

Objekt i statistikproduktionen: roller och relationer

1. Inledning

Terminologin här följer kvalitetshandboken: *Kvalitet för den officiella statistiken – en handbok*. Ibland har en term synonymer och ibland finns ord med likartad betydelse. Det underlättar om alla på SCB använder samma terminologi. En beskrivningsmodell finns genom processkartan i Verksamhetsstödet. Dessförinnan fanns en beskrivningsmodell vid dokumentation (SCBDOK version 3.0, år 2001).

Exemplen nedan är medtagna för att underlätta läsningen. De är valda med ambitionen att vara illustrativa, realistiska och enkla, däremot inte typiska.

1.1. Aspekter på objekt

Objekt – där individ, hushåll, fastighet, juridisk enhet, arbetsställe och kommun är exempel på objekttyper – är på många sätt viktiga i statistikproduktionen och statistiken. Det gäller, för varje statistikproduktion, att hålla reda på

- den eller de **roller** som varje objekt har
- **relationer** mellan såväl roller som objekt.

Ibland behövs **flera led** av objekt för att "nä fram" till ett objekt som ska observeras eller lämna uppgifter.

Tidsaspekter är viktiga. Vissa objekt, som hushåll, är inte konstanta utan ändras över tid, t.ex. vilka individer som tillhör ett visst hushåll.

Relationer mellan objekt kan vara intressanta i sig, t.ex. anställningar (individer av en kommun) och transaktioner (mellan företag). Ibland är det inte självklart vad som är objekt och vad som är variabler knutna till ett objekt. Försäljning och priser avseende varor kan knytas till varorna, till de säljande företagen eller en kombination. En relation kan t.ex. vara att ett hushåll består av en eller flera individer.

I vissa undersökningar ställer en eller flera av punkterna ovan krav på information som måste finnas. Antalet objekttyper har, naturligt nog, betydelse för komplexiteten.

Objekt betraktas sällan fristående; det finns en mängd information om dem. Informationen förekommer ofta i form av variabler, t.ex. att ett objekt ingår i ett urval, att ett objekt är svarsbortfall eller att ett variabelvärde har ändrats i granskning eller tillförts i en maskinell imputering.



I allmänhet är objektmängder, som populationer och urval, viktiga att hålla reda på.

1.2. Utformning i korta drag

Utformningen är det första som görs i en statistisk undersökning och, åtminstone delvis, i en produktionsomgång.

Observationsobjekt och målobjekt är centrala. Ett observationsobjekt är ett objekt som det samlas in uppgifter om. Det finns olika metoder, t.ex. direktinsamling eller hämtning från ett register. Val av uppgiftskällor, en eller flera, är också centralt. Ett skattningsförfarande är steget till statistiska målstorheter från data som avser objekt – observationsobjekt eller målobjekt. Förfarandet beaktar osäkerhetskällor som urval, ramtäckning, bortfall och modellantaganden. Skattningsförfarandet kan använda modeller, och en del av det kan vara på objektsnivå, t.ex. vara en beräkning av målobjekt och målvariabler från observationsobjekt och observationsvariabler.

Intresseobjekten ingår i resonemang om användarbehov när målen sätts, och de ingår i kvalitetsdeklarationen. De diskuteras inte nedan.

Utformningen hanterar bland annat objekttyper, t.ex. individ, arbetsställe och fastighet. Vilken eller vilka objekttyper som ska vara observationsobjekt respektive målobjekt är inte alltid självklart. Valet ska beakta statistikens kvalitet, kostnader och uppgiftslämnarbörda – det gäller att fånga det som statistiken ska beskriva på ett sätt som är effektivt för SCB och som har låg uppgiftslämnarbörda. I avvägningarna ingår vilka objekt (och variabler) som uppgiftslämnarna har information om och vilka objekt (och variabler) som ska ingå i statistiken. Det kan även ingå att precisera en struktur för objekt och variabler; vad som är vad och hur de hänger ihop.

När det finns skillnader mellan observationsobjekt och målobjekt kan det vara bra för transparensen att ha båda sparade, med relevant variabelinformation.

Populationsavgränsningar ingår också i utformningen. Det behöver dock inte vara känt vilka objektmängderna blir. Exempelvis uppstår vissa observationsobjekt, som exporterade varor, i datainsamlingen.

