övertyga läsaren om att det finns mer att säga än vad de två premisserna ovan ger intryck av. Jag hävdar att det inte finns vetenskapliga eller principiella fel i åsikten att vi bör så långt som det är möjligt och kostnadseffektivt dra slumpmässiga urval, och använda design-baserad inferens med kalibrering. Men säger man, vilket Sixten Lundström kanske gör, att den ansatsen är den enda, eller den överlägset bästa, då har man en snäv syn på modern surveyteori. Kanske man ska sörja över den ”splittring” som vi ser i den pågående förgreningen i olika ansatser och metoder i den moderna surveyforskningen. Jag gör det inte.

Det ökande bortfallet i individ- och hushållsundersökningar är oroande. En tanke om bortfall som har vunnit insteg de senaste åren är att rikta in sig på att få en mängd av svar som är av god kvalitet (är ”representativa”), snarare än att få en så hög svarsandel som möjligt, vilket i många fall har visat sig kontraproduktivt (ett av många exempel är Moore et al. (2016)). I min artikel ”Bortfallet ökar i undersökningar” i förra numret av Qvintensen rekommenderade jag artikeln Särndal och Lundquist (2014) som visar på sätt att få svarsmängder av god kvalitet. Tvärtom vad Sixten Lundström skriver, ser man om man analyserar den artikeln lite djupare, att den inte kräver kännedom om inklusionssannolikheterna eller ens ett slumpmässigt urval. Jag höll ett föredrag om det på Baltic-Nordic Conference on Survey Statistics i Helsingfors (Hedlin, 2015). Men det skulle ändå kunna vara så att det är fördelaktigt att börja med slumpmässigt urval, även om det förstås i slutändan blir en icke-slumpmässig mängd av svar, som det alltid blir när man har bortfall. Det är en empirisk fråga som vi, såvitt jag vet, inte har något svar på i dagsläget.

Även om bortfallet är betydande, torde ansatsen slumpmässiga urval, designbaserad inferens, kalibrering fungera ganska bra eller mycket bra. Fast i det enskilda fallet kan vi nästan aldrig veta det säkert. Så är det ju med bortfall: oavsett om bortfallet är 15% eller 50%, eller om vi använder kalibrering med noggrant undersökta hjälpvariabler eller inte, kan vi inte vara helt säkra på att bortfallet inte är snedvridande.

Sixten Lundström skriver ”DH [Dan Hedlin] gör även ett märkligt och farligt utlåtande om svarsandelen och hur den påverkar skattningarna: ’Själv tror jag att det går en mjuk gräns vid ungefär 25-30 % svar. Över den gränsen är chansen ganska »»»

Referenser:

- Arnér, M. och Vännman, K. (2011). Att mäta duglighet hos en tillverkningsprocess. *Qvintensen* 2011:2, sid. 7-9.
- Baker, R. et al. (2013). *Report of the AAPOR task force on non-probability sampling*. The American Association for Public Opinion Research.
- Bethlehem, J. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4(3), 51-60.
- Biemer, P.P. et al. (red.) (2017). *Total Survey Error in Practice*. Hoboken, NJ: Wiley.
- Déville, J.C. (1991). A theory of quota surveys. *Survey Methodology*, 17(2), 163-81.
- Elliot, M.R. och Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, 32(2), 249-264.
- Flygare, A.-M. och Hedlin, D. (2018). Empirical study on the size of nonresponse bias. In *JSM Proceedings, Survey Research Methods Section*. Alexandria, VA: American Statistical Association. 2557-2568.
- Hedlin, D. (2015). *Are we witnessing the end of random sampling in surveys?* BANOCOSS, the Fourth Baltic-Nordic Conference on Survey Statistics, Helsingfors, 24-28 augusti.
- Hedlin, D. (2017). *Is there a 'safe area' where the nonresponse rate has only a modest effect on nonresponse bias despite non-ignorable nonresponse?* Research Report 2017:2, Department of Statistics, Stockholm University.
- Little, R. J.A. (1982). Models for Nonresponse in Sample Surveys. *Journal of the American Statistical Association*, 77, 237-250.
- Lohr, S.L. och Raghunathan, T.E. (2017). Combining Survey Data with Other Data Sources. *Statistical Science*, 32(2), 293-312.
- Moore, J.C., Durrant, G.B. och Smith, P.W.F. (2018). Data set representativeness during data collection in three UK social surveys: generalizability and the effects of auxiliary covariate choice. *Journal of the Royal Statistical Society, series A*, 181(1), 229-248.
- Pfeffermann, D. (2015). Methodological issues and challenges in the production of official statistics. 24th annual Morris Hansen lecture. *Journal of Survey Statistics and Methodology* 3, 425-483.
- Rao, J.N.K. (2003). *Small Area Estimation*. New York: Wiley.
- Smith, T.M.F. (1983a). Comment. Diskussion av ”An Evaluation of Model-Dependent and Probability Sampling Inferences in Sample Surveys” av Hansen, Madow and Tepping. *Journal of the American Statistical Association*, 78, 801-802.
- Smith, T.M.F. (1983b). On the validity of inferences from non-random sample. *Journal of the Royal Statistical Society, Series A*, 146(4), 394-403.
- Smith, T.M.F. (1994). Sample Surveys 1975-1990. An Age of Reconciliation? *International Statistical Review*, 62(1), 5-19.
- Smith, T.M.F. (1997). Social Surveys and Social Science. *Canadian Journal of Statistics*, 25, 23-44.
- Smith, T.M.F. (2001). Biometrika Centenary: Sample Surveys. *Biometrika*, 88, 167-194.
- Statistiska centralbyrån (2008). *Urval – från teori till praktik*. Handbok 2008:1. ISSN 1654 7268.
- Särndal, C.-E. och Lundquist, P. (2014). Accuracy in Estimation with Nonresponse: A Function of Degree of Imbalance and Degree of Explanation. *Journal of Survey Statistics and Methodology*, 2(4), 1-27.
- Weinberg, D.H. et al. (2018). Effects of a government-academic partnership: has the NSF-Census Bureau research network helped improve the US statistical system? *Journal of Survey Statistics and Methodology*, 1-31.
- Zhang, L.-C. (2015). On Proxy variables and categorical data fusion. *Journal of Official Statistics*, 31(4), 783-807.

