

# Qvintensen

NR 1 2026

**TEMA:  
ATT FÖRSTÅ  
STATISTIK**

**VAD ÄR DET EGENTLIGEN  
FRÅGAN OM I ENKÄTER?**

SIDORNA 9-10

**VISUELL  
DATAKOMMUNIKATION  
I ETT POSTSANNINGENS  
TIDEVARV**

SIDORNA 11-14

**VALSTATISTIKENS  
LÅNGA LINJER**

SIDORNA 30-31

**Varför behöver vi prata  
om statistisk litteracitet?**

SIDORNA 6-8

## Innehåll

3	Redaktörens ruta Charlotte Ovesson	28	Journal of Official Statistics – 40 år av utveckling Lilli Japac
4	Föreningen går framåt som ideell förening Nancy Steinbach/styrelsen	30	Valstatistikens långa linjer Richard Öhrvall
5	Reseberättelse från ISCB Stina Zetterström	32	Recension av Daten sind Macht Hans Alberg
6	Varför behöver vi prata om statistisk litteracitet? Hans Alberg	33	Är det verkligen så självklart? Dan Hedlin
9	Vad är det egentligen frågan om i enkäter? Sara Scholtens	34	Polisens trygghetsunder- sökningar och vad SIG-gruppen gör Kjell Elefalk
11	Visuell datakommunikation i ett postsanningens tidevarv Angie Hjort	36	Statistikfrämjandet Ordföranden har ordet
15	t-test och hypotesformuleringar – en kritisk granskning Rolf Sundberg	37	Surveyföreningen Ordföranden har ordet
18	Vad är relevant statistisk information? Göran Nilsson	38	Cramérsällskapet Ordföranden har ordet
20	Stora språkmodeller och survey- undersökningar Leah von der Heyde	39	Föreningen för medicinsk statistik Ordföranden har ordet
24	AI och undersökningar Karin Nilsson	40	Föreläsningsprogram med stor bredd på FMS vårmöte Aaron Levine
26	Hur mycket vatten använder AI? Pontus Wärnestål	42	Anders Martin-Löfs matematiska arbeten Tom Britton, Maria Deijfen, Allan Gut och Filip Lindskog



### SVENSKA STATISTIKFRÄMJANDETS STYRELSE

**Ordförande**  
Nancy Steinbach,  
ordforande@statistikframjandet.se

**Vice ordförande** John Öhrvik

**Kassör** Annika Tillander  
kassor@statistikframjandet.se

**Sekreterare** Samuel Hellman  
sekreterare@statistikframjandet.se

**Ledamöter** Patrik Bernhard och  
Tea Unnebäck

**Representant Surveyföreningen**  
Eva Lagercrantz

**Representant FMS**  
Anita Lindmark

**Representant Cramérsällskapet**  
Jolanta Pielaszkiwicz

**Webbplats** www.statistikframjandet.se



**Ansvarig utgivare**  
Nancy Steinbach

**Redaktör**  
Charlotte Ovesson

**Redaktion**  
Anders Skölleremo  
Hans Alberg  
Marie Linder  
Jesper Rydén

**E-post** qvintensen@statistikframjandet.se

**Produktion**  
Form och redigering: Mezzo Media AB  
Tryckeri: Trydells Tryckeri AB

**Annonser**  
Annonser i Qvintensen bokas med redaktören:  
qvintensen@statistikframjandet.se.  
Annonstskick på hemsida eller per e-post bokas med  
Statistikfrämjandets sekreterare:  
sekreterare@statistikframjandet.se



## Tar du ubåten, cykeln, jetpacken eller rymdraketerna till skolan?

**Folkbildning ligger mig** varmt om hjärtat. Under våren leder jag studiecirkeln *Informerad eller manipulerad?* hos Folkuniversitetet. Cirkeln handlar om AI, desinformation, det digitala informationslandskapet och mänskligt beteende i valrörelsen. Vi genomför en variant av cirkeln med fysiska träffar i Malmö och två digitala. Alla blev snabbt fullteknade.

En färsk undersökning från Indikator Opinion kan ge oss en hint om varför så många har visat intresse för studiecirkeln. Undersökningen visar att 62% av svenskarna är ganska eller mycket oroade att falsk eller vilseledande information sprids om valet. Detta avspeglas i deltagarna i studiecirkeln. Vi pratar mycket om desinformation och trollfabriker, och en hel del om AI. Valrörelsen 2026 blir den första i Sverige där generativ AI kommer att spela en stor roll. Många som deltar i studiecirkeln vill lära sig mer om det.

Vi läser delar av boken *Därför demokrati* av Åsa Wikforss. Där slår hon fast att kunskap är en nödvändig del av en demokrati. När vi ska rösta behöver vi ha korrekt information om huruvida arbetslösheten ökar eller minskar, om bensinpriset går upp eller ner, om brottsligheten ökar eller inte, eller vilka frågor som nu är viktiga för oss. Om vi inte har fakta är risken stor att vi röstar fel – alltså på ett annat sätt än vi hade röstat om vi hade haft korrekt information. Det leder till att valresultatet inte blir som folket vill.

**Här spelar statistiken** stor roll. För att kunna se hur arbetslösheten, bensinpriset och brottsligheten förändras över tid behöver vi statistik. Statistiken ger oss en bild av hur samhället utvecklas och vilka utmaningar som är aktuella. Denna kunskap utgör en viktig grund för oss när vi väljer hur vi ska rösta. För att människor ska kunna tolka och förstå statistik är det nödvändigt med statistisk litteracitet.

När jag berättar att jag är statistiker brukar människor dela med sig av sina tankar och åsikter om statistik. Något som jag ofta hör är Mark Twains ganska uttjatade ”lögn, förbannad lögn och statistik”. Det är tydligt att i stort sett alla människor vet att

det går att luras med statistik. Men vad händer när vi inte har verktyg och metoder för att skilja bra statistik från dålig? Om vi inte kan skilja rimliga tolkningar från orimliga? Om vi inte kan reflektera kring om viktig information utelämnas? Om vi inte kan se om ett diagram är missvisande? Då är risken att vi landar i en cynisk inställning, där all statistik kan avfärdas med ”jag tror inte på den, för man kan visa vad som helst med statistik”.

**I studiecirkeln handlar** en träff om lurig statistik. Jag visar exempel på hur statistik används för att manipulera eller vilseleda. Deltagarna får göra en övning där de i grupp ska ta fram sin egen luriga statistik utifrån ett scenario, till exempel: du arbetar i ett korvstånd. Du vill ta fram statistik som visar att korv är gott och trendigt. Hur kan du göra? Jag har också gjort denna övning med elever i högstadiet. Uppgiften de fick var att utforma en lurig undersökning som visar att många elever cyklar till skolan. De designade denna enkätfråga: hur brukar du ta dig till skolan? Svartalternativen var ubåt, cykel, jetpack och rymdraket.

Att människor har god statistisk litteracitet är viktigt för demokratin. Just nu befinner vi oss i ett valår. Vad kan du som statistiker eller statistikintresserad göra för att öka människors statistiska litteracitet? I vilka sammanhang kan du folkbilda, sprida kunskap och klargöra? Fundera gärna på det.

»Statistiken ger oss en bild av hur samhället utvecklas och vilka utmaningar som är aktuella. Denna kunskap utgör en viktig grund för oss när vi väljer hur vi ska rösta»

CHARLOTTE OVESSON  
Redaktör

# Föreningen går framåt som ideell förening

Mars tog Svenska Statistikfrämjandet ett viktigt beslut om föreningens framtid. Resultatet är tydligt: vi fortsätter den väg vi alltid har gått, som en förening driven av intresset för statistik och av varandra.

Vägskalet handlade om föreningens inriktning och grundläggande idé. Skulle vi fördjupa rollen som mötesplats för alla oss som är engagerade i och för statistiken, eller ta steget mot en mer professionaliserad organisation med tyngdpunkt på yrkesrollen? Medlemmarnas svar var entydigt: vi håller fast vid det som gjort oss till vad vi är, en ideell förening byggd på gemenskap, nyfikenhet och frivilligt engagemang.

Med anor från tidigt 1900-tal har statistiker och statistikintresserade samlats i föreningen. När två stora föreningar slogs samman 2007 och bildade Svenska Statistikfrämjandet skapades en ännu starkare plattform. Den långa historien är ett styrkebesked i sig, vi har gång på gång visat att vi kan anpassa oss, utvecklas och fortsätta vara relevanta för dem som arbetar med eller brinner för statistik.

**Vår uppgift är** att stärka statistikens roll i samhället. Vi vill bidra till forskning och utbildning inom statistik genom att sprida ny kunskap och skapa mötesplatser där akademi och praktik kan mötas. Samtidigt arbetar vi för att statistiska metoder ska användas på ett bra och ansvarsfullt sätt, med fokus på kvalitet, öppenhet och förståelse för hur metoderna fungerar.

Vi skapar också möjligheter för våra medlemmar att utvecklas och utbyta erfarenheter, till exempel genom seminarier, nätverk och andra forum. Där kan man lära av varandra och växa i sin yrkesroll. En viktig del av vårt arbete är att vara en plats för samtal om hur statistik används, men också om de etiska frågor som kan uppstå i en allt mer datadriven värld.

Dessutom vill vi väcka intresse för statistik hos fler. Vi vill visa varför statistik är viktigt i samhällsdebatten och göra den mer begriplig och tillgänglig för alla.

Att vi nu väljer att bygga vidare på de ideella krafterna är ett beslut vi är stolta över. Det innebär att vi sätter tillit

till varandra, till att vi som medlemmar både vill och kan vara med och forma föreningens framtid. Vi vet vad vi vill åstadkomma, och vi vet att kompetensen och engagemanget finns i föreningen.

Framöver kommer föreningens aktiviteter i ännu större utsträckning att skapas och drivas av oss medlemmar. Det är en utveckling som öppnar för både bredd och nytänkande. Webbinarier, arbetsmarknadsdagar, konferenser och andra initiativ uppstår inte av sig själva, de blir till när vi tar initiativ tillsammans. Det är just detta som gör en ideell förening levande, relevant och dynamisk.

**Under 2026 sätter** styrelsen upp nya, tydliga strukturer som gör det lätt att engagera sig och bidra:

- Vi tar fram smidiga rutiner för hur event planeras och genomförs, så att det ska vara enkelt att gå från idé till genomförande, med stöd i varje steg
- Vi kartlägger vad vi kan åstadkomma tillsammans och sätter realistiska men ambitiösa mål: baserat på medlemmarnas kompetens, intressen och tillgängliga resurser
- Vi vill nå ut till er medlemmar för att hitta er som vill komma med förslag till idéer och aktiviteter men även till er som vill driva saker framåt: och skapar former för att matcha idéer med engagemang
- Vi fördjupar samarbeten med andra föreningar för att stärka statistikens synlighet och betydelse i samhället, och skapar fler gemensamma aktiviteter

**2026 kommer att** bli ett år av förnyelse. Ett år där vi tar oss tid att bygga rätt grund för framtiden, organisatoriskt, innehållsmässigt och gemenskapsmässigt. Med den grunden på plats ser vi fram emot att tillsammans skapa nya, inspirerande initiativ som utvecklar både föreningen och statistikens roll i samhället.

NANCY STEINBACH

Ordförande, för styrelsen i Svenska statistikfrämjandet

## Reseberättelse från ISCB i Basel – med fokus på metodutveckling för kliniska prövningar

augusti 2025 deltog jag i *Annual Conference of the International Society for Clinical Biostatistics* (ISCB) i Basel, Schweiz. Konferensen samlar både metodforskare och tillämpade forskare inom biostatistik och klinisk prövningsmetodik och passar mig därför bra, med tanke på att jag jobbar med metodutveckling för kliniska prövningar hos MRC Biostatistics Unit i Cambridge. Syftet med mitt deltagande var att presentera och diskutera min egen forskning och att ta del av aktuell metodutveckling inom biostatistik.

Basel är ett centrum för läkemedelsindustrin, med företag som Roche och Novartis, vilket avspeglades i antalet deltagare (rekord för ISCB!). Det bidrog också till ett tydligt fokus på praktiska tillämpningar inom kliniska prövningar under konferensen, vilket var bra för mig, eftersom jag gärna lär mig mer om praktisk användning av statistiska metoder.

Mitt forskningsprojekt rör hypotesprövning för adaptiva designar i kliniska prövningar. Jag presenterade arbetet *A test for treatment differences using allocation probabilities in response-adaptive clinical trials*. Efter min presentation diskuterade jag mina resultat med flera andra deltagare som var intresserade och arbetade med relaterade ämnen. Det var kul att höra att det finns intresse för adaptiva me-

toder inom kliniska prövningar och att forskningsfrågorna inom projektet är relevanta.

I övrigt var det många intressanta presentationer under konferensen, inom många olika områden inom biostatistik. En särskilt intressant session belyste metodforskarens ansvar att inte enbart redovisa korrekta statistiska antaganden när de utvecklar nya metoder och modeller, utan även tydligt diskutera deras rimlighet och giltighet i praktiska tillämpningar. Detta för att underlätta för tillämpade forskare och i förlängningen bidra till mer tillförlitlig forskning.

Deltagandet i konferensen var möjligt tack vare Wallinska resestipendiet. För juniora forskare kan det vara svårt att finansiera konferensresor, trots deras betydelse för kunskapsutbyte, kompetensutveckling och möjlighet att knyta värdefulla kontakter. Därför vill jag avslutningsvis rikta ett varmt tack till Svenska statistikfrämjandet för tilldelningen av Wallinska resestipendiet, som möjliggjorde mitt deltagande i ISCB 2025!

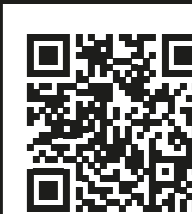
STINA ZETTERSTRÖM  
Research associate,  
MRC Biostatistics Unit,  
University of Cambridge



### Har du flyttat?

Du kan själv ändra dina uppgifter genom att logga in på vår medlemsportal [stat.memlist.se/login/tb7LkmN](https://stat.memlist.se/login/tb7LkmN) eller skanna QR-koden.

Användarnamnet är din mailadress och lösenordet väljer du själv genom att klicka på "Glömt lösenord". Vid frågor, kontakta Samuel Hellman på [sekreterare@statistikframjandet.se](mailto:sekreterare@statistikframjandet.se).



# Varför behöver vi prata om statistisk litteracitet?

“People often underestimate the difficulty of doing this well, due to the ease of doing it poorly.” Så står det i informationen om the Rousseuw Prize for Statistics. Detta stämmer också med Dunning-Kruger effekten, alltså att det finns en tendens att överskatta sin kompetens inom områden där man har begränsade kunskaper.

Sedan millennieskiftet har tillgången på data ökat explosionsartat. Detta har inneburit möjligheter men också utmaningar.

Hans Rosling (1948–2017) drev en kampanj för statistisk litteracitet som fick en hel del uppmärksamhet. Tillsammans med sin son Ola och sonhustru Anna bildade han Gapminder, en icke-vinstdrivande organisation som identifierar missuppfattningar om viktiga samhällsutvecklingar. De har också utvecklat undervisningsmaterial baserat på tillförlitlig statistik. Hans Rosling utsågs till Årets statistikfrämjare 2014 av Svenska statistikfrämjandet. Några kollegor i vår schweiziska systerorganisation har initierat ett projekt i Hans Roslings anda kallat ”Unlocking the Power of Data Initiative”. I skrivande stund har initiativet lagts på is, men det pågår en utvärdering.

Med litteracitet har man traditionellt avsett elementär läs- och skrivkunnskap, det vill säga färdigheter som i princip alla medborgare bör besitta för att ett modernt samhälle skall fungera. Graden av läskunnskap brukar vara nära 100 % i de flesta utvecklade länder. Ribban sitter således inte speciellt högt.

Efter hand har nya begrepp såsom finansiell litteracitet (vilket innebär elementär förståelse för termer såsom ränta, inflation och riskspridning) myntats. Mest adekvat för Qvintensens läsare torde vara begrepp som risklitteracitet, datalitteracitet och statistisk litteracitet. Jag skall inte försöka mig på att skarpt avgränsa dessa begrepp. Jag vill dock nämna att för risklitteracitet finns ett enkelt test med tre frågor kallat The Berlin Numeracy Test.

## Vad är statistisk litteracitet?

Engelskspråkiga Wikipedia har en artikel om statistisk litteracitet som visserligen är klassificerad som ”start-class” men ändå ganska läsvärd. Definitionen av statistisk litteracitet låter rätt rimlig:

”Statistical literacy is the ability to understand and reason with statistics and data. The abilities to understand and reason with data, or arguments that use data, are necessary for citizens to understand material presented in publications such as newspapers, television, and the Internet. However, scientists also need to develop statistical literacy so that they can both produce rigorous and reproducible research and consume it.”

En viktig aspekt av statistisk litteracitet är att ha en förståelse för när man kan dra slutsatser på egen hand och när man bör anlita någon form av expertis. Detta är betydelsefullt att poängtera inom utbildningar av olika slag.

Jag har ett tidigt minne av (brist på) statistisk allmänbildning. När jag gick på mellanstadiet vi-

»Cirka sextio år senare kan man konstatera att brist på statistisk litteracitet fortfarande är utbredd och långt ifrån bara för mellanstadielärare ...»

sade en lärare hur man kan få fram världsrekordet ett visst år genom att interpolera i ett diagram. Jag försökte förklara att rekordet var vad det var tills någon slog det. Läraren förstod inte vad jag menade.

Cirka sextio år senare kan man konstatera att brist på statistisk litteracitet fortfarande är utbredd och långt ifrån bara för mellanstadielärare, vilket man kan läsa

om bland annat i böckerna av Bauer et.al. och Hanne Kjöllér, som återfinns i referenslistan.

**Har intresset för** statistisk litteracitet ökat under 2020-talet? Man skulle kunna tycka att så borde vara fallet med tanke på alla som på egen hand ritade kurvor över smittspridningen under pandemin. När jag tittade på ”pageviews” under statistisk litteracitet i engelskspråkiga Wikipedia fick jag dock inte det intrycket. Grovt skattat var antalet sidvisningar dubbelt så stort i mitten av 2010-talet jämfört med mitten av

2020-talet (vi bör dock beakta att sättet att hitta information ändrats).

Som nämnts ovan skall man vara försiktig med att dra alltför långtgående slutsatser, det kan finnas flera orsaker som skulle behöva analyseras, men en annan indikation på att intresset för statistisk litteracitet har minskat är att SMHI tagit bort sannolikhetsdata från sina prognoser. Jag fick följande kommentar från SMHI:s kundtjänst i mars 2025: ”Vi har tagit bort prognosens säkerhet av tekniska och användarvänlighetsskäl. Prognosen vi visar är alltid den mest sannolika. Vi arbetar på att utveckla nya sätt att förklara prognosens säkerhet. Mer information om prognosernas säkerhet kan du hitta i SMHI:s kunskapsbank.”

Därefter har sannolikhetsdata återinförts för nederbördsprognoser. Man kan konstatera att SMHI valt att presentera osäkerheter något annorlunda än exempelvis norska YR. Det kan finnas anledning att ytterligare studera hur osäkerhetsdata för väderprognoser lämpligen bör presenteras och hur de tolkas av allmänheten. Personligen tycker jag att SMHI:s nya approach, där man redovisar sannolikheten för nederbörd både på tim- och dygnsnivå ger ökad tydlighet.

## Några aktuella exempel

Det går att hitta åtskilliga exempel på hur statistik används i media på ett tvivelaktigt sätt. Nedan ges några exempel som dykt upp på senare tid.

## Är tyska läkare tre gånger effektivare än svenska?

Det har ibland påpekats av bland annat Leif Östling att exempelvis tyska läkare hinner med cirka 2 000 patientbesök per år jämfört med 700 för deras svenska kollegor. Om detta vore ett adekvat mått för jämförelse vore det verkligen anmärkningsvärt. Borde det inte i så fall märkas på exempelvis läkartätheten (vilken om man tittar på antalet praktiserande läkare är något högre i Tyskland) eller medellivslängd (vilken är något högre i Sverige)? I en debattartikel i SvD 2023-01-25 skriver barnläkaren och professorn Jonas F Ludvigsson: ”I Sverige försöker vården maximera hälsan, vi räknar inte patienter som andra räknar pinnar. Det leder till att patienter



Människor behöver ha en god grundförståelse för hur statistik och data fungerar och vilka slutsatser som går att dra. Ökad undervisning i grundskola och gymnasium föreslås för att förstå statistiska resonemang.

med begränsade besvär (exempelvis behov av att mäta blodtryck eller smärta i en armbåge) ofta går direkt till sjuksköterska respektive sjukgymnast för bedömning, eftersom det är mer effektivt än att först passera läkare för ett 5-minuters-besök (ett läkarbesök som annars bättrat på läkarstatistiken men inte gjort någon väsentlig skillnad för patienten).” Det låter som en rimlig delförklaring. Observera att jag inte har tagit ställning till huruvida vården i Sverige eller Tyskland är mest effektiv, bara påpekat att man måste iaktta viss försiktighet när man använder vissa mått för jämförelser.

## Hur mäter man graden av (o)rättvisa i betygs-system?

Ur ”Vi Lärare” 2025-02-25: ”Om det nya betygssystemet redan varit verklighet hade mer än var femte student på de mest attraktiva universitetsutbildningarna inte kommit in. – Det är ett mått på hur orättvist dagens system är, säger utredningssekreteraren Anders Auer.”

Frågan är om det verkligen är ett bra mått? Dels så antar man att det nya systemet är helt rättvist, men uppenbarligen är det svårt att i förväg bedöma om ett system är rättvist eller inte.

Man kan också fråga sig om det går överhuvudtaget att definiera rättvisa? Skall en plats på exempelvis KI ses som belöning för goda insatser i gymnasiet, eller skall man välja ut dem som har bäst förutsättningar att klara första årets tentamina, eller dem som troligen blir bäst läkare? Varför infördes det nuvarande systemet om man inte trodde att det var hyfsat rättvist? Men fränsett detta: då det krävs i stort sett maximalt betyg för att komma in på de mer attraktiva linjerna så kommer ingen in med god marginal. Låt oss ta ett extremt exempel på en utbildningslinje med bara en plats som väldigt många duktiga elever eftertraktar. Med system A är Adam marginellt bättre än Berit. Med system B är Berit aningen bättre. Kan man då säga att system A är mycket orättvist då ingen av dem som antagits med detta system hade kommit in med system B? Låt oss ta den andra extremen

att samtliga behöriga sökanden kommer in, då blir det 100 % överlapp oberoende av hur (o) rättvisa systemen är.

## Coronapandemin

Qvintensen 2020/2 har av naturliga skäl stark fokus på coronapandemin. Bland annat bidrog Åke Wissing med en artikel: ”Vad har forskare och myndigheter för kunskap?”. Under pandemin dryftades en del statistikrelaterade frågor tämligen livligt.

Det var därför intressant att läsa Anders Tegnell's bok ”Tankar efter en pandemi”. Tegnell studerade matematik ett år i USA, där han också träffade sin nederländska fru. Även om han själv håller med om att studierna inte var speciellt krävande indikerar det ändå ett visst intresse för matematik och statistik. Tegnell nämner också att han haft Hans Rosling som föreläsare och inspirerats av honom. Tegnell's mentor Johan Giesecke hade studerat teknisk fysik med inriktning tillämpad matematik innan han sadlade om till läkarbanan. Tegnell nämnde att Folkhälsomyndigheten hade en del kompetenta matematiker och statistiker och nämner specifikt Lisa Brouwers, som enligt LinkedIn är doktor i datavetenskap vid Stockholms universitet. Hon och Tegnell står som medförfattare till en artikel i The Lancet med titeln ”Evaluation and communication of pandemic scenarios”.

Tegnell tar upp flera frågeställningar med bäring på coronastatistik, som dryftats flitigt under pandemin. Exempelvis kom han fram till att överdödlighet var det mest relevanta måttet vad gäller att jämföra situationen i olika länder, trots uppenbara brister (det kan vara svårt att veta hur stor del av överdödligheten som beror på pandemin och hur stor del som beror på andra orsaker, såsom värmeböljor och stokastiska variationer). Eurostat tycks ha kommit till samma slutsats om man tittar på deras dashboard.

**En fråga som** diskuterades i början av pandemin var testning. Folkhälsomyndigheten fick kritik för att man inte prioriterade testning tillräckligt. Jag minns att en dam som brukade ha

## »Ett riskmoment är övertro på AI-genererade resultat, något som bör beaktas på alla utbildningsnivåer.»

starka åsikter i många frågor skrev att ”man kan väl testa även om testerna inte är perfekta?”. Tegnell framhåller att ”testning är inte bara att peta folk i näsan utan en lång resurskrävande process”.

Överdödligheten kan beräknas på lite olika sätt, men allmänt kan sägas att frånsett första vågen under våren 2020 hörde Sverige till de länder som klarade sig bäst. Oberoende av pandemin, men under ungefär samma period, genomgick Folkhälsomyndigheten processen att bli godkänd som statistikansvarig myndighet. Svaret på frågan i Qvintensen 2020/2 – vad har forskare och myndigheter för kunskap? – är troligen ”en hel del” även om det säkert finns förbättringspotential.

**Fler exempel på** tveksamt använd statistik kan vi hitta på den tyska hemsidan ”Unstatistik des Monats”, liksom den svenska ”Statistik och strunt” som underhålls av Christian Lindell.

### Hur går vi vidare?

Under 2020-talet har ett antal ”data literacy charters” tagits fram, bland annat i Tyskland och Schweiz. Dessa lyfter fram ett behov av spritt kunnande inom datalitteracitet. En svaghet med dessa charters är att de avstår från att nämna statistisk litteracitet (vilket dock görs i ett liknande spanskt dokument). Erfarenheten har visat att definitionsfrågor kan vara viktiga. Jag tycker att det vore bra om man inom FENStatS tar fram en ”statistical and data literacy charter”. Begreppen torde vara mycket överlappande.

Bägge nämnda charters förordar ökad undervisning inom datalitteracitet (vilket då bör inkludera statistik) inom hela utbildningsväsendet, inklusive livslångt lärande. Där skulle man kunna lyfta fram några konkreta exempel:

- Inom grundskola och gymnasium för att förstå statistiska resonemang i exempelvis dagspressen, läsa bipacksedlar, förstå väderrapporter.
- Inom akademisk grundutbildning för att förstå statistiska data i forskningsrapporter inom det egna området.
- På forskarutbildningsnivå för att kunna bidra till utveckling av statistiska idéer.

**I ett samhälle där** tillgången till data har ökat explosionsartat behöver vi prata om statistisk litteracitet. Människor behöver ha en god grundförståelse för hur statistik och data fungerar och vilka slutsatser som går att dra. Grundskolans läroplan ger intryck av en ganska hög ambitionsnivå vad gäller statistikfärdigheter, frågan är hur det ser ut i praktiken. Ett riskmoment är övertro på AI-genererade resultat, något som bör beaktas på alla utbildningsnivåer.