1.3. Urvalsförfarande – inkluderar totalundersökningar

Urvalsundersökningar, med urval i ett eller flera led, används särskilt vid direktinsamling. Första ledet i urvalsdragningen är i allmänhet att *dra ett urval* – en delmängd – från den i ramförfarandet avgränsade ramen. De objekt som det ska samlas in uppgifter om, observationsobjekten, ingår ofta i något av dessa led. Vanligen är observationsobjekten kända och utpekade i förväg. I andra, mer sällsynta, fall "uppstår" observationsobjekten genom en beslutsregel, t.ex. hur ett företag ska välja varor att lämna prisuppgifter om.

En totalundersökning har ett motsvarande förfarande i ett eller flera led med *val av objekt(mängder)*, men utan att urval dras. Ramen är ett hjälpmedel för att, såvitt möjligt, nå målpopulationen. Information från mer än ett register kan ingå när valen görs.

Det är fördelar med gemensamma beskrivningar för urvals- och totalundersökningar, generellt och i undersökningsdokumentation. Det gäller t.ex. när båda delarna används i en och samma undersökning (exempel i nästa avsnitt). Benämningen urvalsförfarande används nedan även om totalundersökningar, och totalundersökningar inkluderar registerbaserade undersökningar.

Ramen är utgångspunkt. En undersökning kan ha en eller flera hanteringar av objekten i ramen, se vidare avsnitt 2.2.1.

1.4. Några exempel

Exemplen nedan visar på olika former och grader av komplexitet. Likartade exempel finns i den fortsatta texten.

Exempel 1

Observationsobjekt i en undersökning av individer kan vara individen *eller* hushållet som individen ingår i *eller* alla medlemmar i hushållet. En individ kan lämna uppgifter om sig själv och om hushållet.

Exempel 2

En undersökning kan ha direktinsamling på objektsnivå för vissa variabler och använda registerdata (från ett ställe) för andra variabler.

Exempel 3

En undersökning kan gruppera sina observationsobjekt och samla in data från olika källor, t.ex. använda administrativa data för små företag och direktinsamling för stora företag.

Exempel 4

Det kan finnas skillnader mellan produktionsomgångar, t.ex. kan produktionen för den tredje månaden i ett kvartal skilja sig från de två tidigare månaderna genom att administrativa data har blivit tillgängliga. Urval används då två månader av tre och sedan en registerbaserad totalundersökning. Målpopulationen och målvariablerna är desamma, men objekttyper, objektmängder och skattningsförfaranden är olika.

Exempel 5

Två undersökningar kan ha en gemensam direktinsamling.

- a. Det kan vara en undersökning som har en tilläggsundersökning för en delmängd av objekt.
- b. Det kan vara två undersökningar som görs samtidigt och där det är praktiskt att för gemensamma observationsobjekt ha båda undersökningarnas observationsvariabler i ett formulär.

2. Objekt – aspekter

Detta avsnitt beskriver de i avsnitt 1.1 listade aspekterna.

2.1. Objekttyper

Några exempel på objekttyper är individ, hushåll, juridisk enhet, kommun, fastighet, byggnad och skola. När statistik tas fram är det väsentligt att vara tydlig med vilken eller vilka objekttyper som används. Statistikanvändarna behöver veta vad som utmärker mål-objekten. I datainsamlingen måste det framgå för uppgiftslämnarna vilket eller vilka objekt som den efterfrågade informationen avser.

I individstatistiken dominerar objekttyperna individ och hushåll. Basregistret RTB innehåller i första hand individer men även viss information med anknytning till hushåll.

Det finns relationer mellan hushåll och individer. Ett hushåll består av en eller flera individer, medan en individ ingår i endast ett hushåll. Detta gäller vid varje tidpunkt. Vilket hushåll som en individ ingår i ändras däremot över tid. RTB har inte all information.

Ordet *företag* används för enkelhets skull som en gemensam benämning på flera objekttyper, t.ex. i ord som företagsstatistik och i beskrivningar som "små företag". Objekttypen anges i texten när den är väsentlig för förståelsen. I många fall är det viktigt att vara precis för att minska risken för missförstånd. Det gäller bland annat vid extern kommunikation och då inte minst med uppgiftslämnare.

