

## Undersökning: definitioner, egenskaper, hållpunkter

Huvudsyftet är att ge först mer övergripande och sedan mer specifika definitioner med anledning av pågående förändringar av statistikproduktionen.

### Definition av register och slutligt observationsregister

Ett *register* är en ordnad förteckning över mikrodata eller makrodata, se vidare nedan. Innehållet i ett register bör framgå av en fastställd avsikt.

För ett *slutligt observationsregister* (SOR) finns två möjligheter (precis nedan, i HP3):

- (1) Ett slutligt observationsregister för en *statistikprodukt* – det ska innehålla det datamaterial som behövs för att framställa den redovisade statistiken.
- (2) Ett slutligt observationsregister som utgörs av innehållet i ett *statistiskt register* – detta är utformat för ett eller flera av följande statistiska ändamål:
  - a. ligga till grund för registerbaserad statistik
  - b. vara ram för statistiska undersökningar
  - c. utgöra hjälpinformation t.ex. i gransknings- och skattningsförfaranden.

### Kommentarer

Register definieras ofta som att de ska vara fullständiga eller näst intill fullständiga förteckningar. Ändamålen a-c ovan illustrerar det synsättet. Det gäller inte för ett slutligt observationsregister till en urvalsundersökning.

### Definition av statistisk undersökning

I en *statistisk undersökning* ingår att fastställa informationsbehov, vanligen tillsammans med användare/uppdragsgivare, att designa undersökningen utifrån ändamål med hänsyn till kvalitetskrav och restriktioner (kostnader, uppgiftslämnarbörda, standarder, legala regler etc.), att utföra statistikproduktionsprocessen utifrån den gjorda designen samt att utvärdera slutproduktens kvalitet och produktionsprocess. Uppgifter på mikro- eller makronivå samlas in från interna eller externa datakällor. Uppgifterna bearbetas, varefter den framställda informationen tillgängliggörs eller redovisas internt eller externt. Slutresultatet är statistikvärden med slutligt observationsregister *eller* enbart ett slutligt observationsregister i form av ett statistiskt register. Undersökningen ingår normalt i ett område – som kan skilja sig från den officiella statistikens ämnesområden och statistikområden – och designen beaktar områdets samlade informationsbehov, tillsammans med andra undersökningar. Undersökningens målpopulation(er) är ofta fastställd(a) i förväg, men en målpopulation kan också definieras med utgångspunkt i en eller flera tillgängliga datamängder.

Ett statistiskt register kan ha flera registervarianter som skiljer sig åt vad gäller populationsavgränsningar, variabelinnehåll eller periodavgränsningar. En registervariant kan produceras i flera tidsversioner: realiseringar av en variant för olika referenstider. Vid SCB kallas dessa ofta registerversioner. Både registervarianter och deras tidsversioner kan ha flera framställningsversioner, t.ex. en preliminär och en slutlig framställningsversion. Om inte missförstånd kan uppstå, räcker det att säga version.

### **Kommentarer**

Beskrivningen av registervariant och, ännu mer, av versioner avviker från MetaPlus och även något från nuvarande DOKSTAR.

Vid SCB benämns en statistisk undersökning ofta enbart undersökning, vilket går bra om missförstånd inte uppstår. Däremot ska benämningen *statistisk* undersökning användas enbart när den producerade informationen är statistisk.

Vid översättning till engelska av termen *statistisk undersökning* ska det beaktas dels att när det gäller framställning av statistikvärden är motsvarande term i det europeiska statistiska systemet (ESS) *statistical process*, dels att begreppet *survey* är smalare. I informationsmodellen GSIM motsvaras statistisk undersökning av termen *statistical program*.

### **Definition av undersökning**

Möjligheterna och sätten att framställa statistik utökas. Den traditionella statistiska undersökningen – från ax till limpa i form av en statistikprodukt – ersätts alltmer av separata delar som används tillsammans. Det finns ett behov av att beskriva det arbetssättet med t.ex. datafångst och integration av olika typer av datakällor.