HANS ALBERG

### Referenser

- Alberg, H. (2024) The Importance of Statistical Literacy, FENStatS Newsletter June 2024.
- Bauer, T., Gigerenzer, G., Krämer, W., & Schüller, K. (2022). *Grüne fahren SUV und Joggen macht unsterblich: Über Risiken und Nebenwirkungen der Unstatistik. Gewinner des getAbstract Business Impact Award 2023*. Campus Verlag.
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision making*, 7(1), 25-47. <https://www.cambridge.org/core/journals/judgment-and-decision-making/article/measuring-risk-literacy-the-berlin-numeracy-test/A9B26516D12D48EFA4BD3560E2001E8E>
- Eurostat Recovery Dashboard European Statistical Monitor
- Gerlee, P., Thoreén, H., Joöud, A. S., Lundh, T., Spreco, A., Nordlund, A., ... & Timpka, T. (2024). Evaluation and communication of pandemic scenarios. *The Lancet Digital Health*, 6(8), e543-e544.
- Kjöllér, H. (2020). *Kris i forskningsfrågan: Eller vad fan får vi för pengarna?* Fri Tanke.
- Lindskog, M., Kerimi, N., Winman, A., & Juslin, P. (2015). A Swedish validation of the Berlin numeracy test. *Scandinavian Journal of Psychology*, 56(2), 132-139.
- Rosling, H., Rosling, O., & Rönnlund, A. R. (2018). *Factfulness: ten reasons we're wrong about the world--and why things are better than you think*. Flatiron book
- Spanish Society of Statistics and Operations Research (SEIO). (2025). Let's put an end to statistical illiteracy! [English\\_Manifesto\\_Statistical\\_Literacy.pdf](#)
- Stifterverband (2021) Data Literacy Charter. [www.stifterverband.org/sites/default/files/data-literacy-charter.pdf](http://www.stifterverband.org/sites/default/files/data-literacy-charter.pdf)
- Swiss Academies of Arts and Sciences. (2024). Swiss Data Literacy Charter. Swiss Academies of Arts and Sciences. [doi.org/10.5281/zenodo.11146023](https://doi.org/10.5281/zenodo.11146023)
- Swiss Federal Statistical Office (2025) "Unlocking the Power of Data Initiative"
- Tegnell, A., & Härgestam, E. (2023). *Tankar efter en pandemi: och lärdomarna inför nästa*. Natur & kultur.
- Wissing, Å., (2020) Statistik i Corona-tider. Vad har forskare och myndigheter för kunskap? Qvintensen 2020/2

# Vad är det egentligen frågan om i enkäter?

## SMÅ SKILLNADER I HUR FRÅGOR STÄLLS KAN GE STORA SKILLNADER I STATISTIKEN

**Det är glädjande** att diskussionen om statistisk läskunnighet får allt större uppmärksamhet. Att kunna läsa, tolka och värdera statistik är en grundläggande kompetens i ett samhälle där siffror används för nästan allt, från hälsa och utbildning till politik och ekonomi. Förhoppningsvis inser fler vikten av att kunna bedöma diagram, tolka osäkerhet och genomskåda slutsatser som inte riktigt håller. En aspekt som jag ofta ser hamnar i skymundan är vad som sker innan siffrorna ens blir till – frågorna bakom dem.

Jag fastnade nyligen för en siffra i en artikel: 85 procent av befolkningen uppgavs ha en ohälsosam kost. Det är en sådan uppgift som väcker både oro och nyfikenhet. Hur har man egentligen kommit fram till den siffran? Vad räknas som ”ohälsosamt”?

Jag började följa referenserna bakåt. Artikelnen hänvisade till en studie, som i sin tur byggde på en annan. Där definierades hälsosam kost i termer av medelhavsdiet. Men när jag till slut hittade den ursprungliga studien och enkätfrågorna som låg till grund för resultaten var det bara en fråga som handlade om kost: ”Hur

ofta dricker du läsk eller andra sötade drycker?”. Av de 27 procent som svarade så hade 15 procent uppgett att de *aldrig* drack läsk eller andra sötade drycker.

Plötsligt framstod siffran i ett annat ljus. Från att ha låtit som en bred och allmängiltig bedömning av människors matvanor visade den sig vila på en ganska snäv indikator.

**Mycket av den** statistik vi möter bygger på enkäter. Det gäller inte

bara opinionsmätningar utan också utvärderingar av insatser, studier om hälsa, livsvillkor, attityder och beteenden. Enkäter är i praktiken en av de vanligaste vägarna från mänsklig erfarenhet till siffra. Ändå behandlas de ofta som en enkel transportsträcka: frågor ställs, svar samlas in och sedan finns data redo för analys. Men det är i det här till synes enkla steget som den information vi senare använder som statistik faktiskt tar form. Varje svar påverkas i hög grad av hur frågan är formulerad och tolkad, och därför borde förståelsen



Sara Scholtens.

för frågorna bakom siffrorna vara en självklar del av statistisk läskunnighet.

Att svara på en enkät kan verka trivialt. Formatet är bekant: en fråga, några svarsalternativ, ett kryss. Men i verkligheten ska respondenten översätta något komplext – erfarenheter, minnen och känslor – till ett avgränsat svar. Ofta ska detta ske snabbt, utan möjlighet till förtydliganden. Begreppen kan vara vaga, frågorna kan innehålla dolda antaganden, och ibland finns inget svarsalternativ som riktigt passar.

I ett samtal skulle vi kunna resonera, ställa frågor, kanske ändra oss under tiden. I en enkät finns inte det utrymmet. Frågan står där och svaret ska ges inom de ramar som redan är satta. Ändå visar ögonrörelsestudier att vi sällan läser hela frågan noggrant. Många hoppar direkt till svarsalternativen eller skummar texten, och hjälptexter som förklarar hur man ska tolka en fråga läses nästan aldrig. Det gör svaret till ett komplext beteende där vår tolkning påverkas av vana, kontext och snabba bedömningar.

Inom enkätforskningen beskriver man ofta svaret

# Visuell datakommunikation i ett postsanningens tidevarv

»Att ta ett steg tillbaka och försöka förstå vilka frågor som ligger bakom en siffra kan förändra hur vi tolkar den.»

som en process i flera steg. Frågan ska förstås, relevant information ska hämtas ur minnet, en bedömning ska göras och översättas till ett av de svarsalternativ som erbjuds. Men processen är långt ifrån mekanisk. Den handlar om tolkning, och tolkning sker alltid i ett sammanhang.

**Exemplen är många.** En fråga som *hur trygg känner du dig i ditt område?* kan ge svar som visar att 80 procent känner sig trygga, medan en annan formulering, *hur orolig är du för att utsättas för brott i ditt område?*, kan visa att 40 procent i samma område känner oro. Den första frågan fokuserar på generell trygghet, den andra på upplevd risk och oro. I utvärderingar kan en fråga som *har kursen gjort att du blivit bättre på att samarbeta?* ge att majoriteten svarar att de blivit bättre, eftersom frågan förutsätter förbättring, medan en mer öppen formulering om förändring kan visa ett mer nyanserat resultat. Frågor som *hur nöjd är du med stödet du fått?* påverkas inte bara av innehållet av stödet utan också av relationen till den som ställt frågan och av sociala normer, vilket kan ge stora skillnader mellan olika grupper eller sammanhang. Nöjdhet säger dessutom inte nödvändigtvis något om tillräcklighet, det går att vara nöjd med det stöd man fått och samtidigt ha behov av mer stöd. Detta blir särskilt problematiskt när sådana resultat används för att dra slutsatser om kvaliteten.

Det är inte heller bara själva formuleringen som påverkar svaren. Placering i enkäten, ordningen på frågor och vilka svarsalternativ som erbjuds styr också tolkningen. En fråga om träning kan uppfattas olika beroende på om enkäten inleds med hälsorelaterade frågor eller om den handlar om fritid. Vad som räknas som träning och hur ofta det sker avgör respondenten själv. Små förändringar i formuleringar eller alternativ kan ge påtagliga skillnader i svaren. Det är sällan slumpmässigt, utan konsekvenser av hur frågorna fungerar och hur de förstås.

Det finns också en seglivad föreställning om

att brister i en enkät kan kompenseras i efterhand genom statistisk analys. Om resultaten inte verkar rimliga tänker man sig ofta att det går att justera, vikta eller modellera sig fram till något mer korrekt. Men om frågorna är otydliga, snävt formulerade eller i grunden missar det man vill mäta så hjälper det inte hur avancerad analysen är. Den kan inte återskapa information som aldrig samlades in från början. Statistiken blir i bästa fall en förfining av det material som finns, inte ett sätt att ersätta bristande mätning.

Mot denna bakgrund blir det tydligt att förståelsen för frågorna bakom siffrorna inte är något man bara kan hoppa över. Kompetensen att formulera bra enkätfrågor kräver erfarenhet och vetenskaplig grund. Det är ett hantverk där kunskap om språk, kognition, perception, mentala modeller och mätteknik samspelar för att skapa frågor som ger meningsfulla svar. Den som behärskar detta har inte bara ett verktyg för att samla in data utan också en djupare förståelse för vad statistiken egentligen säger.

För den som möter statistik i vardagen behöver det inte vara så komplicerat. Man behöver inte själv behärska konsten att konstruera enkätfrågor för att bli en mer kritisk läsare av siffror. Ofta räcker det långt att ställa en enkel fråga: vad är det egentligen som har mätts?

Att ta ett steg tillbaka och försöka förstå vilka frågor som ligger bakom en siffra kan förändra hur vi tolkar den. Det gör det möjligt att se begränsningarna, nyanserna och de ibland ganska stora glapp som kan finnas mellan en siffra och det fenomen den påstås beskriva.

Statistisk läskunnighet handlar därför inte bara om att kunna läsa av diagram eller förstå procenttal. Det handlar också om att vara nyfiken på hur siffrorna blivit till. I många fall börjar den förståelsen med något så enkelt som att gå till botten med frågorna bakom dem – och att våga fråga sig: vad är det egentligen frågan om?

SARA SCHOLTENS  
Fil. dr, metodspecialist

## VÄNDNINGEN: ÅKESSON FÅR LÅNGA BEN!

Titla på det här stapeldiagrammet som Sverigedemokraternas officiella Instagramkonto publicerade den 23 januari 2026. Bildtexten lyder: "Jimmie Åkesson har nu det starkaste förtroendet bland partiledarna, för första gången någonsin, enligt SvD/Demoskops mätning." Det här är hela bilden, så som den fortfarande låg kvar på Instagram den 25 april. Inget mer finns att scrolla fram.



Jimmie leder denna opinionsmätning med en procentenhet: 43 % mot 42 %. Om staplarna hade återgivits proportionerligt skulle det motsvara en längdskillnad på 1.024 gånger, inte 1.24 gånger som vi ser i figuren. Det beror på att staplar i ett stapeldiagram jämförs visuellt genom sin längd. För att proportionerna ska stämma måste de därför börja vid noll. Låt mig rätta till det genom att återskapa bilden i full höjd. Om Jimmie står ungefär där staplarna börjar, då måste han få enormt långa ben. ➔

Det här är ett skolboksexempel på det som brukar kallas *lying with charts*, att vilseleda med diagram [4, 7]. Fenomenet fick en stor spridning under Donald Trumps valkampanj 2016, som producerade en rad sådana diagram: de flesta överdrev Trumps ledning,



några tonade i stället ned den. The Washington Post publicerade då en genomgång där man jämförde affischerna och förlängde staplarna, ungefär som jag gjort här, för att tydliggöra manipulationen [13]. Hur hamnade vi här? Hur bör vi kommunicera viktig information när data och diagram, tidigare nästan synonymt till vetenskap och sanning, används för vilseledande propaganda?

I denna artikel gör jag ett försök att besvara dessa frågor genom att lyfta fram en relation mellan presentationer av statistiska bilder, s.k. datavisualisering, och de kulturella strömningar som formar vår syn på hur världen kan förstås och beskrivas. Jag håller mig till den följande förenklade modell av utvecklingsfaser och dess benämningar, som de är sammanfattade hos Ken Wilber [1] på spår av flera socialvetare och utvecklingspsykologer. Jag kallar dem tankebanor och inte utvecklingsfaser för att betona att de existerar samtidigt i samhället såväl som i individer.

- Arkaisk tankebanor: "världen kan förstås och beskrivas genom magi"
- Traditionell tankebanor: "världen kan förstås och beskrivas genom Gud, regler, ordning"
- Modern tankebanor: "världen kan förstås och beskrivas genom vetenskap, observation, experiment"
- Postmodern tankebanor: "världen kan förstås och beskrivas genom perspektiv, konstruerad kunskap, språk, orättvis maktfördelning"

### Datavisualisering enligt den moderna tankebanan: sanning utan perspektiv

Att gestalta information visuellt har följt människan sedan grottmålningarnas tid. Det såg förstås inte ut som dagens diagram, men fyllde samma grundläggande funktioner: att kommunicera, berätta och bevara spår av det som varit. Det moderna sättet att presentera information visuellt brukar spåras till Galileos manuskript från 1612. De framhålls fortfarande som nydanande, kanske därför att vi under en lång period knappt kunde använda fria grafiska inslag mitt

i löptexten, helt enkelt för att det typografiska hantverket gjorde det för dyrt och besvärligt. Nu kan vi nästan göra det igen med datorer, efter sedvanlig brottningsmatch med textverkyget: Saturnus ☉ ←.

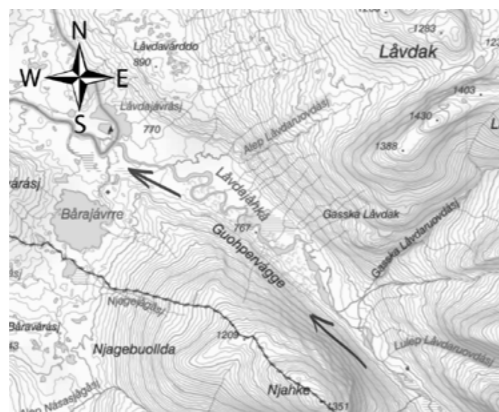
**I takt med** att allt mer data behövde kommuniceras växte också vår repertoar av visuella uttryck. Genom William Playfair, Charles Joseph Minard och Florence Nightingale fick vi många av de former som i dag känns självklara: linjediagram, stapeldiagram, flödeskartor av sankey-typ, solfjäders-diagram och mycket annat. Jag hoppas att du redan ser dessa klassiska verk framför dig och att jag slipper lägga utrymme på att presentera dem i artikeln. Det här sättet att tänka kring datavisualisering när på många sätt sin kulmen i Edward Tufes böcker, särskilt i *The Visual Display of Quantitative Information* [6]. Där spikar han det modernistiska idealet för informationsgrafik: inga dekorativa störningar, inget "chartjunk", en stark norm om sanningsenlig återgivning t.ex. stapellängd ska vara i proportion med mätvärden, och en hög andel informationsbärande bläck i förhållande till den totala mängden bläck.

Det Tufte hade emot sig var inte bara bristande hantverksskicklighet i informationspresentationer och inte heller den växande förekomsten av PowerPoint-estetik och Microsoft Office diagramslop. I bakgrunden pågick också något större: en förändring i vårt förhållande till sanning och mening.

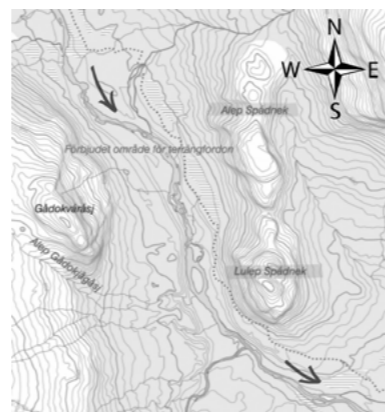
På universitetskurs i vetenskapligt skrivande lärde vi oss att vetenskapliga artiklar bör skrivas på ett neutralt språk och med passiv form som att författaren inte finns. En annan forskare ska kunna följa samma metod och komma fram till samma resultat, alltså till Sanningen som finns där oberoende av forskaren. Men under de senaste decennierna har både forskarens, finansierarens och läsarnas roll i meningsskapandet blivit allt svårare att bortse från. Enligt denna tankebanan, som fick namnet *relativism* eller *postmodernism*, framstår vetenskapen inte längre självklart som en process för att upptäcka kunskap, utan snarare, i bästa fall, som en process för att producera det.



Anders Celsius termometer, museum Gustavianum med upp och ner skala.



Lantmäteriets topografiska kartor över delar av Sareks nationalpark visar två undantag när Alep och Lulep ortnamn inte stämmer med väderstreck och flödesriktning. Författarens anteckningar.



### Datavisualisering enligt den postmoderna tankebanan: perspektiv framför sanning

Ta temperaturmätning som till skillnad från opinionsmätningar är ett fenomen som ändå framstår som ganska objektivt. Vilken skala använder vi? Vi vanliga människor mäter temperatur i Celsius. Molekyler rör sig snabbare, temperaturen känns varmare, alltså borde också siffran bli större. Men titta då på Anders Celsius originaltermometer på Gustavianum i Uppsala. Där är skalan omvänd: 0 vid kokpunkten och 100 vid fryspunkten. På Uppsala universitets webbplats står det: "Det var först året efter Celsius död som Carl von Linné vände skalan rätt ur vårt moderna perspektiv." Varför skriver de "rätt ur vårt moderna perspektiv" och inte "objektivt rätt punkt slut"?

**På lulesamiska betyder** ortnamnen alep "högre, uppströms" och lulep "lägre, nerströms". Samma ord betyder även "västligare" och "östligare", eftersom Luleälven rinner från väst mot öst. I det perspektivet är kopplingen oftast fullt rimlig. Undantagsvist dock kan alep-stället ligga nerströms om det motsvarande lulep-stället som bäckraviner Lävduovdåsj på kartan till vänster, alternativt kan alep-stället ligga nordväst eller till och med direkt norr om lulep-stället. Alep och Lulep Spådnäk på kartan till höger skulle då egentligen heta Nuortap och Oarjep Spådnäk: respektive nordligare och sydligare.

Att försöka lösa upp sambandet mellan uppströms/nerströms och väster/öster i ortnamn skulle vara bekvämt och logiskt men det skulle idag betraktas som ett epistemiskt förtryck — att påtvinga en annan kultur sin egen världsbild. Jag är helt nöjd med kulturminneslagen och att kartor bevarar dessa toponymiska klurigheter. Poängen med detta exempel är att även väderstreck inte slipper vara kontextberoende.

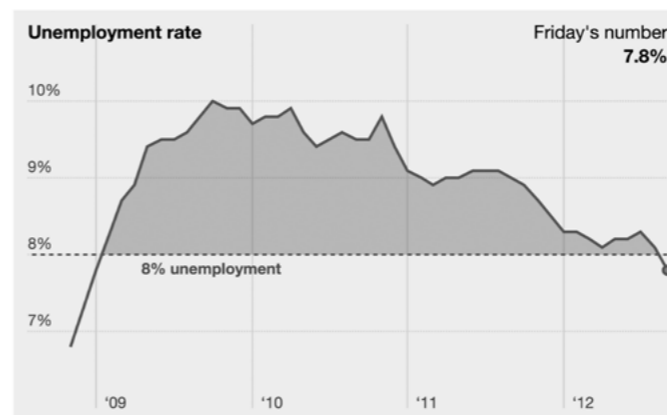
När vår världsbild allt mer framstår som något i stil med en kollektiv överenskommelse, "kulturellt filtrerad och språkligt utformad", har också synen på datapresentation förändrats. Under de år jag själv har varit verksam i fältet har vi fått begrepp som *Data Humanism*,

*Data Feminism* och till och med *Data Vandals*. I denna synvinkel förväntas våra visuella framställningar inte bara redovisa information, utan också granska makt och utmana den, lyfta fram känslor och kroppslighet, ifrågasätta binariteter och hierarkier, synliggöra villkoren för kunskapsproduktionen och öppna för fler perspektiv [5]. Till och med ordet *data*, pluralformen av *datum*, har glidit i betydelse. Från att ha syftat på "det givna" eller "det som är givet" som i "dativ" grammatiskt kasus — givet av vem? Av Gud förstås till att börja med, senare av den tidigare auktoritativa vetenskapen — till något som samlas in, lagras och bearbetas i kunskapsprocessen, en råvara i informationsåldern [3]. Johanna Drucker 2011 [11] lägger fram ett förslag att man helt borde överge ordet *data* till förmån för *capta*, just för att markera att det inte handlar om något som är givet, utan om något som är taget, extraherat, samlat och konstruerat. Betrakta trenden om arbetslöshet i USA under Obama-perioden som är presenterat nedan med två linjediagram. Båda bilder visar exakt samma data och är perfekt faktabaserade, men sanningens ramar lämnar fortfarande ett stort utrymme för grafisk design och tolkning. Den första bildens budskap är grovt sagt "Obama good", den andra "Obama bad".

**Ben Shneidermans 1996** klassiska mantra för visuell informationssökning, "overview first, zoom and filter, then details on demand" [10], vänds idag helt upp och ned, som jag kan observera i dagens produktion av interaktiva webbdiaagram. I stället börjar man med ett enda visuellt element, låter läsaren bekanta sig med det genom annotationer och förklaringar av visuella variabler, och lägger sedan successivt till fler element tills grupper, mönster och kluster blir synliga. Först därefter får helhetsbilden växa fram. Det handlar alltså inte längre om ett top-down-orienterat informationssökande, utan snarare om en bottom-up-baserad visuell teater, där berättandet får bära läsaren fram till överblicken och där skådespelarna (geometriska element) utvecklar en relation med läsaren.

Även noggrannhet av datapresentationer

The rate was above 8 percent for 43 months.

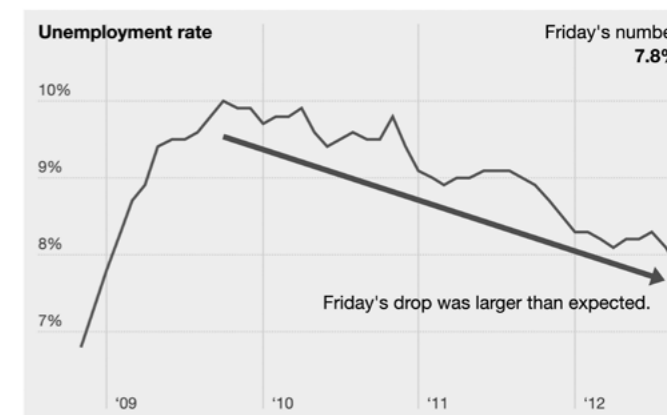


En datavisualisering av jobbberapporten för september 2012, sedd ur demokraternas respektive republikanernas perspektiv. The New York Times datateam visar hur enkla redaktionella grepp kan leda till stora skillnader i inramning och tolkning, som citeras i [5].

får kritik. Osäkerhet är något som inte återges särskilt väl i informationsgrafiker, åtminstone utanför den lilla krets av professionella läsare som faktiskt vet hur den ska tolkas. De typiska exemplen är felstaplar, orkankoner med mittlinjen och linjediagram med felmarginal. Dessa visuella framställningar av osäkerhet avfärdas ofta eller ignoreras av en bredare publik. I stället ger våra diagram, just genom de verktyg vi använder, ett starkt intryck av precision. Vektorgrafikens matematiska exakthet, högupplösta skärmar och trycksaker av god kvalitet får dem att framstå som polerade, exakta, nästan som skurna i metall med hjälp av en datorstyrd tillverkningsmaskin (CNC). Resultatet blir ett sken av auktoritet: kanske inte det sista ordet i frågan, men åtminstone ett mycket självsäkert ord. Dock, hur mycket tvivel finns i data och ska egentligen synas? Opinionsciffror bland annat kan mycket väl ha  $\pm 2...3$  procentenheters osäkerhet, vilket gör 1-procentenhet skillnaden helt betydelselös.

Ur ett postmodernt perspektiv kan just dessa precisa och överblicksorienterade visualiseringar kritiserars för att göra det som Donna

The rate has fallen more than 2 points since its recent peak.



Haraway [12] kallade "the god trick", att tala som om blicken kom från ingenstans och därför såg allt. "Above all else show the data", skriver Tufte. I den snävare tolkning som ofta följer av ett sådant ideal blir svaret att tona ned anspråket på precision: att kanske bara visa felmarginal och ta bort mittlinjen enligt förslag av John Burn-Murdoch nedan, eller låta linjen bli suddig, ojämn, "feminin" eller lätt darrande med hjälp av animation.

**I en bredare mening** handlar invändningen om något ännu större: vem är det egentligen som har rätt att deklarerar sanningen? En omsorgsfull datafeminist enligt [5] skulle här uppmärksamma de ojämlika maktstrukturer som präglar kunskapsproduktionen och försöka skapa en visualisering som bidrar till det som i boken kallas "data for co-liberation".

Men om sanningen inte upptäcks utan produceras, vad är det då för grafik Jimmie ska lägga ut på instagram när han får en (1) procentenhet högre förtroendesiffra än Magdalena Andersson i en opinionsmätning? Exakt.



En av bilderna från 2016 Donald Trump valkampanj [13].

**När världen förstås** genom konstruerad kunskap, språk och kontext framträder också en värld utan förankringspunkter. Medan mina kollegor i akademiska kretsar ägnar sig åt att kontextualisera och hitta på nya mer neutrala och politiskt korrekta ord använder populisterna samma verktygslåda för att vinna röster. "Fake news", "Alternative facts", "Truth Social" När det gäller just stapeldiagram som inte börjar vid noll tycker jag inte längre att vi ska vara generösa och nöja oss med förklaringen att det bara handlar om bristande diagramkompetens. Efter Trumps valkampanj 2016 framstår det snarare som en etablerad metod, en medveten spelbok som visat sig fungera.

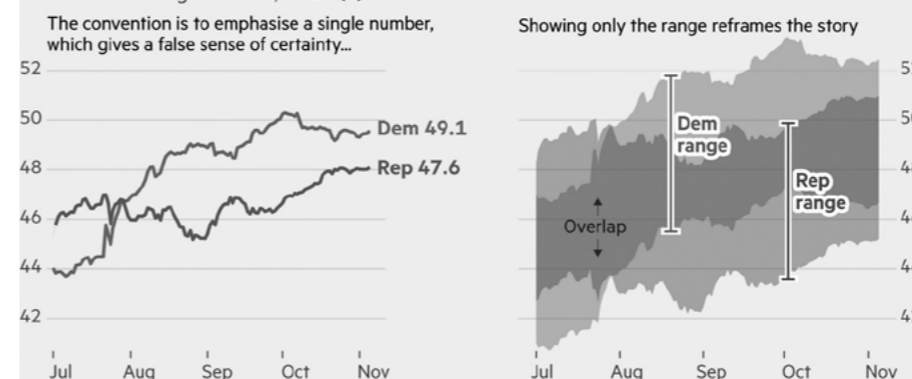
**Tänk efter: bidrar** Jimmies ur-vårt-perspektiv vilseledande diagram till ett högre eller lägre förtroende? Svaret får du om du läser kommentarerna på Instagram. Ifall du gör det, lägg också märke till hur känsla och kroppslig närvaro lyfts fram, precis så som datafeminism [5] föreskriver. Oops.

### Att leta nya tankebanor på en trasig cykel

Nu har vi ringat in terrängen. Frågan är vart vi går härifrån.

### A different representation of polling data could help shape a narrative that more accurately reflects reality and reduces the risk of unpleasant shocks

Presidential voting intention, 2024 (%)



Sources: FiveThirtyEight (polling data); FT poll tracker (averages, ranges) FT graphic: John Burn-Murdoch / @jburnmurdoch © FT

— Världen kan förstås och beskrivas genom integration och paradox.

Den moderna datavisualiseringen trodde på sanning utan perspektiv. Den postmoderna hyllade perspektivet men lyckades inte rädda sanningen. Det vi behöver nu är en praktik som kan hålla båda tankarna samtidigt: varje diagram är gjort, och vissa diagram är ändå fel. Uppgiften för datavisualisering i dag är därför både strängare och svårare. Vi måste erkänna att varje framställning är konstruerad utan att för den skull ge upp våra normer. Vi måste ta kontext på allvar utan att glida över i relativism. Och vi måste försvara en hopplöst gammalmodig men nödvändig tanke: att vissa grafiska framställningar är hederligare än andra. Ett diagram är alltid skapat. Därför är det inte oskyldigt. Men just därför är det heller inte bortom omdöme.

Steget från det moderna till det postmoderna tänkandet i sin kärna består i att erkänna betraktarens roll. Steget från det postmoderna till nästa lager, ofta kallat *metamodern* eller *integrativt* [1], består i att inse att *en människa inte är dennes tankar*. Det gör det möjligt att lossa greppet om de ”ismer” vi annars lätt börjar identifiera oss med eller anklaga våra opponenter vara, som feminist och populist.

**Det mänskliga tänkandet** kan liknas vid en cykel att ta sig fram med till slutsatser och beslut. Ibland rullar den fint, när tanken är klar. Ibland är den gammal och rostig, belagd med föråldrad kunskap och fördomar. Ibland är styret snett i förhållande till framhjulet, så att färdan drar åt ett håll utan att vi riktigt märker det. Wikipedia har en lång lista över kognitiva bias som vi alla har. Se annars ”Thinking fast and slow” av Daniel Kahneman och ”Predictably Irrational” av Dan Ariely. Att inse att man inte är en med sitt tänkande gör det möjligt att inspektera cykeln, laga den, byta ut delar eller välja en hel annan typ av cykel anpassad för en tankebana framför oss. Och ibland går det också bra att stiga av helt, sätta sig ner och ta en paus från själva tänkandet, från detta ständiga menande, tolkande, framför allt vetande.

Vilka diagram kan gestalta en sådan hållning?

Vilket arbetssätt inom datavetenskapen kan hjälpa oss att laga cykeln?



Hans Rosling: Global population growth, box by box, TED Talk från 2010. Finns bl.a. på YouTube.

