
4. KUNSKAPSBILDNINGSPROCESSEN

4.1 Allmänt om kunskapsbildning

En central fråga för statistikteorin är hur kunskap kan utvinnas ur empirin. Vetenskapsteoretiker och filosofer har under lång tid ägnat sig åt den frågan och flera olika ideer, så som skolbildningar, har vuxit fram.

En skolbildning för att skaffa sig kunskap är den som ibland kallas *naiv induktivism*. Här delas kunskapsbildningsprocessen in i fyra steg: Det första steget innebär att man förutsättningslöst observerar och skriver ner alla fakta (gör observationer). I det andra steget beskriver man, klassificerar och analyserar alla observerade fakta. I det tredje steget generaliserar man sina analyser till mer allmängiltiga teorier. I det sista steget, slutligen, gör man förklaringar och prediktioner baserade på dessa teorier. Enligt denna skolbildning börjar kunskapsbildningen i observerandet och går till empiriska generaliseringar. Exempelvis visar studier från Tyskland att i områden där det finns många storkar föds det också många barn. Ju fler områden med hög nativitet och hög storkfrekvens vi träffar på, ju starkare stöd får vi för generaliseringen “det finns ett samband mellan förekomsten av storkar och barnafödande”.

Ett problem med den naiva induktivismen är att vi inte kan observera “allt”. Vi kan bara göra ett ändligt antal observationer och en ändlig mängd observationer kan inte frambringa alla de sanningar vi önskar blottlägga. En möjlig lösning på detta dilemma är att observera endast sådant som är relevant för det problem vi studerar. Enligt detta förslag skulle vi då inte ta hänsyn till exempel tidvattenströmmarna utanför Hollands kust då vi studerar ett visst organisationsproblem i ett företag (såvida vi inte har skäl att tro att tidvattenströmmarna verkligen påverkar det problem vi vill studera). Likaså ledar oss kunskapen att storkar faktiskt inte har något att göra med barnafödande till att överhuvudtaget inte reflektera över något samband mellan nativitet och storkförekomst. Denna i och för sig rimliga tanke för oss dock in i ett nytt dilemma, nämligen att observationerna då är teoriberoende,



Figure 4.1: Vacker kvinna, häxa eller skotte? Tolkningen av observationer beror på observatörens erfarenhetsbakgrund.

dvs. observationerna kan inte längre göras förutsättningslöst. I och med att vi gör en observation måste det också finnas en anledning till varför vi gör den observationen, dvs. vi måste ha någon form av kunskap eller antaganden om verkligheten redan innan vi gör observationen. Ur detta följer också att personer med olika erfarenhetsbakgrund kan tolka samma observation på olika sätt. Figur 4.1 ger ett exempel på detta. Här är observationen en bild där några av oss ser en ung vacker kvinna, medan andra ser en häxa och ytterligare andra en skotte, allt i en och samma observation.

Som alternativ till den naiva induktivismen hävdar filosofen Karl Popper att

kunskap byggs upp genom ett växelspel mellan teori och emperi. Det som kännetecknar en vetenskaplig teori är, enligt Popper, att den skall kunna falsifieras, dvs. kunna förklaras ogiltig, och att det skall vara möjligt att på förhand ange vilken sorts observationer som är nödvändiga att observera för att falsifiera teorin.

Popper hävdar således att observationer alltid är teoriberoende, i så motto att observerandet styrs av någon teori. Denna teori är kopplad till undersökarens kunskap om det verklighetsområde som studeras. Det finns alltså alltid ett syfte med observerandet. Beslutsfattare och planerare behöver data för att kunna värdera olika handlingsalternativ; vetenskapsmän behöver data för att öka förståelsen inom olika forskningsområden; allmänheten behöver data för att kunna delta i samhällsdebatten o s v.

Den metod som Popper har föreslagit för kunskapsbildning på empirisk väg brukar kallas *hypotetisk-deduktiv metod* och beskrivs närmare i avsnitt 4.3. Detta är den metod vi sedan kommer att ansluta oss till och bygga upp statistikteorin kring.

