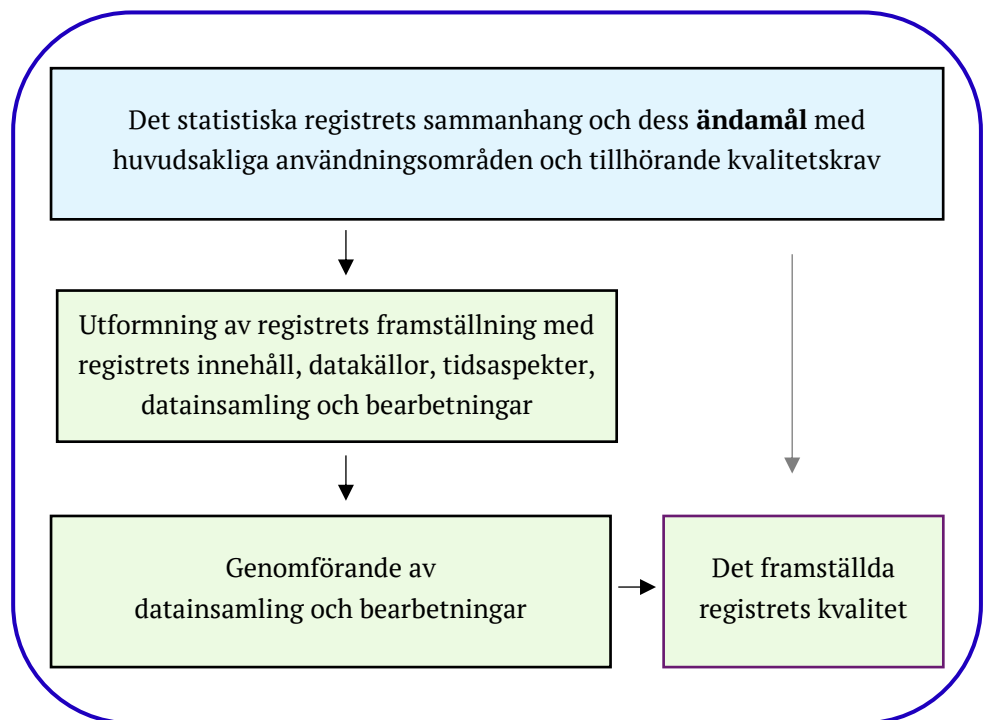


Det statistiska registrets framställning och kvalitet

– en handbok



Inledning

Under 2017 fattades ett flertal beslut vid SCB om dokumentationsformer riktade till användare. Besluten gällde dokumentation avseende statistik och observationsregister, och syftet var att få en större enhetlighet. Flera mallar uppdaterades, bl.a. med hänsyn till mallen för kvalitetsdeklaration av officiell statistik.

Mallen *Beskrivning av registret* (BaR) ersattes, i ett ac-beslut vid PMU, inför 2018 med mallen *Information om registret* (IoR). Under 2018 fattades ett gd-beslut om vilka register som ska dokumenteras enligt denna mall.

I början av 2019 beställde kvalitetschefen en handbok för register som skulle vara analog med *Den officiella statistikens kvalitet – en handbok*. En sådan handbok har nu tagits fram av en arbetsgrupp, som har bestått av Eva Elvers, Annika Gerner och Beatrice Kalnins, med stöd av en styrgrupp och en läsgrupp. Handboken omfattar det statistiska registrets framställning och kvalitet. Dessa aspekter motsvarar dokumentationen i *Statistikens framställning* (StaF) och i *Statistikens kvalitet* som är en del av kvalitetsdeklarationen av officiell statistik.

Handboken innehåller fyra fristående, men samverkande, delar.

- *Kvalitet och andra begrepp för statistiska register*
Syftet är att lägga en grund för de övriga delarna genom att på ett samlat och lättillgängligt sätt redogöra för kvalitet och andra grundläggande begrepp.
- *Anvisningar för det statistiska registrets framställning och kvalitet*
Syftet är att ge korta, övergripande, anvisningar som stöd inför användning av mallen.
- *Mall för det statistiska registrets framställning och kvalitet*
Syftet är att ange vad som enligt mallen ska ingå i en beskrivning av ett statistiskt register.
- *Ordlista*
Syftet är att ge korta förklaringar till centrala begrepp och termer; ett komplement till handbokens första del.

Kvalitet och andra begrepp för statistiska register

1	Inledning.....	2
1.1	Syfte	2
1.2	Ett register och ett statistiskt register	2
1.3	Ett registersystem	2
1.4	Statistikproduktionsprocessen	3
1.4.1	Statistikproduktion allmänt	3
1.4.2	Slutligt observationsregister	3
1.4.3	Framställning av ett statistiskt register	3
1.4.4	Registerbaserad statistikproduktion	4
1.5	Kvalitetsbegreppet för statistiska register	4
1.6	Användningsområden för statistiska register	5
2	Ett cykliskt arbetssätt.....	6
2.1	Registrets ändamål	6
2.2	Utformning.....	6
2.3	Kvalitet i förhållande till ändamål.....	7
2.4	Utvärdering	7
3	Statistiska registers innehåll.....	8
3.1	Objekt, population, variabler, referenstider	8
3.2	Intresse, mål och observation.....	9
3.3	Flera varianter och versioner.....	10
3.3.1	Registervariant och registerversion	10
3.3.2	Planerat och slutligt observationsregister	10
3.3.3	Preliminärt register.....	11
4	Datainsamling.....	11
4.1	Datakällor.....	11
4.2	Objekt och objekt mängder	12
4.3	Mätning (observation).....	12
4.4	Brister i leveranser	12
5	Bearbetningar	12
5.1	Kodning och imputering.....	13
5.2	Granskning	13
5.3	Modellbaserade beräkningar	13
6	De fem huvudkomponenterna	14
6.1	Relevans – ändamål och registrets innehåll	14
6.2	Tillförlitlighet.....	15
6.3	Aktualitet och punktlighet	16
6.4	Tillgänglighet och tydlighet; redovisning	16
6.5	Jämförbarhet och sammanvändbarhet	17

1 Inledning

1.1 Syfte

Den här delen av handboken har tre huvudsakliga syften:

- Motivera och redogöra för några begrepp som är centrala för statistiska register, särskilt kvalitet.
- Beskriva hur kvalitetsbegreppet ligger till grund för utformning och framställning samt användning av statistiska register.
- Lägga en grund för anvisningarna och mallen till *Det statistiska registrets framställning och kvalitet*.

1.2 Ett register och ett statistiskt register

Ett register kan, i det här sammanhanget, ses som en förteckning över objekt. Normalt innehåller registret även variabler. Objekten bör vara identifierbara, och vilka objekt som registret omfattar bör framgå av en fastställd avsikt med registret.

Ett statistiskt register är ett register som är utformat för statistiska ändamål, dvs. för att användas i statistikproduktion. Registret är en slutprodukt. Användare kan vara interna eller externa (eller båda). Det är normalt redan befintliga data som samlas in, t.ex. från en myndighet som har samlat in och bearbetat dessa administrativa data för eget bruk. Även användning av direktinsamlade data förekommer. Det statistiska registret kan, beroende på syfte, innehålla andra objekt och variabler än vad som finns i den eller de datakällor som används vid framställningen. De bearbetningar som görs kan vara omfattande. Inriktningen här är på statistiska register som tas fram återkommande.

1.3 Ett registersystem

SCB:s registersystem består av tre sfärer som har ett basregister var. Basregistren är *Registret över totalbefolkningen* (RTB), *Företagsdatabasen* (FDB) och *Fastighetsregistret* (FR). Därutöver finns en rad register (i allmänhet statistiska register) inom sfärerna med olika ämnesinriktningar. Som namnet säger utgör basregistren en grund för samordningen i systemet. Det finns vissa, standardiserade, variabler och populationer, och det finns utpekade ansvar för dessa, bland annat hos basregistren.

För att registren ska bilda ett system behövs bland annat möjligheter att koppla samman register, vilket görs med hjälp av identifierande variabler och kopplingsvariabler. Basregistren kan kopplas till varandra, och övriga register kan kopplas till ett eller flera av basregistren.

Samordning innebär att information i olika register kan användas tillsammans, på mikro- eller makronivå. En stark form av sam användning av flera register är *integrering*, när ett eller flera nya register utformas.

1.4 Statistikproduktionsprocessen

Ett sätt att beskriva en statistisk undersökning ges av processkartan nedan. För återkommande statistik brukar benämningen produktionsomgång användas, t.ex. finns det vanligtvis fyra produktionsomgångar per år för kvartalsstatistik.

Processerna 1-3 utgör förberedelser av en ny produktionsomgång och tillsammans kallas de för *Utforma*. Processerna 4-7 kan sammanfattas med ordet *Genomföra*. Process 8 innefattar en utvärdering av den genomförda produktionsomgången och återkoppling inför kommande statistikproduktion.

1.4.1 Statistikproduktion allmänt

Datainsamlingen kan, enkelt uttryckt, ske i form av direktinsamling (från enskilda uppgiftslämnare specifikt för undersökningens syfte) eller genom insamling av befintliga data eller vara en kombination av dessa. Befintliga data utgörs vanligtvis av administrativa data, data från ett (eller flera) statistiska register eller data från slutliga observationsregister från andra undersökningar.