Företagsstatistiken använder traditionellt termen enhet snarare än objekt. Företagsdatabasen FDB innehåller dels administrativa enheter, som arbetsställen (AE) och juridiska enheter (JE), dels statistiska enheter som företagsenhet (FE), verksamhetsenhet (VE) och lokal verksamhetsenhet (LVE). En FE består av en eller flera juridiska enheter, och den bedriver en eller flera verksamheter, på en eller flera platser. FE kallas även institutionell enhet, och den förknippas med ekonomiskt beslutsfattande och fullständiga räkenskaper. FE och VE används för institutionell respektive funktionell statistik. För den stora mängden företag råder ett-till-ett-förhållanden mellan de tre objekttyperna JE, FE och VE. Övriga FE har flera VE eller flera JE; endera eller båda delarna.

Ett hushåll med en enda individ kan ingå i både hushålls- och individundersökningar, med hushålls- respektive individvariabler. Ett företag för vilket det råder ett-till-ett-förhållanden mellan JE, FE och VE kan på samma sätt, beroende på situationen, ses som att det är ett JE, ett FE eller ett VE, med respektive variabler.

Det finns även tidsperspektiv att beakta: objekt föds, dör och förändras. Sådan information kommer till SCB, men ofta med fördröjning. Vissa förändringar blir fortare kända än andra. Det är inte bara enskilda objekt som ändras över tid utan även relationer mellan objekt,

t.ex. vilka individer som ingår i ett hushåll eller vilka JE i en FE. Kopplingarna mellan objekt i en sådan struktur är således föränderliga. De behöver hållas uppdaterade, så vitt möjligt.

Det förekommer vidare att ett företag självt vill dela upp sitt uppgiftslämnande så att olika delar av företaget lämnar uppgifter var för sig. Dessa tillkommande objekt är företagets konstruktion, och de kan avvika från de reguljära objekttyperna. SCB behöver hålla reda på dessa tillkommande objekt och vilket företag de tillhör.

Ett annat exempel med flera relationer mellan objekttyper är skola, klass och elev. Ytterligare ett sådant exempel är fastighet, byggnad och lägenhet.

2.2. Objekt – roller i statistikproduktionsprocessen

Detta avsnitt beskriver fyra viktiga roller som objekt kan ha.

- Ramelement
- Uppgiftskälla
- Observationsobjekt
- Målobjekt

Rollerna är formulerade för alla undersökningar: registerbaserade, urval med direktinsamling, flera objekttyper under insamlingen, flera objekttyper i skattningsförfarandet etc. Avsnitt 1.4 ger några exempel och fler följer nedan.

2.2.1. Ramelement

Ramelement är ett objekt i ramen. Det kan finnas en eller flera hanteringar av ramelementen, och varje hantering kan ha ett eller flera led.

Ramförfarandet och ramen är grundläggande i statistikproduktionen. Ramen är en förteckning över objekt som, efter avgränsning med hänsyn till målpopulationen, används i det fortsatta arbetet. Ramen har information som används för att identifiera, avgränsa och i anslutning till datainsamlingen nå uppgiftskällorna, observationsobjekten och, i förlängningen, målobjekten. Att flera undersökningar utgår från en och samma ram, t.ex. en viss version av FDB eller RTB, är ett hjälpmedel i samordning av undersökningar och statistik. Varje undersökning gör sin avgränsning, t.ex. vissa branscher eller ett åldersintervall.

Objekt i ramen kallas, för tydlighets skull, ramelement. Särskilt när det ingår flera led och typer av objekt är det viktigt att hålla isär å ena sidan ramen med dess objekt (ramelementen) och å andra sidan rampopulationen vars objekt har samma typ som målpopulationen. Termen ramobjekt undviks helt.

En undersöknings ram är startpunkt för en eller flera hanteringar. Det kan t.ex. vara direktinsamling och användning av administrativa data. Urvalsdragning kan ingå, i ett eller flera led, varvid vissa objekt

tillhör urvalet och andra inte. I en totalundersökning ingår alla objekt. För registerbaserad statistik används inte alltid termen ram, men det är samma princip: att i ett eller flera led från en förteckning över objekt komma till en objektmängd som har god överensstämmelse med målpopulationen. Ramen innehåller i allmänhet information om bakgrundsvariabler och underlag för kontaktvägar. Om det är praktiskt kan en undersökning använda flera ramar, t.ex. med delvis olika variabler.