Tänk tillbaka på senaste gången en datagrafik faktiskt ingav hopp. Inte på ett triumferande sätt, utan på ett nyfiskt och uppriktigt sätt. Som den där barnsliga förundran inför en oerhört komplex, intrikat och nyanserad process som plötsligt blev begriplig och relaterbar. Kanske känner du igen något av Hans Roslings och Gapminders metod i den beskrivningen. Varför inte luta sig mer åt det hållet?

**Börja med att** undersöka ämnet ordentligt. Tala med experterna, för sådana finns faktiskt. Om din uppgift är att kommunicera vetenskap till en bredare publik, bygg då förenklade modeller som gör ämnet begripligt utan att förlora det som gör det meningsfullt och användbart. Hitta och synliggör dina egna och läsarens bias, missförstånd och felaktiga tankebanor. Erkänn begränsningarna tydligt, omfamna paradoxer och motsägelser. Transparens är den nya neutraliteten. Få människor att skratta åt sina mentala cyklar, präglade av okunskap och hyckleri, men utan att skuldbelägga eller förödmjuka någon grupp. Få med dem i stället. Hjälp dem att laga cyklarna. Inkludera även motståndarens perspektiv och hjälp att överskrida det, snarare än att bara avfärda det som felaktigt. Ta med Ikealådor till presentationen, en fyra meter lång

pekpinne, ett svärd till något cirkusnummer. Var är ditt svärd? Vad är cirkustricket i ditt eget fält?

**Till sist, något** helt orelevant. ”The first thing that makes sense in 2026.” ”I can’t believe I was alive to witness the release of music 2.” ”They didn’t break the internet, they fixed it.” ”The elites don’t want you to know this, but you can just make your own notes.” — Detta är kommentarerna under en föreställning av Angine de Poitrine, en kanadensisk cirkusartad duo som spelar ett sorts ”mikrotonal math rock” och tycks ha tagit över internet när dessa rader skrivs. Tonåringar spelar deras musik på alla möjliga och omöjliga instrument, det är galet kul, men i övrigt helt irrelevant till diskussionen. Frågan är: vad skulle motsvara ett Angine de Poitrine-moment i statistisk kommunikation? Originellt, underhållande, roligt, uppriktigt men inte naivt, nyanserat men ovanligt lättillgängligt. Hoppingivande framför allt. Teckna ned några exempel du kan minnas eller hitta på, för vi behöver fler.

ANGIE HJORT

Utvecklare och föreläsare inom datavisualisering. Encharted Media AB

## Referenser

## Böcker

[1] 2024 — Ken Wilber — A

Post-Truth World: Politics, Polarization, and a Vision for Transcending the Chaos

[2] 2018 — Hans Rosling, Ola Rosling, Anna Rosling Rönnlund — Factfulness: tio knep som hjälper dig förstå världen

[3] 2021 — Jer Thorp — Living in Data: Citizen’s Guide to a Better Information Future

[4] 2020 — Carl Bergstrom, Jevin D. West — Calling Bullshit: The Art of Scepticism in a Data-Driven World

[5] 2020 — Catherine D’Ignazio and Lauren Klein — Data Feminism

[6] 1983 — Edward Tufte — The Visual Display of Quantitative Information

[7] 2019 — Alberto Cairo — How Charts Lie: Getting Smarter about Visual Information

[8] 2011 — Daniel Kahneman — Thinking, fast and slow

[9] 2008 — Dan Ariely — Predictably Irrational: The Hidden Forces That Shape Our Decisions

## Artiklar

[10] 1996 — Ben Schneiderman — The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations

[11] 2011 — Johanna Drucker — Humanities approaches to graphical display. Digital Humanities Quarterly 5 (1), 1-21, 2011.

[12] 1988 — Donna Haraway — Situated Knowledges: The Science

Question in Feminism and the Privilege of Partial Perspective.

[13] 2016 — John Muyskens — Most of Trump’s charts skew the data.

And not always in his favor. Washington Post <https://www.washingtonpost.com/graphics/politics/2016-election/trump-charts/>

# t-test och hypotesformuleringar – en kritisk granskning

På din första kurs i statistik (och kanske bara där?) har du väl fått lära dig att jämföra två normalfördelade stickprovsmedelvärden i ett t-test, samt att t-testa ifall det föreligger (lineär) regression av en variabel på en annan. Vi skall se närmare på dessa båda besläktade procedurer och litteraturen om dem, som kan gagnas av att granskas på ett litet djupare plan.

## Jämförelse av två stickprov

Kanske har du lärt dig jämföra medelvärdena för två oberoende, normalfördelade stickprov såhär, som en liten algoritm (avsiktligt litet löst formulerad här):

0. Formulera  $H_0$ : ”De två normalfördelningarna har ett och samma väntevärde”.
1. Bedöm om stickprovsvarianserna är ungefär lika (eventuellt med ett statistiskt pre-test).
- 2a. Om ”lika varianser” är godtagbart, använd ”Student’s  $t$ -test” av  $H_0$ : ”Samma väntevärde”.
- 2b. Annars, använd ”Welch-testet”, ett slags approximativt  $t$ -test, med färre frihetsgrader.

**Denna elementära situation** kan därmed ha ansetts avklarad, så den har kanske inte ägnats uppmärksamhet i någon kurs på djupare nivå. Det skall vi kompensera för här. Vi skall visa att situationen är ganska intrikat, och att det vanligen finns anledning att ifrågasätta algoritmen.

För drygt ett sekel sedan härledde R. A. Fisher fördelningen för  $t$ -testet, eller tydligare ”Student’s  $t$ -test” (uppkallat efter W. S. Gosset, med pseudonymen Student), som test av hypotesen

$H_0$ : ”De två stickproven kommer från en och samma (normalfördelade) population”.

Testet var designat för att ha styrka mot skillnad i väntevärdena, och  $t$ -teststorheten är (naturligen) differensen mellan stickprovets medelvärden, normerad på lämpligt sätt (med medelfelet). Men lägg nu märke till att Fishers nollhypotes är en annan än algoritmens  $H_0$ , som bara uttalade sig om väntevärdena. Ifall de två stickproven kommer från samma population, så representerar stickprovsvarianserna förstås samma gemensamma populationsvarians (som skattas med den poolade empiriska variansen), och  $t$ -testet är det adekvata testet att använda. *Vi behöver inte*

*göra en extra förutsättning om lika varians, för det ligger redan i formuleringen av nollhypotesen!*

Fisher fick snart anledning betona detta, sedan det börjat ifrågasättas om  $t$ -testet kunde användas när stickprovsvarianserna var (signifikant) olika, och  $t$  ex i Fisher (1934, avsnitt 24.1) kan man läsa:

*It has been repeatedly stated, . . . , that our method involves the “assumption” that the two variances are equal. This is an incorrect form of statement; the equality of the variances is a necessary part of the hypothesis to be tested, namely that the two samples are drawn from the same normal population. The validity of the  $t$ -test, as a test of this hypothesis, is therefore absolute, and requires no assumption whatever.*

**Kritiken att stickprovsvarianserna** kan vara så olika, att de uppenbart inte representerar en och samma fördelning, besvarar Fisher alltså elegant med att i så fall måste det vara en del av avvikelserna från  $H_0$  – dock i en riktning där ett test som utgår från medelvärdedifferensen inte har någon nämnvärd styrka. Det är inte ovanligt, kanske är det rent av typiskt, att en verkkningsfull behandling även påverkar variansen åt ena eller andra hållet. Om behandlingen till exempel höjer de uppmätta värdena, så ökar ofta även variansen, men om den stabiliserar responsten, så minskar variansen.

Ett bättre argument för algoritmen ovan kan då vara att vi inte önskar testa Fishers  $H_0$ , utan just bara algoritmens Welch-hypotes att två normalfördelningar har samma väntevärde, utan något antagande om varianserna ens under  $H_0$ . Det leder oss in på en diskussion av vilken testsituation vi står inför. Den vanligaste och i särklass viktigaste är väl den experimentella, att vi vill testa ifall en behandling eller påverkan har effekt, och ena stickprovet består av på något sätt behandlade eller utsatta individer, och det andra används som kontroll- eller referensgrupp. Om behandlingen

saknar effekt, ja då är vi automatiskt i Fishers nollhypotes.

**I en annan situation**, då vi bara vill jämföra en mätbar storhet i två olika populationer (t ex kvinnor och män), så är inga potentiella behandlingseffekter aktuella. Då kan vi välja att följa algoritmen när vi ska testa. Om vi nu alls ska testa. Är det säkert att det är intressant att påvisa att väntevärdena skiljer sig, ifall vi redan sett att (även) varianserna gör det? När varianserna är lika så kan vi säga att den ena fördelningsfunktionen ligger helt under (eller till höger om) den andra, men det gäller inte om varianserna är olika. Väntevärdesdifferensen är bara en aspekt av ett mer komplicerat förhållande.

Man kan alltså förvänta sig eller åtminstone hoppas att litteraturen framhåller distinktionen mellan de båda testsituationerna inför valet av test. Men så är inte alls fallet, inte ens inom encyklopedin Wikipedia.

**Wikipedia, samt några programpaket**  
Wikipedia har flera relevanta uppslagsord, vilket dock är förvirrande, för de är inte överens. Se här:

*Wikipedia: A/B testing.* Anger att Welch-testet är standard, och att t-testet är ett alternativ.

*Wikipedia: Student's t-test, (two-sample).* Efter att texten sagt att ”Student's t-test” för två stickprov är ett lägestest av  $H_0$ : ”lika medelvärden”, så fortsätter den med följande delvis uppseende-väckande rader:

*All such tests are usually called Student's t-tests, though strictly speaking that name should only be used if the variances of the two populations are also assumed to be equal; the form of the test used when this assumption is dropped is sometimes called Welch's t-test.*

**Under delrubriken** ”similar variances” beskrivs nu t-testet. ”Similar” specificeras där som kvot  $s_1/s_2$  mellan  $\frac{1}{2}$  och 2, dvs den ena stickprovsvariansen får vara 4 gånger så stor som den andra, oavsett stickprovets storlek!

Under delrubriken ”unequal variances” (komplementet till ”similar”) beskrivs det som jag kallat Welch-testet, och där läser man:

*This test, also known as Welch's t-test, is used only when the two population variances are not assumed to be equal.*

Således har de båda testerna olika användningsområden, men inte på grund av olika hypoteser utan olika modellförutsättningar, ungefär som den inledande algoritmen var formulerad. Lagg märke till ”only” i frasen ovan, och jämför med texten nedan för Welch-testet, som helt klart har en annan författare.

*Wikipedia: Welch's t-test.* I den artikeln står det att Welch-testet

*...is an adaptation of Student's t-test, and is*

*more reliable when the two samples have unequal variances and possibly unequal sample sizes.*

Så långt låter det som inledningens algoritm. Men artikeln går ett steg längre och lägger till:

*It is not recommended to pre-test for equal variances and then choose between Student's t-test or Welch's t-test. Rather, Welch's t-test can be applied directly.*

Artikeln rekommenderar alltså att man alltid skall använda Welch-testet. Algoritmen blir därmed helt överspelad, och även Fishers argument för t-testet. I texten refereras till en artikel i en psykometri-inriktad tidskrift (Zimmerman, 2010), som påstår:

*pretest... no longer widely recommended by statisticians, although they persist in some textbooks and software packages.*

Således, det skall fortfarande finnas statistiker och mjukvara som envisas med att rekommendera algoritmen (däribland författaren till den förra Wikipedia-artikeln!), men de flesta av oss statistiker påstås tycka att man alltid skall använda Welch-testet, dvs att Students t-test är onödigt!

**Låt oss även se i några** statistikpaket hur det ser ut (**R**, SAS och Excel):

**R** har tyvärr Welch-testet som default, men om man aktivt ber om t-testet, så får man detta.

SAS PROC TTEST har som default att man får både t-testet och Welch-testet, samt ett F-test som pre-test av varianserna. Förvirrande benämns Welch-testet inte så i SAS, utan ”Satterthwaite”, medan ”Welch” är namnet för en ”option” i GLM för ”flera stickprov” (jfr nedan). Satterthwaites namn brukar annars knytas till formeln för Welch-testets antal frihetsgrader (Wikipedia: *Welch–Satterthwaite equation*).

Excel: Supporten i Microsoft Excel för funktionen T.TEST säger otydligt om dess användning bara:

*för att bestämma om två urval kan antas komma från samma underliggande populationer som har samma medelvärde* (och motsvarande på engelska).

Man undrar bland annat om inte ”samma populationer” automatiskt innebär även lika varians. För att utföra testet skall man välja mellan ”lika varians” och ”olika varians”, så den underförstådda hypotesen avser nog ”lika medelvärden”. Mer information om testen ger inte denna support, men det finns gott om annan Excel-rådgivning på nätet om hur man skall tänka och agera. Den spretar mycket, förstås.

### Konfidensintervall

Slutligen en anmärkning om konfidensintervall för väntevärdesdifferensen. Ett sådant för  $\mu_1 - \mu_2$  utgörs av alla de  $\delta$ -värden som inte förkastas vid test av  $H_0$ : ” $\mu_1 - \mu_2 = \delta$ ”. För dessa test skulle vi behöva veta hur variansen (och frihetsgrader) beror av bl a  $\delta$ . Om vi nu har en behandlingseffekt som syns på såväl väntevärde som varians,

så behöver vi åtminstone i någon mån modellera hur variansen skall tänkas förändras med  $\delta$ , innan vi kan konstruera ett konfidensintervall för  $\delta$ . Detta är förstas vanskligt när vi bara har en enda behandlingsnivå. Även om det inte handlar om en behandlingseffekt, så måste vi vara beredda att modellera sambandet mellan väntevärde och varians, men då finns möjligheten att anta att samband saknas, dvs att de två skattade varianserna är giltiga oavsett väntevärdena. Särskilt om stickproven är små kan en sådan förutsättning dock kännas problematisk. Låter det trassligt? Mjukvaror ger enklare svar, t ex genom att förutsätta att varken  $\sigma_1$  eller  $\sigma_2$  beror på  $\delta$ . Rimligt, eller ej? Avstår du från att granska deras modellering, så ligger väl ansvaret inte bara på mjukvaran? Motsvarande gäller förstas även ifall du anlitar AI för uppgiften.

### Sammanfattning

\* För det första har litteratur och mjukvaror i allmänhet glömt bort möjligheten att välja Fisher-hypotesen  $H_0$ : ”Ingen behandlingseffekt”, eller åtminstone glömt tala om konsekvensen därav, nämligen att rättfärdiga det klassiska t-testet även när de observerade stickprovsvarianserna är olika.

\* För det andra råder en påfallande brist på samstämmighet i litteratur och mjukvaror om vad som är den rekommenderade testproceduren för test av  $H_0$ : ”Samma väntevärde”, från ”använd alltid Welch-testet” till ”välj test efter att ha pre-testat för lika varians”.

### Jämförelse av flera stickprov

Som den första av flera naturliga generaliseringar ser vi kort på en välkänd utvidgning av två stickprov till flera stickprov. Standardtest för att jämföra flera stickprovsmedelvärden är ett inledande F-test inom den lineära ANOVA-modellen för ”ensidig indelning”, av  $H_0$ : ”alla väntevärden lika”. Analogt till t-testet för två stickprov, så skulle man kunna undra ifall vi inte då först borde godta att varianserna är lika, innan vi gör testet? Någon sådan motsvarighet till diskussionen för två stickprov brukar man dock inte se i litteraturen när det handlar om flera stickprov. Ett undantag är programpaketet SAS, med en option ”welch” i PROC GLM, som använder en utvidgning av Welch-testet.

I själva verket räcker det dock att hypotesen i stället formuleras enligt Fisher, för att F-testet skall vara giltigt, dvs  $H_0$ : ”alla stickproven kommer från samma (normalfördelade) population”. Om stickproven representerar olika behandlingar så kan denna  $H_0$  formuleras: ”alla behandlingar är helt likvärdiga”.

### Regression/korrelation

En annan generalisering av ”två stickprov”, som även den fått en annorlunda behandling i litteraturen, är ”beroende mellan två variabler” (beror  $y$  av  $x$ ?). Enklaste statistiska modell för

beroende är ”enkel lineär regression” (ELR) (se teoriruta). Om  $x$  är binär, t ex 0 och 1, så återfår vi ”två stickprov”. Men, som vi skall se, så har litteraturen om regression andra råd att komma med än den om två stickprov.

Om  $y$  beror av  $x$  undersöks typiskt inom ELR-modellen, och då genom att testa hypotesen  $H_0$ : ” $\beta=0$ ” med ett t-test, baserat på skattningen  $b$ . Men vad skall man då göra om data inte ser ut att följa ELR? Data kan se ut att mer motsvara ett krökt samband, och/eller ett samband där även variansen i  $y$  ser ut att variera med  $x$ . Skall man då, enligt algoritmen för ”två stickprov” ovan, först pre-testa, och kräva godtagbar linearitet och konstant varians, och kanske även kräva att  $x$ -värdena utgör ett normalfördelat stickprov (se korrelation nedan), innan man får t-testa  $H_0$ : ” $\beta=0$ ”? Litteraturen föreslår inte det. Typiskt underförstår den bara att ELR-modellen gäller, eller formulerar testet inom den modellen. Det är alltför inskränkt tänkt, men proceduren kan rättfärdigas.

Förhållandet är analogt med det i ”två stickprov”. För att kunna stödja sig på t-testets starka egenskaper inom ELR, och för konfidensintervall etc, så behöver man visserligen godta ELR som modell för data. Men *ELR krävs inte för att t-testet av oberoende skall vara korrekt*, ty den kritiska fördelningen är den som gäller under  $H_0$ : ” $y$  beror inte av  $x$ ”, och då är kraven på linearitet och konstant varians automatiskt uppfyllda. Testet av oberoende har sin dokumenterade styrka mot rent lineärt beroende, men har styrka mot alla former av olineärt beroende som har en lineär komponent i sig (t ex monotont beroende). Den utvidgade giltigheten är praktiskt användbar, men i de vanliga läroböcker och aktuella beskrivningar på webben som jag lyckas finna, så har jag inte sett detta nämnas explicit. Testet används också (default) utan varningar i typiska programpaket för regression, tillsammans med konfidensintervall. Varningar behövs alltså inte heller för testet – däremot för konfidensintervall för  $\beta$ , som inte ens är meningsfulla utanför ELR-modellens linearitet.

När man har påvisat en signifikant lineär komponent genom t-testet, om inte förr, så kan det vara dags att söka en adekvat modell för detta beroende (genom variabeltransformationer eller icke-lineära funktionssamband). Med en ensam  $x$ -variabel kan detta lyckas, medan man i multipel regression typiskt får nöja sig med en approximativ, lineär modell.

För fullständighetens skull, låt oss även se på det med regressionstestet besläktade men i dag mindre vanliga korrelationstestet. Som kanske är bekant, fås ELR-modellen ur bivariat normal för  $(x, y)$  genom betingning på  $x$ . Inom den bivariata normalfördelningen, med stickprovets korrelationskoefficient betecknad  $r$ , så är  $y$  oberoende av  $x$  om  $y$  och  $x$  är okorlelerade. Som Fisher visade redan 1915, så är då  $(n-2) r^2 / (1-r^2)$  kvadraten på en t-fördelat

teststatistika, så även  $r$  kan användas för t-test av oberoende. Hur hänger det ihop? Mycket enkelt – det är samma test, fast uttryckt i olika modeller! Både den bivariata modellens  $r$  och regressionens  $b$  är baserade på kovariansen mellan  $x$  och  $y$ , normaliserad på lämpligt sätt. Men bivariat normal är ett onödigt och oönskat hårt krav – som dock missriktat förekommer bland användare.

### Sammanfattning

Som test av om  $y$  beror av  $x$  tycks standarden vara t-test av hypotesen  $\beta=0$  inom modellen ”enkel lineär regression” (ELR). Emellertid saknas ofta (st?) diskussion av vad man skall/får göra om denna modell inte passar till data, och ifall den behöver göra det. Däri skiljer det starkt från litteraturen om specialfallet ”jämförelse av två medelvärden”. I själva verket, om vi ser en olinearitet eller att variansen varierar med  $x$ , så är det en avvikelse både från ELR och från  $H_0$ : ” $y$  beror inte av  $x$ ”, men det hindrar inte t-testets giltighet under denna  $H_0$ , så skönt!

### Slutord

I samtliga testsituationer ovan tycks litteratur och programvaror vara påfallande omedvetna om det jag kallat Fishers nollhypotes. Detta har i gengäld skapat onödiga validitetsproblem med t-test för modellreduktion, när observerade data inte tycks passa med en klassisk lineär modell. För ”två stickprov” har man genom åren tagit allvarligt på detta, och mer eller mindre starkt (och oftast onödigt) rekommenderat olika alternativa procedurer och test, medan man i de mer generella situationerna ”flera stickprov” och ”regression” inte har bekymrat sig så mycket. I

## TVÅ MODELLER

### Modellen ”två stickprov”

Givet stickprov från  $N(\mu_1, \sigma_1)$  och  $N(\mu_2, \sigma_2)$ . Intresset riktat mot eventuell skillnad i läge,  $\delta = \mu_1 - \mu_2$ .

Om vi föreskriver  $\sigma_1 = \sigma_2$ , så har vi en lineär modell av standardtyp (Sundberg, 2025). Enda rimliga testet av  $H_0$ : ” $\delta = 0$ ” som är ”korrekt” är då t-testet (rätt  $p$ -värde, oberoende av de andra parametrarna).  $H_0$  minskar parameterantalet och den minimala tillräckliga statistikans dimension lika mycket, från 3 till 2.

Om däremot  $\sigma_1 \neq \sigma_2$  tillåts, så är vi utanför klassen av lineära modeller.  $H_0$  reducerar antalet parametrar med 1 (nu från 4 till 3), men inte den minimala tillräckliga statistikan (dim = 4 även under  $H_0$ ). Det är källan till det s k Behrens–Fisher-problemet, som Welch-testet ger en approximativ lösning på. Testets korrekta  $p$ -värde beror på de okända  $\sigma_1$  och  $\sigma_2$ , och testet blir därför bara unge-

**Referenser**

- Cox, D R & Hinkley, D V (1974). Theoretical Statistics. London: Chapman and Hall.
- Cox, D R & Hinkley, D V (1974). Theoretical Statistics. London: Chapman and Hall.
- Sundberg, R (2025). Lineära Statistiska Modeller. Matem inst, Stockholms universitet.
- Zimmerman, D W (2004). A note on preliminary tests of equality of variances. British J. Mathem. and Statistical Psychology.
- Engelska Wikipedia och statistikpaketen R, SAS och Excel, citerade versioner från mars-april 2026.

stället tycks man typiskt ha inskränkt diskussionen till de lineära modellernas teori, utan att bry sig om att diskutera testens faktiska giltighet utanför grundmodellen.

ROLF SUNDBERG  
*Professor em. i matematisk statistik,  
Stockholms universitet.  
2017 års Statistikfrämjare*



får korrekt under  $H_0$ . Se vidare t ex Cox & Hinkley (1974, kap 5).

### Modellen ”enkel lineär regression” (ELR)

Data är  $n$  oberoende observationspar  $(x, y)$ . ELR innebär primärt att  $y$  beror lineärt av  $x$ :  $y = \alpha + \beta x$ , samt att  $y$  därutöver, givet  $x$ , har rent slumpmässigt additivt ”fel” (brus;  $N(0, \sigma)$ ). Variabeln  $x$  får vara systematiskt vald eller slumpmässig. Om  $(x, y)$  är 2-dim normalfördelad, så erhålls ELR-modellen genom betingning på  $x$ . Parametrarna  $\alpha$  och  $\beta$  skattas typiskt med minsta-kvadratmetoden. Skattningen  $b$  av  $\beta$ , normerad med sitt medelfel (std error;  $n-2$  fr-gr) enligt ELR, är teststorhet för t-test av  $H_0$ : ” $\beta = 0$ ”, (eller ” $y$  beror inte av  $x$ ”). Testet ger korrekt  $p$ -värde, och testet är det starkast möjliga inom ELR-modellen. För mer om regressionsmodeller, se t ex Sundberg (2025, kap 3).

Mycket information som ligger till grund för beslut i samhället är av statistisk natur. Därför är det viktigt att medborgarna har förmågan att kunna tolka den och bedöma om den är relevant för den aktuella frågeställningen. Denna förmåga ingår i det som kallas för statistisk litteracitet. Att människor har god statistisk litteracitet bidrar till att minska antalet beslut som baseras på feltolkningar av statistisk information.

# Vad är relevant statistisk information?

Ofta missbrukas statistik genom att man i vilseledande syfte använder information som inte är relevant för den aktuella frågeställningen, eller att man utelämnar relevant information.

Ett exempel är när man har använt sig av ett stickprov för att få information om en population men inte redovisar hur stor osäkerheten är. Att kunna bedöma relevansen i statistisk information är en mycket viktig del av den statistiska litteraciteten. Här skulle Svenska statistikfrämjandet kunna bidra genom att ta fram en sammanställning som visar vad som är relevant information i olika situationer. Nedan presenterar jag några exempel.

**Vi tänker oss** att försvaret ska köpa in material för livremmar till ett stort men känt antal soldater. Materialet köps som metervera. Viken information om soldaternas midjemått är relevant för att man ska kunna bestämma hur många meter som behövs?

Vi kan tänka oss två fall: att livremmarna kapas till individuellt för varje soldat eller att man tillverkar en standardmodell som ska räckta runt midjan på 95 % av soldaterna (resten får banta). I första fallet är det medelvärdet som

är den relevanta informationen medan det i det andra fallet är 95-percentilen.

**Ett annat exempel** är fallet med 33 kvinnor som fått sin livmoder bortopererad i onödan på Akademiska sjukhuset i Uppsala.

Enligt en läkare som uttalade sig på TV var detta inte så allvarligt eftersom medianåldern var över 50 år. Medianen säger ju bara att högst hälften varit under 50 år men inte något om hur många kvinnor som varit i en ålder då en bortopererad livmoder varit en katastrof. En relevant information hade varit att presentera antalet kvinnor i olika åldersintervall (till exempel femårsintervall).

**För att undersöka** om giftrester (från bekämpning av någon parasit) i ett parti äpplen ligger på en ofarlig nivå har man tagit ett stort antal äpplen som mosas och blandas. På blandningen görs sedan några mätningar av gifthalten och medelvärdet beräknas. Genom mosningen och blandningen har man rent fysiskt skapat ett medelvärde och upprepade mätningar på denna blandning ger ingen som helst uppfattning om spridningen mellan äpplen. Vad man som konsument är intresserad av är ju hur stor risk det är att ett enskilt äpple innehåller för mycket gift. Men

det kräver ju att man mäter gifthalten i ett stort antal enskilda äpplen.

**För några år sedan** hade jag en diskussion med en politiker som hävdade att kvinnor diskriminerades vid tillsättning av professurer. Beviset var att medelbetyget i studentexamen var högre för flickor än för pojkar. För varje elev hade man tydligen beräknat medelbetyget för alla ämnen och där efter beräknat medelvärdet för pojkar respektive flickor.

För det första är det tveksamt om medelbetyget för alla ämnen är en relevant storhet när det gäller att bedöma lämpligheten som professor i ett visst ämne. Det är väl snarare betygen i de skolämnena som är relevanta för den aktuella professuren. För det andra är det inte medelvärdet i populationen som är det relevanta statistiska måttet. Det är ju i stället andelen elever i den övre ändan av fördelningen för pojkar respektive flickor och där kan det ju finnas fler pojkar än flickor även om medelvärdet är högre för flickor. Dessa argument bet dock inte på politikern eftersom man redan bestämt vad budskapet skulle vara.

**Mätningar är sällan** felfria. Vid till exempel koncentrationsmätningar i blodprov påverkas resultatet ofta av

andra ämnen i provet än det vi vill mäta. Vi får därför räkna med att det kan finnas provspecifika fel, som naturligtvis kan vara beroende av vilken mätmetod som används. I praktiken används ofta olika mätmetoder för att mäta samma storhet både inom forskning och vid rutinmässiga diagnoser. I sådana fall är det viktigt att visa att överensstämmelsen mellan metoderna är tillräckligt bra för att de ska kunna betraktas som utbytbara.