Figur 1. Statistikproduktionsprocessen

1.4.2 Slutligt observationsregister

Ett slutligt observationsregister för en statistikprodukt innehåller det datamaterial som behövs för att framställa den redovisade statistiken. Det ska även kunna ligga till grund för framtida användning och återskapande av statistiken. Det fastställs efter granskning och röjande-kontroll av statistiken.

Ett observationsregister kan, schematiskt, ses som en eller flera tabeller där varje rad avser ett objekt av en given objekttyp och varje kolumn en viss variabel. Individ, hushåll, fastighet, juridisk enhet, arbetsställe och kommun är exempel på objekttyper. Beskrivningen här bortser från hur register lagras och uppdateras rent tekniskt, t.ex. i form av databaser. Inriktningen är på var i statistikproduktionsprocessen som register används och hur.

1.4.3 Framställning av ett statistiskt register

Processkartan kan användas även vid framställning av statistiska register. Aggregering till statistik ingår inte, men i övrigt ingår utformning och genomförande samt utvärdering och återkoppling. Slutprodukten är inte statistik utan ett slutligt observationsregister. Ett

statistiskt register kan ha flera varianter och versioner, vilket beskrivs nedan, främst i avsnitt 3.3.

1.4.4 Registerbaserad statistikproduktion

I registerbaserad statistikproduktion hämtas data från ett eller flera statistiska register. Ofta görs vissa bearbetningar av dessa data innan efterfrågad statistik tas fram. Statistiken kan t.ex. innehålla nya variabler jämfört med de ingående registren. Den framtagna statistiken grundas på uppgifter i ett slutligt observationsregister.

1.5 Kvalitetsbegreppet för statistiska register

Kvalitetsbegreppet har fem huvudkomponenter. Dessa delas in i underkomponenter och, i förekommande fall, i delkomponenter.

- Relevans
 - Ändamål och informationsbehov
 - Registrets ändamål
 - Registeranvändares informationsbehov
 - Registrets innehåll
 - Objekt och population
 - Variabler
 - Referenstider
- Tillförlitlighet
 - Tillförlitlighet totalt
 - Osäkerhetskällor
 - Täckning
 - Mätning
 - Bortfall
 - Bearbetningar
 - Modellantaganden
 - Preliminärt register jämfört med slutligt
- Aktualitet och punktlighet
 - Framställningstid
 - Frekvens
 - Punktlighet
- Tillgänglighet och tydlighet
 - Tillgång till registret
 - Informationsspridning
 - Dokumentation
- Jämförbarhet och sam användbarhet
 - Jämförbarhet över tid
 - Sam användbarhet med andra register

Kvalitetsbegreppet anger egenskaper och är deskriptivt. Det har ett användarperspektiv. Exempelvis räknas framställningstiden från en referenstidpunkt eller en referensperiods slut till att registret är framställt. Kvalitetsbegreppet har likheter med kvalitetsbegreppet för statistik, men det avser statistiska register, dvs. mikrodata. Därför finns även skillnader, t.ex. vad gäller innebörden av tillförlitlighet. Urval är inte en osäkerhetskälla för statistiska register, eftersom urvalsförfaranden är sällsynta.

Den som ansvarar för det statistiska registret beskriver egenskaperna (kvaliteten). Den som använder registret får därigenom ett stöd för sin tolkning, bearbetning och analys av registrets data, även för att bedöma registrets kvalitet för just den tänkta användningen.

1.6 Användningsområden för statistiska register

Statistiska register har fyra huvudsakliga *användningsområden*. Ett sådant register kan:

- Ligga till grund för registerbaserad statistik.
- Vara ram för statistiska undersökningar.
- Användas i skattningsförfaranden. (Denna typ av registerinformation kallas för hjälpinformation.)
- Användas för granskning av andra register eller direktinsamlade data (mikronivå).

En statistisk undersökning kan använda ett statistiskt register som den huvudsakliga datakällan eller kombinera registeruppgifter med annan information i statistikframställningen.

En ram till en statistisk undersökning används för att identifiera och avgränsa statistikens målobjekt och som källa för att, i vid bemärkelse, nå målobjekten i anslutning till datainsamlingen. Ofta är det en undersökning med direktinsamling för ett urval objekt, men det gäller även när statistiska register ingår i datainsamlingen. I undersökningar med urval används information i ramen för att utforma urvalet, t.ex. för indelning i strata.

I ett skattningsförfarande kan en eller flera variabler från ett statistiskt register användas som hjälpinformation för att minska osäkerheten. Ett exempel är kalibrering för att minska urvalsosäkerhet eller risken för snedvridning på grund av bortfall (ofta båda). Även om det inte är så vanligt kan ett statistiskt register användas som hjälpinformation i syfte att förbättra skattningsars tillförlitlighet också för registerbaserad statistik. Hjälpinformation utgörs då av andra variabler eller möjliggen nyare information.

Vid granskning av mikrodata kan ett statistiskt register vara användbart om det innehåller likartade variabler avseende samma eller en närliggande referenstid.

För varje statistiskt register behövs kunskap om hur det används. Det gäller i första hand användningsområden (ett eller flera) och i andra hand önskemål, behov och krav på registret, främst från de huvudsakliga användarna.

2 Ett cykliskt arbetssätt

2.1 Registrets ändamål

Registeranvändarnas behov och önskemål sammanställs och prioriteras. Även formella krav kan finnas, som förordningar. Det statistiska registrets *ändamål* formuleras, i många fall tillsammans med viktiga användare. Det kan t.ex. finnas användarråd, där frågor om registrets innehåll och andra kvalitetskomponenter kan avhandlas. Det fastställda ändamålet, eller syftet med det statistiska registret, visar vilka informationsbehov som ska tillgodoses. Det kan finnas informationsbehov som är bortprioriterade.

De fyra huvudsakliga användningsområdena beaktas. Basregistren innefattar alla de fyra användningsområdena. För många andra statistiska register är registerbaserad statistikproduktion det enda användningsområdet.

Ändamålet bestämmer kvalitetsmål för registret: för registrets innehåll, dess tillförlighet, dess aktualitet och punktlighet etc. Hur väl målen uppnås beror såväl på indata som på de bearbetningar som görs.

2.2 Utformning

En produktionsomgång för ett statistiskt register inleds med utformning, vilket i korthet innebär att utarbeta hur den kommande omgången av registerframställningen ska genomföras. Ambitionsnivån i utformningen varierar. Det kan t.ex. vara en kort avstämning av att omgången väsentligen liknar den förra, en mindre justering i arbetsrutiner eller en ändring som orsakas av en förändrad datakälla eller av ett nytt användarkrav som bör beaktas. Det kan alternativt vara aktuellt med större ändringar som kräver vidareutveckling av registret eller dess framställning (eller båda). Under utformningen (däribland designen) beaktas tidigare erfarenheter och kvalitetskrav.

Det kan vara motiverat och praktiskt att designa flera statistiska register samtidigt, särskilt om de har gemensamma datakällor eller om de används tillsammans. Det kan t.ex. gälla enhetliga variabeldefinitioner och metoder som bidrar till att registren bildar ett registersystem. Andra samordningsaspekter är datakällor, tidpunkter för uppdateringar och bearbetningsmetoder. Exempelvis bör det för data från Skatteverket som används i flera statistiska register finnas en initial, för SCB gemensam, hantering och inte lokala versioner.

2.3 Kvalitet i förhållande till ändamål

Vissa kvalitetskomponenter är förhållandevis enkla att precisera i förhållande till det statistiska registrets ändamål. Det kan t.ex. gälla registrets innehåll och aktualitet. Därmed inte sagt att kvalitetskraven är enkla att realisera. Andra kvalitetskomponenter kan vara svåra att precisera och att mäta. Det gäller ofta för tillförlitlighet. Det hindrar inte att det är väsentligt att försöka formulera krav och att försöka göra bedömningar eller uppskattningar.

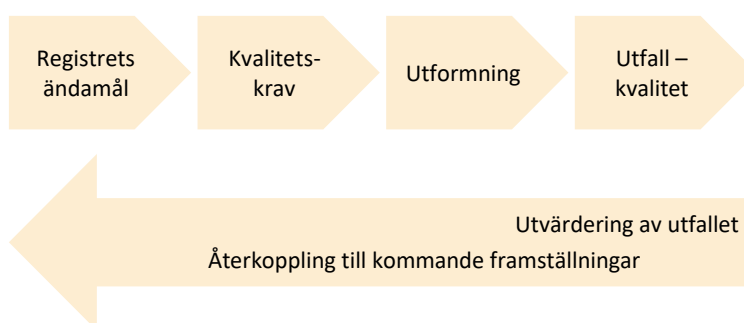
Krav kan ligga på mikro- eller makronivå, dvs. avse enskilda objekt eller aggregat. Om registret används för registerbaserad statistikproduktion är kvaliteten i den statistiken (en eller flera statistikprodukter) utgångspunkt för kvalitetskrav på registret. Det handlar om aggregat och inte om enskilda objekt. Vissa objekt kan dock ha större inflytande än andra.