När urval dras, i ett eller flera led, är det för att sedan på lämpligt sätt nå fram till observationsobjekten (och därefter målobjekt). En skillnad jämfört med totalundersökningar är att information om urvalssannolikheter (motsvarande) behövs när skattningar ska göras. Informationen används även i datainsamling och granskning.

Information om objekt kan tillkomma under statistikproduktionen, t.ex. att de dör eller förändras (avsnitt 2.1). Detta kan innebära att antalet hanteringar och "vägar" i statistikproduktionsprocessen ökar. Om ett urval används under en längre tid, till exempel, kan kompletteringar av urvalet behövas. Vid skattningstillfället finns ofta mer information att beakta än då ramen framställdes, t.ex. om objekt som har fötts eller om relationer som har ändrats.

I en undersökning med ett *cut-off*-förfarande sätts en storleksgräns, under vilken objekten (t.ex. företag eller kommuner) utesluts från direktinsamling; mer i avsnitt 2.3.5.

Vissa objekttyper finns inte i SCB-register. Det finns t.ex. inget register över varor, utan ramen för en undersökning om varor behöver ha en annan objekttyp, t.ex. företag.

2.2.2. Uppgiftskälla

Uppgiftskälla är ett objekt som lämnar uppgifter (om observationsobjekt). Det kan finnas ett eller flera led, och en eller flera objekttyper kan ingå. Det finns ett flertal termer som, beroende på situation, är synonymmer till eller preciseringar av uppgiftskälla.

En uppgiftskälla kan lämna uppgifter om flera observationsobjekt. Det är typiskt i registerbaserad statistik, då "all" information hämtas samtidigt. Även vid direktinsamling kan en uppgiftskälla lämna uppgifter om flera objekt, t.ex. ett företag om alla sina arbetsställen eller om en uppsättning varor.

Vid direktinsamling är termen rapporteringsobjekt vanligt förekommande för den logistiska hanteringen. Inloggningsuppgifter, avprickning och påminnelser kopplar då till rapporteringsobjekt. Flera blanketter som ska besvaras av en uppgiftslämnare kan bilda separata eller ett gemensamt rapporteringsobjekt. Även separata rapporteringsobjekt kan ha gemensamma inloggningsuppgifter, varvid blanketterna skickas in och prickas av var för sig. Ett företag

kan låta ett ombud sköta uppgiftslämnandet, t.ex. en redovisningskonsult. Det är dock fortfarande företaget som har det formella ansvaret (viktigt vid uppgiftsskyldighet). I en företagsundersökning finns det normalt kontaktpersoner som besvarar undersökningen.

Även termen indataleverantör används, i allmänhet för någon som levererar administrativa data.

I hushållsundersökningar är det en individ i hushållet som lämnar uppgifter; det kan vara individen som har dragits i urvalet eller den individ som är mest lämpad. Det förekommer i individundersökningar att en individ lämnar uppgifter om en annan individ. Då används termen *proxy* eller indirekt intervju.

Termen uppgiftslämnare används vid direktinsamling. I företagsundersökningar kan det vara kontaktpersonen.

2.2.3. Observationsobjekt

Observationsobjekt är ett objekt som det samlas in uppgifter om. Flera objekttyper kan vara observationsobjekt i en undersökning.

Det kan vara insamling från ett register eller direktinsamling. Vid direktinsamling måste det framgå tydligt för uppgiftslämnarna vilket eller vilka objekt som den efterfrågade informationen avser. Det gäller inte minst för företagsstatistik, där det finns många objekttyper. Uppgiftslämnarna är ofta mer bekanta med de administrativa enheterna än de statistiska; t.ex. kan JE användas i förklaringar.

Bortfall förekommer, varvid skattningsförfarandet får kompensera för det, t.ex. genom imputering eller justering av uppräkningsstal.

2.2.4. Målobjekt

Målobjekt är ett objekt som ingår i målpopulationen och de statistiska målstorheterna.

Målobjekten kan vara väl kända, t.ex. i en totalundersökning av kommuner.

Ett observationsobjekt som är avsett att vara ett målobjekt kan visa sig vara övertäckning – eller vara övertäckning utan att det upptäcks. Objekt som är undertäckning finns normalt inte i statistikproduktionsregister. Skattningsförfarandet kan kompensera för dem.