Avslutningsvis skall vi här endast nämna ytterligare ett förhållningssätt till kunskapsbildning på empirisk väg, framförd av filosofen Feyerabend. Enligt detta förhållningssätt ifrågasätts om det över huvudet taget är nödvändigt och intressant att diskutera vad som är "Den Vetenskapliga Metoden". Feyerabend hävdar att alla metoder att söka kunskap är tillåtna. Den vetenskapliga metoden (om det finns någon) är varken bättre eller sämre än någon annan metod att söka kunskap.

4.2 Ett klassisk exempel

Vi har tidigare konstaterat att en teori är nödvändig för att avgöra vad som skall observeras. Men hur vet vi att en given teori är en bra beskrivning av de fenomen vi vill studera? Låt oss betrakta ett numera klassiskt exempel för att illustrera hur teorin bestämmer vad som skall observeras och hur rimligheten i en teori kan prövas.

Allmänna sjukhuset i Wien hade på 1840-talet två förlossningsavdelningar. En stor andel av de kvinnor som var inlagda vid första avdelningen drabbades av den allvarliga och ofta dödliga sjukdomen barnsängsfeber. 1844

dog på första avdelningen 260 av 3 157 mödrar, eller 8,2 procent, av denna sjukdom. 1845 var dödstalet 6,8 procent och 1846 var det 11,4 procent. Vid andra avdelningen på samma sjukhus var dödstalen i barnsängsfeber 2,3, 2,0 respektive 2,7 procent för motsvarande år. Det var således en betydande skillnad i dödlighet i samma sjukdom vid två olika avdelningar vid samma sjukhus.

För att kunna minska antalet dödsfall i barnsängsfeber är det nödvändigt att veta orsakerna till sjukdomen, dvs. man måste ha en teori som förklarar orsakerna. En av läkarna vid första avdelningen, Ignaz Semmelweis, ägnade sig åt studier av denna mystiska sjukdom.

Ett förslag till förklaring av sjukdomen rörde procedurerna då sista smörjelsen skulle ges till en döende kvinna. Vid första avdelningen måste prästen passera genom fem salar innan han nådde sjukrummet. Prästens närvaro, som föregicks av att en assistent ringde i en klocka, antogs ha en skrämmande och försvagande effekt på kvinnorna och därmed göra dem mera mottagliga för sjukdomen. På andra avdelningen kunde prästen gå direkt till sjukrummet, utan att några patienter såg honom. En teori var alltså

“Prästens närvaro skrämmar kvinnorna så att de blir mer mottagliga för barnsängsfeber”.

För att pröva denna teori övertalades prästen att gå en omväg och utan att någon ringde i klockan. Därigenom kunde prästen nå sjukrummet i tysthet och utan att observeras av avdelningens övriga kvinnor. Teorin prövades genom att undersöka om dödstalen på första avdelningen minskade. Om man observerade en oförändrad dödlighet efter förändringen av rutinerna vid sista smörjelsen måste teorin om prästens närvaro förkastas. Det visade sig att dödligheten inte minskade efter att rutinerna vid sista smörjelsen ändrats, varför Semmelweis förklarade att den föreslagna teorin inte gav en god beskrivning av verkligheten. Teorin om att prästens närvaro orsakade sjukdomen måste därför förkastas.

Flera andra teorier föreslogs men förkastades efter att motsvarande provningar inte visade på någon minskning av dödstalen när teorierna prövades. En rad dramatiska omständigheter ledde så småningom Semmelweis till att anta att dåligt tvättade händer efter en dissektion är en möjlig förklaring till sjukdomen. Läkarna vid första avdelningen brukade nämligen regelbundet utföra

dissektioner och sedan endast ytligt tvätta sina händer innan de undersökte kvinnorna. Vid andra avdelningen förekom inga dissektioner. Semmelweis föreslog därför teorin

“Ett likämne orsakar barnsängsfeber”.

Om det är en bra teori, så kan barnsängsfeber förebyggas genom att man på kemisk väg förstör det likämne man har fått på händerna vid dissektionerna. Man lät därför alla läkarna tvätta sina händer i en klorkalklösning innan de utförde en undersökning. Dödligheten i barnsängsfeber började då snabbt sjunka och år 1848 var den 1,27 procent på första avdelningen, jämfört med 1,33 på andra avdelningen.