Om registrets används som ram till en statistisk undersökning – ofta men inte nödvändigtvis med direktinsamling – är undersökningens ramförfarande en kravställare. Det kan t.ex. handla om populationsavgränsning, om att kunna dra ett ändamålsenligt urval och att nå uppgiftskällorna (uppgiftslämnarna). Exempelvis kan ett storleksmått vara betydelsefullt för urval till ekonomisk statistik.

Även de som använder ett statistiskt register som hjälpinformation i skattningsförfaranden eller vid granskning av mikrodata kan ha krav, t.ex. avseende tidsaspekter.

2.4 Utvärdering

Registerframställning följer, enligt statistikproduktionsprocessen, en cykel med utformning, genomförande och utvärdering.



Figur 2. Ett cykliskt arbetssätt

Stegen i Figur 2 visar ett cykliskt arbetssätt med betoning på kvalitet. Registrets ändamål, kvalitetskrav, utformning och utfallet av registerframställningen kan ses som fyra länkar i en kedja. Det är utfallet som är viktigast i utvärderingen. Den erhållna kvaliteten kan vara en annan än den som planerades i utformningen. Återkopplingen kan innebära

förslag på mindre eller större justeringar av t.ex. utformning eller arbetsrutiner. Det kan vara justeringar som har ett kortare eller längre tidsperspektiv.

Erfarenheter kan vara till nytta inte enbart för enskilda register utan även för andra, liknande, register. Skillnader mellan utfall och planering kan vara en lärorik erfarenhet. Det gäller också om genomförandet har varit annorlunda än planen: varför och på vilket sätt.

För utvärderingen behövs så kallade *processdata*, dvs. information om hur processerna har fungerat. Det kan exempelvis gälla leveranstider för indata eller vara information om granskningsprocessen, t.ex. avseende signaler om variabelvärden som markerats som avvikande och om antalet vidtagna åtgärder av olika slag.

3 Statistiska registers innehåll

3.1 Objekt, population, variabler, referenstider

Ett register innehåller en eller flera objekttyper. Det finns en mängd objekt av varje objekttyp, en population (eller eventuellt flera) med en uttalad avgränsning. Till varje objekttyp hör en eller i allmänhet flera variabler, t.ex. inkomst eller ålder.

Variabler har referenstider, som är tidsperioder eller tidpunkter. Det kan exempelvis vara inkomst för ett kalenderår eller ålder vid en viss tidpunkt. Även objekt och populationer hänför sig till referenstider.

Det finns oftast fördröjningar innan information, t.ex. om adressförändring, kommer till myndigheter och förs in i dessas register. Rapporteringskrav till en myndighet, liksom efterlevnad av kraven, kan variera mellan företag beroende på dessas storlek och mellan kommuner. Hur ofta objekt och variabler i ett statistiskt register uppdateras kan variera mellan registrets datakällor. Det är viktigt att ha tidsmarkeringar och det kan vara motiverat att ha flera tidsmarkeringar för ett variabelvärde, t.ex. när variabelvärdet registrerades och när detta värde trädde i kraft.

Ett objekt kan förändras över tid, t.ex. kan en individ ändra civilstånd, vilket påverkar variabelvärden och relaterade objekt, som hushåll. Med andra ord är det objektets egenskaper som ändras. Att följa ett objekt över tid kan vara svårt. Det är inte alltid självklart vad som ska följas – i den ekonomiska statistiken finns såväl verksamhet som organisation hos företagen.

En *identifierande variabel* är en speciell typ av variabel som används för att identifiera objekt entydigt, t.ex. personnummer. Detta är ett exem-

pel på en *kopplingsvariabel* som identifierar enskilda objekt eller grupper av objekt entydigt. Kopplingsvariabler används för att sammanvända register.

I produktionen av integrerade register behöver ibland överväganden och ställningstaganden göras för objekt och variabler. Det kan t.ex. gälla vilken objektmängd som ett integrerat registret ska ha när de enskilda registrens objektmängder är olika. Det kan likartat hända att samma variabel förekommer i flera register men med värden som skiljer sig åt på ett sätt som de inte borde göra. Även i detta fall behöver ställning tas till vilken hantering som är lämplig.

3.2 Intresse, mål och observation

Användarnas önskemål, krav och bestämmelser beskrivs genom prefixet *intresse*: intresseobjekt (objekttyp), intressepopulation, intressevariabler etc. Det kan finnas en rad behov och önskemål. En del kan vara alltför omfattande för att tillgodoses, några rentav motstridiga.

Målsättningen för det statistiska registret beskrivs med hjälp av prefixet *mål*: målobjekt (objekttyp), målpopulation, målvariabler etc. Prefixet *mål* signalerar en strävan efter något. Målsättningen är en avvägning mellan å ena sidan användarnas behov och önskemål och å andra sidan vad som rimligen kan åstadkommas med hänsyn till kvalitet och kostnader, däribland vilken information som kan samlas in.

Ett tredje prefix är *observation*. De objekt som det samlas in uppgifter om kallas observationsobjekt, och de variabler som det samlas in värden på kallas observationsvariabler. En variabel i ett administrativt datamaterial är ett exempel på en observationsvariabel.

I avvägningarna av exempelvis lämpliga målvariabler ingår både intressevariabler och observationsvariabler i tänkbara datakällor. Om en observationsvariabel har en något annan definition än en variabel som har pekats ut som intressevariabel, finns det för målvariabeln ett val med två huvudmöjligheter:

- Låt målvariabeln vara lika med observationsvariabeln. Det innebär en brist i relevans.
- Utveckla en modell med antaganden om samband mellan variabler sådan att en målvariabel som är lika med intressevariabeln kan beräknas från observationsvariabeln (en modellbaserad beräkning). Det kan finnas slumpmässiga (stokastiska) inslag i modellen. Det modellbaserade förfarandet innebär en lägre tillförlitlighet.

Den som ansvarar för det statistiska registret kan behöva göra en rad liknande val, vilka kan påverka det slutgiltiga registrets kvalitet på olika sätt. Flera aspekter än relevans och tillförlitlighet ingår, t.ex. jämförbarhet över tid och sammanvändbarhet med andra register.

Det statistiska registersystemet, med sina standardiserade populationer och variabler, innebär såväl en tillgång som begränsningar för de val och prioriteringar som behöver göras.

3.3 Flera varianter och versioner

3.3.1 Registervariant och registerversion

Ett register kan vid behov delas in i flera *registervarianter*, normalt med syftet att strukturera innehåll eller förenkla för användare. En registervariant kan bestå av en delmängd av registret eller utgöra en speciell variant för att täcka specifika behov. Olika registervarianter kan skilja sig åt vad gäller definition av population eller variabelinnehåll.

Det kan t.ex. finnas en variant som avser förhållanden vid kalenderårsskiftet och en variant som inbegriper alla objekt som funnits under kalenderåret oavsett varaktighet och status vid kalenderårets slut. Den första typen är exempel på det som kallas ett *situationsregister*. Den senare typen av register är betydligt svårare att skapa än den förra, eftersom objekt kan finnas under enbart en del av året och objektens egenskaper kan förändras under året. Det innebär i sin tur att det behövs strukturerad information om förändringar under året av objekt och variabler med tidsmarkeringar.

En *registerversion* kan betraktas som en realisering av en registervariant för en viss referenstid (tidpunkt eller tidsperiod). Varje registervariant har åtminstone en registerversion, men normalt flera.

Benämningar som levande register och löpande register förekommer för register som uppdateras ofta, t.ex. RTB:s produktionsdatabas och FDB. Dessa levande register är normalt inte statistiska register i sig – det är däremot de därur framställda registervarianterna.

3.3.2 Planerat och slutligt observationsregister

Under utformningen fastställs det statistiska registrets mål- och observationsobjekt samt dess mål- och observationsvariabler. Då bestäms även registrets *planerade observationsregister* med principer för vilka objekttyper och objekt samt vilka variabler som ska ingå.

Det är en målsättning att under insamling och bearbetningar fylla det planerade observationsregistret med värden. Det planerade observationsregistret innehåller i allmänhet målobjekt, så långt det är möjligt, och målvariablernas värden. Bland annat beroende på vilka användningarna är kan det även innehålla observationsobjekt och observationsvariablernas värden. Om observationsobjekt och målobjekt är objekt av olika typer kan det vara ett skäl att ha med observationsobjekt. Det kan uppstå behov av att kunna göra om bearbetningar och då kan information om insamlade värden behövas. Detta kan i sin tur leda till flera versioner av registret.

Datainsamling och bearbetningar ger variabelvärden som läggs in i observationsregistret. Även information om hur värdet har tillkommit ska ingå, t.ex. vilken imputeringsmetod som har använts.

3.3.3 Preliminärt register

Det finns statistiska register som framställs i flera versioner: först en preliminär version (eller flera) och sedan en slutlig. Ändamålet med det prelimära registret kan vara att framställa preliminär statistik.