»»» god att bortfallet inte snedvrider alltför mycket...”. Den försiktigt uttryckta hypotesen grundar sig på en enkel iakttagelse av ett teoretiskt uttryck som Bethlehem (1988) visade. Uttrycket beskriver den relativa biasen orsakad av bortfall inom grupper av svarande, och jag valde att betrakta biasen som en funktion av medelvärde av svarssannolikheterna. Det skulle föra för långt att gå igenom det i Qvintensen. Den intresserade hänvisas till Hedlin (2017). Vid tidpunkten för Qvintensens publicering kommer jag att ha lagt en uppdaterad version av artikeln på ArXiv.org.

Sixten Lundström skriver: ”DH fortsätter med ’...under gränsen [25-30% svarsandel] blir det snabbt mycket sämre, om inte korrelationen mellan svarssannolikhet och undersökningsvariabeln är väldigt liten’. Det är en oanvändbar kommentar eftersom svarssannolikheten aldrig är känd och även om den vore känd så ingår det många undersökningsvariabler i en undersökning.”

Svarssannolikheten är förstås inte känd, men den kan skattas med metoder som man lär sig på grundutbildningen i statistik. Det som däremot är omöjligt att skatta är korrelationen mellan svarssannolikheter och undersökningsvariabeln, och möjligen syftade Sixten Lundström på korrelationen och inte svarssannolikheterna. Ann-Marie Flygare och jag genomförde en undersökning som var så upplagd att vi faktiskt kunde skatta korrelationen för två undersökningsvariabler. Vi presenterade resultaten på konferensen Joint Statistical Meetings, se Flygare och Hedlin (2018). I Hedlin (2017) visas att om man inte har en nära ideal situation så växer den relativa biasen mycket snabbt när bortfallet ökar bortom 70-75%. Naturligtvis olika för olika undersökningsvariabler, men att situationen skulle vara nära ideal är inte mycket att hoppas på, så rejäl bias får man räkna med för samtliga undersökningsvariabler.

I sin andra debattartikel i det här numret av Qvintensen, (se under *Debatt 2*), nämner Sixten Lundström mitt namn och undrar i det sammanhanget om kurser på universiteten om sannolikhetsurval kommer att upphöra. Jag utgår från att frågan är retoriskt menad.

Urvalsundersökningar görs i enormt många sammanhang och det hör till den professionella statistikerns grundläggande yrkeskunskaper att ha viss kännedom om urvalsundersökningar. Magnus Arnér och Kerstin Vännman berättar på ett underhållande sätt om ett exempel på en urvalsundersökning i industrin (Arnér och Vännman, 2011). Läs den artikeln, den är rolig och belyser

ett fenomen som tyvärr är alltför vanligt så fort något komplext översätts till ett enstaka tal.

Sixten Lundström skriver också i sin artikel under *Debatt 2* att: ”Innan universitetet lämnar metoder publicerade under ett århundrande bör man kunna hänvisa till artiklar publicerade i välrenommerade tidskrifter som visar att det finns bättre metoder.” Det leder fram till den intressanta frågan, att om slumpmässiga urval, design-baserad inferens och kalibrering fungerar lika bra som något annat, även vid bortfall (förutsatt att urvalen och svars-mängderna är stora), varför söker då forskare på surveyområdet, statistiska centralbyråer, privata företag och organisationer som AAPOR efter alternativ? Skälet är ökande svårigheter och kostnader med slumpmässiga urval.