Om man vill undersöka överensstämmelsen mellan två metoder bestämmer man koncentrationen i ett antal prov från olika patienter med båda metoderna. Från en sådan studie kan man, om skillnaderna inte är nivåberoende, beräkna gränser som förväntas omsluta 95 % av de skillnader man kan få mellan metoderna. Sådana gränser kallas överensstämmelsegränser och är relevanta för att bedöma överensstämmelsen mellan metoderna. Tyvärr har det blivit vanligt att använda korrelationskoefficienten som mått på överensstämmelsen. Att en korrelationskoefficient inte är ett relevant mått på överensstämmelsen visades av statistikerna Bland och Altman (1986) i en artikel i The Lancet.

**Ett exempel på hur** vilseledande det kan vara är en artikel som publicerades 2023 (Siegbahn et al.). Författarna har jämfört ett nytt instrument för diagnos av hjärt- och kärlsjukdomar med en etablerad metod och tolkat en korrelationskoefficient på 0,87 som god överensstämmelse.

I artikeln finns dock uppgifter som gör det möjligt att beräkna överensstämmelsegränser

för relativa skillnader. Dessa visar att kvoten mellan resultaten med de båda metoderna kan variera från 1/3 till 3. Detta är förmodligen helt oacceptabelt för de flesta tillämpningar.

Med anledning av de tänkbara konsekvenserna skickade jag mejl till några av författarna och påtalade den felaktiga slutsatsen utan att få något svar. Eftersom fem av författarna är från Uppsala universitet anmälde jag artikeln till Nämnden för prövning av oredlighet i forskning som beslutat att författarna inte gjort sig skyldiga till oredlighet. Som oredlighet i akademisk forskning räknas nämligen bara fabricering, förfalskning och plagiering. Användningen av en korrelationskoefficient som ett mått på överensstämmelse mellan mätmetoder strider mot den etiska kod som Svenska statistikfrämjandet antagit. Den säger att när det gäller användning av statistiska metoder och mått ska man välja de metoder och mått som ger en så korrekt bild av verkligheten som möjligt.

**Jag anser att** Svenska statistikfrämjandet bör arbeta för att få den egna koden accepterad även inom akademisk forskning. Detta blir dock inte lätt. En anledning kan vara systemet med kollegial granskning (peer review). Har ett missbruk av statistik blivit tillräckligt utbrett är det troligt att även granskarna gjort sig skyldiga till det och de kommer då inte bara att acceptera utan även kräva att den olämpliga statistiska metoden används. Detta är en nackdel med att forskarvärlden granskar sig själv. Inbland är det därför som granskande journalister avslö-

jar felaktiga slutsatser. Ett exempel är Andrew Wakefields påstående i en artikel i The Lancet 1998 att MPR-vaccinet orsakade autism. Till slut drog The Lancet tillbaka artikeln.

Tyvärr har dock media vanligen en okritisk rapportering av forskningsresultat. Jag har ibland meddelat tidningar när de spridit allvarliga statistiska feltolkningar men hittills inte fått någon respons. Jag har även försökt få in debattartiklar som belyser problemet. För en person utan en position (politiker, professor eller representant för en organisation) är det dock inte lätt.

GÖRAN NILSSON  
Statistiker



## Referenser

- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurements. *Lancet* 1986; February: 307-310
- Siegbahn, A., Eriksson, N., Assarsson, E., Lundberg, M., Ballagi, A., Held, C., ... & Wallentin, L. (2023). Development and validation of a quantitative Proximity Extension Assay instrument with 21 proteins associated with cardiovascular risk (CVD-21). *Plos one*, 18(11), e0293465.

# Stora språkmodeller och surveyundersökningar

## – VAD BEHÖVER VI TÄNKA PÅ FÖR GOD DATAKVALITET?

**A**rtificiell intelligens (AI) förändrar snabbt hur vi lever och arbetar. Särskilt datanära yrken, som statistik och surveyforskning, brottas med både möjligheter och begränsningar när det gäller hur AI kan användas i deras arbetsprocesser. I centrum av diskussionen står avvägningar mellan effektivitet och kvalitet. För att kunna bedöma vilken roll AI kan spela i surveyforskningen och hur teknologin påverkar datakvaliteten behöver man först förstå hur den fungerar.

### Vad är stora språkmodeller och hur fungerar de?

När man talar om AI idag syftar man oftast på stora språkmodeller (large language models, LLM:er) – teknologin bakom chattbotar som ChatGPT. LLM:er är en form av generativ AI som är skapad för att bearbeta och generera mänskligt språk, baserad på samband mellan ord (så kallade embeddings) och statistiska sannolikheter.

För att möjliggöra detta tränas LLM:er på stora, men selekterade textmängder från internet, till exempel böcker, Wikipedia och sociala medier (*pre-training*). Därefter finjusteras de för att kunna följa instruktioner (*instruction tuning*) och får ofta återkoppling på sina svar för att bättre anpassas till mänskliga preferenser (*alignment, RLHF*). När en LLM används omvandlas användarens input (en prompt) till numeriska representationer (*tokens*). LLM:en analyserar relationerna mellan dessa tokens utifrån vad den lärt sig under träningen och *föresäger sedan steg för steg den (mest) sannolika nästa token*, givet både prompten och det som redan har genererats.

Resultatet blir en text i naturligt språk (en

*completion*). Denna text genereras genom att ord väljs från en sannolikhetsfördelning över möjliga nästa ord. Fördelningen påverkas inte bara av prompten, utan också av så kallade *hyperparametrar*. Med dessa kan man till exempel styra om endast de mest sannolika orden ska beaktas (*top-p*, vilket innebär att sannolikhetsfördelningen beskärs), eller hur stor vikt som ska ges åt mindre sannolika alternativ (*temperature*, vilket modifierar fördelningens form) – båda dessa parametrar påverkar hur varierad texten blir.

**Prompting och finjustering.** En central aspekt i arbetet med LLM:er är hur man ger dem instruktioner. LLM:er kan användas med olika typer av så kallad *prompting*. Vid *zero-shot prompting* får LLM:en en uppgift utan ytterlig information, det vill säga utan definitioner eller exempel på önskat resultat. Vid *few-shot prompting* ger man LLM:en både en instruktion och några exempel på prompter och motsvarande önskade completion för att leda den. *Fine-tuning* innebär att LLM:en tränas vidare på en uppgiftspecifik datamängd av *prompt-completion-par* innan den används. På det sättet finjusterar man de interna vikterna, vilket brukar förbättra LLM:ens prestation på den aktuella uppgiften.

**Tillgång.** LLM:er kan nås på olika sätt, i fallande ordning efter användarvänlighet: via *chattbaserade gränssnitt* (till exempel ChatGPT), via *Application Programming Interfaces* (API:er) eller genom egen drift (*self-hosting*). API:er gör det möjligt att använda LLM:er via webbplattformar eller programmeringsspråk som R eller Python, med större kontroll över vissa hyperparametrar, utan att man behöver egna dataresurser eller avancerade program-

meringskunskaper. Dessa krävs däremot vid *self-hosting*, som i gengäld ger maximal kontroll över modellen.

**Typen av modeller.** En viktig skillnad mellan olika LLM:er, nära kopplad till hur de kan nås, är om de är öppna eller slutna. Öppna modeller erbjuder en viss grad av *transparens* vad gäller sina vikter (*open-weight*, till exempel DeepSeek eller OpenAI:s GPT-OSS-serie) och träningsdata (*open-source*, till exempel Meta:s Llama och Mistral). De kan laddas ner fritt (till exempel via plattformar som Hugging Face) och ger användaren större kontroll, eftersom de kan installeras och köras på servrar som användaren själv har tillgång till.

Open-source- och open-weight-LLM:er minskar därmed behovet av att förlita sig på tredjepartsservrar (vilket är viktigt för integritet) och skyddar mot oförutsedda modelluppdateringar (vilket underlättar reproducerbarhet). Samtidigt har inte alla forskare tillgång till de dataresurser och den tekniska kompetens som krävs för att köra öppna modeller (plattformar som LM Studio försöker överbrygga detta).

I kontrast till detta är *proprietära modeller*, som OpenAI:s GPT, Google:s Gemini och Anthropic:s Claude, oftast *slutna "black boxes"* med begränsad insyn i hur de är uppbyggda. De erbjuder användarvänliga chattgränssnitt och API:er, vilket gör det enklare att replikera tidigare studier (åtminstone vad gäller arbetsprocessen). Chattgränssnitt är ofta gratis upp till en viss gräns, men innebär kostnader per token (det vill säga beroende på längden på prompten och svar), vilket kan bli dyrt vid stora datamängder eller mer komplexa uppgifter.

Tillgång till proprietära API:er kräver vanligtvis en prenumeration. Mot bakgrund av dessa

frågor om tillgänglighet och transparens, liksom integritet och reproducerbarhet (se nedan), har det därför framförts argument för att enbart använda open-source-LLM:er i forskning (se Palmer et al., 2023; Spirling, 2023).

För en mer detaljerad översikt över LLM:er för social- och beteendevetenskap, se GESIS Guides to Digital Behavioral Data (Holtdirk et al., 2025).

### Möjligheter med LLM:er i surveyforskningen

Moderna LLM:er är i regel flerspråkiga – även om de ofta är optimerade för engelska och dessutom fungerar i allmänhet bättre ju större ett språk är. Detta har bland annat med träningsdatan att göra, som oftast består av mest engelska texter. Vikter och anpassningsprocessen spelar också en roll. Dessa LLM:er fungerar som allmänna verktyg som inte nödvändigtvis kräver avancerade tekniska kunskaper. Denna egenskap gör dem särskilt attraktiva för en bred grupp av forskare. Samtidigt kräver de, ur ett datakvalitetsperspektiv, en viss AI-kompetens – och, som vi kommer att se, ibland även tekniska färdigheter för att ge bästa möjliga resultat. [CO1.1]

Generellt kan man identifiera tre huvudsakliga roller för LLM:er i surveyforskningsprocessen: som *forskningsassistenter*, *intervjuare och respondenter*. Det innebär att det finns potentiella användningsområden för LLM:er i alla delar av surveyprocessen, där de kan bidra till att minska fel som påverkar datakvaliteten, samtidigt som de ofta är mer effektiva än människor.

**Före datainsamlingen** kan LLM:er fungera som forskningsassistenter genom att stödja utvecklingen av forskningsdesignen, samt utveckling, testning, förbättring och översätt-

ning av material som används i en survey. Detta inkluderar (delar av) surveyinstrumentet (till exempel frågor, svarsalternativ, instruktioner och experimentella stimuli) samt annat material som används vid rekrytering av potentiella respondenter (till exempel inbudsbrief, rekryteringsinlägg och informationsmaterial).

**Under datainsamlingen** kan LLM:er fungera som intervjuare genom att genomföra helt standardiserade eller adaptiva, samtalsbaserade intervjuer. De kan då dynamiskt administrera frågor baserat på tidigare svar, generera nya följdfrågor i realtid och ge exempel för att underlätta respondenternas förståelse.

Som respondenter kan LLM:er generera syntetiska surveydata, oftast genom att simulera mänskliga respondenter via *persona-baserade prompts* som motsvarar specifika demografiska eller attitydmässiga profiler. Dessa så kallade *silicon samples* kan användas för att komplettera eller delvis ersätta datainsamling från människor, men också i samband med pretesting (före den huvudsakliga datainsamlingen) eller imputering (efter datainsamlingen).

*Silicon samples* är en av de mest omdiskuterade tillämpningarna av LLM:er inom surveyforskningen. Medan tidiga studier visade att vissa LLM:er kunde spegla aggregerade attitydmönster hos människor visar senare forskning att dessa resultat inte generaliserar över olika populationer, samband och forskningsdesigner. Detta väcker frågor kring kvaliteten på data som genereras med *silicon samples*.

**Efter datainsamlingen** kan LLM:er återigen stödja surveyforskare genom att (*för*)bearbeta, analysera och sammanfatta data. Förbearbetning omfattar till exempel att digitalisera ljud- och

pappersbaserade data och att strukturera ostrukturerade data från sociala medier. LLM:er kan också bidra till att identifiera svar av låg kvalitet och om intervjuer avviker från instruktionerna. Bearbetningsuppgifter för LLM:er inkluderar klassificering av text-, bild- och ljuddata från öppna svar eller andra datakällor. Eftersom LLM:er är en teknik inom naturlig språkbehandling (natural language processing, NLP), är detta inte bara en av de vanligaste och mest framgångsrika tillämpningarna, utan också en vidareutveckling av tidigare maskininlärningsbaserade klassificeringsmetoder. Samtidigt finns det viktiga begränsningar att beakta (se nedan). Under dataanalysen kan LLM:er användas för att skriva programkod och möjligtvis för att sammanfatta kvantitativa och kvalitativa data samt resultat.

En kommande systematisk översikt av användningen av LLM:er inom surveyforskningen sammanfattar det nuvarande empiriska kunskapsläget, inklusive både framgångar och misslyckanden (working paper: Buskirk et al., 2025).

### Överväganden gällande datakvalitet när man använder LLM:er

Även om de tidigare nämnda tillämpningarna kan göra surveyforskningen mer effektiv och samtidigt minska mänskliga fel, kan LLM:er också leda till *egna, svår-förutsägbara fel* och potentiellt förstärka befintliga bias i hur olika populationer och centrala begrepp förstås. Bias i LLM:ers output kan komma både från *modell-designen* – vad de tränades på och hur de har anpassats – och från *forskningsdesignen*, inklusive modellval, prompting och hyperparametrar. Dessa typer av bias kan direkt påverka kvaliteten på surveydata som samlas in, genereras, bearbetas och analyseras med hjälp av LLM:er.

Det gäller bland annat täckning, validitet, generaliserbarhet, reliabilitet, reproducerbarhet och transparens.

**LLM-design.** LLM:er är byggda på data från internet. I motsats till surveydata skapas sådana data inte för att producera generaliserbara mätningar, utan de är ”upphittade”. Detta kan äventyra både representativa och mätmässiga aspekter av datakvalitet i surveyforskningen.

**Representationsfel** kan uppstå på grund av att *träningsdata* för LLM:er är obalanserade och sannolikt inte speglar den mångfald av attityder och beteenden som finns i mänskliga populationer. Två mekanismer kommer fram: För det första påverkar den *digitala klyftan* – det vill säga systematiska skillnader i vilka som har tillgång till, använder och bidrar på olika sätt till de internetplattformar som används i träningen av LLM:er – sammansättningen av ”urvalsramen” av potentiella texter som representerar människor i förhållande till målpopulationen. För det andra påverkar det *icke-slumpmässiga urvalet av texter* till träningskorpusar sammansättningen av det faktiska ”urvalet” av träningstexter i förhållande till urvalsramen.

Därutöver kan bias introduceras genom de personer som *annoterar* träningsdata och som *anpassar* LLM:erna, särskilt med avseende på deras personliga bakgrund. Systematiska skillnader mellan en texts avsedda betydelse och annotatörens tolkning kan uppstå på grund av bristande kontextförståelse. Detta kan leda till att vissa tolkningar överrepresenteras i LLM:er eller till felaktig återgivning av mänskliga attityder och beteenden.

Som en följd av dessa typer av bias i LLM:ers uppbyggnad är det sannolikt att deras output inte representerar alla grupper eller individer på ett likvärdigt sätt, vare sig i omfattning eller kvalitet. Stereotyper om grupper kan reproduceras. Generellt har LLM:er visats uppvisa både kulturell och psykologisk bias, inklusive en tendens att återspegla eller utgå från så kallade WEIRD-nor-

mer (Western, Educated, Industrialized, Rich, and Democratic), luta politiskt åt vänster och vara biased till nackdel för icke-engelskspråkiga och marginaliserade grupper. Denna brist på mångfald och variation skapar problem när LLM:er används för att efterlikna respondenter (inklusive i multivariata analyser), klassificera och analysera svar eller vid design och pretesting av enkäter (inklusive power-analyser).

Slutligen innebär träningsdatans tidsbundenhet att LLM:er inte automatiskt är uppdaterade med aktuella förändringar, till exempel i språkbruk eller i globala politiska, ekonomiska och sociala förhållanden. Eftersom mänskliga attityder och beteenden förändras i takt med sådana externa händelser kan detta leda till både representations- och mätproblem när LLM:er används i surveyforskning, eftersom LLM:er kan generera output baserat på föråldrade uppfattningar om attityder och beteenden.

Även ur ett *mätperspektiv* speglar träningskorpusar för LLM:er inte nödvändigtvis faktiska mänskliga preferenser, eftersom beteenden online inte alltid är giltiga indikatorer på attityder eller beteenden – något som det välkända exemplet med Google Flu Trends visade (Lazer et al., 2014). Där försökte forskare ”nowcasta” säsongsinfluensans utveckling baserad på hur ofta människor sökte efter influensarelaterade termer på Google. Även om de initiala trenderna följde uppskattningar från hälsomyndigheter, började algoritmen senare överskatta influensans utbredning med mer än det dubbla. En förklaring var förändringar i hur människor använde Google, delvis på grund av föreslagna sökningar och delvis eftersom vissa sökningar klassificerades som influensarelaterade trots att de inte var det. Idag påverkas till exempel data från sociala medier i hög grad av plattformarnas algoritmer och funktioner, och användare kan skapa innehåll strategiskt eller oavsiktligt, och uttrycka sig på andra sätt än de skulle göra offline – men LLM:er lär sig på online-information.

Mätproblem uppstår också när man tänker på vad en LLM:s output faktiskt är: den be-

tingade sannolikheten att tidigare ord följs av ett visst nästa ord. Även om outputen liknar mänskligt språk är det oklart om den också representerar (och därmed kan approximera) mänskliga kognitiva processer, vilket väcker frågor om validitet. Mätkvaliteten kompliceras ytterligare av att den genererade texten ibland inte överensstämmer med vad de underliggande numeriska sannolikhetsvärdena (så kallade log-probs) skulle indikera. Det är alltså inte alltid den numeriskt mest sannolika fortsättningen som faktiskt genereras. Därmed spelar det roll för inferensen om forskare tolkar textoutputen direkt eller i stället arbetar med de underliggande sannolikheterna.

**Forskningsdesign.** Datakvaliteten i LLM-assisterad surveyforskning kan också påverkas av forskarens val. Det mest uppenbara är valet av det specifika LLM:en, vilket i sin tur kan bero på faktorer som tillgänglighet, användargränssnitt, hastighet, kostnader och användningsbegränsningar. Varje LLM bygger på en unik kombination av träningsdata, anpassningsprocesser, vikter och övergripande modellarkitektur, och skiljer sig åt i hur de är optimerade – och därmed i sin prestation – för olika språk och uppgifter. LLM:er kan också ha olika standardvärden för hyperparametrar, som forskare kan tro ”ska” användas utan att justera dem, vilket kan leda till mindre optimala och svårjämförbara resultat. Därför kan olika LLM:er prestera olika väl även när de används för samma uppgift inom surveyforskningsprocessen. Frågan är alltså inte bara *om* en LLM kan utföra en viss uppgift, utan *vilken* LLM som kan det. Detta innebär en utmaning för *generaliserbarheten* i slutsatser om vilka uppgifter som kan stödjas av LLM:er, liksom för att formulera rekommendationer om bästa praxis.

Vidare kan variationen i modellernas **hyperparametrar** påverka datakvaliteten i LLM-assisterad surveyforskning. Ett exempel är avvägningen mellan reliabilitet och variation: genom att sänka temperaturen kan slumpmäs-

sigheten i LLM:ers output minskas, vilket ökar reliabiliteten. Samtidigt minskas variationen inom grupper till en nivå som sannolikt inte motsvarar mänskligt data. Om modellen till exempel ställs inför två möjliga svarsalternativ – exempelvis två svar på en attitydfråga eller två kategorier i en textklassificering – där det ena har sannolikheten 0,51 och det andra 0,49, kommer en LLM med mycket låg temperatur alltid att välja alternativet med sannolikheten 0,51. I verkligheten skulle dock nästan hälften av fallen tillhöra den andra kategorin. Även om experiment med olika hyperparametrar kan ge insikter om hur de kan optimeras, är deras exakta påverkan på den datagenererande processen i LLM:er svår att avslöja, vilket skapar utmaningar för validiteten.

Slutligen utgör LLM:ers *känslighet för hur en prompt formuleras* ytterligare en utmaning för datakvaliteten i LLM-assisterad surveyforskning. Både val och ordning av ord och svarsalternativ i prompten kan påverka outputen, på sätt som inte nödvändigtvis motsvarar mänskliga ordnings- och formuleringseffekter. Dessutom leder mer information i prompten (till exempel mer detaljerade kategoribeskrivningar vid kodning av öppna svar eller mer information om de respondenter som ska simuleras) inte alltid till bättre outputkvalitet. LLM:ers kontextfönster är begränsade, och modellerna har ibland svårt att bevara all information lika väl i längre prompts. Till exempel tenderar vissa att ”glömma” mitten av prompten. Därför blir valet av prompting- eller finjusteringsansats avgörande för mätkvaliteten.

LLM- och forskarberoende faktorer påverkar tillsammans *reliabilitet och reproducerbarhet*. Den probabilistiska naturen i hur LLM:er genererar text skapar utmaningar. Detta kan delvis hanteras genom att forskare sänker temperaturparametern mot 0, vilket gör modellen mer deterministisk i valet av det mest sannolika svaret (hur detta påverkar validiteten diskuterades tidigare), samt genom att sätta ett seed: att man sätter fast startvärdet för en slumpgenerator (till

exempel vid slumpmässigt urval från en fördelning), så att samma ”slumpmässiga” resultat kan återskapas varje gång koden körs. Detta är vanligt i simuleringsstudier inom statistisk programmering. [CO2.1]

Trots detta kan avvikelser från det ursprungliga resultatet fortfarande uppstå. En annan faktor som påverkar reliabilitet och reproducerbarhet rör främst proprietära LLM:er som inte kan laddas ner och köras lokalt. Dessa är föremål för så kallad *deprecation*, vilket innebär att leverantörer kan välja att avveckla äldre modeller. Detta kan hindra forskare från att använda dem på längre sikt eller från att upprepa tidigare analyser, både sina egna och andras. Men också om leverantörer tvärtom *uppdaterar* befintliga modeller innebär det inte nödvändigtvis förbättrad prestanda för den aktuella uppgiften. Eftersom varje LLM har en egenartad datagenererande process är det inte säkert att arbetsflöden som validerats för en viss modell kan överföras till en uppdaterad version eller efterföljare.

Särskilt proprietära LLM:er väcker också frågor om *transparens*. Även om full insyn i tränings- och anpassningsprocesser är ovanlig även för öppna LLM:er, erbjuder slutna modeller inte ens information om de vikter som används. Detta gör det svårt att förstå den datagenererande processen, identifiera och tillskriva bias samt att hantera dessa eller bedöma modellens lämplighet för ett visst syfte.

För en mer utförlig diskussion om datakvalitetsaspekter vid användning av LLM:er i surveyforskningen, se von der Heyde (2025).

### Slutsats

I form av LLM:er har AI potentialen att förändra hela surveyforskningsprocessen. Samtidigt har deras tillämpningar inom området och de utmaningar de medför ännu inte utvärderats och hanterats på ett systematiskt sätt. Den stora mängden möjliga (och observerade) fel väcker frågor om generaliserbarheten i enskilda studier som visar på framgångsrik användning av LLM:er i specifika surveyuppgifter och kon-

### Referenser

- Buskirk, T. D., Keusch, F., Heyde, L. von der, & Eck, A. (2025). *More Parameters Than Populations: A Systematic Literature Review of Large Language Models within Survey Research* (arXiv:2509.03391). arXiv. <https://doi.org/10.48550/arXiv.2509.03391>
- Holtdirk, T., Saju, L., Fröhling, L., & Wagner, C. (2025). *Overview of Large Language Models for Social and Behavioral Scientists (GESIS Guides to Digital Behavioral Data, 16)* (Version 1.1). GESIS - Leibniz-Institute for the Social Sciences. <https://doi.org/10.60762/GGDBD25016.1.0>
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176), 1203–1205. <https://doi.org/10.1126/science.1248506>
- Palmer, A., Smith, N. A., & Spirling, A. (2023). Using proprietary language models in academic research requires explicit justification. *Nature Computational Science*, 4(1), 2–3. <https://doi.org/10.1038/s43588-023-00585-1>
- Spirling, A. (2023). Why open-source generative AI models are an ethical way forward for science. *Nature*, 616(7957), 413–413. <https://doi.org/10.1038/d41586-023-01295-4>
- von der Heyde, L. (2025, October 6). *Who Counts? The Potentials and Pitfalls of Using LLMs in Survey Research*. First Workshop on Bridging NLP and Public Opinion Research. <https://openreview.net/forum?id=ww2KqnPLdK>

texter. Dessutom är LLM:ers styrkor och svagheter olika viktiga beroende på vilken uppgift inom surveyforskningsprocessen det pratas om. Därför behöver varje användning av LLM:er i surveyforskningen utvärderas i relation till den specifika uppgiften och kontexten där de ska tillämpas, för att säkerställa god datakvalitet (*fit-for-purpose*). Med tillräcklig kunskap om deras möjligheter och begränsningar, samt med mänsklig övervakning och validering, är det möjligt att LLM:er kan integreras med mer etablerade metoder för att ge en bättre bild av hur människor tänker och agerar.

LEAH VON DER HEYDE  
Forskare inom AI och  
surveymetoder hos  
GESIS (Tyskland)



AI håller på att förändra survey- och analysarbete i grunden – men inte nödvändigtvis på de sätt som dominerar det offentliga samtalet. Under de senaste tre åren har vi på Demoskop arbetat systematiskt med att integrera generativ AI i undersökningsprocesser, med fokus på kvalitet, säkerhet och metodologisk kontroll.

Arbetet har gett tydliga insikter om möjligheter och risker, men har också väckt en fråga som snabbt blivit strategisk för branschen: om AI tar hand om nybörjaruppgifterna, hur lär sig då nästa generation statistiker det hantverk som behövs för att kunna göra välgrundade bedömningar?

# AI och undersökningar

## – ERFARENHETER OCH LÄRDOMAR FRÅN TRE ÅRS SYSTEMATISKT UTVECKLINGSARBETE

### *AI är inte nytt – men generativ AI förändrar arbetssättet*

AI i bred mening är inte nytt i undersökningsbranschen. Metoder som i dag ofta samlas under etiketten maskininlärning har länge använts i analysarbete, särskilt för prediktion och mönsterigenkänning – till exempel klustring/segmentering, klassificering och modellering av stora och komplexa datamaterial. Den typen av metodkompetens är väl förankrad och har levererat nytta under lång tid.

Det som nu förändrar förutsättningarna är den generativa AI:n. Den kan generera, tolka och omforma text, kod och analys på ett sätt som tidigare krävde en mer specialiserad manuell arbetsinsats. Det innebär inte bara att vissa moment går snabbare, utan att arbetsflöden kan organiseras på ett kvalitativt annorlunda sätt – vilket i sin tur påverkar roller, kompetenskrav och former för kvalitetssäkring.

### *Från experiment till systematik: en kontrollerad arbetsmiljö*

När vi började undersöka hur stora språkmodeller kan användas i undersökningsarbete präglades fältet av snabb utveckling, höga förväntningar och begränsad tydlighet kring riskerna. En central lärdom tidigt var att hållbar användning kräver mer än enskilda verktyg. Den kräver en kontrollerad miljö där olika modeller kan utvärderas, jämföras och användas på ett sätt som är spårbart och går att styra.

Avgörande för att detta ska fungera i praktiken är kontinuerliga och strukturerade tester. Ingen ny modell eller ny tillämpning bör tas i produktion utan att prestanda jämförs mot ett känt facit. I grunden handlar detta inte om något

radikalt nytt, utan om att tillämpa statistiskt tänkande på ett nytt slags verktyg: att pröva, mäta, dokumentera och dra slutsatser utifrån systematisk utvärdering. Ett praktiskt resultat av detta arbetssätt är också att kompetensutveckling blir en integrerad del av produktionen, där utvärderingen görs öppet och det skapas en gemensam förståelse för både möjligheter och fallgropar.

### *Regelverken: bekanta ramar, nya tillämpningar*

Undersökningsverksamhet är sedan länge omgiven av tydliga etiska och juridiska ramar, vilket är en förutsättning för legitimitet. Samtidigt tillför generativ AI ett nytt lager av regulatorisk och metodologisk komplexitet. EU:s AI-förordning (AI Act) trädde i kraft 2024 och börjar gälla successivt, med vissa bestämmelser från 2025 och merparten från 2026. Förordningen skärper kraven på bland annat transparens, riskklassificering och mänsklig tillsyn vid användning av AI-system.