Teorin om likämnen beskrev sålunda verkligheten på ett relevant sätt och betraktades som bra. Senare har teorin modifierats och nyanserats. Den teori som är aktuell idag är att det är mikroorganismer, snarare än ett “likämne”, som förorsakar sjukdomen. Vi vet också att farliga mikroorganismer finns på andra ställen än i lik, vilket innebär att det är nödvändigt att tvätta händerna även efter andra aktiviteter än dissektioner.

Denna historia ger oss ett intressant exempel på prövning av en teori. En prövning innebär att vi jämför teorin med verkligheten. Om teorin och verkligheten inte överensstämmer (som teorin om prästernas skrämseleffekt) förkastas teorin. Överensstämmer teorin och verkligheten någorlunda bra (som för teorin om ett likämne) håller vi tills vidare fast vid teorin. Vid ett senare tillfälle, då vi har fått mera kunskap om problemet, kan vi formulera en bättre och kanske mer nyanserad teori som då får ersätta den gamla.

Semmelweis använde sig av fler teorier som var och en kunde anses relativt rimliga utifrån de föreställningar om omvärlden som förekom i mitten av 1800-talet. I syfte att pröva teorierna gjorde Semmelweis observationer vilka ledde till att teorierna måste överges, utom den teori som bygger på förekomsten av ett likämne. Efterhand som kunskaperna inom mikrobiologin har utvecklats har “likämnesteorin” fått ge plats åt teorier baserade på existensen av bakterier.

4.3 Den hypotetisk-deduktiva metoden

I den hypotetisk-deduktiva metoden används logiska argument för att pröva en eller flera hypoteser. Låt oss först betrakta ett exempel från deckarvärlden där det gäller att pröva hypotesen att den mystiske betjänten mr B knivmördade greven sir G. Om mr B mördade sir G måste mr B ha befunnit sig på mordplatsen vid tiden för brottet. Två fall kan nu inträffa, antingen befann sig mr B på mordplatsen eller så gjorde han det inte vid den aktuella tiden. Om mr B kan uppvisa ett trovärdigt alibi att han inte befann sig på mordplatsen kan han frias från misstankarna. Om han däremot befann sig på mordplatsen ger detta förvisso ett indicium, men det bevisar på intet sätt mr B's skuld i mordfallet. Det kan fortfarande vara den om möjligt ännu mer mystiske chauffören mr C som är sir G's baneman.

I det logiska schemat som karaktäriserar den hypotetisk-deduktiva metoden identifierar vi först en logisk implikation: "Om mr B är mördaren, så måste mr B ha befunnit sig på mordplatsen vid brottstidpunkten". Låter vi H beteckna hypotesen att mr B är mördaren och I beteckna att mr B befann sig på mordplatsen vid brottstidpunkten, kan vi skriva den logiska implikationen som $H \rightarrow I$.

Vi identifierar sedan en av två möjliga observationssatser, nämligen "mr B befann sig vid mordplatsen vid brottstillfället" eller "mr B befann sig inte på mordplatsen vid brottstillfället". Med logiska symboler betecknar vi dessa satser med I respektive $\bar{I}$.

De möjligheter till slutsatser vi har beror nu på vilken observationssats vi har. Om vi har observationssatsen $\bar{I}$ (mr B befann sig inte på mordplatsen vid brottstidpunkten) kan vi tillämpa *modus tollens* regeln från logiken och dra slutsatsen $\bar{H}$, dvs att hypotesen måste vara falsk (mr B är ej mördaren). Den logiska strukturen kan då sammanfattas i följande schema

$H \rightarrow I$ Implikation

$\bar{I}$ Observation

$\bar{H}$ Slutsats

Om vi å andra sidan har observationssatsen I (mr B befann sig på mord-

platsen vid brottstillfället) kan vi tyvärr inte säga något bestämt om H . Att observera I innebär antingen att H är sann, eller att det finns någon annan hypotes, säg H' (t ex "sir G ropade till sig betjänten mr B när han såg sig hotad av mördaren"), som är sann, där H' har samma implikation som H , dvs $H' \rightarrow I$. I detta fall kan vi sålunda inte avgöra om hypotesen H är sann eller falsk.