Framställningstiden – tiden från referenstidpunkten eller referensperiodens slut till färdigställandet – är kortare för en preliminär version än för den slutliga versionen, men tillförlitligheten är normalt lägre. Den som framställer registret har kortare tid för sina bearbetningar till den preliminära versionen än till den slutliga. Normalt finns det skillnader också i indata. Det kan vara en datakälla som inte ingår i den preliminära versionen eller att administrativa data är preliminära i något avseende. Det finns då skillnader i framställning och i kvalitet mellan den preliminära och den slutliga versionen av det statistiska registret.

Det preliminära och det slutliga registret kan vidare ha delvis olika innehåll. Det innebär att det finns två varianter av registret med var sin registerversion, en preliminär och en slutlig.

4 Datainsamling

4.1 Datakällor

Ett statistiskt register använder en eller flera datakällor (i allmänhet främst administrativa data), var och en med sitt innehåll, sin tillförlitlighet i enskilda uppgifter, sin aktualitet etc. Det kan finnas ytterligare datamaterial som används som stöd vid exempelvis granskning, bland annat inom det statistiska registersystemet.

En viktig skillnad mellan användning av befintliga data och direktinsamling är påverkansmöjligheterna. Till ett statistiskt register är det vanligen redan befintliga data som samlas in, t.ex. från en myndighet som har samlat in och bearbetat dessa administrativa data för eget bruk. SCB kan i vissa fall ha visst inflytande på sådana data.

Varför och hur dessa administrativa data har samlats in och bearbetats har betydelse för det enskilda statistiska registret och för registersystemet som helhet. Denna information är viktig för att förstå på vilket sätt data bör hanteras och bearbetas.

I allmänhet finns någon form av överenskommelse med varje uppgiftskälla som levererar uppgifter till registret: om leveransens innehåll och dess tidpunkt.

4.2 Objekt och objektmängder

Ett statistiskt register kan användas som ram (se avsnitt 1.6). I statistikproduktionsprocessen är ramförfarandet viktigt. Även i framställningen av statistiska register behövs detta slags överväganden, som leder fram till observationsobjekt och lämpliga datakällor samt, i förlängningen, det statistiska registrets målobjekt och målpopulation.

Vissa statistiska register använder en ram, ofta ett basregister. För andra statistiska register finns inte denna möjlighet. I valet av datakällor ingår vilka de planerade observationsobjekten är; objekttyp och objektmängd. Detta ingår i de överenskommelser som görs med uppgiftskällorna. Bearbetningarna leder sedan vidare till målobjekt.

4.3 Mätning (observation)

Om administrativa data har samlats in genom ett formulär, till exempel, ger formuläret information om mätningen och hur uppgiftslämnaren kan ha tolkat och besvarat frågan (motsvarande) i formuläret. Det är önskvärt att också känna till vilka granskningsrutiner som uppgiftskällan använder. Bedömningar kan då göras av vilka mätfel, dvs. avvikelser från det sanna värde som efterfrågas, som kan ha uppstått i den insamlingen och vilka som kan återstå efter den granskning och bearbetning som skett hos uppgiftskällan. De senare, dvs. de återstående mätfelen, är viktigare för registerframställningen.

4.4 Brister i leveranser

I den överenskommelse som i allmänhet finns med den uppgiftskälla som levererar uppgifter till registret ingår leveransens innehåll och dess tidpunkt. Avvikelser från överenskommelsen är en brist som kan uppstå och som behöver bevakas. Det kan gälla hela eller delar av en leverans.

Om data saknas helt för ett objekt föreligger objektbortfall. Det uppstår när det är känt att objektet ska ingå i en viss objektmängd, men variabelvärden för det objektet saknas. Det kan också vara så att vissa variabelvärden har funnits men bedömts vara av så låg kvalitet att de har tagits bort. Objektbortfall kan förekomma för både observationsobjekt och målobjekt.

Partiellt bortfall innebär att ett eller flera men inte alla variabelvärden saknas för ett objekt. Det finns således användbara, men inte fullständiga, data för objektet.

5 Bearbetningar

Med *bearbetningar* menas bearbetningar av insamlade uppgifter, där t.ex. kodning och granskning samt imputering av variabelvärden och andra modellbaserade beräkningar kan ingå.

5.1 Kodning och imputering

Variabler kan tillföras i bearbetningarna. I *kodning* används regler för att kategoriindela variabelvärden. Det kan vara klasser som införs i registret, t.ex. ålders- eller storleksklasser. Kodningen kan gälla bransch som bildas utifrån en uppsättning regler.

Vid bortfall kan *imputering* användas när det är möjligt, motiverat och rimligt. Det innebär att motsvarande tomma celler i observationsregistret fylls med beräknade variabelvärden. En tillhörande markering införs som anger att värdet är imputerat och med vilken metod. En annan möjlighet att hantera bortfall i observationsregistret är att föra in informationen *uppgift saknas*.

5.2 Granskning

Dubblettkontroll är ett exempel på granskning inom ett register. Jämförelser med andra register och datakällor i syfte att hitta omotiverade skillnader är ett annat exempel på granskning. I allmänhet kombineras granskningskontroller med en åtgärd. Att eliminera dubletter är ett exempel och att eliminera eller reducera inkonsistenser är ett annat exempel. Granskning sker normalt flera gånger under registerframställningen som en del av bearbetningarna.

Objekt och variabelvärden ändras över tid. En individ kan t.ex. ändra civilstånd eller emigrera. En juridisk enhet kan utöka sina arbetsställen eller ändra verksamhet. Viss sådan information kommer regelbundet från administrativa källor, medan det kan dröja innan andra förändringar upptäcks.

Ambitionsnivån i granskningen ska sättas i relation till registrets ändamål och de prioriterade användningarna. Detta gäller även metodval. Det kan vara svårare att ställa krav på mikronivån än på en makronivå. Hur mycket arbete som ska läggas på enskilda objekt varierar, beroende på objektets relativa betydelse. Kvalitetskrav varierar mellan variabler.

5.3 Modellbaserade beräkningar

I vissa fall är det möjligt att beräkna variabelvärden utan osäkerhet, t.ex. en persons ålder en viss dag utifrån personens födelsedatum. Det är emellertid vanligt att osäkerhet ingår när ett objekts variabelvärde beräknas, t.ex. vid imputering. En sådan beräkning baseras vanligen på ett antagande om likhet eller samband. När en modell används för beräkningar av exempelvis variabelvärden, bör modellen formuleras explicit för att tydliggöra antaganden och möjliggöra bedömning av osäkerheter. Termen härledning används också. Ofta innebär den en beräkning utan osäkerhet, men det måste inte vara så.

En målvariabel kan behöva beräknas från en eller flera observationsvariabler. Det kan vara ekonomiska variabler som ska "räknas om", t.ex.

omsättning från en till en något annan tidsperiod, från brutto till netto eller från observationsobjekt till målobjekt av annan typ.

Även objekt kan behöva skapas med hjälp av en relation eller modell. Det kan handla om att dela upp ett objekt i flera (t.ex. en företagsenhet i flera verksamhetsenheter) eller att föra samman flera objekt till en ny objekttyp (t.ex. individer som kopplas ihop till ett hushåll).

Det kan finnas täckningsbrister som registerproducenten bör kompensera för, om det är möjligt. En sådan brist kan t.ex. bestå i att objekt saknas i något av de register som ingår i framställningen. Att uppgift om lägenhetsnyckel saknas innebär t.ex. att hushåll inte kan skapas på reguljärt sätt, utan att modellantaganden behöver göras.

Modellbaserade beräkningar kan således användas dels för att reducera konsekvenser av andra osäkerhetskällor, dels för att ge det statistiska registret ett delvis nytt innehåll, vad gäller objekt eller variabler (eller båda).

Utmärkande för beräkningar som baseras på modellantaganden är att ämneskunskap krävs, ofta även metodkunskap till hjälp för bedömning av rimlighet och konsekvenser.

6 De fem huvudkomponenterna

Här görs en kort genomgång av kvalitetsbegreppet som komplement till den tidigare texten. Det är viktigt att se på kvalitet ur användarnas perspektiv.

6.1 Relevans – ändamål och registrets innehåll

Kvalitetskomponenten relevans har de två underkomponenterna *ändamål och informationsbehov* samt *registrets innehåll*. Ändamålet har en särställning jämfört med övriga kvalitetskomponenter. Det framgår bland annat av Figur 2.

Det statistiska registrets innehåll beskrivs genom målobjekt, målpopulation, målvariabler samt dessas referenstider. I allmänhet ska dokumentation innehålla även information om t.ex. observationsvariabler, observationsobjekt av annan typ än målobjekten, samt information om metoder som används vid t.ex. modellbaserade beräkningar.

Från FDB hämtas underlag till ramar fyra gånger om året; dessa underlag är situationsregister. En sådan ram kan bidra till samordning av statistiska undersökningar som använder ramen. En undersökning kan från FDB hämta aktualiserad information om t.ex. adresser eller objekts aktivitetsstatus inför datainsamlingen och fortsatt under insamling och bearbetningar. Detta är ett vanligt förfarande för både individ- och

företagsbaserade undersökningar, som utgår från RTB respektive FDB när ramen skapas.

