

Allting varierar – därför statistik

Karin Dahmström, juni 1999

- Hur kan vi förstå verkligheten?

Verkligheten är komplex. Vi har svårt att överblicka och se sammanhangen i allt som är och sker. Det gäller våra egna liv med familj och arbete, men som individer i samhället ingår vi också i större sammanhang. Då tillhör vi grupper, vars åsikter och attityder är av intresse för beslutsfattare och politiker. Och går vi vidare, till andra samhällsbildningar, kan vi försöka se verkligheten i det internationella perspektivet. Och då tycks det som om "allt hänger ihop". Är det ekonomisk kris någonstans i världen, så påverkas även de ekonomiska förutsättningarna här hemma, men hur kan vi förstå alla samverkande faktorer?

Det mesta förändras hela tiden. Och vill vi veta vad som kommer att hända, vilka beslut vi bör fatta, kan vi aldrig förutsäga eller bestämma detta med fullständig säkerhet. Osäkerheten, variationen i framtiden, kan vi dock försöka mäta med hjälp av olika prognosmodeller, som i bästa fall kan vägleda oss om vad vi kan förvänta oss.

- Variation - statistikerns arbetsmaterial

Den variation som vi alla ser och är del av kan man med hjälp av statistiska metoder få viss kontroll över. Statistik kan just uppfattas som metoder för att beskriva och förklara variationen i slumpmässiga försök, situationer där vi inte med full säkerhet kan säga vad som kommer att inträffa. Denna variation kan uppträda i många dimensioner och sammanhang. Vi ser variation i tiden, t ex när vi studerar antalet födda barn, antingen när vi jämför olika år bakåt i tiden eller när vi ser hur många barn som föds varje månad under ett år. Då ser vi också att det finns en säsongvariation: det föds flest barn under vårmånaderna och minst antal under hösten. Vi kan också studera variationen i rummet. Om man ser hur granarna i ett naturligt skogsområde är "placerade", dvs har växt, står de ej i något symmetriskt mönster. Konkurrensen om livsutrymme har i stället medfört en till synes slumpmässig placering, för vilken statistiska modeller har utarbetats.

Variationen finns också i hög grad hos oss själva. Ur statistisk synpunkt - och från opinionsinstitutens synvinkel - är våra olika beteenden, åsikter och attityder intressanta. Men även alla faktiska, varierande egenskaper hos oss och vår omgivning, dvs hur verkligheten ser ut, är statistikerns arbetsmaterial. Då är det inte enbart en beskrivning i tabeller och diagram av variationen som efterfrågas, utan lika väsentligt är att analysera och förklara variationen. Vilka bakomliggande faktorer ökar risken för att drabbas av en viss cancerform? Och hur påverkas aktiekurserna i olika typer av företag av t ex ränteläget och av dollarkursen?

- Som man frågar får man svar, eller ... ?

Men det finns problem med variationen som statistikern måste försöka se bakom och förstå. Det är här som annars det välkända, "dåliga" ryktet om att "allt kan bevisas med statistik" har sin grund. Det gäller t ex att veta hur en undersökning är gjord; visst kan man med hjälp av frågorna få de svar man vill ha och visst kan man få fram förklaringar som kan verka trovärdiga vid första anblicken, men som logiskt har förbisett något avgörande.

Nästan dagligen kan vi läsa om opinionsundersökningar av olika slag. Ofta framhålls då att en annan undersökning på samma tema har kommit till andra resultat; det tycks som om vi som svarar på alla dessa frågor inte kan bestämma oss, vi svarar olika hela tiden. Denna variation kan till stor del förklaras av att frågorna inte är exakt lika formulerade och att svaren samlas in på olika sätt. I vissa undersökningar är det t ex någon intervjuare som ringer upp och i andra får vi en skriftlig enkät att besvara. Även om vi inte gärna vill tillstå det, förstår vi inte alltid vad som avses med frågan och om frågan rör någon händelse bakåt i tiden, har vi svårt med minnet. Vi vill heller inte erkänna beteenden eller åsikter som inte uppfattas som socialt accepterade, och vi vill hellre instämma i ett påstående än att tycka motsatsen. "Svaret" kan också bli att vi struntar i undersökningen, vi blir ett "bortfall". Alla dessa tillkortakommanden både hos den som utarbetar ett frågeformulär och den som skall svara på det, bidrar till att det finns variationer i svaren som är svåra att mäta och tolka. En viktig första förutsättning för en kvalitetsmässigt godtagbar undersökning är att statistikern, i samarbete med den som vill ha en undersökning utförd, utifrån ett så objektivt och kritiskt/ neutralt perspektiv som möjligt planerar hela arbetet från början till slut och sedan håller fast vid detta.

- Variation och samband - falskt eller verkligt?

Visst har vi alla hört och läst om s k nonsenssamband. Ett klassiskt sådant är att när antalet personer i England som hade radiolicens på 20- och 30-talen ökade, visade det sig samtidigt att allt fler patienter var

intagna på mentalsjukhus. Och nu på 1990-talet ser vi att antalet personer som har Internetabonnemang ökar samtidigt som antalet barn som drabbas av allergier ökar. Kan vi därför dra slutsatsen att en ökning av den ena seriens värden skulle orsaka en förändring i den andra? Eller har vi missat något i logiken? Här rör det sig om två tidsserier som båda hade en likartad långsiktig utveckling - uppåtående - men som inte hade det minsta med varandra att göra rent sakligt; sådana serier finns det gott om. Men hur kan man skilja dessa falska samband från utvecklingar där den ena serien faktiskt orsaksmässigt följer på den andra? Kan man t ex visa att variationen i den ena serien följer variationen i den andra med en viss tidsförskjutning? Här, liksom i all tolkning av siffror och samband, är sunt förnuft något av det viktigaste man både som producent och som användare av statistik bör vara utrustad med!