För verksamheter som producerar underlag för beslut är detta mer än en efterlevnadsfråga. Det handlar om metodologisk integritet. Om data är känslig, eller om resultaten kan påverka viktiga beslut, blir kontroll över hur data behandlas och hur slutsatser uppstår en kvalitetsdimension i sig – inte en sidofråga.

### *Vad AI gör bra – och var det krävs professionellt omdöme*

I praktiken har AI-stöd förändrat flera arbetsmoment. Kodning av öppna svar, som tidigare var tidskrävande manuellt arbete, kan ofta göras snabbare och mer konsekvent med stöd av språkmodeller. Detsamma gäller vissa former

av kvalitetsgranskning av enkätformulär, stöd vid analysutkast och första versioner av rapporttexter. När rutinmoment avlastas frigörs tid för sådant som är svårare att automatisera: metodutveckling, kvalitetskontroller, fördjupade analyser och integration av andra datakällor med surveydata.

Samtidigt är det avgörande att förstå vad generativ AI inte är: den är inte en sanningsmaskin. Den kan ”hallucinerar”, dvs producera till synes trovärdiga uppgifter utan empirisk grund. Den kan reproducera bias från träningsdata. Den kan vara känslig för hur instruktioner och frågor formuleras – på ett sätt som påminner om, men inte är identiskt med, klassiska problem kring frågekonstruktion och mätfel. Att känna igen dessa mönster, och att veta hur de ska hanteras, har blivit en kärnkompetens.

### *Statistikerns kompetens: viktigare än någonsin, men mer sårbar att bygga*

Här uppstår en paradox. Ju mer AI kan automatisera, desto viktigare blir statistikerns kärnkompetens – men desto svårare kan det bli att förvärva den.

Traditionellt har många av de moment som nu automatiseras fungerat som en del av upplärningen, bland annat kodning, datarensning, första analysutkast och arbete med ”stökig” rådata. Det är ofta genom just dessa repetitiva moment som professionellt omdöme byggs upp. Om de stegen försvinner eller komprimeras riskerar man att få en generation som kan producera resultat snabbt, men har svårare att avgöra när resultaten inte håller.

Det innebär att lärprocessen behöver designas mer explicit än tidigare. För att kunna bedöma kvaliteten i ett AI-genererat resultat

måste man veta vad man letar efter: rimlighetskontroller, antaganden, bias, mättens begränsningar, modellernas känslighet och var felkällor typiskt uppstår. Den typen av kompetens kräver fortsatt gedigen statistisk träning – och en medveten koppling mellan metodkunskap och praktisk tillämpning i en AI-assisterad miljö. I förlängningen aktualiserar detta även behovet av en nära dialog med lärosätena om hur statistisk metod, dataförståelse och kritisk granskning kan tränas när arbetsprocesser förändras.

### *Slutsatser*

Tre års systematiskt arbete med generativ AI i undersökningsprocesser pekar mot en tydlig slutsats: värdet ligger mindre i snabb implementering och mer i metodisk prövning. Genom att testa, mäta och dokumentera går det att ringa in vilka uppgifter AI löser väl, vilka som kräver noggrann tillsyn och var kvaliteten ännu inte är tillräcklig för att kunna användas i produktion utan betydande risk.

Generativ AI i surveyresearch är varken frälsare eller hot. Det är ett kraftfullt verktyg som kräver statistisk kompetens, etisk eftertanke och institutionell vaksamhet. För verksamheter som tar ansvar för kvaliteten i beslutsunderlag är ett aktivt och metodiskt förhållningssätt inte valfritt – det är en förutsättning för fortsatt trovärdighet.

### *Generativ AI i survey- och statistikarbete – praktiska hållpunkter*

#### 1) När AI ofta fungerar bra

- Koda öppna svar och samla dem i teman – men med manuell kontroll (och ibland dubbelkodning).
- Granska enkätfrågor språkligt: hitta otydliga

formuleringar, ledande frågor och svarsalternativ som kan misstolkas.

- Ta fram första utkast till analysplan eller rapport: tabeller att ta fram, möjliga förklaringar och saker att kontrollera.
- Hjälpa till med dokumentation: metodtexter, kodlistor och enklare sammanfattningar (som alltid ska granskas).

#### 2) Vanliga risker

- Påhittade uppgifter: AI kan hitta på siffror eller fakta som låter rimliga.
- Snygga men felaktiga slutsatser: texten kan bli övertygande även när den är fel.
- Snedvridningar: AI kan ge systematiska fel, till exempel i hur svar tolkas eller grupperas.
- Känslig för instruktioner: små ändringar i frågan till AI kan ge olika svar.
- Sekretessrisk: data kan hamna fel om man använder tjänster utan kontroll på lagring och åtkomst.

#### 3) Minsta nivå av kvalitetssäkring

- Testa mot facit innan det används skarpt: jämför med manuellt arbete eller annan känd referens.
- Dokumentera hur ni gjorde: vilken modell, version, instruktion (prompt), datum och vad som ändrats i efterhand.
- Gör rimlighetskontroller: stämmer nivåer, fördelningar, bortfall och resultat mellan grupper?
- Tydligt ansvar: bestäm vem som godkänner AI-resultatet.
- Stickprovskontroll löpande: kontrollera ett urval regelbundet, särskilt vid öppna svar.

#### 4) Dataskydd (extra viktigt i känsliga projekt)

- Lägg inte in respondent- eller kunddata i

verktyg där ni inte vet hur data lagras och används.

- Begränsa åtkomst och logga: det ska gå att se vem som gjort vad.
- Tänk igenom konsekvenserna av fel: vad händer om AI ger ett missvisande resultat?
- Implementera en AI-policy och var öppen i rapporter: skriv om AI har använts, till vad och vilka kontroller som gjorts.

#### 5) Kompetens: när rutinarbete automatiseras

- Säkerställ grundträning: juniora medarbetare behöver fortfarande lära sig datarensning, kodning och vanliga felkällor.
- Träna granskning: att kunna bedöma AI:s output är en färdighet som måste övas.
- Dialog med lärosäten: statistik och metod måste fortsatt läras ut med fokus på mätfel, bias, dokumentation och kritisk prövning relaterat till AI-utvecklingen.

KARIN NELSSON  
VD Demoskop



# Hur mycket vatten använder AI?

## ETT STATISTISKT PROBLEM UTAN ENKLA SVAR

Det är lätt att tro att generativ artificiell intelligens (AI) är något immateriellt. En fråga in, ett svar ut. Vi pratar om kod och om "molnet" – men vatten?

Stora datacenter kan förbruka 19 miljoner liter per dygn, vilket motsvarar vattenförbrukningen i en ort med 10 000–50 000 invånare – beroende på klimat (Yañez-Barnuevo, 2025).

Hur hänger det här ihop?

Generativ AI kräver enorm beräkningskapacitet. Först i träningsfasen, där stora språkmodeller tränas på gigantiska datamängder under veckor eller månader. Träningen körs på tusentals GPU:er som drar el. El blir värme. Värme måste kylas bort.

Sedan kommer användningen – det som kallas "inferens", alltså det ögonblick då en färdigtränad modell tar emot en fråga ("prompt") och genererar ett svar i realtid. Till skillnad från traditionell mjukvara är generativ AI beräkningsintensiv även i inferensfasen. Uppskattningsvis står inferens för 80–90 procent av energianvändningen för AI-tjänster (O'Donnell et al. 2025).

Till skillnad från mycket annan mjukvara är generativ AI beräkningsintensiv även när modellen är färdig. Varje användarfråga innebär realtidsberäkningar i stora nätverk – och när miljoner människor gör detta samtidigt blir energiuttaget enormt.

Mer beräkning betyder mer elförbrukning. Den nya generationens AI har på kort tid gått

från att knappt märkas till att uppskattningsvis stå för omkring 15–20 procent av datacentrens globala energianvändning från 2022 till 2024 (International Energy Agency, 2024).

Mer elförbrukning betyder mer värme. Mer värme kräver mer kylning. Och kylning kräver ofta vatten.

Vatten används direkt i datacenter, särskilt i system och kyltorn där värme förs bort genom avdunstning. Vatten används också indirekt via själva elproduktionen. Resultatet är att AI:s vattenpåverkan inte bara handlar om enstaka träningskörningar, utan om ett permanent flöde kopplat till global digital konsumtion – lokalt förankrat i specifika avrinningsområden.

**WUE: ett elegant mått som inte räcker**  
Statistiker är tränade att göra komplexitet hanterbar: definiera systemet, ange vad som ingår, vara explicita med antaganden och räkna. Resultatet blir ett tal – helst jämförbart och reproducerbart.

I datacenterbranschen är det dominerande vattenmättet Water Usage Effectiveness (WUE), uttryckt som kubikmeter vatten per megawattimme ( $m^3/MWh$ ). För en statistiker är det attraktivt: en tydlig kvot som kan jämföras mellan anläggningar.

Problemet är att WUE systematiskt lämnar utanför sådant som ofta är avgörande när frågan gäller AI:

- indirekt vattenanvändning kopplad till elproduktion
  - vatten i leverantörsled och byggnation
  - hur vatten ska fördelas mellan olika tjänster i en delad infrastruktur
  - effekter på lokala samhällen och ekosystem
- WUE mäter alltså endast intern effektivitet – inte systempåverkan.

Inom vattenförvaltning skiljer man mellan:

- **vattenuttag:** vatten som tas från en källa
- **vattenkonsumtion:** vatten som inte återförs (t.ex. avdunstning)

Ett datacenter med kyltorn tar in vatten, avdunstar en stor del och släpper tillbaka resten. Ett luftkylt system tar nästan inget vatten på plats – men kräver mer el, och elproduktion kan i sin tur ha ett vattenavtryck.

Alltså: vad ska egentligen stå i täljaren – uttag, konsumtion, på plats, eller även indirekt?

Här blir det tydligt att energi och vatten inte är oberoende målvariabler. Karimi et al. (2022) jämförde två kyltekniker för datacenter i Phoenixområdet: luftkylda kylaggregat respektive vattenbaserad kylning. Det vattenkylda systemet hade i genomsnitt 13 procent bättre energieffektivitet, men vattenavtrycket var ungefär fyra gånger högre än för det luftkylda.

Optimerar man för energieffektivitet med avdunstande vattenkylning tenderar vattenanvändningen att öka. Optimerar man för låg vattenanvändning tenderar energibehovet att öka. Det finns ingen gratis lunch. Val av teknik innebär alltid en förskjutning mellan resurser.

WUE är alltså inte ett neutralt – eller oproblematiskt – mått. Det är beroende av hur vi definierar systemgränser.

**Systemgränser: samma data, olika slutsatser**

Ett enkelt sätt att synliggöra komplexiteten är att vara tydlig med systemgränser. Tre typfall:

- A: endast drift på plats
  - B: drift på plats + vatten kopplat till elproduktion
  - C: hela livscykeln (servrar, byggnation, leverantörsled)
- Poängen är inte att en gräns är "rätt", utan att



AI kräver stora datamängder vilket leder till ökad energianvändning och i dess följd ökad vattenförbrukning. Bilden: Microsoft Azure datacenters i Pangborn, Wenatchee, Washington

slutsatsen kan vändas när gränsen flyttas. I ett traditionellt elnät kan indirekt vatten via el stå för en betydande del av total konsumtion. I elnät med mer vind och sol är den delen lägre. Ett datacenter kan därför se "vatteneffektivt" ut lokalt men inte nödvändigtvis ur ett systemperspektiv.

**Geografin gör kvoten normativ**

Anta att två datacenter förbrukar 1 miljon kubikmeter vatten per år. De är alltså volymmässigt identiska.

Men ponera nu att det ena ligger i ett vattenstressat område (till exempel i Arizona, som är en torr region med överutnyttjat grundvatten), medan det andra ligger i ett vattenrikt område (till exempel i Oregon, precis vid Columbiaflodens avrinningsområde).

Den fysiska volymen är densamma – men konsekvenserna är radikalt olika.

Detta gör att endast volymmättet (WUE) är systematiskt missledande: ett bolag kan sätta sina mest vattentörstiga anläggningar i torra regioner och ändå se "bra" ut. Vattenstressviktning (eng: *scarcity weighting*) avslöjar den verkliga skadan och inverterar ofta rankingen av anläggningar – en lågvolymsanläggning i ett vattenstressat område kan alltså vara skadligare än exempelvis en högvolymsanläggning vid ett

vattenrikt avrinningsområde vid kusten.

Utan platsspecifik viktning är jämförelser mellan operatörer i hög grad meningslösa ur ett faktiskt miljöperspektiv.

När vi tillämpar vattenstressviktning kan påverkan skilja sig med en faktor två eller tre, trots identiska kubikmeter.

Det innebär att måttet kodar in:

- Hydrologiska förhållanden
  - Konkurrerande vattenanvändning
  - Ekologiska trösklar
  - Förvaltningsregler och tilldelningssystem
- Vi har då gått från en enkel och trevlig statistikkvot till ett inbäddat samhälls- och naturvetenskapligt antagande.

**Generativ AI och allokeringens problemet**

När vi frågar "hur mycket vatten använder generativ AI?" uppstår ett särskilt problem: datacenter är delad infrastruktur. Samma GPU-kuster kan användas för AI-träning, AI-inferens, andra molntjänster och vanlig klassisk beräkning.

Allokering kan göras på minst två rimliga sätt:

- **Utnyttjandebaserad allokering:** fördelning efter faktisk resursanvändning – till exempel GPU-tid eller energiförbrukning. Om "AI"

står för 30 procent av resursutnyttjandet under en period tilldelas den 30 procent av vattenanvändningen.

- **Tjänstebaserad allokering:** fördelning efter levererad tjänst – till exempel antal förfrågningar, genererade svar eller intäkter. Då kan en AI-tjänst med många korta svar få en annan profil än en mer sällan använd tjänst som kräver tung beräkning per tillfälle.

Båda metoderna är försvarbara, men de ger olika svar – särskilt när kapacitetsutnyttjandet varierar, när träning och inferens blandas, och när en stor del av infrastrukturen står i beredskap.

**Årsmedelvärdet döljer toppbelastning**

Vattenförvaltning är säsongberoende. I vattenstressade regioner är det ofta sommarens uttag som är kritiska, liksom låga flöden och icke-linjära grundvattenreaktioner. Ett årsmedelvärde kan därför vara gravt missvisande: påverkan under toppmånader kan vara mycket större än vad ett jämnt medelvärde antyder.

Låt oss återvända till det vattenstressade området i Arizona där grundvattennivåerna redan sjunker. Jordbrukare har bevattningsrestriktioner. Regionen investerar i vattenbesparande åtgärder. Samtidigt etableras ett stort datacenter

## »Många vattenstressade regioner erbjuder samtidigt billig mark, generösa incitament och ett starkt politiskt intresse av investeringar. Om vatten i beslutsunderlaget behandlas som en genomsnittlig och relativt billig insatsvara kan dessa områden framstå som rationella val.»

(kanske under parollen att det skapar jobb i regionen).

Volymmässigt kanske anläggningen står för en relativt liten andel av det totala uttaget i regionen – kanske några procent. På pappret kan det se marginellt ut.

Men vad händer under en varm och torr sommar?

- Kylbehovet i datacentret ökar
- Avdunstningen ökar
- Samtidigt är det just då som hushåll och jordbruk också behöver mest vatten
- Miljöflöden i vattendrag riskerar att underskrida kritiska nivåer

Ett årsmedelvärde fångar inte huruvida belastningen sammanfaller med perioder av maximal stress.

Och för invånarna i området är frågan inte hur många liter per kilowattimme som används. Frågan är: påverkar detta våra familjers vattenförsörjning, vårt jordbruk, våra kostnader och våra framtida tillgångar?

Datacenter byggs i en rasande fart i olika typer av klimat och regioner. I norra och mellersta Sverige finns exempelvis en rad datacenter där vårt kyliga klimat och tillgång till billig el brukar anföras som skäl till etableringen (Norling & Gunnars, 2024). Men en förvånansvärt stor del av världens datacenter byggs i torra områden där vattentillgången redan är begränsad. Varför byggs då datacenter så ofta i just sådana vattenstressade regioner?

Det kan verka motsägelsefullt, men ur just ett statistiskt perspektiv är det logiskt. Lokalisering av datacenter bygger i hög grad på optimering mot mätbara variabler:

- Låga markpriser
- Tillgång till billig el
- Skatteincitament

- Politiskt välkommande kommuner/regioner (ofta ur ett nästan desperat ekonomiskt behov)
- Tillgång till redan etablerad infrastruktur
- Torrt klimat som minskar korrosionsproblem och ibland kylbehov under delar av året
- Handläggningstid för tillstånd(!)

Dessa faktorer är kvantifierbara och kan vägas samman i beslutsmodeller. De är jämförbara, stabila och går att stoppa in i kalkyler.

Vattenstress däremot är svårare att operationalisera. Den är säsongsberoende, påverkas av klimatförändringar, beror på juridiska tilldelningssystem och uppvisar tröskeeffekter snarare än linjära samband. Den låter sig inte lika enkelt översättas till en enhetlig kostnad per kubikmeter. Det som är svårt att mäta riskerar därför att få mindre tyngd i beslutsfattandet.

Många vattenstressade regioner erbjuder samtidigt billig mark, generösa incitament och ett starkt politiskt intresse av investeringar. Om vatten i beslutsunderlaget behandlas som en genomsnittlig och relativt billig insatsvara kan dessa områden framstå som rationella val. Det är inte nödvändigtvis så att man aktivt söker vattenbrist – men om vatten antas vara homogent och stabilt underskattas riskerna systematiskt just där de är som störst.

När vi presenterar ett genomsnittligt vattenmätt utan geografisk kontext behandlar vi vatten som om en kubikmeter vore densamma överallt och alltid.

Det ironiska är alltså att just förenklade statistiska mått kan ha bidragit till den problemställning vi nu försöker analysera. Det som framstår som rationell optimering inom modellens gränser kan i ett bredare systemperspektiv innebära ökad regional sårbarhet och fördelningskonflikter.

### Referenser

- O'Donnell, J. & Crownhart, C. (2025). We did the math on AI's energy footprint. Here's the story you haven't heard. *MIT Technology Review*. <https://www.technologyreview.com/2025/05/20/1116327/ai-energy-usage-climate-footprint-big-tech/> [Hämtad 2026-04-05]
- International Energy Agency. (2024). Electricity 2024. Analysis and forecast to 2026. <https://www.iea.org/reports/electricity-2024> [Hämtad 2026-04-05]
- Karimi, L., Yacuel, L., Degraft-Johnson, J., Ashby, J., Green, M., Renner, M., Bergman, A., Norwood, R., & Hickenbottom, K. L. (2022). Water-energy tradeoffs in data centers: A case study in hot-arid climates. *Resources, Conservation and Recycling*, 181, 106194.
- Norling, A. & Gunnars, L. (2024). Microsoft, Facebook, Google och Amazon: Här ligger datajättarna i Sverige. Arbetet, 2024-02-22. <https://arbetet.se/2024/02/22/microsoft-facebook-google-och-amazon-har-ligger-datajattarna-i-sverige/> [Hämtad 2026-04-05]
- Yañez-Barnuevo, M. (2025). Data centers and water consumption. *Environmental and Energy Study Institute*, 55. <https://www.eesi.org/articles/view/data-centers-and-water-consumption> [Hämtad 2026-04-05]

### Precision utan sanning

Det går alltså att pressa fram ett tal: "x liter vatten per AI-förfrågan". Med två decimalers noggrannhet dessutom.

Men utan att redovisa systemgränser, uttag kontra konsumtion, indirekta effekter via el, geografisk vattenstress, säsongvariation och fördelningseffekter riskerar vi att producera precision utan sanning eller mening.

Statistiker måste ständigt påminna sig om att det inte räcker att beräkningen är korrekt. Vi måste också synliggöra vilka antaganden som styr svaret – och de komplexa konsekvenser som döljs när vi beskriver världen som kvoter.

PONTUS  
WÄRNSTÄL



» Utöver oss chefredaktörer har vi ett stort internationellt nätverk av forskare som arbetar ideellt med JOS.»

# Journal of Official Statistics (JOS) – 40 år av utveckling

Under 2025 firade *Journal of Official Statistics (JOS)* sitt 40-årsjubileum. Sedan starten har tidskriften haft som mål att vara en bred och central plattform för forskning och metodutveckling inom officiell statistik.

JOS grundades 1985 av Lars Lyberg vid SCB. I det första numret beskrivs målet på följande sätt:

”Producenter av officiell statistik använder, eller bör använda, tillförlitliga metoder från många olika discipliner (statistik, ekonomi, datavetenskap, samhällsvetenskap etc.) och från flera specialområden inom dessa discipliner. Därför kommer den nya *Journal of Official Statistics* att omfatta ett bredare metodområde än de flesta andra statistiska tidskrifter.”

Denna inriktning gäller fortfarande i dag, och JOS är numera en internationellt erkänd vetenskaplig tidskrift med läsare över hela världen. JOS är dessutom helt gratis tillgänglig och öppen för alla intresserade att läsa.

Sedan 2023 drivs JOS genom ett samarbete mellan SCB, Statistikcentralen i Finland och Statistisk sentralbyrå i Norge. Vi är i dag fem chefredaktörer som representerar de tre statistikbyråerna – Suad Elezovic, Ingegerd Jansson, Henri Luomaranta, Li-Chun Zhang och undertecknad. Utöver oss chefredaktörer har vi ett stort internationellt nätverk av forskare som arbetar ideellt med JOS. Vi har 40 Associate Editors från olika länder, och varje artikel granskas noggrant av två till tre oberoende granskare. Därutöver har vi en Book

Review Editor som ansvarar för att nya, intressanta böcker inom vårt område recenserar.

JOS ges ut fyra gånger per år och fortsätter att utvecklas tack vare det engagerade nätverk som står bakom tidskriften.

För att uppmärksamma 40-årsjubileet gavs ett särskilt specialnummer ut i september 2025. Temat för numret är “The New Era of Official Statistics — Challenges and Research Needs” (volym 41, nummer 3). Specialnumret inleds med en översikt över numrets artiklar. Därefter följer en artikel av Risto Lehtonen som sammanfattar hur tidskriften har utvecklats och vilken roll den har haft under sina första fyrtio år.

Huvuddelen av numret är framåtblickande. Ledande forskare inom olika metodområden som är relevanta för officiell statistik har bidragit med korta artiklar där de beskriver de viktigaste utmaningarna inom sina respektive områden. Resultatet är ett mycket spännande och inspirerande nummer. Totalt innehåller specialnumret 22 artiklar som tillsammans täcker centrala utvecklingsfrågor inom officiell statistik – från nationella statistiksystem till datainsamling, statistikproduktion, inferens och analys.

Vi hoppas att specialnumret ska inspirera till fortsatt forskning och utveckling. Och vem vet – när JOS fyller femtio år kanske det är dags att blicka tillbaka på detta specialnummer och se hur långt vi har kommit på vår gemensamma resa.

LILLI JAPEK  
Chefredaktör JOS

## Om *Journal of Official Statistics (JOS)*

*Journal of Official Statistics (JOS)* är en internationell, refereegranskad vetenskaplig tidskrift med fokus på forskning och metodutveckling inom officiell statistik. Tidskriften är öppet tillgänglig och välkomnar bidrag från forskare och praktiker som arbetar med frågor relevanta för officiell statistik. Senaste numret hittar du här: <https://journals.sagepub.com/toc/jofa/42/1>

I samband med JOS 40-årsjubileum gavs ett specialnummer ut med temat *The New Era of Official Statistics — Challenges and Research Needs*. Du hittar specialnumret här: <https://journals.sagepub.com/toc/jofa/41/3>

För den som är intresserad av att skicka in ett manuskript till JOS finns information om inriktning, kriterier och instruktioner för författare på tidskriftens webbplats.

# Valstatistikens långa linjer

Äntligen är det valår. Den 13 september är det åter dags för den svenska demokratins högtidsdag, valdagen. Då är det inte bara ett val som äger rum, utan hela 311: ett riksdagsval, 20 regionfullmäktigeval och 290 kommunfullmäktigeval.

Att ha val till nationell, regional och lokal nivå samma dag är internationellt sett mycket ovanligt. Denna speciella svenska ordning är ett resultat av de partipolitiska överläggningar som föregick övergången till enkammarsdag 1971 (Erlingsson et al. 2022, s. 45f).

Bakgrunden var att framför allt Socialdemokraterna ville behålla det så kallade ”kommunala sambandet” mellan riks- och kommunpolitiken. Under den tidigare tvåkammarsdagens tid utsågs nämligen den första kammarens ledamöter indirekt av landstingen och vissa större städer, vilket medförde att kommunalvalen hade en viss påverkan på rikspolitiken. Argumentet var att en gemensam valdag skulle bidra till att kopplingen mellan nationell och lokal politik bibehölls.

En följd av den gemensamma valdagen är att den kommunala politiken i viss mån hamnar i skuggan av den rikspolitiska debatten. Å andra sidan bidrar den till att hålla uppe valdeltagandet i lokala och regionala val. För oss som intresserar oss för val innebär det dessutom att vi får många rösträkningar att följa under valnatten.

**Varje valtillfälle är i sig** intressant att studera, men för många analyser är det de långsiktiga trenderna som ger djupare inblick i hur den svenska politiska opinionen formats och utvecklats. Vi kan därför vara tacksamma över att Sverige har en lång tradition av officiell valstatistik, med Statistiska centralbyrån (SCB) som ansvarig producent.

Serien sträcker sig ända tillbaka till 1872 års andrakammarval. Den historiskt bevandrade

kanske frågar sig varför valstatistiken inleddes just då och inte 1866, då det första valet till den nya tvåkammarsriksdagen ägde rum. Avsikten var mycket riktigt att SCB skulle ta fram officiell valstatistik redan 1866, men försöket var inte lyckosamt. Som myndigheten själv uttryckte det: ”*Omständigheter af olika slag vållade, att de insända uppgifterna varken blefvo så fullständiga eller så tillförlitliga som önskligt varit*” (SCB 1873, s. 2).

Det nesliga bakslaget medförde att inget nytt försök gjordes i valet 1869, men till 1872 års val var efterfrågan så pass stor att SCB åter tog sig an uppgiften och den gången gick det betydligt bättre. Därmed inleddes en tradition av utförlig officiell valstatistik som lever än i dag.

**Över tid har valstatistiken** utvidgats och utvecklats. I SCB:s redogörelse över valresultaten i 1911 års riksdagsval ingick för första gången mer ingående studier av valdeltagandet i olika grupper av röstberättigade. Dessa studier var inledningsvis totalundersökningar, men efter att först ha testats i 1942 års kommunalval användes urvalsundersökningar fullt ut från och med 1944 års riksdagsval.

Att manuellt registrera information om valdeltagande från röstlängder är ett krävande arbete, så övergången till att nyttja urvalsmetodik innebar stora besparingar. Med tiden har dock nya tekniker för digitalisering utvecklats, vilket har bidragit till att SCB från och med 2018 års val åter studerar valdeltagandet i hela den röstberättigade befolkningen.

Även om hanteringen numera är enklare än när allt registrerades manuellt är det ändå ett omfattande arbete, eftersom röstlängder från landets över 6 000 valdistrikt ska samlas in och skannas. Möjligen kan det arbetet bli lättare framöver. I årets val kommer försök med digitala röstlängder att genomföras i ett mindre antal vallokaler.

Om försöken faller väl ut och digitala röstlängder på sikt kan användas i alla vallokaler skulle det medföra flera fördelar. Inte minst skulle det öppna nya analysmöjligheter, eftersom det då blir möjligt att studera vid vilken tidpunkt rösthandlingen sker.

**Valstatistiken omfattar inte bara** valresultat och valdeltagande, utan även statistik över de politiskt valda. Det innefattar information om vilka som kandiderar i val och vilka som sedermera blir invalda i riksdag, region- och kommunfullmäktige. Dessutom har SCB sedan 2003 tagit fram statistik över vilka som är förtroendevalda i kommuner och regioner. Statistiken över de politiskt valda utgör bland annat värdefullt underlag till analyser av hur väl representerade olika befolkningsgrupper är i olika politiska organ på olika nivåer.