Det tredje huvudsakliga användningsområdet för statistiska register är att ge underlag till skattningsförfaranden. Ett exempel vid kalibrering är RTB. Hjälpinformationen till skattningsförfarandet kan vara knuten till ramen eller komma från en senare variant eller version av registret.

Det fjärde huvudsakliga användningsområdet för statistiska register är att vara stöd vid granskning, där jämförelse görs t.ex. av direktinsamlad information med administrativa data.

Basregistren används brett. Många andra register har mer specifika syften, ofta för registerbaserad statistikframställning.

6.2 Tillförlitlighet

För många registeranvändare är det tillförlitligheten totalt som är intressant. En uppdelning på osäkerhetskällor kan vara ett stöd och ge viss information till registeranvändare. Det kan vara lättare för registerproducenten att uttala sig om enskilda osäkerhetskällor än om helheten. Det är inte uppenbart vilka tillförlitlighetsmått som bör användas – det kan variera mellan användningar. I vissa fall har enskilda värden betydelse, t.ex. när registret används som underlag vid granskning på mikronivå. I andra fall är det tillförlitligheten i aggregat som är viktig, t.ex. när registret används för statistikframställning.

För det statistiska registret finns en målsättning i utformningen: målobjekt och målpopulation samt målvariabler, alla med referenstid. I praktiken är målvariablernas värden ofta behäftade med osäkerhet. Även för objekt finns osäkerheter, t.ex. om de är rätt konstruerade och om de tillhör målpopulationen. Det finns ofta variabler som ger information om objekts status, t.ex. om en företagsenhet ska ses som aktiv.

Enkelt uttryckt beror tillförlitligheten i det statistiska registret av två faktorer: tillförlitligheten i insamlade data samt tillförlitligheten i de bearbetningar som görs under registerframställningen. Kraven på tillförlitlighet (och andra kvalitetskomponenter) ska, åtminstone i princip, ha formulerats efter prioritering av registeranvändarnas behov, som Figur 2 visar. Vid utformningen tillgodoses, om möjligt, kraven och i utvärderingen mäts och bedöms utfallet för kvaliteten.

Fördröjningar är en viktig faktor att beakta i bedömningar av tillförlitlighet – hur lång tid det kan ta innan en faktisk ändring når det statistiska registret.

I arbetet med bedömning och beskrivning av tillförlitlighet bör de fem osäkerhetskällorna ingå – om de är tillämpliga och vilken betydelse var och en kan tänkas ha för *tillförlitlighet totalt*. Bedömningar bör göras

både i absoluta och i relativa termer. Mätning kan relateras till den ursprungliga mätsituationen med hänsyn till använt formulär (mätinstrument) och primär användning av den lämnade uppgiften. Även granskning och gjorda bearbetningar har betydelse för återstående mätfel. Brister i täckning – det vill säga skillnader mellan registrets objektmängd och dess målpopulation – kan bero på fördröjningar i information och på andra osäkerhetskällor, t.ex. mätning och modellantaganden som kan leda till att ett objekt placeras i fel population. Vissa bedömningar av täckningsbrister kan göras i efterhand när information har tillkommit. Att information tillkommer i efterhand kan i viss mån gälla även bortfall; objekt som saknar värden, se avsnitt 4.4.

Modellantaganden hör i vissa fall ihop med en annan osäkerhetskälla, som t.ex. bortfall, och i sådana fall ska de redovisas i beskrivningen av den osäkerhetskällan. I andra fall är det fristående modellantaganden, t.ex. för att skapa nya variabler. De modeller som används för omräkningar bör studeras med avseende på hur goda anpassningar de utgör och hur goda skattningar de innebär med hänsyn till approximationer och eventuella stokastiska (slumpmässiga) inslag i modellen.

Preliminära register bör jämföras med slutliga versioner för att få information om hur stora skillnader det är i olika avseenden och för att se om det finns förbättringsmöjligheter för informationen i det preliminära registret.

Mätningar och bedömningar av det statistiska registrets tillförlitlighet kan göras från olika utgångspunkter och med olika medel. Några exempel har getts ovan. När registret används för statistikproduktion kan, och bör, registerproducent och registeranvändare ha ett informationsutbyte om krav och bedömningar. Registeranvändare kan tillföra information som registerproducenten inte har. Riktade kvalitetsstudier kan göras för specifika kvalitetsfrågor.

6.3 Aktualitet och punktlighet

Användare som ska ta fram registerbaserad statistik eller använda registret som ram har önskemål om korta framställningstider. Detta ska vägas mot de fördröjningar som finns innan uppdateringar når registret och mot den tid som behövs för granskning och övriga bearbetningar.

Att ta fram en preliminär version av ett register kan vara ett sätt att tillmötesgå behov av kortare framställningstid, om än med lägre tillförlitlighet.

6.4 Tillgänglighet och tydlighet; redovisning

Tillgänglighet avser var, och på vilket sätt, användare får tillgång till det statistiska registret, i preliminär och i slutlig form.

För externa användare finns flera möjligheter: t.ex. tillgång till av-identifierade mikrodata för egna bearbetningar i MONA (*Microdata Online Access*) eller beräkningar som SCB kan göra på uppdrag.

Interna användare har tillgång till observationsregistret, i dess slutliga form och i preliminär version när det är aktuellt. Behörighet behövs.

Tydlighet avser information om de egenskaper registret har. Det kan vara olika typer av tillgänglig dokumentation, exempelvis översiktlig information om registret, detaljerad information om variabler och värdemängder, t.ex. i MetaPlus, samt information i form av rapporter om genomförda kvalitetsstudier och om kvalitetsindikatorer som är användbara för olika syften.

6.5 Jämförbarhet och sammanvändbarhet

Jämförbarhet och sammanvändbarhet påverkas såväl av använda begrepp, definitioner och klassifikationer som av de metoder som används. Detta gäller både datakällor och bearbetningar.

Om en variabeldefinition ändras påverkas jämförbarhet över tid. I vissa fall kan det vara möjligt att göra omräkningar för att överbygga t.ex. tidsseriebrott. Även sammanvändbarhet med andra register påverkas, om inte definitionsändringen är samordnad. I definitioner ingår även målpopulationer – olikheter reducerar möjligheter att göra jämförelser och att sammanvända register.

Olika metoder kan, men måste inte, innebära systematiska skillnader som i så fall reducerar jämförbarhet och sammanvändbarhet. De olika metoderna kan t.ex. gälla granskning av administrativa data. Det kan vara granskningsmetoder och åtgärder som ändras över tid, som skiljer sig mellan två register eller som skiljer sig mellan variabler. Effekter av olika metoder för t.ex. granskning kan vara svåra att mäta. Bedömningar bör eftersträvas, med tanke på såväl kvalitetsbeskrivning som kvalitetsförbättring.

Anvisningar för det statistiska registrets framställning och kvalitet

Inledning

Det främsta syftet med *Det statistiska registrets framställning och kvalitet* är att ge förståelse för det statistiska registret och därmed underlätta användningen. Beskrivningen vänder sig i första hand till de kända huvudanvändarna.

Anvisningarna har, utöver denna inledning och en avslutning, i likhet med mallen tre kapitel:

- *Registrets framställning*
- *Registrets kvalitet*
- *Allmänna uppgifter*

De tre kapitlen kompletterar varandra. Det finns några mindre överlapp för att underlätta läsningen, se avslutningen. Beakta helheten när information placeras och delarna skrivs. Texten bör vara överskådlig. Hänvisningar kan användas om det är praktiskt, men beskrivningar ska vara självbärande. Tablåer kan användas i stället för, eller som komplement till, löpande text om det är lämpligt för läsaren. Placera dock längre sammanställningar i en bilaga eller möjligen separat med en hänvisning. Avsnitten är numrerade med en kombination av bokstav och siffror, där bokstaven anger vilket av de tre kapitlen det är: F, K respektive U.

När det finns både en preliminär och en slutlig version av registret är ofta en gemensam beskrivning att föredra framför två separata. Den slutliga versionen är ett slutligt observationsregister i registerframställningen. Om det är två mycket olika registervarianter kan det vara praktiskt med separata beskrivningar.

När ett statistiskt register används, helt eller delvis, för statistikframställning använder statistikproducenten (en eller flera) information i beskrivningen av registret. Statistiken har, med enstaka undantag, ett eget slutligt observationsregister, t.ex. med några ytterligare variabler. Detta slutliga observationsregister ska dokumenteras.

Disponera texten med tanke på både läsare i närtid och arkivering. Detta gäller även bilagor och hänvisningar till andra dokument som ska finnas tillgängliga. Tänk också på både externa och interna läsare, om det är tillämpligt.

Innehållet i mallen förutsätter kunskaper motsvarande det som ingår i handbokens del *Kvalitet och andra begrepp för statistiska register*.

Mallens försättsida

Av försättsidan ska tydligt framgå: textens versionsnummer och datum, vad det statistiska registret heter, om beskrivningen avser en variant, vilken produktkod registret har samt vilken eller vilka referenstider registret avser. Produktnamnet kan kompletteras med en förkortning av namnet.

Mall och layout

Mallen anger strukturen med numrerade rubriker och tillhörande angivet innehåll. Därutöver får onumrerade underrubriker och bilagor användas vid behov.