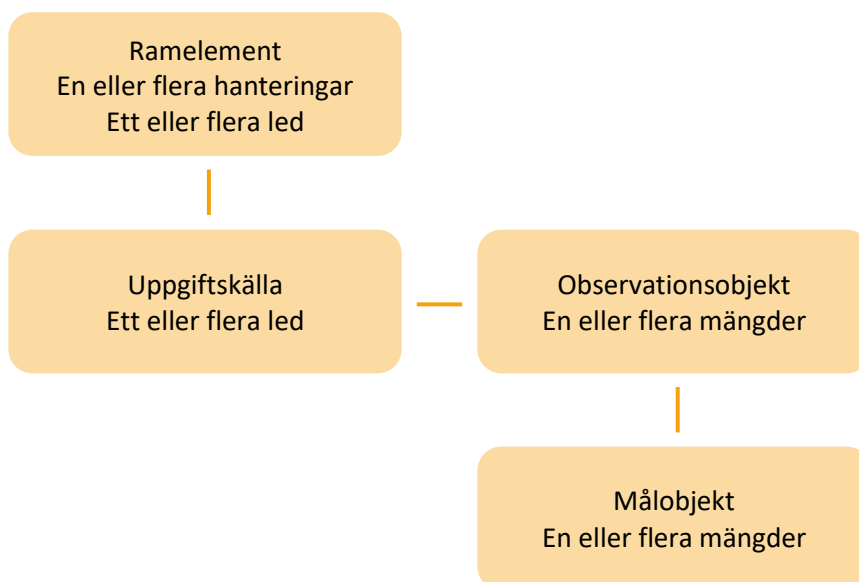
2.3. Relationer – mellan och inom roller

2.3.1. Inledning

Figuren nedan visar de fyra roller som beskrivs i avsnitt 2.2. De tre strecken indikerar en ordningsföljd i genomförandedelen av statistikproduktionsprocessen. De visar samtidigt relationer. Det finns relationer inte bara mellan roller, utan även inom roller. Det sistnämnda gäller vid förfaranden i flera led.

Inom varje roll kan det finnas flera mängder av objekt, och normalt finns det då relationer mellan objekten i dessa mängder.

Ramförfarandet pekar ut samtliga roller. Skattningsförfarandet handlar främst om de två sista rollerna, men det behöver information också om ramen och urvalsförfarandet. Det bör även, om möjligt, beakta att objekt förändras över tid (se avsnitt 2.1).



Avsnitten 2.3.2–2.3.4 beskriver de relationer som strecken indikerar. De är en fördjupning och utvidgning av informationen i avsnitt 2.2. Avsnitt 2.3.5 kommenterar skattningsförfarandet. Exempel finns även i avsnitt 2.4.

2.3.2. Ramelement och uppgiftskälla

Det är vanligt att det sista ledet i en hantering av ramelement innehåller objekt som överensstämmer med uppgiftskällorna eller observationsobjekten eller båda.

Exempel på ramelement och uppgiftskälla:

- En uppsättning JE med Skatteverket som uppgiftskälla för alla.
- En uppsättning VE från urvalsförfarandet, där varje VE är uppgiftskälla för sig själv.
- En uppsättning VE från urvalsförfarandet, med FE som uppgiftskälla för sina VE.

2.3.3. Uppgiftskälla och observationsobjekt

Uppgiftskällan lämnar uppgifter om observationsobjekt, ett eller flera. Dessa observationsobjekt kan vara av en eller flera objekttyper.

Exempel på uppgiftskälla och observationsobjekt:

- Det kan vara identitet mellan uppgiftskälla och observationsobjekt, t.ex. en individ som lämnar uppgifter om sig själv. Detta fall, med *samma* objekt, är vanligt förekommande.
- En individ kan lämna uppgifter om sitt hushåll eller om individer i detta hushåll (eller båda). Ett företag kan lämna uppgifter om sig självt eller om sina arbetsställen (eller båda).
- Ett företag kan lämna uppgifter om varor som företaget exporterar (okända i förväg).