Som en del av framställningen av valstatistiken skapas register över valdeltagande och politiker, vilka i sig är mycket värdefulla. De är fantastiska datamaterial för analyser, inte minst när de kombineras med andra register som finns tillgängliga. Tillgången till sådana material är ett skäl till att data om svensk lokal demokrati på senare år har legat till grund för artiklar i de mest prestigefyllda internationella statsvetenskapliga tidskrifterna (se t.ex. Dancygier et al. 2025; Oskarsson et al. 2021; Folke et al. 2016).

**Den högkvalitativa valstatistiken** och andra registerdata ger en konkurrensfördel för forskare som intresserar sig för svenska förhållanden. Det är dock av vikt att dessa datamaterial är väl strukturerade, dokumenterade och arkiverade. Dokumentation är sällan den mest lustfyllda delen av statistikproduktionen (jag vet, jag har själv ägnat mig åt det), men värdet är stort, särskilt när äldre material ska användas och de som tog fram statistiken inte längre är i tjänst.

Arkiveringen har dessutom fått förnyad betydelse i takt med att det blivit enklare att digitalisera historiskt material. Forskare har exempelvis nyligen sammanställt och tillgängliggjort en databas med information om kön och parti för alla valda till kommunfullmäktige 1919–2018 (Bawati et al. 2025). Det hade inte varit möjligt om inte SCB hade arkiverat underlag från framställningen av den officiella valstatistiken hos Riksarkivet.

**Så långt om valstatistikens historia** och värdet av långa tidsserier – vad finns då att säga om de val som stundar i höst? Mandatfördelningen i riksdagen och vem som får bilda regering är nog de resultat som flest är spända inför. Det finns dock mycket annat att hålla ögonen på. Utifrån mina forskningsintressen är det två saker som jag först och främst kommer att titta närmare på när rösterna räknats klart.

För det första är jag nyfiken på hur högt valdeltagandet blir. I förra valet sjönk andelen röstande i riksdagsvalet med 3 procentenheter till 82,2 procent. Visserligen är det svenska valdeltagandet fortsatt mycket högt internationellt sett, men nedgången 2022 kom efter att valdeltagandet hade stigit i fyra val i rad. Höstens val kommer att visa om det var en tillfällig nedgång eller möjligen början på en längre nedåtgående trend.

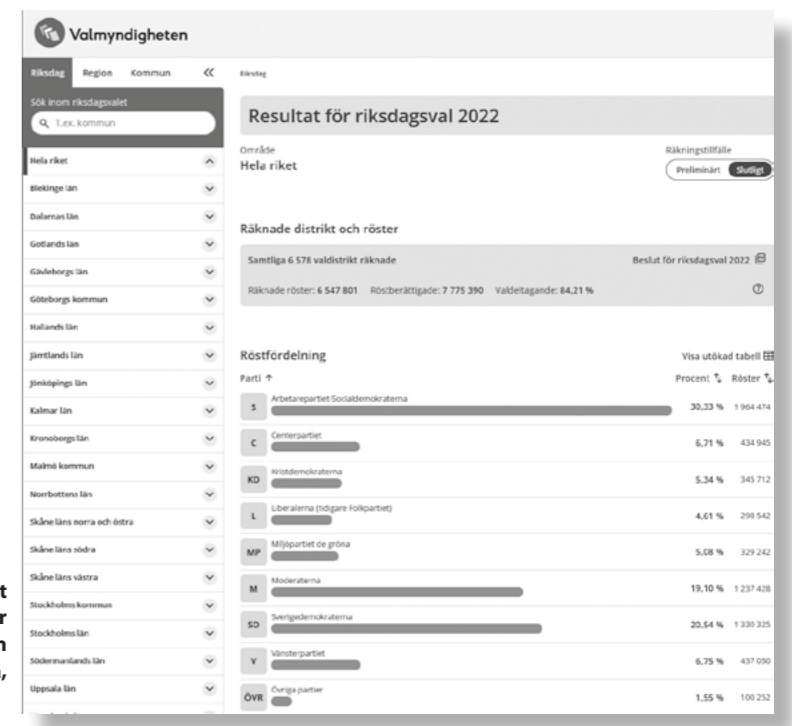
För det andra vill jag se om det finns en tydligare skillnad i partival mellan stad och land än i tidigare val. Under flera år har jag tillsammans med kollegor arbetat i ett forskningsprojekt, finansierat av LF Forskningsstiftelse, där vi undersökt om de ökande politiska spänningar mellan stad och land som kunnat konstateras i många andra länder också går att se i Sverige. Fenomenet fick särskilt stor uppmärksamhet 2016, i efterdyningarna av presidentvalet i USA och folkomröstningen om Brexit i Storbritannien, men har även observerats på andra håll.

**Det är dock inte självklart** hur man ska studera den frågan. I äldre svensk statistik redovisades uppgifter länge efter indelningen i stad och land, som framför allt byggde på dåtidens kommuntyper: landskommuner, städer och köpingar. I dag har vi inte längre olika sorters kommuner, och i den officiella statistiken används inte längre indelningen stad och land. Inte heller finns det någon allmänt vedertagen definition av begreppen. Ofta räknas de områden som SCB klassificerar som tätorter som stad, medan övriga områden betraktas som landsbygd. Tätortsdefinitionen, som fastställdes av de statistikansvariga myndigheterna i de nordiska länderna 1960, innebär att ett område med sammanhängande bebyggelse och minst 200 invånare räknas som tätort.

Om sådana områden betraktas som stad innebär det bland annat att Rusksele tätort, med sina 221 invånare i Lycksele kommuns utkant, klassificeras som stad. Samtidigt betraktas delar av Ekerö kommun som landsbygd, trots att de ligger endast en kort bilresa från centrala Stockholm. Det är nog inte alla som tycker att en sådan klassificering fångar begreppens innebörd.

**I projektet har vi i stället** utgått från andra definitioner av stad och land, bland annat en egen klassificering och den kommungruppsindelning som Tillväxtverket ansvarar för. När vi använder den senare för att studera långsiktiga trender finner vi på det hela taget inte samma polarisering i partival mellan stad och land som har observerats i en rad andra länder. Tvärtom visar ett sammantaget mått en minskad skillnad över tid. Samtidigt kan vi konstatera att de partier som utgör det nuvarande regeringsunderlaget har gått från att ha starkare stöd i städer till att numera ha starkare stöd på landsbygd. Det omvända gäller för partierna i oppositionen. Dessa skillnader ökade dessutom i det senaste valet 2022.

Den 13 september i år är det inte bara ett val som äger rum – utan 311. Bilden från Valmyndighetens hemsida, val.se



Efter valdagen den 13 september kommer vi att veta om den trenden håller i sig och om Sverige rör oss mot den polarisering mellan stad och land som vi har sett på andra håll i världen.

RICHARD ÖHRVALL

## Referenser

- Bawati, Abrar, Moa Frödin Gruneau, Josefine Magnusson & Johanna Rickne, 2025. "The Swedish historical municipal council database", *Electoral Studies*, 95: 102921.
- Dancygier, Rafaela, Sirus H. Dehdari, David D. Laitin, Moritz Marbach & Kåre Vernby. 2025. "Emigration and radical right populism", *American Journal of Political Science*, 69: 252–267.
- Folke, Olle, Torsten Persson & Johanna Rickne. 2016. "The Primary Effect: Preference Votes and Political Promotions", *American Political Science Review*, 110(3): 559–78.
- Oskarsson, Sven, Christopher T. Dawes, Karl-Oskar Lindgren & Richard Öhrvall. 2021. "Big Brother Sees You, but Does He Rule You? The Relationship between Birth Order and Political Candidacy", *The Journal of Politics*, 83(3): 1158–1162.
- Statistiska centralbyrån. 1873. Bidrag till Sveriges officiella statistik. R: *Valstatistik I, Berättelse rörande riksdagsmannavalen år 1872*. Stockholm: P. A. Norstedt & söner.
- Öhrvall, Richard, Gissur Ó. Erlingsson & Emanuel Wittberg (2024). "Geografisk polarisering i röstningsbeteende: Nationella och lokala skillnader", *Statsvetenskaplig tidskrift*, 126(5): 983–1004.

# Recension av *Daten sind Macht* av Katharina Schüller

I nom Föreningen för industriell statistik beklagade vi understundom beslutsfattares bristande statistiska insikter. Det handlade inte så mycket om matematiska bevis eller R-programmering utan mer om förståelse för vilka möjligheter och fallgropar som insamling av data ("den nya oljan") innebär, speciellt i kombination med artificiell intelligens. En viktig aspekt är också att beslutsfattare bör ha tillräckliga kunskaper för att ställa de relevanta frågorna.

Boken "Daten sind Macht" av Katharina Schüller kom ut hösten 2025 och riktar sig explicit till beslutsfattare ("Führungskräfte"). Den torde kunna bidra till att förbättra beslutsfattares förståelse. Naturligtvis bör även andra grupper kunna ha nytta av boken, exempelvis de som behöver argumentera med ledningsgrupper.

Namnet Katharina Schüller borde inte vara helt obekant för många av Svenska Statistikkonsultbolaget STAT-UP i München, vilket hon lett sedan dess. Hennes bok innehåller en hel del beskrivningar av exempel hon mött i sin konsultverksamhet samt en del intervjuer med framstående experter (såsom Gerd Gigerenzer, som vid ett tillfälle hamnade på listan över världens 100 främsta "thought leaders" och som bidrog till en artikel i Qvintensen, och Georges-Simon Ulrich, som bland annat varit Schweiz chefsstatistiker).

Katharina grundade 2003 statistikonsultbolaget STAT-UP i München, vilket hon lett sedan dess. Hennes bok innehåller en hel del beskrivningar av exempel hon mött i sin konsultverksamhet samt en del intervjuer med framstående experter (såsom Gerd Gigerenzer, som vid ett tillfälle hamnade på listan över världens 100 främsta "thought leaders" och som bidrog till en artikel i Qvintensen, och Georges-Simon Ulrich, som bland annat varit Schweiz chefsstatistiker).

**Boken innehåller inga** formler eller programmeringsexempel men däremot en del historiska tillbakablickar och filosofiska resonemang. Som exempel på det förra kan nämnas referens till Florence Nightingales ikoniska diagram som påverkade beslutsfattare så att de släppte till mer resurser till fältsjukvården.

Några centrala budskap i boken är:

- Data är inte neutrala utan kan presenteras på ett sätt som passar avsändaren. Det är därför viktigt att förstå avsändarens agenda när man tolkar data.
- Vikten av att ställa de rätta frågorna. Ibland kan man bli kritiserad för att man "bara ställer massa frågor men levererar inga svar". Det finns också det ironiska talesättet att en dåre kan fråga mer än tio lärda kan besvara. Emellertid torde det vara en viktig egenskap, inte minst för beslutsfattare, att ställa de rätta frågorna.

- Utan klart problem, ingen tydlig lösning. Resultatorienterade människor vill gärna komma till lösningar så snabbt som möjligt, men risken finns då att man lyfter fram sin käpphäst som inte alls löser det faktiska problemet.

**Några andra noteringar** som jag gjorde vid genomläsningen:

Ett intressant exempel på filosofiskt resonemang är en diskussion om begreppet "rättvisa". I boken används det engelska begreppet "fairness". Katharina påpekar (helt korrekt som jag ser det) att en del fundamentala frågeställningar kring begreppet är olösta, och förmodligen kommer så att förbli.

Ett exempel på fallgrop är att tro att stora datamängder per automatik leder till bättre information. EU-kommissionen skickade ut en enkät om sommardag till ett stort antal EU-medborgare. Av dem som svarade var 80 % för avskaffande av sommardag. Nu kan man mistänka att de som inte brydde sig lät bli att svara i stor utsträckning, medan de som var starkt emot sommardag tog tillfället i akt att svara.

I boken nämns ett exempel på att en begränsad undersökning i ett perifert land som Sverige kan få stort genomslag, på ett sätt som inte var tänkt. Novus skickade en enkät till cirka 1 000 utlandsfödda i Sverige. Av dessa angav 183 att de flytt krig och förtryck och av dessa svarade nästan 80 procent att de besökt sitt hemland efter flytten till Sverige. Av detta drog Elon Musk den generella slutsatsen att "nästan 80 procent av de asylsökande semestrar i sina hemländer, så vari ligger deras skyddsbehov?". Det framgår inte när och varifrån dessa personer flytt, men man kan mistänka att "de 183" huvudsakligen är personer som flytt för länge sedan, exempelvis från det forna Jugoslavien eller Chile, och som är säkra att besöka i dagsläget. Ytterst få nyanlända flyktingar torde kommit med i Novus panel.

**En aspekt som nämns** i förbifarten är behovet av strikt terminologi. I ett av exemplen nämns att en försäljningsavdelning kan lägga in en annan betydelse av ordet "kund" än en finansavdelning. Min personliga reflexion är att terminologiarbete prioriterats ned under de senaste årtiondena vilket lett till en del onödiga diskussioner och missförstånd. Som exempel kan nämnas att Tekniska Nomenklaturcentralen och dess efterföljare har blivit styvmoderligt behandlade.

Det påpekas att en stor del av arbetet i da-

taprojekt går åt till att korrigera ("tvätta") data. Exempel på hur data blivit förvrängda och därmed vilseledande om man okritiskt försöker dra generella slutsatser ur dessa hittar man under rubriken "Notizen aus der Praxis: Verwahrloste Verwaltungsdaten", ("försummade administrativa data"). Exemplet är specifika för Tyskland men motsvarande exempel borde finnas även i Sverige.

**Ett intressant praktiskt** exempel som nämns är ett containerföretag som tycks ha problem med att hålla reda på var containrarna håller hus på grund av brister i datahanteringen. Tanken var bland annat att man skulle förse containrarna med sensorer för insamling av data som sedan kunde användas som underlag för optimering. Detta påminner om en av mina käpphästar. När jag studerade teknisk fysik på KTH på 70-talet var operationsanalys en stor grej. Under mina yrkesverksamma år (1976–2015) stötte jag knappt på begreppet. I Wetterbergs bok "Ingenjörerna" finns ett kort avsnitt om operationsanalys, men enligt denna verkar historien sluta på 70-talet. Kan brister i datahantering vara en orsak till att operationsanalys inte fått det genomslag som man kanske förväntade sig?

Sammanfattningsvis kan man säga att det är en mycket läsvärd bok, som gärna skulle kunna översättas till engelska. Det nämns ganska många namn på såväl historiska som nutida personer. Det hade varit trevligt med ett appendix med korta biografier av dessa.

HANS ALBERG

## Referenser

- Bauer, T., Gigerenzer, G., Krämer, W., & Schüller, K. (2022). *Grüne fahren SUV und Joggen macht unsterblich: Über Risiken und Nebenwirkungen der Unstatistik. Gewinner des getAbstract Business Impact Award 2023*. Campus Verlag
- Godhe, A.-L., Magnusson, P., & Sofkova Hashemi, S. (2020). Adequate Digital Competence: Exploring revisions in the Swedish national curriculum. *Educare*, (2), 74–91. <https://doi.org/10.24834/educare.2020.2.4>
- Wetterberg, G. (2020). *Ingenjörerna*. Albert Bonniers Förlag.

# Är det verkligen så självklart?

**Anna Torrång skrev i sin krönika i Qvintensen 2014/3 att "det som känns så självklart för mig behöver inte vara självklart för andra." Det är ju alldeles – självklart. Men varför är det så svårt att förstå?**

När mina barn var små och uppgivna inför något som verkade alldeles för svårt påminde jag dem om hur lärande känns. "När du lärde dig att simma tyckte du att det var jättesvårt. Nu är du som en fisk i vattnet." Man glömmer hur svårt det var i början.

Vi människor har olika erfarenheter. Många kan inte räkna procent. Om något har varit 80 och nu är 160, hur många procentökning är det? Om du höjer på ögonbrynet åt det, tänk då på något du är dålig på.

Min mattelärare i högstadiet, Dorith Nilsson, sa att man ska ta "förändring delat med ursprunget". Och gånger 100 så att det blir procent. Hon upprepade det så ofta att det nu känns – självklart.

UK Statistics Authority skriver att det är "important that statistical literacy is not viewed as a deficit that needs to be fixed". Producenter av statistik ska fokusera på vad de gör för att underlätta för läsare

**»Underförstått: gnäll inte över att folk är okunniga, gör något själv i stället.»**

med vitt skilda bakgrunder och kunskaper, säger de. Underförstått: gnäll inte över att folk är okunniga, gör något själv i stället.

Jag gnällde inför ett gäng vetenskapsjournalister. Enligt sajten FiveThirtyEight hade Trump 35% sannolikhet att vinna över Clinton i

valet 2016. Vi vet nu att han gjorde det, fast då kom det som en stor överraskning. Jag sa i intervjuer att han hade jättestor chans. Det gick inte hem. "Men 35% är ju mindre än 50%" var det en som sa. Jag klagade inför vetenskapsjournalisterna. Varför begriper de inte vad jag säger? En av dem sa att jag ska säga "en gång av tre". Inte 35%.

Vad betyder detta 35% egentligen? En gång av tre vinner Trump, två gånger av tre vinner Clinton.

**Råd 1:** Tänk efter vad det du vill kommunicera egentligen betyder, och säg det. Säg inget annat.

Vetenskapsjournalisterna förstod att "35%" kan vara svårt att nå fram med. Men statistiker kan se ett problem med "en gång av tre". Vilka tre gånger? Valet 2016 hände bara en gång.

**Råd 2:** Tänk efter vilka du kommunicerar till, och anpassa din kommunikation till målgruppen.

När jag var redaktör för Qvintensen slog det mig ofta att statistiker gärna vill ha med detaljer. Även om de inte behövdes. En sorts bokföringsattityd. I stället för att bara skriva "mötet ägde rum i oktober", eller "i höstas", skulle datumen vara med.

**Råd 3:** Förenkla budskapet. Är detaljerna nödvändiga för din kommunikation?

Journalisten Louise Kristoffersson ringde mig för ett antal år sedan om att föreningen Arkitekturupproret hade utsett Nya Hotell Continental till Sveriges näst fulaste nya byggnad. Är det represen-

tativt för vad folk tycker? Jag har problem med det oklara begreppet representativ. Jag försökte väl säga att om man frågar folk som har en viss inställning kan man inte generalisera till andra som har en annan inställning. Hon lyssnade på mitt krångliga svar, och gjorde en analogi i sin artikel. Det är som att fråga personer i Friskis & Svettis om vad de tycker om motion.

**Råd 4:** Tänk ut en analogi eller en liknelse. Kan dock falla platt om mottagaren inte tycker att liknelsen är lik.

Det här är svårt att förklara: "sannolikheten för en händelse A givet B är inte samma sak som sannolikheten för en händelse B givet A". Ännu värre blir det om man väljer ordet betinga i stället för "givet". Men så här kan man göra. Johan Bring och Lars Rönnegård diskuterar problemet att bedöma sannolikheten för att en person är under eller över 18 år om en bild med magnetkamera har visat att hon har ett moget knä. Frågan är juridiskt viktig för personer som saknar papper som svenska myndigheter godkänner, och gränsen är skarp: under eller över 18 år. "Tänk er att alla deltagare [i de studier det gäller] står i samma rum. Vi ber alla barn kliva fram och räknar sedan alla med moget knä. Andelen i denna grupp med moget knä är P(Moget knä|Barn). Det är samma sak som sannolikheten att missta ett barn för en vuxen. Nu ber vi barnen återgå till de vuxna. Sen ber vi alla med moget knä kliva fram. Andelen barn i denna grupp är P(Barn|Moget knä)."

**Råd 5:** Måla upp en bild.

Om jag inte missminner mig hade Lars Rönnegård pratat med sin fru om hur man skulle förklara "sannolikheten för en händelse A givet B".

**Råd 6:** Prata med någon som inte har fastnat i "men det är ju självklart".

DAN HEDLIN



FOTO: EVA DALIN

## Referenser

- Bring, Johan & Rönngård, Lars. Åldersbedömningar – en statistisk utmaning. *Folkvett 2018:1*. <https://www.vof.se/wp-content/uploads/2018/06/Aldersbestamningar-en-statistisk-utmaning.pdf>
- Office for Statistics Regulation. 2023. *Statistical Literacy – it's all in the communication*. <https://osr.statisticsauthority.gov.uk/publication/statistical-literacy-its-all-in-the-communication/#:~:text=The%20review%20recommends%20that%20producers%20of%20statistics,%20Using%20simple%2C%20easy%20to%20understand%20language>

# Polisens trygghetsundersökningar och vad SIG-gruppen gör

## Polisens trygghetsmätningar – stärker polisens och kommuners arbete för ökad trygghet sedan 1990-talet

Polisen har genomfört lokala trygghetsmätningar i egen regi sedan mitten av 1990-talet. Initiativet togs av polishögskolans forskningsenhet initierat av de nuvarande professorerna Per-Olof Wikström (Cambridge) och Marie Torstensson Levander (Malmö universitet) efter förebild av amerikanska kriminologers epokgörande forskning. Det genomfördes sedan av rikspolisstyrelsen under många år, som ett inslag i verksamhetsstyrningen. Tanken var att mäta effekter av polisarbete, och av de flesta av de dåvarande 21 polismyndigheterna, för operativ planering. Detta utvecklades särskilt inom polismyndigheten i Skåne och de andra tre polismyndigheterna inom det som då var Syd samverkan (föregångaren till Polisregion Syd) [2].

I ett mycket stort antal kommuner har kontinuerliga mätningar av polisens genomförts över lång tid, i vissa fall närmare 30 år. Därtill har andra aktörer med samma enkätfrågor genomfört helt jämförbara trygghetsmätningar i stor skala, ofta oberoende av polisens. Idag år 2026 har 289 kommuner åtminstone en gång deltagit i undersökningen under perioden, och över 150 kontinuerligt över ett antal år. Antal undersökta kommuner kan uppskattas till minst 2 200 geografiska områden om varje år räknas separat. Minst en och halv miljon personer i Sverige har någon gång svarat på enkäten. År 2025 tillkom över 220 000 nya respondentsvar, merparten från sex av sju polisregioners samtliga kommuner (Bergslagen, Mitt, Nord, Syd, Väst och Öst).

**Arbetet med trygghetsmätningar** innefattar numera mycket mer än bara själva mätningen. Det är en underrättelsebaserad arbetsmetod – ett koncept – för att stärka polisens och kommunernas lokala arbete, reducera utsatthet för brott och utveckla samhällets möjligheter att förbättra tryggheten i localsamhället. Detta tillvägagångssätt har utvecklats över lång tid av de tidigare polismyndigheterna och numera av polisregioner i samverkan med kommuner.

**”Att alltid inse att testet på polisens effektivitet är frånvaron av brott och oordning, och inte de synliga bevisen på polisens aktivitet i arbetet mot problemen.”**

SIR ROBERT PEEL, ÅR 1829 [1]

**Teorin bakom** trygghetsmätningen kan enklast förklaras som ett försök att undersöka nivåer av utsatthet för brott och ordningsstörningar med dess konsekvenser byggd på den så kallade nolltolerans teorin och Broken Windows teorin. Trygghetsmätningens upplägg bygger i grunden på dessa hypoteser. Frågeformuläret ställer frågor kring vilka problem den boende uppfattar i sitt bostadsområde som redan inträffat i utemiljön, eller som observeras vid passage. Det är också frågor kring egen utsatthet för brott, respondentens allmänna oro för brott, hans mer konkreta känsla av otrygghet inkluderat konsekvenser av detta samt en bedömning av polisens engagemang i bostadsområdet. Det är då också möjligt att bedöma nio delkomponenter: utemiljön, ordningsstörningar, trafikrelaterade problem, utsatthet för brott, allmän oro för brott, mer konkret otrygghetskänsla, beteendeförändringar, polisens lokala agerande och social tillit som starkt antas påverka varandra. Detta är i huvudsak en formativ undersökning, dvs indikatorerna är i sig trygghetsmiljön och utgör en problemkartläggning. Data och metoder har upprepade gånger utvärderats av akademien och har av internationella världsledande forskare ansetts som en av de mest intressanta longitudinella studierna i Europa. Till exempel deltog kriminologipristagaren professor Tom Tyler, Yale University, i Polisregion Syds panel vid Stockholms Criminology Symposium 2024.

**En Särskild intressegrupp skapas tillsammans med Statistikfrämjandet** Hösten 2025 tog Polisregion Syd ett initiativ till att starta en särskild intressegrupp (SIG) inom Statistikfrämjandet, vilket stöttades av Statistikfrämjandet. Polisregion Syd, som bevarat majoriteten av individrådata, var mycket

intresserade av att dessa data kom till särskild nytta i samhället, framför allt för forskare inom olika ämnesområden, särskilt då statistikvetenskap är ett relativt utforskat ämnesområde för polisens. Då data är anonymiserad och på inget sätt kan spåras till individer, samt att resultaten är helt officiellt kommunicerade, så fanns det inga hinder att erbjuda denna möjlighet för högskolor och universitet. Linköpings (LiU) och Stockholms universitet (SU) var intresserade.

Linköpings universitets kandidatprogram för statistik och analys bidrar med två studier. En kandidatuppsats [3] handlar om att undersöka vilka faktorer som hänger ihop med att bli utsatt för brott, med fokus på det som i polistermer kallas mängdbrott, det vill säga våld, stöld och skadegörelse. Studien bygger på enkätdata från 2005 till 2025, där ett urval av år används för jämförelser. Studien använder information om individers bakgrund, bostadsort, upplevd trygghet och erfarenheter av utsatthet för brott. Flernivåmodeller används för att studera hur både individuella faktorer och skillnader mellan områden kan påverka risken att utsättas för mängdbrott. Dessutom används faktoranalys för att sammanfatta vissa enkätfrågor till mer övergripande mått. Målet är att bidra till bättre förståelse av vilka faktorer som är kopplade till brottsutsatthet och hur dessa varierar mellan olika individer och områden.

**En andra kandidatuppsats** från LiU:s studenter [4] undersöker sambandet mellan mängdbrott och människors upplevda trygghet i samhället. Syftet är att identifiera vilka faktorer som påverkar känslan av oro och otrygghet samt hur dessa samband kan förklaras genom både sociala, miljömässiga och individuella faktorer. Studien analyserar bland annat oro för brott,

konkret otrygghet och tidigare brottsutsatthet. Den huvudsakliga metoden är strukturell ekvationsmodellering (SEM) som analyserar komplexa samband mellan observerade och latent variabler. Metoden valdes eftersom den kan hantera mätfel och analysera flera samband samtidigt genom så kallad ”multigroup SEM”. För att bygga modellen används först explorativ faktoranalys (EFA) där variabelernas naturliga gruppering undersöks. Därefter genomförs en konfirmativ faktoranalys (CFA) för att testa om faktorstrukturen passar datan och fungerar som mätmodell.

**Studenter från** Stockholms universitet, Statistiska institutionen, bidrar med tre studier. En viktig studie undersöker hur bortfall i enkätdata kan hanteras med statistiska metoder för att förbättra skattningarna. Tidigare undersökningar har visat marginella skillnader mellan olika typer av viktningar i relation till polisiär operativ användning. Generella svårigheter är om svarande i exempelvis yngre åldersgrupper som är relativt få i relation till åldersgruppen på något sätt skiljer från den mycket stora majoritetsgrupp som inte svarar. Här råder idag stor oklarhet och många olika åsikter bland sociologer och kriminologer. Polisens i Region Syd följer idag SOM-institutets forskningsens rekommendationer.

**En annan uppsats [5]** kartlägger Polisregion Syd och hur geografiskt närliggande områden påverkar varandras upplevda otrygghet. I Statistiska Centralbyråns statistikdatabas finns det möjlighet att inhämta information om ett antal typer av befolkningsstatistik uppdelat efter demografiskt statistikområde (DeSO). Genom tillgång till 157 155 observationer och 1127 unika DeSO mellan 2022–2025 skapas möjligheten till en unik statistisk analys. Denna uppsats utgår från frågan “om du går ut ensam sent en kväll i området där du bor, känner du dig då trygg eller otrygg?” Syftet är att försöka fånga en allmän otrygghet hos respondenterna. Utfallet för frågan är dikotomt vilket underlättar för de valda statistiska modellerna, samt att bortfallen för denna fråga uppgår till 13.4% (24

259 observationer) vilket är betydligt lägre än normalt för frågor i denna typ av studie.