All slutsatsdragning sker under förutsättningen att alla andra antaganden i modellen är korrekta. När t. ex. Semmelweis prövade en hypotes om att ett likämne orsakar barnsängsfeber förutsatte han bl. a. att likämnet försvinner om man tvättar händerna i en klorkalklösning. De antaganden som förutsätts i prövningen men inte explicit provas brukar kallas hjälphypoteser. I de situationer vi betraktar här, är modellen i övrigt de hjälphypoteser som kommer till användning. Vi betecknar hjälphypoteserna med M . En mer fullständig logisk struktur för prövningsproblemet kan därmed sammanfattas i följande schema:

$$H \wedge M \rightarrow I$$

$$\bar{I}$$

$$\bar{H} \vee \bar{M}$$

Först har vi en logisk implikation, $H \wedge M \rightarrow I$. Symbolen $\wedge$ betyder här "och", så att implikationen säger att om både hypotesen H (likämnet orsakar barnsängsfeber) och hjälphypotesen M (likämnet försvinner om man tvättar händerna i en klorkalklösning) är sanna så inträffar I (proportionen kvinnor med barnsängsfeber minskar). Den därpå följande observationssatsen är densamma som tidigare, I eller $\bar{I}$. Om observationssatsen I inträffar kan vi på grund av de argument vi har redogjort för tidigare, inte dra några slutsatser. Att observera en minskning av dödstalen bevisar inte att det finns ett likämne som orsakar barnsängsfeber, det kan ju finnas något annat, t.ex. bakterier, som orsakar barnsängsfeber. Vi kan alltså inte med logisk giltighet acceptera en hypotes. Om observationssatsen $\bar{I}$ inträffar (som i schemat ovan) blir slutsatsen $\bar{H} \vee \bar{M}$. Här betyder symbolen $\vee$ "logiskt eller" vilket ibland beskrivs som "och/eller". Slutsatsen innebär alltså att antingen är hypotesen H falsk (likämnet orsakar barnsängsfeber) eller så är hjälphypotesen M falsk (likämnet försvinner inte då man tvättar händerna

i en klorkalklösning) eller så är både H och M falska. Vi kan alltså med logisk giltighet förkasta en hypotes förutsatt att eventuella hjälphypoteser är sanna.

4.4 Prövning av modeller av slumpmässiga försök

När vi arbetar med sannolikhetsmodeller uppstår ytterligare en komplikation i samband med den hypotetisk-deduktiva metoden. De händelser vi observerar är slumpmässiga och inträffar inte med säkerhet, utan med en viss sannolikhet. Antag att t. ex. hypotesen H och händelsen I har formulerats så att I inträffar med mycket stor sannolikhet om H är sann, och följaktligen att $\bar{I}$ inträffar med mycket liten sannolikhet om H är sann. Låter vi, liksom tidigare, M beteckna de hjälpantaganden vi behöver göra, blir det logiska schemat

$$\begin{array}{c} H \wedge M \xrightarrow{p} I \\ \bar{I} \\ \hline [p] \\ \bar{H} \vee \bar{M} \end{array}$$

där $\xrightarrow{p}$ indikerar att I följer med stor sannolikhet om $H \wedge M$ är sann och $[p]$ indikerar att antingen har något osannolikt inträffat (dvs $\bar{I}$ faktiskt har inträffat trots att $H \wedge M$ är sann) eller så är hypotesen $H \wedge M$ falsk. Eftersom osannolika händelser inträffar sällan kan vi med "praktisk säkerhet" dra slutsatsen $\bar{H} \vee \bar{M}$. Förutsatt att hjälphypotesen M är sann förkastar vi med "praktisk säkerhet" hypotesen H .

Låt oss ta ett exempel. När ett barn föds blir det antingen en pojke eller en flicka. Betrakta det slumpmässiga försöket "ett barn föds och vi noterar barnets kön". En modell av detta försök är

$$\begin{aligned} \Omega &= \{\text{pojke, flicka}\} \\ P(\text{pojke}) &= 0,5, P(\text{flicka}) = 0,5 \end{aligned}$$

dvs det är lika troligt att det blir en pojke som en flicka.