En följdfråga är om man kan identifiera områden eller situationer där man tills vidare bör hålla sig till slumpmässiga urval, trots ökande kostnader, medan man i övrigt kan tänka sig andra ansatser. Om urvalen är små, torde det idag vara okontroversiellt att man bör använda modellberoende inferens (se Rao 2003, kap. 5) eller bayesiansk inferens (Rao 2003, kap. 10). Bland andra Deville (1991) har varit inne på att det finns en vattendelare mellan officiell statistik och statistik som används av en enstaka kund. Officiell statistik används i hela samhället, av många olika aktörer. Vid publiceringen känner man inte till alla sätt som statistiken kan komma att användas på. När däremot en kund köper en undersökning av en statistikproducent, har kunden ett precist syfte och kan komma överens med producenten om metoder och kvalitet för just den undersökningen. Smith (1983b) påpekar också att det är viktigt att hela samhället har förtroende för officiell statistik, och därför måste metoderna vara allmänt accepterade. I det andra fallet räcker det med att kunden har förtroende för producenten. Naturligtvis är det också en fråga om hur mycket en undersökning får kosta.

Jag skrev i det förra numret av Qvintensen att min känsla är att synen på slumpmässigt urval kommer att ändra sig under de närmaste årtiondena. Nu är jag benägen att byta åsikt: jag tror att synen på slumpmässigt urval *redan* har ändrat sig. Det handlar mer om man som surveystatistiker vill nöja sig med de tankar man själv redan har tänkt, eller om man är nyfiken på andras tankar.

DAN HEDLIN, STOCKHOLMS UNIVERSITET

LUNDSTRÖM:

Kommentar till Hedlin

■■■ Även på min tid fanns det alternativ till designbaserad inferens, men DH, som är mer uppdaterad av det som hänt sedan dess, konstaterar ändå att designbaserad inferens är ”den överlägset mest vanliga på statistiska centralbyråer världen över”. Det stora och ökande bortfallet är ett allvarligt problem i designbaserad inferens och därför är det bra att kunniga statistiker forskar på andra alternativ. Här tycker vi inte olika. Men DH bemöter inte huvudskälet till att jag reagerade på inläggen i Qvintensen. Jag upprepar det:

Resultaten från opinionsinstitutens undersökningar tar större och större plats i samhället och massmedia presenterar skattningarna som fakta. För att skapa tillit till sina skattningar använder opinionsinstituten begrepp från surveyteorin och jämförelse mellan skattningar och valresultat. Statistikanvändarna förstår inte att instituten använder/beräknar signifikans, svarsandel, viktning, etc. på ett felaktigt sätt. DH:s beskrivning av alternativen till design-baserad inferens kommer också att ge instituten begrepp och referenser som ger kunderna en känsla av att metoderna står på vetenskaplig grund. Det skall naturligtvis inte hindra DH att göra en sådan genomgång, men jag hade önskat att DH tydligt markerat att även om instituten använder icke-sannolikhetsurval så är de inte baserade på forskarnas metoder. I stället finner man mitt i DHs professionella text: ”Kvoturval med webbenkät används idag av bland andra Ipsos.”

DH gör en tydlig skillnad mellan metoder för officiell statistik och metoder för institutens statistik. I officiell statistik är det viktigt med allmänt accepterade metoder. ”När däremot en kund köper en undersökning av en statistikproducent har kunden ett precist syfte och kan komma överens med producenten om metoder och kvalitet för just den undersökningen”, skriver DH. Detta är en förhoppning långt ifrån verkligheten. Kunden har ingen möjlighet att se att kvaliteten är mycket lägre än det institutet påskiner. Men DH menar att det räcker med ”att kunden har förtroende för producenten”. I Exempel 2 i mitt inlägg har Aftonbladet och DN naturligtvis förtroende för Inizio resp. Ipsos och presenterar skattningarna som fakta och baserade på vetenskapliga principer. När fler liknade fall kommer till allmänhetens kännedom är det stor risk att tilltron till vetenskap och fakta eroderar, vilket även drar med sig officiell statistik. Samhället där ingen tror på någonting fruktar jag.

Slutligen vill jag återigen kommentera DH:s uppseendeväckande uppfattning att när svarsandelen är över 30 % ”är chansen ganska god att bortfallet inte snedvrider alltför mycket.” Alla bortfallsanalyser jag gjort med hjälp av en mängd registervariabler säger att det är felaktigt.