Den tredje studien från SU har två dimensioner – en teoretisk och en empirisk del. Uppsatsen har som mål [6] att koda Hoffs algoritm själv, för att sedan byta apriori-fördelning, vilket i bayesiansk teori, som Hoff använder, är det antagande om korrelationsmatrisen vi gör innan vi sett data, för att sedan kunna uppdatera matrisen i kombination med datans likelihood, för att få fram den resulterande posteriorfördelningen. Genom att byta apriori-fördelning från “Inverse-Wishart”-fördelningen till “LKJ”-fördelningen, är förhoppningen att kunna formulera ett nytt sätt för forskningsvärlden att estimerar korrelationer för komplexa ordinaldata på ett effektivt och kreativt sätt. För att göra det här arbetet meningsfullt utöver den nya teoretiska infallsvinkeln, är ett kriminologiskt perspektiv baserat på teorin Broken Windows central.

En av studierna utvecklar en *Crime Trends Cross Analysis* för fyra sydliga län (Skåne, Blekinge, Kalmar, Kronoberg). Analysen omfattar länsvisa trendvisualiseringar, könsuppdelade analyser (män, kvinnor, totalt) samt ett sammansatt *Problem Index*. Datamaterialet består av medborgarundersökningar 2015–2025 och täcker 14 indikatorer inom fem kategorier. Detta möjliggör tydlig år-för-år-trendanalys och direkt jämförelse mellan länen. Resultaten bör ge polisanställda, kommuner, beslutsfattare och forskare större insikter över utvecklingen och standardiserade dataunderlag för fortsatt uppföljning [7].

SIG välkomnar fler studenter, forskare och andra aktörer som är intresserade av att förstå data från Polisregion Syds trygghetsmätningar.

KJELL ELEFALK

Senior rådgivare, tidigare utvecklingsdirektör på rikspolisstyrelsen

## Referenser

- [1] Sir Robert Peel's Principles of Law Enforcement, 1829
- [2] Heidenberg: Balanced Scorecard, University of Stockholm; Grönlund: Balanced Scorecard, Örebro Universitet, Sweden; Kujanpää och Virta: Polisens Balanced Scorecard, Polissskolan i Tammerfors, Finland
- [3] N. Petersen, O. Clareus, LIU
- [4] S. Castman, F. Forss, LIU
- [5] J. Sellers och L. Bergström Larsson, SU
- [6] H. Tornberg, SU
- [7] A. Banerjee P. Banerjee S. P. Dessai Shuang-Lin. C, Hyper Island



## ORDFÖRANDEN HAR ORDET

# Nu har vi en inriktning för framtiden!

Årsmötet har gått av stapeln! I mars träffades vi både i riktiga livet och virtuellt för årsmötet, denna gång besökte vi Uppsala.

Årets föreläsare dominerades av vinnarna av Wallinska resestipendiet. Chen Ma höll en djuplodad presentation av sitt arbete med *Iterative Feature Selection and Unsupervised Learning for High-Dimensional Complex Data*. Hannes Waldetoft tog oss med på en resa i *Skattning av hatbrottsmotiv i polisanmälningar med hjälp av prediktionsmodeller baserade på textdata*. Jonas Bjermo berättade om *Algorithms for optimizing model-based incomplete block designs*. Alla tre ska ut internationellt och presentera sina arbeten, och det ska bli intressant att få höra mer om deras erfarenheter framöver.

Vi fick dessutom äran att lyssna till Måns Magnusson som berättade om en elegant och

**»Med detta vägval kommer vi inte att arbeta på som vanligt. I stället innebär vägvalet att föreningen behöver fler som ställer upp på ideell basis för att genomföra aktiviteter...»**

kvalitetssäkrande lösning för *Iterative data curation – a new paradigm for large-scale data quality?* Jag hoppas vi får återse Måns framöver även om tekniken tycktes ta (an)stöt av hans monolog!

Årsmötet detta år var något utöver det normala. Framtidsvalet stod framför oss – åt

vilket håll vill föreningen att organisationen ska? Svaret kan ni läsa om i detta nummer av Qvintensen, att vi fortsätter som ideell organisation. Några kanske undrar om detta kommer att förändra något i hur styrelsen och föreningen arbetar, och någon kanske gör ett antagande att vi arbetar på som vanligt.

Med detta vägval kommer vi inte att arbeta på som vanligt. I stället innebär vägvalet att föreningen behöver fler som ställer upp på ideell basis för att genomföra aktiviteter som exempelvis organisera och delta i arbetsmarknadsdagar, webinarier och mycket annat. Jag hoppas verkligen att medlemmarna är redo för detta!

**Till årsmötet hade** valberedningen på grund av detta vägval arbetat med att säkerställa en bra sammansättning av den kommande styrelsen – oavsett utfall. Vi avtackade Magnus Pettersson hans insatser och engagemang över åren. Magnus har fört med sig en positiv och nyfiken inställning till allt som görs och även drivit det internationella arbetet i FENStatS med ackrediteringen. Att vi har haft en stor framgång med att

etablera detta system på EU-basis ser jag som och en enorm personlig prestation. Magnus fortsätter att vara aktiv medlem i föreningen via SIG Väst och som fortsatt representant för ackrediteringen hos FENStatS.

Vi tackade även av Mattias Strandberg som har varit en bärande person i allt Statistikfrämjandet har gjort. Mattias har arbetat för

föreningen i över tio år och lämnar ett stort tomrum efter sig. När jag träffade honom för första gången introducerade han sig som en person ”som gillar att få saker gjorda”. Så riktigt det visade sig vara! Nu får resten av styrelsen rycka upp sig och få saker gjorda utan honom.

Som ny sekreterare välkomnade vi Samuel Hellman, till vardags arbetande på Folkhälsomyndigheten. Jag ser fram emot att lära känna Samuel framöver!

Vi har även flera från styrelsen som fortsätter – Annika som kassör, John som vice ordförande, Tea och Patrik som ledamöter – jag är så glad att ni är med på vår nya resa!

Det finaste priset vi har, Årets statistikfrämjare, gick till SOM-institutet. Sedan 1986 har de arbetat för att öka förståelsen av svenskarnas tankar och åsikter om samhället, opinionen och media. I december 2025 lanserade de ett nytt öppet dataverktyg för att tillgängliggöra denna guldgruva till datamängder.

**Missade du årsmötet** kan du titta på de olika sessionerna på vår hemsida.

Nu ser vi framåt med spännande aktiviteter framöver! Vi vill fortsätta att lyfta Qvintensen och har därför utvecklat en ny plats på vår hemsida för att läsarna ska hitta enskilda artiklar på ett enkelt sätt. Gå in på hemsidan och kolla in detta.

Det är tryggt att se en ambitiös plan forma sig – var med och påverka utvecklingen för vår framtid du med!

NANCY STEINBACH  
Ordförande Statistikfrämjandet

## ORDFÖRANDEN HAR ORDET

# Två årsmötesdagar att reflektera över

Surveyföreningen hade i år traditionsenligt årsmöte dagen efter Statistikfrämjandet. Jag reflekterar här över dagarna i omvänd ordning.

### Surveyföreningens årsmöteskonferens

Först ut på konferensen var årets Tore Dalenius-talare Dan Hedlin, Stockholms universitet, som talade under rubriken *Mina möten med Tore Dalenius och mina tankar om statistikers roll*.

Förra året fanns inga nomineringar till Surveyföreningens uppsatspris. I år hade styrelsen glädjen att läsa tio uppsatser. Vinnare och talare på konferensen var Max Johansson, som presenterade *Estimating Mode Effects by Three Methods in The Swedish Survey on Income and Living Conditions*. Uppsatsen ingick i hans masterexamen i statistik vid Uppsala universitet. Tre väl kvalificerade uppsatser kom på en delad andraplats. Författarna i bokstavsordning var: Åsa Elmén, Stockholms universitet, Isa Jansson, Örebro universitet, och Simon Thernström, Uppsala universitet.

Konferensens tredje föredrag behandlade ett aktuellt ämne: *Är opinionsinstitutet en färskock?* Johan Martinsson, opinionschef Demoskop och docent i statsvetenskap, presenterade sin studie i detta ämne.

### Surveyföreningens årsmöte

På årsmötet valdes en styrelse med åtta medlemmar som anges intill med roller efter årsmötet och det konstituerande styrelsemötet.

En av punkterna på årsmötet var en stad-

### Surveyföreningens årsmöte valde följande representanter för 2026:

Felix Cassel, kassör  
Larisa Constantin, webbansvarig  
Johan Eklund, ordlistan  
Eva Elvers, ordförande  
Åsa Greijer, sekreterare  
Eva Lagercrantz, representant i Statistikfrämjandets styrelse  
Sebastian Lundmark, vice ordförande och seminarieansvarig  
Sara Scholtens, ledamot

geändring. Rollen för sektionsrepresentanten i Statistikfrämjandets styrelse diskuterades i höstas i den ordförandegrupp som bildades i samband med Statistikfrämjandets initiativ till modernisering. Ett dokument togs fram av främst de tre sektionsordförandena. Rollen handlar om att vara en brygga mellan den egna sektionens styrelse och Statistikfrämjandets styrelse samt att vara en aktiv ledamot i Statistikfrämjandets styrelse. Det visade sig att Surveyföreningens stadgar saknade en beskrivning av denna representant – att den finns och hur den utses. Nu har stadgarna utvidgats.

### Statistikfrämjandets årsmöte och föreningarnas relationer

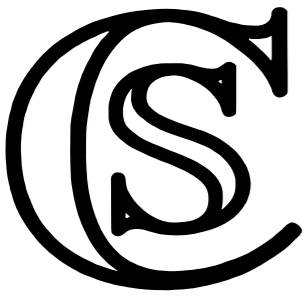
Statistikfrämjandets årsmöte med vägval var välbesökt och intensivt. Att vägvalet med klar

majoritet blev fortsatt ideell förening var en lättnad för mig, både som ordförande för Surveyföreningen och som privatperson. När Statistikfrämjandet bildades genom samgående av två ideella föreningar var jag ordförande i den ena: Statistikersamfundet.

Surveyföreningen var en sektion inom Statistikersamfundet men då inte en egen juridisk person. Det innebar oklarheter i ansvarsfrågor. I anslutning till samgåendet formulerades stadgarna om, och Surveyföreningen blev en egen juridisk person. Surveysektionen är, liksom de andra sektionerna, både en egen ideell förening och en sektion inom Statistikfrämjandet. Den som är medlem i Surveyföreningen är också medlem i Statistikfrämjandet med allt vad det ger.

**De senaste åren** har Statistikfrämjandet arbetat med tankar avseende framtida organisation och ekonomi. Detta arbete har även påverkat sektionerna. När vägvalet nu är gjort hoppas jag att det blir mer fokus på den utåtriktade verksamheten och att inte energikrävande ”störningar” uppstår. Surveyföreningen vill gärna se mer samarbete med Statistikfrämjandet och med de två andra sektionerna. När detta skrivs i april har de tre sektionsordförandena träffats digitalt, och ordförandegruppen med fyra ordförande har haft ett möte i, och om, den nya situationen.

EVA ELVERS  
Ordförande Surveyföreningen



## ORDFÖRANDEN HAR ORDET

# En viktig mötesplats för alla med intresse för sannolikhetsteori och statistik

ett nytt verksamhetsår har tagit sin början, och Cramérsällskapet fortsätter att vara en viktig mötesplats för alla med intresse för sannolikhetsteori och statistik runt om i landet. Vår verksamhet bärs, precis som tidigare, av ett starkt ideellt engagemang, vilket vi ser som en av våra största tillgångar.

Vid årsmötet den 25 mars, som hölls digitalt via Zoom, valdes en ny styrelse för det kommande året. Jag vill tacka för förtroendet att få fortsätta som ordförande. I styrelsen ingår även Malin Palö Forsström (sekreterare), Anders Lundquist (kassör) samt ledamöterna Jolanta Pielaszkiwicz, Johan Koskinen och Johan Lindström. I samband med årsmötet tackade vi också Tatyana Turova och Fan Yang Wallentin för deras värdefulla insatser under sina år i styrelsen.

Den 13 april arrangerade vi digitalt vår årliga vårkonferens för att uppmärksamma och hedra vinnaren av Cramérpriset för bästa doktorsavhandling 2025. Årets pristagare, Federica Milinanni, disputerade vid Institutionen för matematik vid Kungliga Tekniska högskolan (KTH) med avhandlingen *Large Deviation Analysis of Markov Chain Monte Carlo Methods and Algorithmic Advances for Uncertainty Quantification*.

*tion in Neuroscience*. Huvudhandledare var Pierre Nyquist.

**Konferensen bjöd på** ett starkt vetenskapligt program med flera framstående talare. Förutom pristagaren själv medverkade Christophe Andrieu (University of Bristol), Andi Wang (University of Warwick), Chang-Han Rhee (Northwestern University), Jun Yang (Köpenhamns universitet) samt Daniele Avitabile (Vrije Universiteit Amsterdam). Föredragen knöt an till pristagarens forskningsområde och behandlade såväl teoretiska som algoritmiska aspekter av Markovkedje-Monte Carlo (MCMC), inklusive konvergensanalys, algoritmdesign och osäkerhetskvantifiering i komplexa modeller. I sitt avslutande föredrag gav Federica Milinanni en presentation av de centrala delarna av sin avhandling, med särskilt fokus på hur teorin för stora avvikelser kan användas för analys av MC-MC-metoder.

Mötet blev mycket uppskattat och samlade ett engagerat deltagande.

**Under det kommande** året planerar vi att fortsätta med våra uppskattade webinarier två gånger per termin. Dessa ger en möjlighet att ta del av aktuell forskning

inom statistik och matematisk statistik i Sverige, och vi välkomnar gärna förslag på talare och teman. Ett annat viktigt fokus för sällskapet just nu är planeringen av det kommande höstmötet. De senaste årens möten har bjudit på engagerade och livliga diskussioner om statistikens roll i en omvärld där datadrivna metoder, AI och maskininlärning utvecklas snabbt. Ambitionen är att även årets möte ska bli lika givande, och vi ser fram emot att snart kunna återkomma med mer information.

Det gångna året har också präglats av viktiga diskussioner om framtiden för vår moderförening, Statistikfrämjandet. Efter en längre tids utredning ställdes Statistikfrämjandets årsstämma inför ett vägval mellan en fortsatt ideell verksamhet och en mer professionaliserad inriktning. Cramérsällskapets styrelse har genom hela processen tydligt förespråkat den ideella linjen, och vi välkomnar att årsstämman fattade ett entydigt beslut i denna riktning.

Avslutningsvis vill jag tacka alla er som bidrar till sällskapets verksamhet – genom deltagande, idéer eller egna initiativ – och hjälper till att skapa ett levande och relevant forum för statistik och matematisk statistik i Sverige.

Jag ser fram emot ett nytt år med givande möten och inspirerande samtal.

JIMMY OLSSON  
Ordförande Cramérsällskapet



## ORDFÖRANDEN HAR ORDET

# Det går bättre för projekt som lyssnar på statistiskt litterata kollegor

Detta är min första ordförandespalt – tack för förtroendet! Mitt ursprung är en avhandling i matematisk statistik inom matematisk statistik/bioinformatik från Uppsala Universitet 2005. Sedan dess har jag mest jobbat som konsult i life science, men jag har också haft 4 års akademiska forskartjänster i bioinformatik (Uppsala Universitet och WEHI i Melbourne Australien), och jobbat på Big Pharma i 3 år (CSL Behring, Melbourne Australien).

Förutom under utlandstjänsterna har jag bott och bor utanför Uppsala sedan länge, med man och 4 barn (2 nu utflugna barn och 2 på tröskeln), och jag jobbar sedan 5,5 år som konsult på Cytel där jag också ansvarar för vårt Uppsalakontor. För 10 år sedan satt jag som sekreterare för Statistikfrämjandet, som jag uppskattat mycket genom åren, men nu är det dags att dra mitt strå till stacken även för FMS.

**FMS är för mig** en stabil och trygg professionell familj – man kan dyka upp på möten ett par gånger per år, och möts alltid av bekanta kollegor, utvecklande ämnen och intressanta diskussioner. Jag ser också FMS som en viktig brygga mellan utbildning och yrkesliv. Vi fångar upp unga statistiker med intresse för det medicinska området och ger dem en kontinuerlig tillhörighet med möjlighet till personlig utveckling. Det är särskilt viktigt då upplärningen till statistikeryrket fortsätter långt efter att man lämnat grundutbildningen, och man dessutom som statistiker inte alltid har lokala kollegor av samma profession att lära av på sin arbetsplats. Vårt att värna om tycker jag – både FMS verksamhet och dess tillgänglighet!

**Qvintensens tema för** det här numret, statistik litteracitet, sätter fingret på ett par fenomen jag noterat genom åren. Jag har 20 års erfarenhet som statistiker inom life science, mestadels som konsult i läkemedelsbranschen. I mitt jobb möter jag väldigt många olika projektgrupper från olika företag, institutioner och ibland myndigheter. Min första (helt subjektiva) reflektion:

Det går märkbart bättre för de projekt som har, lyssnar på och litar på statistiskt litterata kollegor.

**Vad betyder det** till exempel att vi ser effekt endast i en subgrupp? Jo, vi som läser Qvintensen vet nog att tolka det som att huvudresultaten är negativa, att vi har fått ett uppslag om en alternativ patientgrupp där det kan finnas effekt, men att det lika gärna kan vara ett falskt positivt resultat, så vi skulle inte satsa vår samlade förmögenhet på en kommande studie på subgruppen bara sådär. Det finns CMO:s som hör detta och förstår precis, medan andra springer i väg till VD och drar i gång en ny stor studie på subgruppen utan eftertanke (det borde heta förtanke!) – för att sedan se miljonerna ramla ner i intet. De data som redan finns i ett projekt utgör viktig evidens. Ibland utgör de 100% av underlaget för beslut om nästa steg i projektet. Inte undra på att det gör skillnad för projektet att man förstår och kan tolka dessa!

Kapitel två i min statistiskt litterata skola handlar sedan om att planera vilka data som behöver samlas in för att kunna dra ytterligare användbara slutsatser i framtiden.

Kommunikationen med både CMO och VD här ligger såklart även på oss statistiker, vi har en viktig ledarskapsroll i att få fram våra budskap i varje projektsituation.

**Jag vill slå** ett slag för den möjlighet vi har att sprida kunskap om statistisk litteracitet. Detta är min andra reflektion. Den får bli ett tips att ta med sig för idag:

Passa på att flika in en ordentlig förklaring till resultaten eller statistiken du visar närhelst du får chansen. Det är i regel uppskattat i stunden även om det tar några extra sekunder av presentationstiden, och det tjänar definitivt ett högre syfte!

Jag och min statistikerkollega Anna Törner har ett nyhetsbrev på LinkedIn, *Statistik är mer än siffror*, som syftar till att förklara statistik för icke-statistiker. Där kan ni ta del av vår ansats att öka människors statistiska litteracitet.

INGRID LÖNNSTEDT  
Ordförande FMS

## FMS årsmöte valde vid vårt vårmöte följande representanter för 2026:

Ingrid Lönnstedt, Cytel (ordförande – nyval)

Adnan Noor Baloch, Göteborgs Universitet (kassör)

Joshua Entrop, Red Door Analytics (sekreterare)

Anna Johanson, Karolinska Institutet (ledamot)

Erik Bülow, Göteborgs Universitet (ledamot)

Marie Linder, Karolinska Institutet (ledamot – nyval, Qvintensen-representant)

Anita Lindmark, Umeå Universitet (ledamot, SSF, nyval)

Carl-Fredrik Burman, AstraZeneca (EFSPi-representant – nyval)

Jonas Häggström, HDC (EFSPi-representant)

samt Henrik Imberg och Cecilia Lundholm (revisorer) och Gino Almondo, Caroline Dietrich och Aaron Levine (valberedning)

Vi tackar Aaron Levine, Calliditas Therapeutics Company, och Carl-Fredrik Burman, AstraZeneca, för värdefulla insatser i styrelsen under tidigare år. Tack också till avgående valberedning Annika Dominicus, Gabriel Granäsen och Anna Oksanen.

# Föreläsningsprogram med stor bredd på FMS vårmöte

FMS vårmöte ägde rum den 25:e mars 2026 på Pingstkyrkan i centrala Uppsala. Ungefär 30 medlemmar deltog i 6 föreläsningar. Talarna bjöds in från statliga myndigheter, akademien, och industrin.

Programmets första föreläsning hölls av Johanna Holm från Socialstyrelsen med titeln "Sweden's implementation of EHDS2 – what is happening and what will the future hold?" European Health Data Space, förkortat EHDS, är ett EU-initiativ som ämnar etablera en gemensam struktur för användning och utbyte av elektronisk hälsodata inom EU. Dess primära syfte är att förbättra tillgång till patientdata inom vården i EU. EHDS sekundära syfte är att göra hälsodata tillgänglig för forskning, innovation, officiell statistik, och policyutformning. Det sekundära syftet var föreläsningens huvudfokus.

Dataanvändare kommer att kunna söka bland EHDS datakällor via en centralkatalog eller nationella kataloger tillhandahållna av medlemsstatens kontaktmyndighet, så kallade Health Data Access Bodies (HDAB). Tillgång till specifika datakällor ges när projektet blir godkänt av HDAB. Dataanvändare får tillgång till data i en säker miljö (safe processing environment). Projektets utfall rapporteras tillbaka till HDAB och EU när resultaten är färdiga. I Sverige är Socialstyrelsen utnämnd som HDAB och Statistiska Centralbyrån har ansvar för att ta fram säkra miljöer.

**Xiaoyang Du, doktorand** hos Karolinska Institutet, presenterade preliminära forskningsresultat med titeln "Cost-effectiveness of Artificial Intelligence-based Risk Profiling for Guiding Treatment of Patients With Early Breast Cancer". Xiaoyang bad publiken att inte dela resultat från presentationen då de är just preliminära. AI-verktyget har utvecklats med hjälp av data från svenska cancerregister och kan komma att användas som ett komplement till existerande genuttrycksanalyser. Genut-

trycksanalyser används redan idag för att guida behandling av bröstcancer.

**Helena Ahlin, statistiker** på Swedish Orphan BioVitrum (SOBI), diskuterade alternativa sätt att planera kliniska prövningar när klassiska upplägg inte är lämpliga i sin presentation "Beyond Randomization: Statistical Innovation for Rare Disease Trials". SOBI fokuserar på läkemedelsutveckling inom sällsynta sjukdomar. Patientgrupperna är små och saknar ofta behandlingsalternativ vilket gör det opraktiskt att randomisera patienter till placebo och uppnå statistisk signifikans mellan behandlingsgruppen och placebogruppen.

Alternativen som presenterades var enarmade studier, adaptiva studiedesigner, cross-over-studier, enpatientstudier, basketstudier, bayesianska studiedesigner, och studier med externa kontrollarmar. Jämförelser kan göras mellan behandlingsgruppen i studien och externa eller historiska data. Några principer att följa är hur likvärdiga patientpopulationerna är i de olika datakällorna, hur pålitliga datakällorna är, och i vilken utsträckning regulatoriska myndigheter accepterar evidensen som skapas. Många statistiska verktyg finns för att redogöra för osäkerheter som kan uppstå. Bayesian dynamic borrowing kan hantera bias när populationerna inte är helt lika, simuleringar kan stärka designbeslut, och sensitivitetsanalyser kan beskriva robusthet mot nödvändiga antaganden.

**Chamika Porage, doktorand** på Uppsala universitet, presenterade sin forskning inom kausal inferens i sin presentation "Prognostic

Score Methods for the Estimation of the Average Causal Effect". Metoden bygger vidare på en metod som heter single prognostic score där möjligheten att hantera variabilitet i behandlingseffekten mellan individer saknas. Metoden som föreslås heter full prognostic score (FPGS) och skiljer sig från single prognostic score i och med att två prediktioner görs för varje patient, en där patienten ingår i behandlingsgruppen och en där patienten ingår i kontrollgruppen. Så länge alla störfaktorer mäts och alla patienter har möjlighet att hamna i behandlings- eller kontrollgruppen kan FPGS estimeras.

FPGS kan estimeras med både klassiska statistiska metoder och med maskininlärningsmetoder. Simuleringsstudier visade att linjär regression fungerade bra i linjära sammanhang och att maskininlärningsmetoder fungerade

bra när datastrukturerna var mer komplicerade. FPGS tillåter estimering av den genomsnittliga kausala effekten (*average causal effect*) i epidemiologiska studier.

**Alessandro Gasparini presenterade** "From Methods to Maintainability: The Statistical Software Revolution in the Pharmaceutical Industry and Minimum Viable Good Practices for High Quality Statistical Software Packages". Läkemedelsindustrin står inför ett skifte på grund av större datamängder, mer komplex forskning och ökande kostnader. Öppen källkod kan vara ett sätt att arbeta snabbare och mer öppet. Företag och forskare kan dela metoder, hitta fel tillsammans och enklare visa hur verktygen fungerar för myndigheter. Samtidigt finns hinder, som krav på dokumentation, versionshantering och validering, osäkerhet kring att dela kod, olika

arbetsätt mellan avdelningar och risken att små projekt blir svåra att underhålla. Lösningarna handlar om smartare validering, bättre samarbete och att bygga upp tydliga rutiner och roller för mjukvaruutveckling inom läkemedelsindustrin.

Alessandro presenterade också grundregler för att skapa pålitliga och långsiktigt hållbara mjukvarupaket. Dessa inkluderar tydlig dokumentation, fungerande exempel för användarna, samt tester som säkerställer att funktionerna verkligen fungerar. Funktioner ska vara korta och begripliga, kod ska följa en enhetlig stil och paketets livscykel ska skötas med få beroenden och ordnad versionshantering. Tillsammans gör detta utvecklingen mer kontrollerad och skapar programvara som blir lättare för både användare och utvecklare att lita på och bygga vidare på.

**Ollie Östlund presenterade** Uppsala Clinical Research Center (UCR) med verksamhetsidé "Improving Human Health in Sweden and Worldwide." UCR är en akademisk forskningsorganisation som ger expertråd om klinisk forskning till både akademiska och privata aktörer. Konceptet stora enkla prövningar ställdes mot pragmatiska prövningar där det förstnämnda ämnar beskriva effekter i olika sammanhang, medan det sistnämnda ämnar beskriva effekter inom samma sammanhang. UCR har varit drivande i fenomenet registerbaserade randomiserade kliniska prövningar. Exempel på dessa är TASTE, DETO2X och REDUCE, vilka visar att stora enkla prövningar kan genomföras snabbt, kostnadseffektivt och med hög extern validitet. Dessa studier utnyttjar nationella register för uppföljning och därmed minskar den praktiska belastningen på kliniker, samtidigt som de möjliggör mycket stora patientunderlag.

Sammantaget bjöd dagen på den stora bredden inom metodutveckling och tillämpningar i medicinsk statistik, samt värdet av kunskapsutbyte mellan myndigheter, akademi och industri.

AARON LEVINE

Principal Biostatistician, Calliditas Therapeutics



**Anita Lindmark  
ny ledamot  
i FMS styrelse**

**Grattis till ditt nya uppdrag som styrelseledamot i FMS och representant i Statistikfrämjandets styrelse, Anita! Vem är du?**

– Jag heter Anita Lindmark och är lektor och docent i statistik vid Enheten för statistik, Handelshögskolan, Umeå universitet. Jag disputerade år 2016 men har undervisat i statistik i olika former och omfattning sedan 2009, då jag anställdes som amanuens vid sidan av mina kandidatstudier. Jag gillar kombinationen av forskning och undervisning i mitt arbete och att jag hela tiden har möjlighet att utvecklas inom båda områdena.

**Vilka områden inom statistiken är du särskilt intresserad av?**

– Mina forskningsintressen är främst inom kausal inferens och registerstudier. De senaste åren har jag framför allt arbetat med metoder för kausal mediationsanalys, och hur dessa kan tillämpas för att ta reda på mer om orsaker till ojämlikheter i vård och utfall efter stroke. Jag har använt data från bland annat Riksstroket, det svenska strokeregistret.

**Vad ser du mest fram emot i ditt nya uppdrag?**

– Jag tycker att det känns väldigt spännande att få vara med i det förändringsarbete som kommer att ske i och med Statistikfrämjandets vägval. Jag hoppas att vi kommer att kunna hitta arbetsformer och samarbeten som leder till ökad tydlighet och större engagemang.