2.3.4. Observationsobjekt och målobjekt

Observationsobjekt och målobjekt kan ha samma objekttyp. I allmänhet skiljer sig objektmängderna dock åt, t.ex. för att det finns brister i täckning eller för att det är en urvalsundersökning. Täckningsbrister kan, i vissa fall, ha blivit delvis eller helt kända vid skattningstillfället.

Observationsobjekt och målobjekt kan ha olika objekttyper. Det kan t.ex. vara FE som observationsobjekt och VE som målobjekt. Information fördelas i detta exempel ned från den högre FE-nivån till den lägre VE-nivån, typiskt med en fördelningsnyckel baserad på ett modellantagande om likhet. Det kan vara praktiskt att ha såväl FE som VE i det slutliga observationsregistret.

2.3.5. Kommentarer till skattningsförfarandet

Skattningsförfarandet fokuserar normalt direkt på de statistiska målstorheterna, dvs. på en aggregerad nivå. Det kan ibland vara praktiskt att genomföra skattningar med hjälp av modellantaganden på mikronivå, t.ex. vid imputering för partiellt bortfall och i vissa modellbaserade skattningar, som när variabelvärden för en FE "fördelas ned" till variabelvärden för dess verksamhetsenheter.

Vid urval behöver skattningsförfarandet information om urvalssannolikheter (motsvarande). Informationen förs via objekten i urvalet vidare till observations- och målobjekt. Den kan vara variabelberoende, t.ex. om flera urval förekommer i undersökningen.

Ett förfarande med *cut-off* (som ofta, oegentligt, kallas *cut-off-urval*) innebär att t.ex. små företag utesluts från direktinsamling. En skattning görs typiskt med hjälp av information i ramen om dessa företag och modellantaganden om likhet med en grupp något större företag.

2.4. Exempel med roller och relationer

Exemplen har likheter med tidigare givna exempel, som i avsnitt 1.4. De är delvis förenklingar. Syftet är att visa på huvuddragen i några olika situationer. Exemplen har både likheter och olikheter.

Exempel A: en undersökning med individ som målobjekt

Ramen kommer från RTB, och ett urval individer dras. Individernas ålder har samma avgränsning som målpopulationen. Individer i urvalet är uppgiftskälla i första ledet (den SCB vänder sig till) och observationsobjekt. Skattningsförfarandet avser målpopulationen med dess målobjekt (individer). Det använder urvalet och kompenserar för bortfall och för skillnader mellan rampopulation och målpopulation.

Exempel B: en undersökning med kosthushåll som målobjekt

Ramen kommer från RTB, och ett urval individer dras, med åldersbegränsning. Individer i urvalet är uppgiftskälla i första ledet för det kosthushåll som det ingår i; kosthushållet är observationsobjekt. En annan kosthushållsmedlem kan vara faktisk uppgiftskälla (andra ledet). Skattningsförfarandet utgår från urvalet och kompenserar för antalet valbara kosthushållsmedlemmar, för bortfall och för skillnader mellan rampopulation och målpopulation.

Exempel C: en registerbaserad undersökning om bostadshushåll

Ramen kommer från RTB, där det finns information om bostadshushåll: individer som är folkbokförda i samma lägenhet. Urvalsförfarandet tar med alla skapade bostadshushåll (den sista december). RTB är uppgiftskälla, noga räknat i sista ledet; dessförinnan har RTB hämtat information från Skatteverket och Lantmäteriet. Observationsobjekten är desamma som objekten i urvalet. Målobjekt är faktiska bostadshushåll. Skattningsförfarandet baseras på observationsobjekten.

Exempel D: en undersökning med VE som målobjekt

Ramen kommer från FDB med information om FE, VE och JE. Ett VE-urval dras, med *cut-off*. Företagsenheter som har flera VE ingår med de VE som tillhör målpopulationen (bland de objekt som totalundersöks). Alla så erhållna VE utgör urvalet. Även information från tidigare år och andra undersökningar kan vara till nytta i avgränsningar av objekt och populationer samt i urvalsförfarandet.

I direktinsamlingen är FE uppgiftskälla för sina VE (en eller flera). Det finns ofta kontaktpersoner på företaget sedan tidigare. Observationsobjekt är VE, ibland dock med undantag så som avsnitt 2.3.4 beskriver.

Information från ramen används för den modellbaserade skattningen för små företag (under gränsen för *cut-off*), där information från direktinsamlingen om utvecklingstal från föregående år ingår.