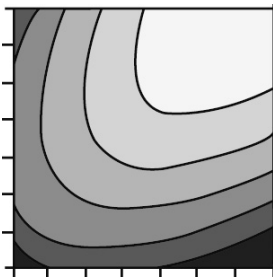
SIXTEN LUNDSTRÖM

HEDLIN:

Slutreplik

■■■ I mitt första svar på Sixten Lundströms två repliker fokuserade jag på vad han skrev om min artikel i Qvintensen nr 3/2018, som handlade om bortfall och inte om undersökningsinstitutens metoder. Jag hade dock svårt att identifiera resonemanget i Sixten Lundströms replik med titeln ”Kommentarer till ’Bortfallet ökar i undersökningar’” och därför inledde jag mitt första svar med att beskriva vad jag uppfattade som hans resonemang. Jag trodde väl för mycket när jag utgick från att det fanns ett tydligt sådant i Sixten Lundströms replik. Först nu när jag läser hans replik 2 förstår jag (se stycket efter ”Men DH bemöter inte huvudskälet till att jag reagerade på inläggen i Qvintensen. Jag upprepar det”) att han bara vill att jag ska säga att ”institutet” gör fel. Det kan jag inte, för jag vet inte vad alla gör och jag tänker inte sänka mig till dålig argumentation. ”Institutet” är alla olika, de använder vitt skilda metoder, även slumpmässiga urval med någon variant av kalibrering. Dessutom har flera av dem har genomfört väsentliga förändringar i sina metoder de senaste åren. Men Sixten Lundström vet. Trots att han skriver i *Kommentarer till surveyartiklar i Qvintensen, nr 3, 2018* att ”Jag hade gärna skaffat mig kunskap om institutens metoder, men dokumentationen är så bristfällig att det inte är görligt. Jag har via mejl och telefon försökt få mer information men misslyckats.”

DAN HEDLIN, STOCKHOLMS UNIVERSITET



ORDFÖRANDEN HAR ORDET

Är Indstats verksamhet allmännyttig?

Syftet med detta inlägg är att uppmanna till en diskussion om allmännyttan med den verksamhet, som bedrivs av Föreningen industriell statistik (i fortsättningen förkortat "Indstat"). Diskussionen kan även beröra Statistikfrämjandet, Cramérsällskapet, FMS och Surveyföreningen.

Anledningen till inlägget är ett rutinskattebesked från Skattemyndigheten beträffande beskattning av Indstat för året 2018, och Skatteverkets syn på nyttan med föreningens verksamhet. Det märkliga är att Skatteverket anser föreningen vara *ideell, men inte allmännyttig*.

Med en *ideell* förening kännetecknas av att den har ett ideellt ändamål eller bedriver ideell verksamhet. Med ideellt ändamål menas att syftet inte är att främja medlemmarnas ekonomiska intressen. Religiösa föreningar och idrottsföreningar är exempel på föreningar med ideella ändamål.

Med en *allmännyttig* ideell förening menar Skatteverket en förening som uppfyller följande fyra krav:

1. **Ändamålskravet.** Föreningens syfte ska vara till nytta för allmänheten. (Lag 2013:960).
2. **Verksamhetskravet.** Minst 90 procent av föreningens verksamhet ska bidra till att uppfylla föreningens syfte.
3. **Fullföljdskravet.** Minst 80 procent av föreningens inkomster ska användas till att uppfylla föreningens syfte
4. **Öppenhetskravet.** Föreningen ska vara öppen för alla som vill bli medlemmar

Ur skattesynpunkt betyder Skatteverkets be-

dömning att föreningen Indstat är deklarationspliktig, vilket i sig inte är något problem. Det som är förvånande är Skatteverkets uppfattning att föreningen inte är allmännyttig. Indstat befrämjar utbildning och vetenskaplig forskning inom statistik, vilket jag anser bör tolkas som en allmännyttig verksamhet. Även om tolkningsfrågan saknar praktisk betydelse för beskattning, så har Skatteverkets lagtolkning en bäring på hur nyttan med vår verksamhet uppfattas.

Vår uppfattning är att Indstat uppfyller alla kraven 1, 2, 3 och 4 ovan. Vi ska nu titta närmare på frågan hur föreningen uppfyller krav nr 1, ändamålskravet. Skatteverket exemplifierar på följande sätt hur en förenings syfte kan vara till nytta för allmänheten: "Ett syfte som är till nytta för allmänheten kan till exempel vara idrott, kultur, politik eller religion. Föreningar vars syfte är att främja medlemmarnas ekonomiska intressen har inte ett syfte som är till nytta för allmänheten."

Syften som är till nytta för allmänheten är enligt Skatteverket t.ex.

- idrott, kultur eller religion
- omsorg om barn och ungdom
- social hjälpverksamhet, sjukvård eller miljö-
vård
- politik eller Sveriges försvar och krisberedskap i samverkan med myndighet
- utbildning och vetenskaplig forskning.