I föregående avsnitt sa vi att prövning av en modell innebär att jämföra modellen med verkligheten. För att studera verkligheten i det här fallet kan vi skaffa oss uppgifter om könet på ett antal nyfödda barn.

Antag att 834 födslar har ägt rum i en kommun under ett år. Även om det är lika sannolikt att det blir en pojke som en flicka då ett barn föds kommer det inte att födas precis lika många pojkar som flickor. I det här fallet kräver vi alltså inte att det måste födas precis 417 pojkar och 417 flickor under ett år för att vi skall tycka att modellen är bra. Vi tycker heller inte att det är särskilt konstigt om 420 av de 834 födda barnen är pojkar. Men om det föds 450 pojkar börjar vi nog bli lite misstänksamma mot modellen.

Någonstans mellan 417 och 450 pojkar av 834 födda barn finns alltså en gräns för vad vi anser vara rimligt om modellen ger en bra beskrivning av verkligheten. I de närmast följande kapitlen skall vi utvidga våra kunskaper i sannolikhetssteori så att det blir möjligt för oss att med hjälp av modellen beräkna hur sannolikt det är att så och så många barn är pojkar av ett visst antal nyfödda barn. Om det visar sig att det som har inträffat var, enligt modellen, mycket otroligt säger vi att antingen har något mycket otroligt inträffat eller så är vår modell felaktig.

Redan på 1700-talet intresserade man sig för modeller av könet hos nyfödda barn. 1720 studerade Arbuthnot just den modell som vi har satt upp här. Han sa att om det är lika troligt att det föds en pojke som en flicka så måste det vara lika troligt att det föds fler pojkar än flickor sammanlagt under ett år, som att det föds fler flickor än pojkar under samma år. Arbuthnots hypotes kan alltså formaliseras som

H : Det är lika troligt att det blir ett "pojkar" som ett "flickor".

där "pojkar" är år det föds fler pojkar än flickor och "flickor" är år det föds fler flickor än pojkar. För att pröva modellen kontrollerade han födelselängder från London. Han hade tillgång till födelselängder från 82 år. Händelsen

I : Bland de observerade 82 åren är det ungefär lika många "pojkar" som "flickor"

är då en händelse som inträffar med rimligt stor sannolikhet om H är sann.

Den exakta sannolikheten för att I skall inträffa beror givetvis på vad vi menar med "ungefär lika många".

Det visade sig att under alla de 82 observerade åren hade det fötts fler pojkar än flickor, dvs. alla observerade år var "pojksår". Arbuthnot ansåg, med rätta, att det var alltför många "pojksår" för att tro att modellen ger en god beskrivning av verkligheten. Med hjälp av sannolikhetsteorin som presenteras i de kommande kapitlen kan man beräkna att om modellen är sann, så är sannolikheten c:a $2.068 \cdot 10^{-25}$ att det blir "pojksår" 82 år i följd. Motsvarande sannolikhet för händelsen "41 pojksår och 41 flickår" är c:a 0.0878. Jämför vi sannolikheterna för dessa två händelser genom att bilda kvoten mellan sannolikheterna, får vi

$$L \approx \frac{2.068 \cdot 10^{-25}}{0.0878} \approx 2.35 \cdot 10^{-24},$$

dvs. händelsen att vi skulle få 82 "pojksår" i följd är en mycket otrolig händelse om vår modell är sann. Alltså har antingen något mycket otroligt inträffat eller så är vår modell inte bra. Med "praktisk säkerhet" drar vi slutsatsen att det inte är lika troligt att det föds en pojke som en flicka då ett barn föds.

4.5 Kunskapsbildningsprocessen

Vi har tidigare i detta kapitel allmänt diskuterat kunskapsbildning och några viktiga begrepp i samband med kunskapsbildning. Vi skall nu använda våra kunskaper i sannolikhetsteori för att mer systematiskt diskutera hur modeller av slumpmässiga försök kan användas i kunskapsbildningsprocessen.

Det första steget i kunskapsbildningsprocessen är att sammanfatta våra teorier till en modell av det vi studerar. Modellen formuleras här i sannolikhetsteoretiska termer. Med hjälp av modellen kan vi sedan göra förutsägelser om framtida händelser.