Och i en studie av långtidsarbetslösa personer fann man att svenska män och kvinnor hade längre arbetslöshetstid än utländska personer. Då är det lätt att dra den (vanligaste) slutsatsen att svenskar "ratar" erbjudna jobb, medan utländska medborgare inte är så "kräsna". Men om man också tar hänsyn till åldern hos de arbetslösa, visade det sig att de svenska medborgarna var mycket äldre än de utländska i det här materialet och att det är åldern som framför allt påverkar möjligheten att kunna komma tillbaka i arbetslivet efter lång arbetslöshetstid. Detta visar vikten av att försöka hitta alla faktorer och förklaringar till variationen innan man drar de definitiva slutsatserna.

- Vad kan vi förvänta oss för värden?

Vi själva varierar alltså på olika sätt. Här får vi inte glömma vad som kan kallas den naturliga variationen. Vi väger olika, är olika långa redan när vi föds. För att kunna uppmärksamma när någon mätning är ovanlig, måste vi veta vilka värden som är vanligast. Mätningar på människor kan röra just vikt och längd, men också alla typer av t ex blodprover för att avgöra olika sjukdomstillstånd samt tester för kunskap och lämplighetsbedömningar. Utifrån mätningar av ett stort antal personer kan man få fram vad som kan anses vara normala värden, dvs erhålla ett naturligt variationsområde för en viss analys.

Att bedöma vad man kan vänta sig för mätvärden är ett centralt problemområde inom statistiken. Ofta måste man lita till ett urval av individer i stället för att göra en undersökning av samtliga i den population som man är intresserad av att undersöka. Man vill dock ändå kunna dra slutsatser från mätningarna i urvalet till alla man skulle ha velat mäta. Vad kan vi säga om variationen, om osäkerheten i sådana slutsatser? Om vi har gjort urvalet helt slumpmässigt, kan vi med hjälp av sannolikhetskalkyl bedöma vilka värden vi vanligen kommer att få i ett sådant urval. Vi kan därmed också sätta gränser för värden som vi bedömer vara ovanliga. Utifrån sådana överväganden kan vi våga oss på att ge mått för variationen i urvalet; vi har här bara ett urval att utgå ifrån. Ett sådant mått är ett intervall, den s k statistiska felmarginalen. Resultat som vi anser vara vanliga utifrån givna förutsättningar, ligger inom felmarginalen, andra inte. Om vi t ex i en politisk opinionsundersökning först antar att ingen förändring i sympatierna för ett visst parti har ägt rum sedan föregående mätning, kan vi sedan avgöra om den faktiska förändringen bedöms som liten och ligger inom felmarginalen eller om den är så stor att den inte hör till de vanliga värdena; i det senare fallet är förändringen "statistiskt säkerställd".

- Kvalitetskontroll med statistik

Vem vill ha ruttna apelsiner? Eller färdigpackad mat som inte väger vad vi betalar för? Här kan vi använda statistiska metoder för att bedöma när andelen apelsiner som bör kasseras är större än vad som är acceptabelt och när vikten på skinkpaketet ligger utanför den tillåtna variationen. Denna kvalitetskontroll bygger också på vad vi kan förvänta oss om vi undersöker ett slumpmässigt urval av t ex apelsiner. Men slutsatserna är trots allt osäkra. Vi kan nämligen i exemplet med apelsiner tala om att vi kan fatta fel beslut både utifrån konsumentens och producentens synvinkel, eftersom vi inte har praktisk möjlighet att undersöka hela partiet. Om vi inte kasserar ett parti med stor faktisk andel ruttna apelsiner är detta inte bra för konsumenten. Men om vi i stället bedömer att partiet innehåller för många dåliga apelsiner och så inte är fallet, är det till nackdel för producenten. Att tillfredsställa båda parterna blir en avvägning mellan kostnad och vilka risker de är beredda att ta.

- Hur många blir vi?

Osäkerheten i framtiden till sist, försöker vi hantera med hjälp av prognoser. Här måste vi analysera utvecklingen både bakåt och framåt i tiden. Hur kan vi beskriva och förklara vad som har skett fram till nu? Finns det någon modell som fångar in variationen som varit och som vi tror kanske fortsätter oförändrat eller som fortsätter med vissa modifieringar? Risker är här att det händer oväntade saker som helt kullkastar prognosen. Nya skatteregler, oförutsedd konkurrens, nya produkter kan vara några sådana faktorer. Ett sätt att hantera osäkerheten är att göra alternativa prognoser med olika antaganden om den omgivande verkligheten i framtiden. En typ av prognoser kan dock göras med större säkerhet än många andra, nämligen befolkningsprognoser. Vi vet rätt väl hur många personer som förväntas dö varje år och i vilken ålder. Något större osäkerhet har vi för hur många som kommer att födas varje år och ännu osäkrare blir vi för hur många personer som blir invandrare till respektive utvandrare från landet. Trots allt kan vi med alternativa prognosvärden hantera osäkerheten relativt väl.

- Variationen ger oss kunskapen

Statistik är inte alltid den exakta vetenskap som ett stort antal siffersammanställningar i t ex Statistisk Årsbok ger sken av. Vi möter variation och osäkerhet, som vi trots modeller och analyser inte alltid lyckas fånga och förstå. Men statistiken erbjuder oss ändå möjligheten att med hjälp av sannolikheter och sunt förnuft uttala oss om och försöka förstå verkligheten. Kunskapen kommer från variationen!