Föreningar som gynnar medlemmarnas ekonomiska intressen har inte ett syfte som är till nytta för allmänheten. Exempel på sådana föreningar är:

- fackföreningar och branschföreningar
- samlarföreningar
- jaktföreningar och fiskeföreningar

- vägföreningar och vattenföreningar
- stugföreningar, villaföreningar och hyresgästföreningar
- bryggföreningar och hamnföreningar.

Indstat har som syfte att befrämja utbildning och vetenskaplig forskning, inte att gynna medlemmarnas ekonomiska intressen. Därför är det förvånande att se Skatteverkets besked att föreningen inte uppfyller ändamålskraven.

Indstats engagemang och verksamhet befrämjar statistikanvändning för att förbättra verksamheten inom alla branscher, discipliner och läroinstitutioner utan undantag, på samma sätt som t.ex. matematik gör. Indstats verksamhet hjälper till att förbättra och säkra kvalitén inom exempelvis miljövard, omsorg, sjukvård, säkerhet, utbildning, vetenskaplig forskning, och annan likvärdig verksamhet. Detta bör vara tillräckliga argument för att klassa föreningen som uppfyllande ändamålskraven.

Indstat har under åren 2010 – 2018 publicerat en rad artiklar om statistikens allmännyttighet vilka publicerats i tidskriften Qvintensen, utgiven av Svenska Statistikfrämjandet. Syftet med artiklarna är att informera om nyttan med användning av statistisk metodik i många olika sammanhang, inte bara inom industrin. Med detta vill föreningen visa att den statistiska metodiken kan sägas uppfylla kraven för att kallas allmännyttig.

Många har säkert noterat att den pågående digitaliseringen och uppstarten av den nya industrirevolutionen ("Industri 4.0") driver på statistikens användning och utveckling. Vi kan konstatera att den pådrivande kraften och behovet är större än tillgängliga resurser och beredskap helt klarar av. Glädjande är att den

Referenser

Skatteverket (2013). Skatteregler för ideella föreningar.

<https://www.skatteverket.se/foretagochorganisationer/foreningar/ideellforening>

statistiska metodiken verkar gå mot en renässans. Detta är en förutsättning för att industrisatsning på Industri 4.0 ska kunna lyckas. Att säkerheten mot intrång på nätet ska kunna förbättras, att Sveriges försvar ska kunna fungera bättre, att vi ska kunna göra bättre riskanalyser, förbättra beredskapen, prediktera miljöförändringar, väder, klimat, och anpassa oss till oväntade, men sannolika, händelser är sådant som föreningen ser som mål.

Statistikutveckling är nödvändigt för att lyckas med det viktiga maskinlärandet och för att hantera den behövliga, men icke helt ofarliga artificiella intelligensen. Det är bl.a. statistisk metodik som ligger bakom framgångarna med språkigenkänning inom telefoni och datateknik, och det är statistik som geofysiker och meteorologer använder för att t.ex. förutsäga väder, farliga jordbävningar och tsunami. Många grenar av forskning inom t.ex. teknik, medicin, biologi, kemi, skulle aldrig kunna genomföras utan statistikens hjälp. Det är bevisligen en allmännyttig metodik!



GÖRAN LANDE

FOTO: LINA LANDE



Från Surveyföreningens årsmöte

Den 22 mars ägde Surveyföreningens årsmöte rum i Göteborg. Före det formella årsmötet presenterades fyra föredrag.

Anders Holmberg höll det traditionella Tore Dalenius-föredraget. Den när gången med utgångspunkt i Tores nästan 250-sidiga stencilerade rapport *Ends and Means of Total Survey Design* från 1974, vilken är en rapport från hans forskningsprojekt "Fel i undersökningar". Anders fann mycket av Tores synsätt och begrepp fortfarande aktuellt. Tore trycker på vikten av kontroll och evalvering av olika moment i en survey, och hans mål är "a survey design aiming at an economic balance between sampling and non-sampling errors in a survey, a Total Survey Design".

Elina Lindgren talade om forskning kring mittalternativ i svarsskalan, när man frågar om respondenters värderingar och andra tyckanden. Ett problem är att ett mittalternativ kan tolkas på olika sätt av olika respondenter. Ett annat problem är om man över huvud taget ska ha ett mittalternativ eller ej.

Axel Tandberg från ERM (Etiska Rådet för Marknadsundersökningar) redogjorde för pågående arbete med etiska problem i samband med surveys.

Årets uppsatsvinnare var Rikard Gard med uppsatsen "En skattningsmetod med hjälp av machine learning". Ett hedersomnämmande gavs också till Jannis Kalpouzos för uppsatsen "Indicators for evaluating the nonresponse bias in the Swedish Labour Force Surveys 2007 – 2013".