Nästa steg i kunskapsbildningsprocessen är att jämföra modellen med verkligheten. Detta går till så att vi gör observationer på verkligheten och jämför de faktiskt inträffade händelserna med vad som har förutsagts av modellen. Om observationerna står i konflikt med modellen, dvs. modellen har förutsagt att det observerade är mycket otroligt, kan detta ha två orsaker:

antingen har något mycket otroligt faktiskt inträffat eller så är vår modell fel och har därför gett felaktiga förutsägelser. Med "praktisk säkerhet" förkastar vi då vår modell, dvs. vi tror att den är dålig. Följaktligen blir nästa steg att revidera teorin om detta är nödvändigt och att revidera modellen så att den bättre stämmer överens med verkligheten.

Om observationerna inte står i konflikt med modellens förutsägelser, behöver detta inte innebära att vår modell är en bra beskrivning av verkligheten. Det kan ha varit tillfälligheter som gjorde att modellens förutsägelser och observationerna överensstämmer. Det kan t.ex. vara så att modellen endast stämmer under de betingelser som rådde just då observationerna gjordes etc. Om förutsägelser och observationer överensstämmer säger vi att vi inte kan förkasta modellen, vilket skall tolkas som att vi hittills inte har observerat något som tyder på att modellen är dålig. En viktig förutsättning för att använda en modell för att göra förutsägelser (prognoser), förklaringar etc, är att modellen inte är förkastad. Vi säger att modellen skall ha ett empiriskt stöd. En annan förutsättning är att modellen har ett teoretiskt stöd, dvs. den är baserad på väl etablerade teorier om det studerade problemområdet. Då vi observerar något som står i konflikt med modellen måste den förkastas och bör ej användas för att göra förutsägelser, förklaringar etc. Vi säger att en sådan modell saknar eller har ett svagt empiriskt stöd.

En viktig slutsats från detta resonemang är att en analys av data endast kan föreslå att vi förkastar en modell. Det är en felaktig uppfattning att observationer kan "bevisa" att en modell är sann. Uttryck som "Det är vetenskapligt bevisat att..." eller "Det är statistiskt bevisat att..." saknar grund i empiriska vetenskaper.

Vi har här delat in kunskapsbildningsprocessen i fem steg:

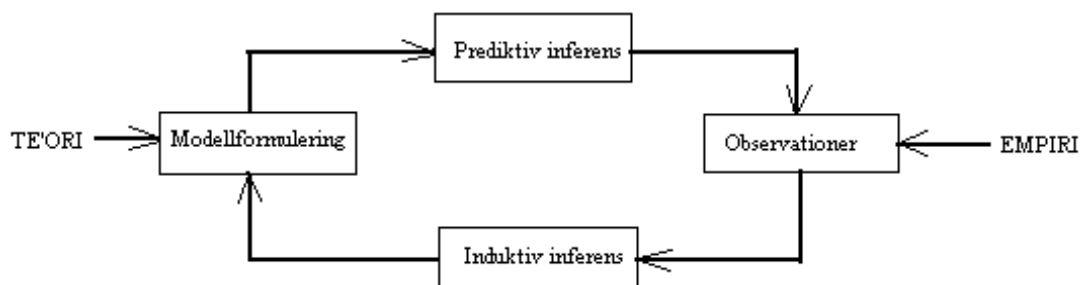
- modellformulering
- prediktiv inferens
- empiriska observationer
- induktiv inferens
- revidering

Med prediktiv inferens menas här att konsekvenser av modellen bestäms.

Vanligen innebär det att sannolikheter för händelser beräknas under förutsättning att modellen är en bra beskrivning av verkligheten. För att markera att det är verkligheten vi observerar har vi kallat observationerna empiriska observationer. Det steg där förutsägelseorna jämförs med empiriska observationer har vi kallat induktiv inferens. Induktion betyder att man från vissa observationer försöker formulera mer allmänna lagar. I vårt fall betyder alltså induktion att från att ha observerat utfall från slumpmässiga försök, vill vi precisera en given modell eller avgöra om en modell är en bra beskrivning av verkligheten.

Det sista steget, revidering, utgör samtidigt det första steget i nästa cykel, dvs. den nya modell vi får efter revideringen ger nya implikationer och måste på nytt provas med nya empiriska observationer.