CHARLOTTE OVESSON

# Anders Martin-Löfs matematiska arbeten

## Introduktion

Anders Martin-Löf gick bort 25 februari 2026, strax innan sin 86:e födelsedag. Anders avlade civilingenjörsexamen i teknisk fysik vid KTH i början av 60-talet och fortsatte därefter forska med inriktning mot optimeringslära och statistisk mekanik. Efter forskningsvistelser bland annat vid MIT och Rockefeller University i New York, disputerade han i maj 1973 vid KTH med Frank Spitzer som opponent. Han innehade därefter docenttjänster vid Uppsala universitet och Stockholms universitet, innan han lämnade akademien för att bli chefsaktuarie vid Folksam. Efter några år återvände han och tillträdde Cramérprofessuren i Försäkringsmatematik och matematisk statistik vid Stockholms universitet, där han stannade till sin pensionering. Anders var mycket brett kunnig inom matematiken och dess tillämpningar. Han var prestigelös och drevs att ett genuint intresse för de problem han studerade. Det gjorde att han förblev aktiv även efter pensioneringen och han var en flitig besökare vid seminarier inom många olika matematiska områden. Här beskriver vi några områden där Anders bidrog med viktiga resultat.

## Statistisk mekanik

Under Anders tidigaste år som forskare arbetade han till största delen inom statistisk mekanik, där sannolikhetsteori och statistik kombineras för att studera makroskopiska egenskaper hos fysikaliska system som är uppbyggda av ett stort antal mikroskopiska enheter. En gas består till exempel av ett så stort antal molekyler att en fullständig beskrivning av systemet är utom räckhåll. Vill man studera makroskopiska egenskaper hos gasen, såsom tryck, temperatur och entropi, är ett möjligt angreppssätt att i stället betrakta molekylnas beteende som slumpmässigt. Slumpmässigheten ärvs då av de makroskopiska storheterna och, med hjälp av stora talens lag eller generaliseringar kända som ergodssatser, kan det sedan ibland visas att den makroskopiska storheten beskrivs väl av sitt medelvärde. Andra probabilistiska metoder kan också användas för att härleda ekvationer som beskriver relationer mellan olika makroskopiska variabler. Slumpmässighet används således för att hantera brist på information – oavsett om

verkligheten i en djupare mening är slumpmässig eller inte, så kan sannolikhetsteoretiska modeller användas för att beskriva fysikaliska fenomen och dra slutsatser om dessa.

En av de mest grundläggande modellerna inom statistisk mekanik är Ising-modellen. Den beskriver ferromagnetism och baseras på spin-variabler (+1 eller -1) associerade med noderna i ett lattice. Ett sannolikhetsmått, känt som Boltzmann-måttet, konstrueras sedan på spin-konfigurationerna. Betrakta till exempel det kvadratiske latticet  $Z^2$  och låt  $\omega = \{\omega_x\}_{x \in Z^2}$  vara en spin-konfiguration på dess noder, med  $\omega_x \in \{+1, -1\}$ . Interaktionen mellan grannoder  $x$  och  $y$  beskrivs av koefficienter  $\{J_{xy}\}_{|x-y|=1}$  och energin hos en konfiguration (utan externt fält) ges av  $H(\sigma) = -\sum_{|x-y|=1} J_{xy} \omega_x \omega_y$ . Sannolikheten för en given konfiguration beror av en parameter  $\beta > 0$ , känd som den inversa temperaturen, och ges av

$$\rho(\omega) = e^{-\beta H(\omega)} Z_{\beta}^{-1},$$

där  $Z_{\beta}$  är en normaliserande konstant känd som partitionsfunktionen. Om  $J_{xy} > 0$  för alla  $x, y$  så blir energin liten för konfigurationer där grannoder tenderar att ha samma spin och sådana konfigurationer premieras då av sannolikhetsmålet. Parametern  $\beta$  beskriver hur stark preferensen är. Modellen är känd för att uppvisa en fasövergång i  $\beta$ : För  $\beta > \beta_c$  ses en ferromagnetisk fas där sannolikhetsmassan koncentreras till de konfigurationer där alla spin har samma värde (+1 eller -1), för  $\beta < \beta_c$  ses en oordnad fas där massan är spidd över många olika typer av konfigurationer.

Anders visade en rad resultat för olika versioner av Ising-modellen. Samtliga är publicerade i *Communications in Mathematical Physics*, som är den ledande tidskriften inom matematisk fysik. Anders samarbetade med väletablerade personer inom området som Giovanni Gallavotti och Joel Lebowitz. Exempel på resultat inkluderar rigorösa argument för kopplingen mellan en ferromagnetisk fas och spin-korrelationer som inte går mot 0 med avståndet mellan noder [4]. Specifikt visade Anders, tillsammans med Gallavotti, att man för höga temperaturer ser korrelationer som inte går mot 0, medan man för låga temperaturer får asymptotiskt okorrelerade spin. I ett annat arbete [9] stu-



Anders Martin-Löf gick bort 25 februari 2026, strax innan sin 86:e födelsedag.

derade han Ising-modellen på en torus med olika typer av randvillkor och visade att det under vissa förutsättningar uppstår två regioner, en med positiv magnetisering och en med negativ. Han studerade också kurvan som separerar dessa regioner. Slutligen undersökte han i [11] egenskaper hos den så kallade fria energin vid låga temperaturer och visade centrala gränsvärdessatser för olika storheter i samma regim.

Anders intresse för statistisk mekanik kulminerade i en mycket välskriven översikt av ämnet som publicerades av Springer Lecture Notes in Physics 1979. I översikten hos MathSciNet kan man läsa: "The author has written very well presented notes for a mathematically oriented introduction to the subject. He says they are intended to bridge the gap between the treatment in physics textbooks and the in the moder presentations of mathematically rigorous results. He has succeeded admirably in his intention."

## Försäkringsmatematik

Anders var innehavare av Cramérprofessuren i Försäkringsmatematik och matematisk statistik under 20 års tid. Liksom sin föregångare Harald Cramér kombinerade Anders en djup förståelse för verkliga försäkringsfrågeställningar, från sin tid som chefsaktuarie på Folksam, med att vara en briljant matematiker. De försäkringsmatematiska artiklar som Anders skrev är publicerade i tidskriften *Scandinavian Actuarial Journal*.

Artiklarna kännetecknas av en förhållandevis lättläst framställning av den matematiska analysen. Det framgår tydligt att Anders inte såg ett värde i en alltför teknisk inramning av sina originella idéer. Målet var inte att bygga

den mest allmänna teorin, men snarare att låta enkla och eleganta argument övertyga läsaren om att de frågeställningar han studerade och de lösningar han förespråkade är de mest naturliga. I linje med denna strävan efter avskalad enkelhet betraktade Anders typiskt enkla stokastiska modeller i exempel som illustrerade de lösningsmetoder han presenterade. Detta visar på en pedagogisk skicklighet och ett självförtroende hos någon som inte behöver bevisa sina matematiska färdigheter.

Liksom Harald Cramér och andra framstående matematiker med intresse för verklighetens problem såg Anders försäkringsverksamhetens praktiska utmaningar som en utgångspunkt för att formulera matematiska frågeställningar vars lösningar kräver kreativitet och kombination av tillämpad matematik av olika slag. Kombinationen av sannolikhetsteori och kontrollteori för dynamiska system är ett återkommande tema i Anders försäkringsmatematiska forskning.

Anders intresserade sig för teori och tillämpningar av optimal styrning av stokastiska processer, ett tema som passar mycket naturligt till frågeställningar inom försäkring. I [10] studeras optimal styrning av Markovkedjor. Mer specifikt studeras hur en försäkringstagare ska agera givet att försäkringsbolaget tillämpar ett bonussystem, där en rapporterad skada innebär en premiehöjning. Artikeln tillhandahåller en beslutsregel, framtagen med hjälp av Markovska beslutsprocesser och dynamisk programmering, som anger vilka skador som försäkringstagaren bör rapportera trots att det innebär en försäkring i form av en övergång till ett sämre tillstånd i bonussystemet.

Ett försäkringsbolags (ekonomiska) tillstånd kan be-

Anders Martin-Löfs artiklar om smittspridning blev internationellt uppmärksammade och blev startskottet för forskningsområdet smittspridningsmodellering vid Stockholms universitet.



skrivs som ett dynamiskt system i många tidsindexerade variabler såsom skadestånd, premienivå, storlek på reserver, nivå på överskott etc. Genom att specificera hur dessa variabler uppdateras och beror av tidigare värden fås ett stokastiskt dynamiskt system. Eftersom vissa av variablerna kan väljas kan systemet styras. Genom att beskriva vad som är önskvärt (stabila premier, stort överskott, etc.) kan ett optimalt stokastiskt kontrollproblem formuleras. I [13] beskriver Anders hur tekniker från linjär styrteori kan säkerställa systemets stabilitet och optimala styrning.

Som beskrivits ovan hade Anders ett stort intresse för och djup kunskap om teoretisk fysik vilket naturligt gjorde att han kunde se naturliga kopplingar mellan detta ämnesområde och försäkringsmatematik. I [16] härleds asymptotiska egenskaper för den s.k. ruinsannolikheten, en klassisk storhet inom försäkringsmatematik. Anders argumenterar i artikeln för att entropibegreppet och termodynamik i allmänhet kan möjliggöra effektiv analys av klassiska frågeställningar inom försäkringsmatematik.

Traditionell matematisk livförsäkringsanalys involverar en mängd förenklande antaganden och implementeringen kräver diverse ad-hoc-regler för bland annat hantering

av överskott till följd av att verkligheten är mer komplex än den modell som antagits. I [15] kritiserar Anders den traditionella analysen och visar på hur en mer realistisk modell tillsammans med teori för dynamiska stokastiska system kan erbjuda ett bättre alternativ.

Baserat på en kurs som Anders gav för Svenska Aktuarietjänsterna skrev Anders översiktartikeln [18] som visade på hur användbar optimal kontrollteori för Markovkedjor är för att analysera frågeställningar inom försäkring. Nu, mer än tre årtionden senare, kan man konstatera att Anders har haft rätt i sin återkommande betoning av detta ämnesområde. Optimal kontrollteori, Markovska beslutsprocesser och dynamisk programmering har blivit standardbegrepp inom en stor del av modern försäkringsmatematik.

Anders organiserade Harald Cramér-symposiet 1993 som syftade till att belysa Harald Cramérs (1893-1985) mångsidiga och mycket inflytelserika forskning inom talteori, sannolikhets-teori, statistik och försäkringsmatematik. I samband med detta sammanställde Anders de två samlingsvolymerna [21] och [22] som publicerades av Springer 1994. Flera av de ingående arbetena

är svåråtkomliga idag och många har visat tacksamhet över Anders värdefulla redaktörsarbete som möjliggjorde samlingsvolymerna.

### Epidemimodellering

Ett annat av Anders forskningsområden var smittspridningsmodellering, eller som det då kallades, epidemimodellering. Den grundläggande modellen kallas SIR (susceptible – infectious – recovered) som beskriver en situation där individer först är mottagliga. Om de smittas blir de smittsamma, och när smittperioden är till ända blir de immuna och förblir så under resten av utbrottet. I stokastiska epidemimodeller, som Anders intresserade sig för, så betraktar man en stor men ändlig population av storlek  $n$ , och smittade personer antas ha smittsamma kontakter med andra individer oberoende av varandra, med sannolikhet  $p_n = 1 - e^{-\lambda/n}$ , där smittsamma kontakter resulterar i smitta om den andra personen är mottaglig, men inte annars. Det förväntade antalet smittsamma kontakter är således  $(n-1)p_n \approx (n-1)\lambda/n \approx \lambda$ . Inledningsvis kommer nästan alla sådana kontakter vara med mottagliga individer, så  $\lambda$  blir också genomsnittligt antal personer som man

smittas i inledningen av utbrottet, något som brukar kallas det basala reproduktionstalet och skrivs  $R_0$ , så  $R_0 = \lambda$  för denna modell. De två fallen  $\lambda \leq 1$  eller  $\lambda > 1$  har kvalitativt olika beteende: Om  $\lambda \leq 1$  kommer utbrottet inte kunna ta fart, men om  $\lambda > 1$  så kan ett stort utbrott ske (om än inte säkert – av en slump kan utbrottet likväl dö ut tidigt).

Men kommer alla smittas om  $\lambda > 1$ ? Nej. Ju fler som smittas under utbrottet, desto fler av de smittsamma kontakterna kommer att ”slösas” på redan smittade. Om andelen mottagliga har sjunkit till  $s < 1$  så kommer en smittad person ju i genomsnitt smitta  $\lambda_s$  individer. Så när  $s < 1/\lambda$ , dvs när fler än  $1 - 1/\lambda$  har smittats (och är immuna mot att smittas), så kommer en smittad person i genomsnitt smitta färre än en person och då kommer smittspridningen avklunga. Av denna anledning sägs flockimmunitet ha uppnåtts när andelen  $1 - 1/\lambda$  har smittats.

Anders första (och viktigaste) artikel [1] inom smittspridning skrevs 1980 tillsammans med Bengt von Bahr, Anders företrädare på Cramérprofessuren. I denna artikel analyserar de denna smittmodell (inkl vissa generaliseringar) rigoröst. Redan innan fanns heuristiska resonemang om att inledningen av epidemin asymptotiskt liknar en

förgreningprocess, där levande (smittsamma) individer föder (smittar) ett Poissonfördelat antal barn (individer) med väntevärde  $\lambda$ , och de härleder även fördelningen för totalt antal födda barn i denna förgreningprocess (som har positiv sannolikhet att växa obegränsat om  $\lambda > 1$ ). I samma artikel visar de även det svårare resultat som säger att det totala antalet individer  $X_n$  som smittas (under förutsättning att epidemin tar fart) är asymptotiskt normalfördelat:  $n^{-1/2}(X_n - n\lambda) \rightarrow N(0, \sigma^2)$ . Beviset går ut på att härleda i vilket läge en sorts slumpvandring korsar en rät linje, och bygger till viss del på ett tidigare arbete av Anders [11]. Resultatet fick stor uppmärksamhet eftersom man tidigare endast heuristiskt hade antagit motsvarande resultat för andelen  $z$  som kommer smittas, samt att ett explicit uttryck för variansen  $\sigma^2$  erhöles. I artikeln gör de även den viktiga kopplingen mellan ett smittoutbrott och en slumpgraf (antalet smittade har samma fördelning som antal individer som hänger samman med indexfallen i en Erdős-Renyi graf). I en senare artikel [17] generaliserar Anders resultaten till en mer allmän modell där en smittad individ har  $Y$  smittsamma kontakter, för en godtycklig slumpvariabel  $Y$  och där individerna väljs likformigt i populationen – i föregående artikel antogs  $Y$  vara  $\text{Bin}(n, 1 - e^{-\lambda/n}) \approx \text{Po}(\lambda)$ .

Som påpekats var båda dessa artiklar uppmärksammade internationellt, och Anders blev ofta inbjuden till workshops om smittspridningsmodellering. Resultaten var också mycket viktiga för avdelningen för Matematisk statistik på Matematiska institutionen vid SU, eftersom de var startskottet för forskningsområdet smittspridningsmodellering som lever än idag vid institutionen. Först ut att fortsätta i området var Gianpaolo Scalia Tomba och Åke Svensson som då var doktorander och senare blev lektorer och professorer (Gianpaolo flyttade senare till Rom). Generationen efter dessa bestod av Håkan Andersson och Tom Britton, som även de innehaft/innehar anställning på avdelningen. Efter det har ytterligare ca 10 personer doktorerat i forskningsområdet.

Senare i karriären skrev Anders ytterligare några artiklar inom smittspridningsmodellering. I [19] studerar Anders en ”nästan kritisk” epidemi, med  $\lambda = 1 + c/n^{-1/3}$ , dvs man smittar aningen mer än en person i genomsnitt. Asymptotiskt visar det sig då att cirka  $cn^{2/3}$  kommer smittas, och i bevisen använder Anders att smittprocessen konvergerar mot en sk Airy-funktion. I Anders sista artikel om smittspridning [2] studerar han, tillsammans med Tom och Svante Janson, en smittomodell där smittspridningen sker på ett (socialt) nätverk. Mer precist härleds hur många som kommer smittas om personer har vaccinerats i förväg, enligt olika vaccinationsstrategier.

### Sankt Petersburg-spelet

Anders sista vetenskapliga arbeten handlade om Sankt Petersburg-spelet. Anders hade intresserat sig för detta redan i ett par tidigare arbeten och återvände till ämnet i slutet av sin karriär, då i samarbete med Allan Gut, professor emeritus i Uppsala. Det klassiska Sankt Petersburg-spelet definieras, i en uppdaterad version av Daniel Bernoullis originalformulering, som följer: Peter kastar ett symmetriskt mynt tills han får krona. Om detta inträffar i omgång  $k$ , betalar han Paul  $2^k$  dukater. Vad är värdet av spelet för Paul? Om  $X$  betecknar Pauls vinst, så gäller att

$$P(X = 2^k) = \frac{1}{2^k}, \quad k = 1, 2, \dots$$

Med en dubblingsstrategi, är Pauls nettovinst alltså alltid 1. Vän-

tevärdet för  $X$  är dock oändligt och ger alltså ingen vägledning kring vad Paul bör vara villig att betala för att delta i spelet. Detta är känt som Sankt Petersburg-paradoxen, eftersom det bygger på antagandet att Paul har ett oändligt kapital att tillgå. En variation är att låta insatsen vara en funktion av antalet spel, vilket leder till den välkända Feller-lösningen [3]: Om  $X, X_1, X_2, \dots$  är oberoende stokastiska variabler fördelade enligt ovan, och  $S_n = \sum_{k=1}^n X_k, n \geq 1$ , så gäller att

$$\frac{S_n}{n \log_2 n} \xrightarrow{p} 1 \quad \text{då } n \rightarrow \infty,$$

där  $\log_u(\cdot)$  betecknar logaritmer med avseende på basen  $u > 0$ . Mer om spelets historik finns i Anders artikel [14] från 1985. Eftersom  $E(X) = \infty$ , så argumenterar Feller för att  $n \log_2 n$  är ett rimligt pris att betala för att delta i  $n$  spel, och motiverat av detta visade Anders en mer precis gränsvärdessats i [14]. Med hjälp av ett sofistikerat förnyelseargument och introduktion av Poissonprocesser visar han att, om  $T_N = (S_N - N \log_2 N) \sqrt{N}$ , så gäller, för  $N = 2^n$ , att  $T_N$  konvergerar i fördelning mot en oändligt delbar och semistabil gränsfördelning med den explicita karaktäristiska funktionen  $\varphi(t) = e^{g(t)}$ , där

$$g(t) = \sum_{k=-\infty}^0 2^{-k} (\exp\{it2^k\} - 1 - it2^k) + \sum_{k=1}^{\infty} 2^{-k} (\exp\{it2^k\} - 1).$$

Baserat på detta resultat härleds en tumregel för vad Paul bör vara beredd att betala för att delta i spelet, med den egenskapen att sannolikheten att han går bankrutt inom  $N = 2^n$  omgångar är liten.

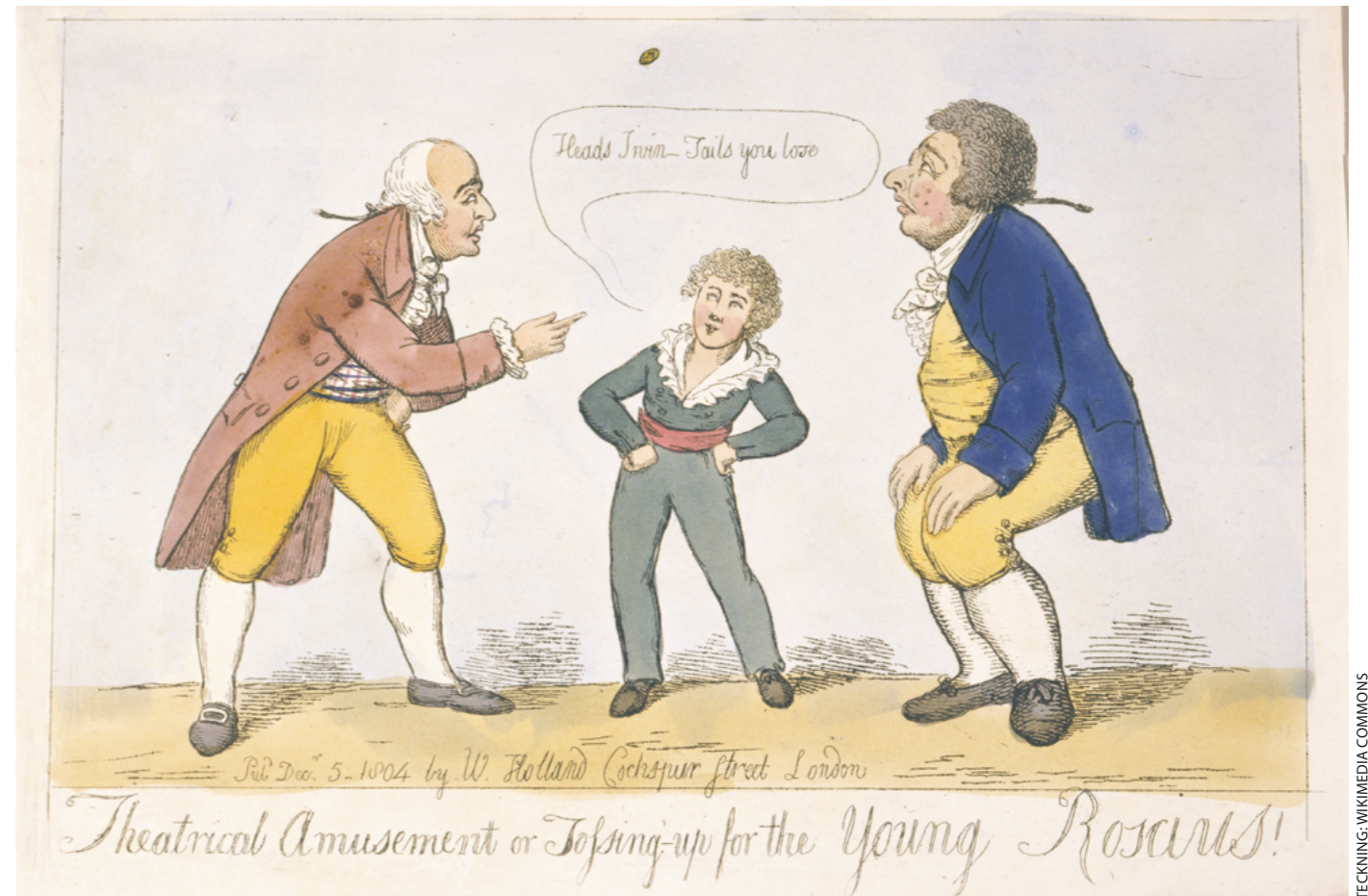
Senare, i [20], studerade Anders två modifieringar av spelet. Den första är trunkering, där Paul har ett ändligt kapital  $2^a$ . Eftersom sannolikheten att överleva en omgång är  $1 - 2^{-a}$ , så följer det att antalet omgångar innan spelet är över,  $M_a$ , är geometriskt fördelat med parameter  $2^{-a}$ . Det medför att  $2^{-a} M_a$  konvergerar i fördelning mot en standard exponentialfördelat variabel  $U$  och att det för nettovinsten,  $V_a = M_a - (2^a - 1)$ , gäller att  $2^{-a} V_a$  konvergerar mot  $U - 1$  när  $a \rightarrow \infty$ . I den andra modifieringen introducerar Anders ränta på kapitalet och visar en gränsvärdessats av samma slag som ovan.

Anders samarbetade därefter i flera artiklar med Allan Gut. Bland annat studerade de i [5] asymmetriska versioner av spelet, där fördelningen för vinsten ges av

$$P(X = sr^{(k-1)/\alpha}) = pq^{k-1}, \quad k = 1, 2, \dots, \quad \text{där } s = 1/p \text{ och } r = 1/q.$$

I [7] studerade de en situation där det största värdet tas bort så att den relevanta storheten blir  $S_n - \max_{1 \leq k \leq n} X_k$ , så kallad max-trimming. Eftersom variablerna är diskreta är maximum inte nödvändigtvis unikt, vilket gör att man behöver introducera  $\tau_n =$  antal maximala observationer efter  $n$  omgångar, och den modifierade summan  $S_n - \tau_n \cdot \max_{1 \leq k \leq n} X_k$ . I [6] studeras på motsvarande sätt min-trimming och extrem-trimming, och i [8] behandlas en kombination av trimming och asymmetri.

TOM BRITTON,  
MARIA DEIJFEN,  
ALLAN GUT,  
FILIP LINDSKOG



Anders Martin Löfs sista vetenskapliga arbeten handlade om Sankt Petersburg-spelet. Anders hade intresserat sig för detta redan i ett par tidigare arbeten och återvände till ämnet i slutet av sin karriär, då i samarbete med Allan Gut, professor emeritus i Uppsala. Brittisk teckning från 1800-talet.

### Referenser

- [1] von Bahr, B. och Martin-Löf, A. (1980). Threshold limit theorems for some epidemic processes. *Advances in Applied Probability* **12**, 319-349.
- [2] Britton, T., Janson, S. och Martin-Löf, A. (2007). Graphs with specified degree distributions, simple epidemics, and local vaccination strategies. *Advances in Applied Probability* **39**, 922-948.
- [3] Feller, W. (1945). Note on the law of large numbers and "fair" games. *Ann. Math. Statist.* **16**, 301-304.
- [4] Gallavotti, G. och Martin-Löf, A. (1972). Surface tension in the Ising model. *Communications in Mathematical Physics* **25**, 87-126.
- [5] Gut, A. och Martin-Löf, A. (2013). Generalized St. Petersburg games revisited. *arXiv:1506.09015*.
- [6] Gut, A. och Martin-Löf, A. (2015). Extreme-trimmed St. Petersburg games. *Statistics and Probability Letters* **96**, 341-345.
- [7] Gut, A. och Martin-Löf, A. (2016). A maxtrimmed St. Petersburg game. *Journal of Theoretical Probability* **29**, 277-291.
- [8] Gut, A. och Martin-Löf, A. (2018). An asymmetric St. Petersburg game with
- [9] Martin-Löf, A. (1972). On the spontaneous magnetization in the Ising model. *Communications in Mathematical Physics* **24**, 253-259.
- [10] Martin-Löf, A. (1973). A method for finding the optimal decision rule for a policy holder of an insurance with a bonus system. *Scandinavian Actuarial Journal* **1**, 23-29.
- [11] Martin-Löf, A. (1973). Mixing properties, differentiability of the free energy and the central limit theorem for a pure phase in the Ising model at low temperature. *Communications in Mathematical Physics* **32**, 75-92.
- [12] Martin-Löf, A. (1979). Statistical mechanics and the foundations of thermodynamics. *Springer Lecture Notes in Physics* **101**.
- [13] Martin-Löf, A. (1983). Premium control in an insurance system, an approach using linear control theory. *Scandinavian Actuarial Journal* **1**, 1-27.
- [14] Martin-Löf, A. (1985). A limit theorem that clarifies the 'Petersburg paradox'. *Journal of Applied Probability* **22**, 634-643.
- [15] Martin-Löf, A. (1986). A stochastic theory
- [16] Martin-Löf, A. (1986). Entropy, a useful concept in risk theory. *Scandinavian Actuarial Journal* **3-4**, 223-235.
- [17] Martin-Löf, A. (1986). Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *Journal of Applied Probability* **23**, 265-282.
- [18] Martin-Löf, A. (1994). Lectures on the use of control theory in insurance. *Scandinavian Actuarial Journal* **1**, 1-25.
- [19] Martin-Löf, A. (1998). The final size of a nearly critical epidemic, and the first passage time of a Wiener process to a parabolic barrier. *Journal of Applied Probability* **35**, 671-682.
- [20] Martin-Löf, A. (2008). An analysis of two modifications of the Petersburg game. *Stochastics* **80**, 175-179.
- [21] *Harald Cramér – Collected Works I*. Springer Collected Works in Mathematics, Springer Berlin, Heidelberg, 1994. Editor: A. Martin-Löf.
- [22] *Harald Cramér – Collected Works II*. Springer Collected Works in Mathematics, Springer Berlin, Heidelberg, 1994. Editor: A. Martin-Löf.

## 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 [www.statistikframjandet.se](http://www.statistikframjandet.se) och läs mer i högerspalten under "Bli medlem". Har du frågor kan du kontakta [sekreterare@statistikframjandet.se](mailto:sekreterare@statistikframjandet.se).

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.