Surveyföreningens nya styrelse ser ut så här:

- Ordförande: Marcus Berg, Odyssey (nyval)
 - Vice ordförande: Gösta Forsman, Uppsala (omval)
 - Kassör: Ann-Marie Flygare, Örebro universitet (omval)
 - Webbansvarig: Ying Fu Xie, SCB Stockholm (omval)
 - Ledamot: Rikard Gard, SCB Örebro (nyval)
 - Ledamot: Johan Martinsson, Göteborgs universitet (omval, ansvarig opinion)
 - Ledamot: Mats Nyfjäll, Statisticon (omval)
 - Ledamot: Karin Schneller, Kantar Sifo (omval)
 - Ledamot: Åke Wissing, Wissing & Co. (tidigare ordförande)
- (Sekreterarfrågan ännu inte löst.)

Vinnare av MarC Awards 2018

Den 20 november 2018 firades i Stockholm undersökningsbranschens galakväll med utdelande av MarC Awards. ("MarC" står för Market research and Customer insight.) Närmare 100 personer från olika delar av undersökningsbranschen hade samlats för att fira vinnarna i årets åtta priskategorier. Bakom eventet stod branschorganisationen SMIF, tillsammans med SÖK, SMUF, ESOMAR, Surveyföreningen och Statistikfrämjandet. Vinnare och motiverings-texter följer här (ordagrant hämtat från SMIF:s hemsida):

ÅRETS UNDERSÖKNINGSPROJEKT

Vinnare: Mats Elzén och Liza Pettersson från Novus med studien Vad tycker människorna som bor i Sveriges mest utsatta områden?

Motivering: Liza Pettersson och Mats Elzén på Novus har på uppdrag av The Global Village genomfört en unik studie i landets 61 utsatta områden. De gjorde 2000 intervjuer med möjlighet att svara på flera olika språk. Studien har totalt 76 olika nationaliteter representerade.

ÅRETS KÖPARE

Vinnare: Martin Düring, Avdelningschef, Insikt & Analys AMF.

Motivering: Martin har framgångsrikt kombinerat olika datakällor och discipliner för att maximalt utnyttja företagets samlade information och kunnat skapa en fördjupad och mer nyanserad bild av kunden. Han kallar sig själv för "jobbig" då han ställer höga krav på leverantörerna.

ÅRETS NYKOMLING

Vinnare: Odyssey.

Motivering: Odyssey är ett analysföretag med rötter i traditionell research parat med avancerad dataanalys och strategisk rådgivning. De har skapat en nytänkande cross-over mellan nya möjligheter och beprövade metoder.

ÅRETS STATUSHÖJARE

Vinnare: SOM-institutet och Valforskningsprogrammet.

Motivering: Valforskningsprogrammet och SOM-institutet är två av Sveriges viktigaste aktörer när det kommer till opinions-undersökningar och att förstå svenskarnas inställning i en rad politiska och aktuella samhällsfrågor.

ÅRETS OPINIONSUNDERSÖKARE

Vinnare: Demoskop.

Motivering: Demoskop har sedan förra valet utvecklat sin metod och, efter ett gediget analysarbete av metoder och källor, skiftat från att basera sin väljarbarometer på enbart telefonintervjuer till att komplettera dessa med webbintervjuer.

ÅRETS KVALITETSUTVECKLARE

Vinnare: Nicklas Källebring, Ipsos.

Motivering: Nicklas Källebring på Ipsos är kanske en av de skickligaste metodutvecklarna och kvalitetsutvecklarna i Sverige. Han har jobbat på flera olika ställen och var han än varit har kvalitets- och metodutveckling stått i centrum för hans arbete.

BRANSCHGÄRNINGEN FÖRETAG

Vinnare: Norstat.

Motivering: Norstat är ett av Sveriges största undersökningsföretag som levererar pålitliga undersökningar till både branschkollegor och företag som gör egen research. De utvecklar sina metoder och sitt sätt att arbeta i en tid när undersökningsbranschen är under stort tryck.

BRANSCHGÄRNINGEN INDIVID

Vinnare: Åke Wissing

Motivering: Åke Wissing på Wissing & Co. har länge varit en av de mest tongivande undersökarna i Sverige. Med en bakgrund som statistiker har han varit aktiv i Statistikfrämjandet och Surveyföreningen. Åke vurmar för kvalitet parat med nytänkande i tider av stor utveckling av statistik.



Coop utnyttjar en kombination av data avseende försäljning, kunders beteende i butiken, samt kunders vanor och preferenser för att fatta bättre affärsbeslut.

Kvalitetsseminariet 2019

Surveyföreningens kvalitetsseminarium 2019 ägde rum den 7 februari på Westmanska Palatset i Stockholm. Årets tema var: *Nya möjligheter för surveystatistiker, genom användande av big data och andra okonventionella datakällor*. Arrangör var Surveyföreningens ordförande Åke Wissing, och konferencier var Karin Nelsson, styrelseordförande i SMIF. Först ut var tre föredragshållare med en mer metodstatistisk inriktning (Lilli Japac (SCB), Dan Hedlin (Stockholms universitet) och Thomas Laitila (Örebro universitet). Därefter kom fem föredragshållare från den kommersiella världen med exempel från marknadsanalys och marknadsföring; se nedan.