Kunskapsbildningsprocessen utgör således ett kretslopp såsom har indikerats i figur 4.2. En konsekvens av detta är att vi aldrig kan säga att en modell eller en teori är sann eller att den har gett det slutliga svaret. De teorier och modeller som vi har idag kommer förmodligen att i framtiden förkastas till förmån för nya, bättre teorier och modeller. Framstegen består alltså inte i att vi uppnår sanningen, utan att vi (kanske) kommer närmare den.



Figur 4.2 Kunskapsbildningsprocessen.

4.6 En undersöknings olika steg

I figur 4.3 visas schematiskt de viktigaste stegen i en undersökning. De fem stegen är problemformulering, planering, datainsamling, analys och rapportering.

STEG 1: PROBLEMFORMULERING

STEG 2: PLANERING

STEG 3: DATAINSAMLING

STEG 4: ANALYS

STEG 5: RAPPORTERING

Figur 4.3 De viktigaste stegen i en undersökning

Att formulera problemet och planera en undersökning är de viktigaste stegen. En dålig problemformulering kan leda till en bra lösning på fel problem, något som kan kosta mycket möda och sällan är användbart. Problemformuleringen behandlas närmare i avsnitt 4.7.

En dålig planering kan innebära att kostnaderna för undersökningen blir avsevärt större och till och med att undersökningsresultaten blir helt spolerade. Vid planeringen bestämmer man sig för vilken typ av data man skall samla in, om undersökningen skall vara en total- eller en urvalsundersökning, hur data skall samlas in, dvs man väljer datainsamlingsmetod och hur data skall analyseras. Datainsamlingen kan ske med t. ex. olika mekaniska anordningar (för att mäta fysikaliska storheter som längd och vikt), psykologiska test (för att mäta t. ex. intelligens och personlighet), olika register med tidigare insamlade data (dataregister) och frågeformulär för brevenkät, telefonintervju eller besöksintervju. Anordningar för att utföra mätningar kallar vi för mätinstrument.

Att skaffa sig ett urval kan ske på många olika sätt. Förutom att själv välja ut vilka individer som skall ingå i urvalet, kan tekniker som bygger på slumpmässiga försök användas. Att välja urvalsteknik är en viktig del i planeringssteget. Lika viktigt är det att välja analysteknik, trots att analysen sker först i det fjärde steget. Olika analystekniker ställer nämligen olika krav

på data och datainsamlingen, eller tvärt om: olika datainsamlingsmetoder ger olika förutsättningar för analys.

Det tredje steget, datainsamling, innebär att observationerna samlas in. För urvalsundersökningar dras urvalet. Skall man själv utföra mätningarna sker det i detta steg. Datainsamlingen är det steg som vanligen tar den längsta tiden och är mest kostsamt. För att nedbrinka tid och kostnader för undersökningen är det därför viktigt att ha en god organisation vid datainsamlingen.

I analyssteget sammanfattas de gjorda observationerna och jämförs med de uppställda teorierna. Detta sker med hjälp av de inferensmetoder som valdes i planeringssteget. Några av de vanligaste inferensmetoderna som därvid kommer till användning behandlas i kapitlen 10-19.

Den indelning vi nu har diskuterat anger undersökningsarbetets logiska struktur. I praktiken kan mycket väl olika steg pågå samtidigt och man kan även behöva gå tillbaka till ett tidigare steg flera gånger under arbetet.

För att en undersökning skall bli framgångsrik är det viktigt att undersökningens alla steg kan genomföras väl. För att lyckas med detta måste undersökaren vara förtrogen med både statistikteori och med det problemområde man undersöker.

4.7 Problemformulering

Utgångspunkten för en undersökning är oftast ett antal frågeställningar med anknytning till ett visst problemområde. Det är inte ovanligt att dessa frågeställningar är relativt diffusa. I problemformuleringsfasen försöker man därför precisera formuleringen av frågeställningarna. Allmänt gäller att ju mer precisa frågeställningarna formuleras desto större möjligheter har man att ge precisa svar. Vagt formulerade frågeställningar leder ofta till vaga och oprecisa slutsatser. Den resulterande problemformuleringen i denna fas utgör utgångspunkten för alla övriga faser i undersökningen och är ur den synvinkeln undersökningens viktigaste fas.