Det statistiska registrets framställning

Nedan ges en kort sammanfattning över vad som ingår i F-kapitlet om det statistiska registrets framställning. F-kapitlet består av tre huvuddelar:

- *F1 Registrets sammanhang*
- *F2 Utformning av framställningen*
- *F3 Genomförande*

Huvuddelarna F2 och F3 är indelade i avsnitt. Mer detaljerade anvisningar om respektive avsnitt finns i mallen.

F1 Registrets sammanhang

Den första huvuddelen handlar om det statistiska registrets plats i SCB:s statistiska registersystem och om registrets huvudsakliga användning. Registersystemet består av tre sfärer som innehåller ett basregister var. Basregistren är grundvalen för hela registersystemet och har till huvudsaklig uppgift att ansvara för viktiga objekt-mängder och ha kopplingsvariabler till andra basregister. Övriga statistiska register ska kunna kopplas till ett basregister via standardiserade kopplingsvariabler.

F2 Utformning

I den andra huvuddelen ska huvuddragen av de principer och metoder som har använts vid utformningen av registret beskrivas. Det första avsnittet omfattar registrets innehåll, datakällor samt tidsaspekter. Därefter kommer ett avsnitt som handlar om insamlingsförfarandet; vilka metoder som används, vilka leverantörer som levererar data till registret, mätning samt brister i leveranser. Det sista avsnittet som ingår i utformning av framställningen är vilka bearbetningar som görs inklusive granskning. Här tas följande upp: kodning, dubblettkontroller, rimlighetskontroller, imputeringar, modellbaserade beräkningar samt jämförelser med andra register och datakällor.

Observera att detta avsnitt omfattar utformningen av framställningen, själva genomförandet tas upp under *F3 Genomförande*.

F3 Genomförande

Den tredje huvuddelen i F-kapitlet handlar främst om att beskriva utfallet från de registerproduktionsprocesser som framställer och fastställer det slutliga observationsregistret, samt eventuella preliminära versioner. Beskrivningen av utfallet kan göras i form av kvantitativ information, t.ex. processdata från olika bearbetningar.

Om det har förekommit några avvikelser från planeringen ska även dessa beskrivas, t.ex. varför avvikelser har skett och vad det har haft för konsekvenser.

Det statistiska registrets kvalitet

Nedan ges en kort sammanfattning över vad som ingår i K-kapitlet om det statistiska registrets kvalitet. K-kapitlet består av fem huvuddelar:

- *K1 Relevans*
- *K2 Tillförlitlighet*
- *K3 Aktualitet och punktlighet*
- *K4 Tillgänglighet och tydlighet*
- *K5 Jämförbarhet och sammanvändbarhet*

Varje huvuddel är indelad i avsnitt. Mer detaljerade anvisningar om respektive avsnitt finns i mallen.

Huvuddelarna motsvarar de fem huvudkomponenterna i kvalitetsbegreppet. Var och en av dessa omfattar en eller flera underkomponenter, som i sin tur kan ha delkomponenter.

Kvalitetsbegreppet har likheter med kvalitetsbegreppet för statistik, men det avser statistiska register, dvs. mikrodata. Därför finns även skillnader, t.ex. vad gäller innebörden av tillförlitlighet.

Genomgående gäller att kvaliteten ska beskrivas som egenskaper och utan värdering. Information ska ges utifrån användarnas perspektiv. Interna och externa användare kan ha delvis olika behov, och båda ska beaktas.

K1 Relevans

Huvudkomponenten *Relevans* handlar om hur väl registret motsvarar användarnas behov. Relevansen bedöms utifrån i vilken grad registerproducentens val av målpopulation och målvariabler stämmer överens med de huvudsakliga användarnas intressen.

Underkomponenten *Ändamål och informationsbehov* handlar om det statistiska registrets huvudsakliga syften och användningsområden. Utgångspunkten är de huvudsakliga användarnas informationsbehov utifrån registerproducentens kännedom.

Underkomponenten *Registrets innehåll* handlar om registrets målpopulation och målvariabler. Dessa sätts i relation till intressepopulationer och intressevariabler definierade utifrån användarnas önskemål. Beskriv relationerna mellan målobjekt, observationsobjekt och de huvudsakliga användarnas intresseobjekt. Förklara de skillnader som finns.

Beskriv målvariablerna översiktligt. Långa listor och detaljerade definitioner placeras med fördel separat med hänvisning. Beskriv relationerna mellan registrets målvariabler, observationsvariabler och de huvudsakliga användarnas intressevariabler, med tonvikt på de skillnader som finns. Ange även vilka variabler som används på olika sätt i statistikproduktionen samt identifierande variabler som används som kopplingsvariabler vid integrering med andra register.

Beskriv registrets referenstider. Det kan finnas flera referenstider, t.ex. en tidsperiod för variabler och en tidpunkt för målpopulationen. Det kan finnas skillnader mellan variabler och det kan finnas skillnader mellan objekt och variabler.

K2 Tillförlitlighet

Huvudkomponenten *Tillförlitlighet* handlar om osäkerheten i registrets målobjekt och i målvariablernas värden. Enkelt uttryckt beror registrets tillförlitlighet dels på tillförlitligheten i indata och dels på de bearbetningar som registerproducenten gör.

Underkomponenten *Tillförlitlighet totalt* handlar om tillförlitligheten utifrån registrets huvudsakliga användningsområden, t.ex. att ligga till grund för registerbaserad statistik eller att användas för granskning av andra register eller direktinsamlade data. Gör här en sammantagen bedömning av det statistiska registrets tillförlitlighet, med hänsyn till konsekvenser av osäkerhetskällorna nedan.

Underkomponenten *Osäkerhetskällor* handlar om fem vanligt förekommande källor till osäkerhet, och dessas påverkan på tillförlitligheten. Det är vanligen i första hand

tillförlitligheten totalt som är intressant för användarna. En uppdelning på osäkerhetskällor har i första hand ett producentperspektiv, eftersom det underlättar för registerproducenten att bedöma, beskriva och påverka tillförlitligheten. En presentation av osäkerhetskällorna var för sig kan dock ge även användarna värdefull information, t.ex. om hur respektive osäkerhetskälla i registret påverkar motsvarande osäkerhetskälla i den statistik som registret används till (om tillämpligt).

Beskriv i avsnitt *K.2.2.1 Täckning* konsekvenser av täckningsbrister för registrets tillförlitlighet. Det kan finnas flera målpopulationer att beskriva. Ange, om möjligt, kvantitativa mått på täckningsbristerna med förklaring. Ett sådant mått kan t.ex. uppskattas genom identifierad övertäckning. Beskriv om det finns delpopulationer som har särskilt stora täckningsbrister.

Avsnitt *K.2.2.2 Mätning* handlar om konsekvenser för registrets tillförlitlighet av mätfel, det vill säga skillnader mellan observerade och sanna variabelvärden. Bedöm och beskriv konsekvenserna av mätfel för registrets tillförlitlighet, både systematiska och slumpmässiga effekter. Använd information i avsnitt *F2.5 Bearbetningar med granskning*, utan att upprepa men så att texten blir självbärande.

Avsnitt *K.2.2.3 Bortfall* handlar om konsekvenser för registrets tillförlitlighet av bortfall. Konsekvenser av bortfall kan vara systematiska eller slumpmässiga (eller båda). Effekterna kan skilja sig mellan registrets användningsområden. Beskriv osäkerheten i registret dels utifrån objektbortfall och dels utifrån partiellt bortfall.

Avsnitt *K.2.2.4 Bearbetningar* handlar om konsekvenser för registrets tillförlitlighet av bearbetningar, t.ex. kodning, granskning och ändringar samt integrering av olika register och andra datakällor. Misstag kan förekomma t.ex. under granskning, i programmering och vid överföring av information. Det kan finnas erfarenheter och kunskap från tidigare produktionsomgångar och de åtgärder som då vidtagits.

Beskriv i avsnitt *K.2.2.5 Modellantaganden* konsekvenser för tillförlitligheten till följd av de modeller som använts i framställningen av registret. Det kan t.ex. vara modeller för "omräkningar" till målvariabler från observationsvariabler.

Underkomponenten *Preliminärt register jämfört med slutligt* handlar om revideringar, främst storlek och riktning, mellan de versioner som framställs; en eller flera preliminära och det slutliga registret. Den jämför således de successivt framtagna versionerna av registret, och den avser därmed inte tillförlitligheten i respektive version.

K3 Aktualitet och punktlighet

Huvudkomponenten *Aktualitet och punktlighet* handlar om registrets framställningstid, frekvens (periodiciteter) och punktlighet. Det kan vara bra att kort beskriva de avvägningar som har gjorts vid valet av referenstid, om det inte står i avsnitt K1 om relevans. Registrets framställningstid är kalendertiden mellan referenstidpunkten eller referensperiodens slut och den tidpunkt då registret är färdigställt.

Förklara eventuella förändringar av framställningstiden under en tidsperiod bakåt och även framåt om det är planerat. För en användare som hämtar information från registret vid en godtyckligt vald tidpunkt behöver det framgå hur långt tillbaka senast tillgängliga referenstids slut kan ligga.