LILLI JAPAC tog upp problem kring användning av nya datakällor i statistikproduktionen. Det kan handla om att utnyttja traditionella system som t.ex. transaktionsregister, kundregister, kreditkortsdata. Det kan också handla om att utnyttja sensordata rörande t.ex. väder, trafik, övervakning, mobiltelefoner och bilar, samt även satellitbilder. Vidare finns data från sociala nätverk, t.ex. Facebook, Twitter, bloggar, foton, videofilmer, Youtube och internetsökningar. Data från smarta elmätare var ett annat exempel. Allt detta väcker frågor om kvalitet, samt även juridiska och etiska frågor.

DAN HEDLIN redovisade tankar inspirerade av en artikel av Xiao-Li Meng i The Annals of Applied Statistics, 2018. Tyvärr förstår jag inte Xiao-Li Mings tankar eller hans begreppsapparat, så jag har inget att säga om detta.

THOMAS LAITILA talade om användning av nya datakällor i officiell statistikproduktion. För att t.ex. producera statistik om hushållens utgifter kommer man kanske att kombinera administrativa data, bankkortsdata, surveydata och big data. Det blir en blandning av designbaserad och modellbaserad ansats. En ännu obesvarad fråga är: Hur ska man då redovisa kvalitet och osäkerhet?

Så kom då exemplen från kommersiell verksamhet.

FREDRIK TÖRN (Coop) berättade om att man i syfte att fatta bättre affärsbeslut utnyttjar en kombination av data avseende försäljning, kunders beteende i butiken, samt kunders vanor och preferenser.

GUNNAR EHRNBORG (Ericsson) hade funnit att kunders självrapportering om hur mycket tid per dygn som ägnas åt e-post, internetsökning, tv, sociala medier, datorspel, och liknande stämmer dåligt med de resultat man erhållit genom mer objektiva mätmetoder.

CECILIA YDREMARK (Tobii) framhöll "eye tracking" (studium av hur blicken rör sig över en yta) som en intressant teknik när man vill studera genomslagskraften hos olika typer av visuell reklam, t.ex. i tidningar eller på tv.

TOBIAS SJÖQVIST (Odyssey) talade om fördelar med en typ av marknadsföring som (om jag förstått rätt) utnyttjar data från olika datakällor tillsammans, som tidigare bara studerats var för sig.

SOPHIE HEDESTAD (Meltwater) framhöll vikten av att tillgodogöra sig den stora mängd data som finns "online", att hålla sig uppdaterad om händelser och trender i omvärlden, samt att få feedback från de egna kunderna.

Slutligen några kritiska synpunkter. Personer inom marknadsföringsbranschen använder sig gärna av en engelskinspirerad yrkesjargong som är rätt obegriplig för oss utomstående. Åtskilligt av innehållet i de fem sistnämnda föredragen hade jag svårt att förstå av sådana språkliga skäl. Som statistiker hade jag också önskat mig mer fokus på själva metodiken, samt på kvalitén hos de data som marknadsfolket använder sig av. Och naturligtvis också kvalitén hos de slutsatser som dras utifrån dessa data.

JAN WRETMAN

FRIMIS 2018

2 018 års Frimisseminarium ägde rum den 28 november med temat *Ett föränderligt surveylandskap – möjliga strategier bortom traditionella sannolikhetsurval*.

Det så kallade Frimisseminariet är vid det här laget en tradition. Det är en heldagssammankomst som arrangeras årligen (på Frimurarhotellet i Linköping) sedan 1995 (med undantag för åren 1998, 2002 och 2017). Målsättningen har redan från början varit att ta upp aktuella metodfrågor kring marknads- och opinionsundersökningar, samt vara ett forum där akademiska forskare och företrädare för privata och offentliga undersökningsorganisationer kan träffas och diskutera gemensamma frågor. Gösta Forsman har hela tiden varit drivande.

Seminariet 2018 inleddes med att Lars Lyberg gav en översikt över hur "surveylandskapet" förändrats under hans långa tid i undersökningsbranschen. Under "guldåldern" 1960 – 1975 var sannolikhetsurval det självklara idealet när det gällde urval. Bortfallet var lågt, jämfört med läget idag. Redan bortfall på 10 – 15 procent ansågs oroväckande stort. Jakten att minimera tid och kostnad går inte heller att jämföra med dagens läge. Ur ett nutida perspektiv var det en lycklig tid. Perioden efter 1975 karakteriseras bland annat av ökat bortfall och ökade kostnader, men också av mer avancerad teknologi för datainsamling och förhoppningar om att kunna utnyttja nya datakällor. Att genomföra traditionellt sannolikhetsurval har visat sig allt svårare, och intresse har börjat riktas mot utveckling av kostnads- och tidseffektiva metoder i samband med icke-sannolikhetsurval. Som några få exempel på nya datakällor kan nämnas administrativa datasystem, sociala media, GPS och satelliter. Utvecklingen är här bara i sin början.