Ett undersökningsproblem skall ställas utifrån ett bestämt syfte och på ett så precist sätt att det går att belysa med hjälp av de metoder vi har till vårt förfogande. En metod att formulera problem går ut på att först besvara

frågorna vad, vilka, var, hur, när och varför. När dessa frågor är besvarade på ett tillfredsställande sätt försöker man formulera undersökningsproblemet i en enda mening. Följande exempel illustrerar en tillämpning av denna metod för problemformulering.

Exempel 4.1

Antag vi har frågeställningen “Finns någon löneklyfta mellan män och kvinnor?” Vi vill nu bearbeta frågeställningen så att den kan bilda utgångspunkten för en undersökning. Vi försöker först besvara följande frågor:

Vad är löneklyfta (vad menas med begreppet)?

Vilka skall undersökas (någon speciell yrkesgrupp, branch, utbildningsgrupp etc)?

Var skall vi genomföra undersökningen (i hela riket, i någon speciell region eller en viss arbetsplats)?

Hur? Är vi intresserade av att beskriva begreppet eller söker vi även förklaringar? Vilka förklaringar kan i så fall finnas?

När? Vilken tidsperiod vill vi undersöka?

Varför vill vi undersöka just det här? Vad är syftet med undersökningen? Motivera valet av undersökningsproblem.

Ett undersökningsproblem skulle kunna formuleras som “Får män och kvinnor, anställda vid företaget xxx olika lön för samma arbete enligt nu gällande lönesättning? ■

Vi låter problemformuleringsfasen även innefatta en modellformulering och en populationsavgränsning. Modellformuleringen syftar till att skapa en (förenklad) bild av den del av verkligheten vi vill studera. Modellen utformas så att den baserar sig på och sammanfattar vår teoretiska kunskap om problemområdet. Därmed består modellen av ett antal teoretiska begrepp och en beskrivning av hur de teoretiska begreppen är relaterade till varandra. För att kunna besvara undersökningens frågeställningar måste de teoretiska begreppen operationaliseras, dvs. översättas till konkreta mätbara begrepp.

I statistiken kallas detta också för att man definierar undersökningens variabler.

En annan viktig del av problemformuleringen är att avgränsa och definiera den mängd individer som undersökningen skall avse. I statistiken kallas detta för att man definierar undersökningens population. När man gör dessa definitioner ställs man ofta inför mycket svåra avvägningar mellan vad man vill undersöka och vad man kan undersöka.

Problemformuleringen sammanfattas schematiskt i figur 4.4 och illustreras i följande exempel.

PROBLEMFORMULERING

Precisera undersökningsproblemet

Formulera en modell

Definiera undersökningens variabler

Definiera undersökningens population

Figur 4.4 Problemformuleringens viktigaste beståndsdelar.

Exempel 4.1 (forts)

Vår kunskap om problemområdet säger oss att lönen för en viss person beror av en lång rad faktorer. De viktigaste är:

- tjänsteställning (högre tjänster medför högre lön),
- utbildning (högre utbildning medför högre lön),
- anställningstid (längre anställningstid medför högre lön),
- bransch (expansiva branscher ger ofta högre lön för att locka till sig en större mängd arbetskraft) och
- geografiskt område (i områden med "överhettad" arbetsmarknad konkurrerar olika företag om arbetskraften genom att erbjuda högre löner).

Med hjälp av våra teoretiska kunskaper har vi här infört olika teoretiska begrepp, tjänsteställning, utbildning etc, samt deras relation till begreppet

lön. När vi studerar löneskillnader mellan män och kvinnor är det sålunda viktigt att ta hänsyn till eventuella skillnader i tjänsteställning, utbildning etc. ■

För att kunna genomföra en jämförelse måste vi också operationalisera våra begrepp. Begreppet tjänsteställning kan t. ex. operationaliseras genom att ett antal tjänstekategorier definieras.

Att avgränsa undersökningspopulationen är relaterat till “vilka”-frågan. I exempel 4.1 har vi att välja mellan geografiska områden, branscher, företag och/eller delar av företag.