K4 Tillgänglighet och tydlighet

Huvudkomponenten *Tillgänglighet och tydlighet* handlar om användarnas möjlighet att få tillgång till registret och den tillhörande dokumentationen. Beskriv de distributionskanaler som används för att göra registret tillgängligt för interna respektive externa användare. Interna tekniska detaljer ska inte tas med. Beskriv även hur informationsspridning till interna respektive externa användare går till.

K5 Jämförbarhet och sam användbarhet

Huvudkomponenten *Jämförbarhet och sam användbarhet* handlar om möjligheterna att göra jämförelser över tid och om möjligheterna att på olika sätt använda registret tillsammans med andra register.

Jämförbarhet över tid påverkas, som all sam användbarhet, av använda definitioner och metoder. Om dessa förändras över tid kan det ge upphov till brott i tidsserier. Beskriv nya, ändrade och borttagna variabler jämfört med föregående version av registret. Förändringar i samhället kan också ha betydelse, t.ex. ändringar i regelverk. Beskriv konsekvenser av eventuella förändringar, och åtgärder som vidtagits för att mildra effekterna av tidsseriebrott.

För att beskriva sam användbarhet i övrigt behöver en avgränsning göras av vilka register som det kan finnas intresse att använda detta statistiska register tillsammans med. Gör denna bedömning med tanke på registrets ändamål och användares informationsbehov.

Allmänna uppgifter

Kapitlet om allmänna uppgifter benämns U-kapitlet och handlar främst om vilka lagar och förordningar som behandlingen av uppgifter i registret har att förhålla sig till. Dessutom beskrivs registrets historik samt kontaktuppgifter till viktiga områden som är kopplade till registret, såsom registerproducent, mikrodatautlämnade och statistikupdrag.

Avslutning

Det statistiska registrets ändamål är en utgångspunkt för såväl registret i sig som dokumentationen av registret.

Innan texten skrivs är det bra att ha en överblick över strukturen med dess tre kapitel och underrubriker. Det finns några beröringspunkter mellan kapitlen. Tre exempel på parvisa beröringspunkter ges nedan. Syftena inom varje par är något olika. Exempelvis är avsnitt K1.2 mer omfattande och detaljerat än avsnitt F2.1, vars främsta syfte är att underlätta läsningen av F-kapitlet.

- { F1 Registrets sammanhang
- { K1.1 Ändamål och informationsbehov

- { F2.1 Det statistiska registrets innehåll i stora drag
- { K1.2 Registrets innehåll

- { K5.1 Jämförbarhet över tid
- { U5 Historik

DET STATISTISKA REGISTRETS FRAMSTÄLLNING OCH KVALITET

<Registrets namn>

Om det är en registervariant (eller flera) som beskrivs, ange vad den här beskrivningen omfattar. Det är särskilt viktigt om det finns flera beskrivningar avseende samma register (t.ex. FDB eller RTB).

Produktkod

Ange produktkod.

Referenstid

Ange referenstid, dvs. vilken tid registret avser. Det kan vara en period eller en tidpunkt (datum). Om det finns flera referenstider, ange en huvudsaklig och beskriv detaljer i avsnitt K1.2.3 *Referenstider*. Angivelsen ska passa ihop med registervarianten.

Innehåll

Det statistiska registrets framställning	4
F1 Det statistiska registrets sammanhang	4
F2 Utformning av framställningen	4
F2.1 Det statistiska registrets innehåll i stora drag	4
F2.2 Datakällor	4
F2.3 Tidsaspekter	5
F2.4 Insamlingsförfarande	5
F2.4.1 Datainsamlingsmetoder och leverantörer	5
F2.4.2 Mätning	5
F2.4.3 Brister i leveranser	5
F2.5 Bearbetningar med granskning	5
F2.5.1 Kodning	6
F2.5.2 Dubblettkontroller	6
F2.5.3 Rimlighetskontroller	6
F2.5.4 Imputeringar	6
F2.5.5 Modellbaserade beräkningar	6
F2.5.6 Jämförelser med andra register och datakällor	7
F3 Genomförande	7
F3.1 Kvantitativ information	7
F3.2 Avvikelser från utformningen	7
Det statistiska registrets kvalitet	7
K1 Relevans	7
K1.1 Ändamål och informationsbehov	7
K1.1.1 Registrets ändamål	7
K1.1.2 Registeranvändares informationsbehov	8
K1.2 Registrets innehåll	8
K1.2.1 Objekt och population	8
K1.2.2 Variabler	8
K1.2.3 Referenstider	8
K2 Tillförlitlighet	9
K2.1 Tillförlitlighet totalt	9
K2.2 Osäkerhetskällor	9
K2.2.1 Täckning	9
K2.2.2 Mätning	9
K2.2.3 Bortfall	9
K2.2.4 Bearbetningar	10
K2.2.5 Modellantaganden	10
K2.3 Preliminärt register jämfört med slutligt	10
K3 Aktualitet och punktlighet	10
K3.1 Framställningstid	10
K3.2 Frekvens	10
K3.3 Punktlighet	11

K4	Tillgänglighet och tydlighet.....	11
K4.1	Tillgång till registret	11
K4.2	Informationsspridning.....	11
K4.3	Dokumentation	11
K5	Jämförbarhet och sam användbarhet	12
K5.1	Jämförbarhet över tid	12
K5.2	Sam användbarhet med andra register.....	12
Allmänna uppgifter	12
U1	Sekretess och personuppgiftsbehandling.....	12
U2	Bevarande och gallring	13
U3	Uppgiftsskyldighet	13
U4	EU-reglering och internationell rapportering.....	13
U5	Historik	14
U6	Kontaktuppgifter	14

Det statistiska registrets framställning

F1 Det statistiska registrets sammanhang

Beskriv översiktligt det statistiska registrets plats i registersystemet genom att t.ex. ange vilken sfär (Fastighet, Företag eller Individ) registret tillhör och kopplingen till basregistret. Om registret är en variant av ett basregister (FR, FDB, RTB) ska detta framgå.

Ge kortfattad information om registret som helhet, särskilt om den här beskrivningen avser en specifik variant av registret.

Nämn registrets huvudsakliga användningar utan de detaljer som finns i kvalitetsbeskrivningen. Ange om det används inom ett eller flera av de fyra huvudsakliga användningsområdena för statistiska register.

- 1) Ligga till grund för registerbaserad statistik.
- 2) Vara ram för statistiska undersökningar.
- 3) Användas i skattningsförfaranden.
- 4) Användas för granskning av andra register eller direktinsamlade data.

Ange om andra speciella användningar finns, värt att nämna är uppdragsverksamhet.

F2 Utformning av framställningen

Avsnittet är inriktat på utformning och de principer som har använts i utformningen. Genomförandet beskrivs i avsnitt F3 *Genomförande*.

Ingen text behövs på den här nivån.

F2.1 Det statistiska registrets innehåll i stora drag

Ge en kortfattad beskrivning av det statistiska registrets innehåll för att framställningskapitlet ska kunna läsas fristående, men utan att ge de detaljer som finns i kvalitetsbeskrivningen. Se exempel nedan.

”Utbildningsregistret (UREG) är ett individregister som innehåller uppgifter om personens högsta utbildning och ligger till grund för statistik över befolkningens utbildning.”

Om flera registervarianter tas fram och ingår i den här beskrivningen, ange skillnaderna mellan de olika varianterna. (En registervariant kan bestå av en delmängd av registret eller utgöra en speciell variant för att täcka specifika behov. Olika registervarianter skiljer sig åt i fråga om population(er) eller variabelinnehåll. I t.ex. RTB finns det ett antal registervarianter såsom Folkbokförda personer 31 december, Födda, Döda, Hushåll, Inrikes flyttningar och Utvandringar)

F2.2 Datakällor

Detta avsnitt handlar om förfarandet som leder fram till observationsobjekt och lämpliga datakällor samt, i förlängningen, det statistiska registrets målobjekt och målpopulation. Det blir bl.a. ett underlag för avsnittet om täckning i kvalitetsbeskrivningen.

Beskriv översiktligt vilka objekttyper som ingår. Om integrering av två eller flera register ingår, ange utformningen av objekt mängden i det nya registret (ett eller flera) om de

enskilda registrens objektmängder är olika. Beskriv vilka principer som gäller om samma variabel förekommer i flera register.

F2.3 Tidsaspekter

Beskriv i stora drag tidsaspekter vad gäller t.ex. indatakällor, uppdateringar, preliminära och slutliga versioner samt frysta versioner.

Ange registrets bryttidpunkt med hänsyn till sent inkomna uppgifter samt motiv.

F2.4 Insamlingsförfarande

Ingen text behövs på den här nivån.

F2.4.1 Datainsamlingsmetoder och leverantörer

Beskriv hur insamlingen går till hos den primära uppgiftskällan, t.ex. Skatteverket, Lantmäteriet eller Försäkringskassan. Det kan också vara en SCB-intern leverantör.

Redogör för eventuella lagar och regler som styr insamlingen. Om länkning till dokument eller webbsidor görs, tänk på att länkar kan vara instabila över tid. Viktiga dokument kan behöva bifogas även som bilaga.