Daniel Thorburn talade därefter om användning av bayesianska metoder i surveysammanhang. Bayesiansk slutledningsteori är ett

alternativ till "klassisk" statistisk slutlednings-teori. Vagt uttryckt skulle man kunna säga att det är en matematiskt formulerad teori för hur subjektiva förhandsförmodanden i kombination med därefter observerade data leder till modifierade förmodanden. Säg att man i surveysammanhang vill skatta andelen individer med en viss egenskap i en population. Vid bayesianskt tänkande utgår man från vad man i förväg tror om populationens utseende. Därefter insamlas data genom en undersökning. Detta leder slutligen fram till en reviderad föreställning om populationens utseende. Här kommer Bayes sats till användning; därav namnet "bayesiansk" metodik. Det matematiska i den bayesianska teorin hänger ihop med att all osäkerhet formuleras i sannolikhetsstermer, ibland en blandning av subjektiva och frekventistiska sannolikheter. Matematiskt tenderar det hela ofta att bli komplicerat. Det framhölls att man med ett bayesianskt angreppssätt (i samband med surveys) skulle kunna bilda sig uppfattningar om bland annat bortfallets effekt, bortfallskorrektoners egenskaper, hjälpvariablers användbarhet och skattningars felmarginaler. Man kan fråga sig: vems förhandsförmodanden är det som man ska utgå från? Svaret är att det finns metodik även för fall där de subjektiva förhandsförmodandena är mycket vaga.

Anders Holmberg berättade om utnyttjande av transaktionsdata inom officiell statistikproduktion. Hans exempel var hämtat från Statistisk Sentralbyrå, Oslo, närmare bestämt deras undersökning om hushållens utgifter. I stället för att på traditionellt vis låta hushåll föra dagbok över utgifter vill man här försöka utnyttja transaktionsdata, som är tillgängliga genom ny datateknik. Det kan vara data avseende kortbetalningar och girotransaktioner och dessutom kassadata och scannerdata från detaljhandeln. En fördel kan då vara att man snabbt och billigt får tillgång till stora mängder data. Man slipper en kostnadskrävande sur-

veyapparat för datainsamling och bearbetning. Dessutom slipper hushållen en avsevärd uppgiftslämnarböroda. Det hela befinner sig än så länge på experimentstadiet. Anders poäng var att en statistikproducent i framtiden kan behöva kombinera traditionella och okonventionella data i samband med surveyer.

Lars Lyberg återkom med information från en konferens med titeln "BigSurv18 – Big Data Meets Survey Science" 25-27 oktober i Barcelona. Konferensen kommer också att resultera i en monografi med samma titel. Monografins innehåll planeras i stora drag bli följande:

- The new survey landscape
- Total error and data quality
- Big data in official statistics
- Combining big data with survey statistics: methods and applications
- Combining big data with survey statistics: tools
- Regulations, ethics, privacy

Nicklas Källebring redogjorde för en studie av svenska väljarbarometrar från 1944 till idag. Det rörde sig om prognoser inför tre typer av val: folkomröstningar (5 stycken), val till Europaparlamentet (5 stycken) och riksdagsval (23 stycken). I stort har mätningarna haft likartad träffsäkerhet hela tiden. Ingen påtaglig förbättring eller försämring har iakttagits. (Se vidare Nicklas Källebrings artikel på sid. 16-17 i detta nummer.)

Seminariet avslutades med att en panel av partisympatiundersökare diskuterade erfarenheter från årets undersökningar i samband med riksdagsvalet. Deltagare var Toivo Sjören (Sifo), Nicklas Källebring (Ipsos), Karin Nelsson (Inizio), Liza Pettersson (Novus) och Peter Santesson (Demoskop).

JAN WRETMAN



Perioden efter 1975 karakteriseras bland annat av mer avancerad teknologi för datainsamling och förhoppningar om att kunna utnyttja nya datakällor. Som några få exempel på nya datakällor kan nämnas administrativa datasystem, sociala media, GPS och satelliter.

Bli medlem i Svenska statistikfrämjandet

Svenska statistikfrämjandets syfte är bland annat att främja sund användning av statistik som beslutsunderlag och att väcka och sprida intresse för statistik i samhället.

För att bli medlem, gå till <http://www.statistikframjandet.se> och läs mer i högerspalten under "Vill du bli medlem?". Har du frågor kontakta Mattias Strandberg 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.