Undersökningsdatabasen (UDB) tillhandahåller en gemensam struktur för tillämpning av statistikproduktionsprocessen (processkartan) med dess omgångar. UDB är en grundförutsättning för hantering av metadata och kopplingar till gemensamma tjänster inom statistikproduktionsprocessen. En *undersökning* i UDB är antingen en statistisk undersökning enligt ovan eller utvalda fristående delar av statistikproduktionsprocessen. Såväl design som utvärdering ingår alltid.

En statistikprodukt skapas av en statistisk undersökning, som kan använda information från andra undersökningar. En statistisk undersökning kan skapa noll (ingen statistik), en eller flera statistikprodukter. En undersökning i UDB som inte utgör en statistisk undersökning skapar ingen statistikprodukt.

### **Kommentarer**

Även om aggregering till statistikvärden inte ingår vid framställning av ett statistiskt register kan processkartan över statistikproduktionsprocessen användas också för att beskriva registerframställningen.

Processkartan är lämplig också för att beskriva produktionen av ett *bearbetat observationsregister* (BOR i HP2 nedan) och av ett *situationsregister* som uppdateras löpande. Dessa register är inte slutliga observationsregister. En eller flera varianter av situationsregistret kan vara ett BOR. En eller flera andra varianter av situationsregistret kan vara ett SOR. I detta fall framställer en statistisk undersökning ett statistiskt register med ett eller flera av de ändamål som listas i definitionen.

## Egenskaper hos undersökningar

De fem huvudpunkterna nedan är exempel på egenskaper hos undersökningar. Egenskaperna är inte varandra uteslutande. Sådana egenskaper kan användas för att ytterligare beskriva olika typer av undersökningar.

- Vilka slags data som används (ofta flera), t.ex.
  - direktinsamlade data
  - register från andra myndigheter
  - data från webbsidor eller mobiltelefondata
  - data från andra undersökningar.
- Omfattningen av insamlade data, särskilt om urval görs (till skillnad från en totalundersökning) och då med information om urvalet, t.ex.
  - sannolikhetsurval eller inte
  - tidsaspekter på urvalet, som panel, samordning över tid.
- Delar av statistikproduktionsprocessen som är relevanta, t.ex.
  - insamling till bearbetat observationsregister (se vidare nedan)
  - integrering av flera bearbetade observationsregister med vidarebearbetningar till ett nytt bearbetat observationsregister
  - framställning av statistik från bearbetade observationsregister.
- Vad som framställs, t.ex.
  - bearbetat observationsregister
  - situationsregister
  - statistiskt register
  - slutligt observationsregister och statistikvärden.
- Undersökningens frekvens, t.ex.
  - varje år, kvartal eller månad (regelbundet återkommande)
  - intermittent
  - enbart en gång.

## Mikro- och makrodata med struktur

SCB har en strukturerad datalagring för mikrodata och statistikvärden.

### **Input/output och mikro/makro för undersökningar**

En undersökning samlar in mikrodata, statistikvärden eller andra makrodata (t.ex. aggregat från en objekttyp till en annan); i de två

senare fallen utgör de insamlade värdena undersökningens mikronivå. En undersökning framställer mikrodata eller statistikvärden.

### **Observationsregister**

Ett observationsregister med *mikrodata* är en datamängd vars innehåll ofta, schematiskt, kan ses som en eller flera tabeller där rader och kolumner innehåller objekt respektive variabler.

Ett observationsregister med *makrodata* (oftast statistikvärden) är uppbyggt av grupper av objekt som ofta är baserade på indelningsvariabler, statistiska mått och variabler som måtten appliceras på. Detta observationsregister kan härledas i ett eller flera led från observationsregister med mikrodata, där objekten och variablerna ingår.

### **Hållpunkter**

Hållpunkterna är i undersökningsdesignen utvalda avstämningspunkter i statistikproduktionsprocessen med väldefinierade och metadata-beskrivna datamängder. Dessa kan (åter)användas för datautbyte, inom och mellan undersökningar, i en automatiserad produktionsprocess och för utveckling av statistiken. Regler för datautbyte, registerutdrag, arkivering, gallring med mera finns och tillgodoses genom dokumentation av data i hållpunkterna, t.ex. hur de är designade eller konfigurerade. Kvalitetskrav kan finnas.

Hållpunkterna har vissa likheter med så kallade *steady states* vid CBS i Nederländerna och *quality gates* vid ABS i Australien.

### **HPO indata**

Indata enligt överenskommelse med extern part. De är inkommen handling och sparas i inkommande format eller motsvarande. De är dokumenterade med metadata.

### **HP1 tillrättalagda indata**

Indata som är ”inskick” har tillrättalagts. Format och värdemängder bör vara anpassade till SCB. De sparas så, av SCB ogranskade data.

### **HP2 bearbetat observationsregister (BOR)**

Ett bearbetat observationsregister är oftast ett observationsregister från en datainsamling (datafångst). Data är sammanställda, och de kan vara något bearbetade (t.ex. flaggade värden, skapade variabler). Det kan vara flera datamängder med olika strukturer. En annan möjlighet är att BOR framställs genom integrering av flera datamängder (ofta BOR) och vidarebearbetningar. Ett BOR är antingen en mellanpunkt i en undersökning eller slutpunkten för en undersökning som inte är en statistisk undersökning. Ett BOR är således inte ett SOR. Ett BOR används normalt av flera undersökningar. Varianter, tidsversioner och framställningsversioner kan finnas.

Kraven på dokumentation av BOR är mindre omfattande än för SOR.

### **HP3 slutligt observationsregister (SOR)**

Ett SOR framställs av en statistisk undersökning. Det finns två fall (som beskrevs redan i inledningen).

- (1) Ett slutligt observationsregister för en *statistikprodukt* ska innehålla det datamaterial som behövs för att framställa den redovisade statistiken. Detta datamaterial är ofta mikrodata med objekt och variabler i en eller flera tabeller (jämför observationsregister). Statistikvärden förekommer också – enbart eller i kombination med mikrodata. Det slutliga observationsregistret kan innehålla mer information än vad den publicerade (motsvarande) statistiken kräver, t.ex. för att även annan statistik ska kunna tas fram. Numerisk information som behövs i skattningsförfarandet (som parametervärden och uppräkningsstal) kan ingå i det slutliga observationsregistret eller lagras separat.
- (2) Ett slutligt observationsregister med innehållet i ett *statistiskt register*, som är utformat för att tillgodose ett eller flera av följande statistiska ändamål som har specificerats:
  - a. ligga till grund för registerbaserad statistik
  - b. vara ram för statistiska undersökningar
  - c. utgöra hjälpinformation t.ex. i gransknings- och skattningsförfaranden.

Detta register kan ha varianter, tidsversioner och framställningsversioner. I produkt databasen (PDB), där indelningen i produkter är delvis ekonomiskt motiverad, kan det statistiska registret ingå i en registerprodukt eller vara en del av en statistikprodukt.

Kraven på dokumentation är strikta. Dokumentationsmallarna är MetaPlus (krav för statistiken ges av HP5 nedan) respektive DOKSTAR.

### **HP4 statistikvärden**

Statistikvärden, det vill säga skattade statistiska (mål)storheter.

Kraven på dokumentation formuleras i dialog mellan leverantör och mottagare.

### **HP5 publicerade eller levererade mikro- eller makrodata**

Publicerade eller levererade statistikvärden (skattade statistiska målstorheter) eller levererade mikrodata (från HP3, avidentifierade). Av sekretesskäl är data eventuellt ytterligare bearbetade jämfört med HP4.

Kraven på dokumentation är strikta.