Beskriv hur insamlingen går till hos datainsamlingsavdelningen, om det är tillämpligt. Det gäller t.ex. för skolregistren.

Beskriv hur insamlingen från andra går till hos registerproducenten på SCB, t.ex. utifrån leverantörskontakter med Skatteverket, Lantmäteriet och Försäkringskassan eller kontakter med datainsamlingsavdelning eller annan registerproducent på SCB. Ange registren på för läsaren lämpligt sätt.

Om registerproducenten på SCB använder direktinsamling, beskriv metoderna för datainsamlingen. Beskriv även kontaktstrategin, t.ex. hur och när uppgiftslämnarna ska kontaktas, om det är motiverat. Det kan vara information om när datainsamlingen ska göras i förhållande till referensperioden.

F2.4.2 Mätning

Bilagga, om lämpligt, externa leverantörers insamlingsformulär för administrativa data.

Beskriv, vid direktinsamling, frågeformuläret samt något om konstruktionen. Lägg frågeformuläret som bilaga.

F2.4.3 Brister i leveranser

Beskriv hur brister som upptäcks vid leverans (t.ex. beställda data som saknas eller ej beställda data som levereras) motverkas och hanteras.

F2.5 Bearbetningar med granskning

Nedanstående avsnitt handlar om att beskriva i korta drag vilka bearbetningar inklusive granskning som görs, i allmänhet i flera led. Det första utförs hos respektive myndighet eller organisation (uppgiftskälla), och resultatet är de bearbetade uppgifter som sedan hämtas in till SCB. Länka till myndighetens webbplats om det finns relevant information där. Tänk på att länkar kan vara instabila över tid. Viktiga dokument kan behöva bifogas

även som bilaga. Därefter följer, med inriktning på principer, bearbetningen och granskningen hos datainsamlingsavdelningen (om tillämpligt) och hos registerproducenten på SCB.

De sex avsnitten F2.5.1-F2.5.6 handlar om utformning. Genomförandet sker ofta parallellt eller iterativt, och ordningen kan vara något annorlunda.

Kvantitativ och mer detaljerad information redovisas i avsnitt F3 *Genomförande*.

Ingen text behövs på den här nivån.

F2.5.1 Kodning

Beskriv kodningen i form av berörda variabler, eventuella klassifikationer och vald metod.

F2.5.2 Dubblettkontroller

Beskriv de viktigaste dubblettkontrollerna. Kontroller kan göras för exakta dubblettrader, men också för att hitta dubletter på t.ex. personnummer där andra variabler skiljer sig åt. Beskriv hur dubletter hanteras.

F2.5.3 Rimlighetskontroller

Beskriv de viktigaste rimlighetskontrollerna som görs, för kvantitativa och kvalitativa variabler. Ange om rimlighetskontrollerna görs för indata eller i framställningen av det slutliga observationsregistret. Ett exempel på en rimlighetskontroll är att sammanräknad inkomst inte kan ha ett negativt värde. Detta avsnitt handlar om rimlighetskontroller inom registret, medan granskning med hjälp av andra register beskrivs i avsnitt F2.5.6 *Jämförelser med andra register och datakällor*.

Beskriv principerna för de viktigaste ändringar som görs för att åtgärda t.ex. icke valida variabelvärden eller motsägelser mellan variabelvärden.

F2.5.4 Imputeringar

Beskriv principer för de imputeringar som görs till följd av objektbortfall, partiellt bortfall och till följd av granskning.

F2.5.5 Modellbaserade beräkningar

Beskriv översiktligt och principiellt de modellantaganden som används i modellbaserade beräkningar. Sådana beräkningar (härledningar) bygger på att antaganden görs, t.ex. om det finns samband mellan variabler. Beräkningsreglerna kan vara komplicerade, och det är normalt behäftat med osäkerhet. Vissa beräkningar, som härledning av sammanräknad inkomst, kan vara enkla och utan osäkerhet.

Beskriv översiktligt vilka variabler som är baserade på modeller och härledningar. Hänvisa till bilaga eller separat dokumentation (t.ex. MetaPlus) om det är många variabler.

F2.5.6 Jämförelser med andra register och datakällor

Beskriv översiktligt vilka jämförelser som görs med andra register och källor i granskningssyfte. Granskningen kan t.ex. omfatta vissa variabler och egna modellbaserade beräkningar som jämförs med andra register eller datakällor för att hitta skillnader och orsaker. Beskriv åtgärder som vidtas.

F3 Genomförande

Avsnittet handlar om implementering av utformningen för registrets framställning.

I nedanstående avsnitt är det viktigt att tänka på vad som är intressant dels för externa och interna läsare, dels på kort sikt, dels för arkivering.

Ingen text behövs på den här nivån.

F3.1 Kvantitativ information

Beskriv utfallet från registerframställningen.

Ge kvantitativ information om t.ex. populationer och bortfall. Beskriv, om det är tillämpligt, processdata för t.ex. kodning, granskning och imputering.

Processdata ger information om hur registerframställningen har fungerat.

Bilagor med tabeller kan användas om det är praktiskt.

F3.2 Avvikelser från utformningen

Beskriv eventuella påtagliga avvikelser från planeringen. Ange orsakerna och förklara konsekvenser.

Om inga påtagliga avvikelser har gjorts från planeringen, skriv: Inga påtagliga avvikelser har gjorts från utformningen.

Det statistiska registrets kvalitet

K1 Relevans

Relevansen bedöms utifrån hur väl registerproducentens val av målpopulation och målvariabler stämmer överens med de huvudsakliga användarnas intressen. Beakta både externa och interna användare. Om det underlättar den fortsatta läsningen, ge här en kort sammanfattning av avsnitten K1.1 och K1.2.

K1.1 Ändamål och informationsbehov

Ingen text behövs på den här nivån.

K1.1.1 Registrets ändamål

Det finns fyra huvudsakliga användningsområden för statistiska register. Ett sådant register kan:

- 1) Ligga till grund för registerbaserad statistik.
- 2) Vara ram för statistiska undersökningar.
- 3) Användas i skattningsförfaranden.
- 4) Användas för granskning av andra register eller direktinsamlade data.

Beskriv de huvudsakliga syftena och användningsområdena för registret, så att det blir en inledning till avsnitt K1.2 *Registrets innehåll*.

K1.1.2 Registeranvändares informationsbehov

Beskriv översiktligt registeranvändarnas informationsbehov utifrån registrets huvudsakliga användningsområden. Gör det på ett sådant sätt att det ger förståelse för de intressepopulationer och intressevariabler som beskrivs i avsnitt K1.2.1 *Objekt och population* respektive K1.2.2 *Variabler*.

K1.2 Registrets innehåll

Ge en övergripande, sammanfattande, beskrivning av registrets målpopulation och målvariabler här, utan de detaljer som kommer i avsnitten K1.2.1–K1.2.3.

K1.2.1 Objekt och population

Beskriv registrets målobjekt: en eller flera objekttyper. Beskriv relationerna mellan olika typer av målobjekt om det finns flera. Ange, om tillämpligt, vilka andra objekttyper som användaren själv kan ta fram genom härledning från de objekt och variabler som finns i registret.

Beskriv relationerna mellan målobjekt och de huvudsakliga användarnas intresseobjekt, samt mellan målobjekt och observationsobjekt som är av annan typ och förklara de skillnader som finns.

Beskriv registrets målpopulation, en eller flera.

Beskriv relationerna mellan mål- och intressepopulationer och ge en kort motivering till de skillnader som finns.

K1.2.2 Variabler

Beskriv målvariablerna översiktligt. Ange vilka typer av variabler som registret innehåller, t.ex. inkomstvariabler och befolkningsvariabler. Gruppera gärna variablerna.

Beskriv relationer mellan målvariabler och intressevariabler samt mellan målvariabler och observationsvariabler, med tonvikt på skillnader.

Ange, om det är motiverat, vilka variabler som vanligen används som hjälpvariabler till statistiska undersökningar. Ange också identifierande variabler som används för integrering med andra register. Om registret används för granskning av andra register eller direktinsamlade data, ange vilka variabler som ofta används i detta syfte.

K1.2.3 Referenstider

Beskriv registrets referenstider. Om flera referenstider förekommer i registret, beskriv skillnaderna mellan dem. Ibland är det praktiskt att beskriva referenstider tillsammans med objekt och variabler. Hänvisa i så fall härifrån till de avsnitten.

K2 Tillförlitlighet

Ingen text behövs på den här nivån.

K2.1 Tillförlitlighet totalt

Ge en sammanfattande beskrivning av registrets tillförlitlighet med hänsyn till de viktigaste användningarna, externa såväl som interna. Ta hänsyn till tillförlitligheten i insamlade data och gjorda bearbetningar. Beakta de sammantagna konsekvenserna av osäkerhetskällorna nedan. Beskriv tillförlitligheten för objektmängder och variabler var för sig.

Det finns inga självskrivna osäkerhetsmått. Om registret används för t.ex. statistikframställning är registeranvändaren intresserad av information som påverkar skattningsförfaranden och av underlag för att bedöma statistikens kvalitet. Bedöm vad som är mest betydelsefullt för registrets huvudsakliga användningar och lyft fram det.

Kontrastera om möjligt registret mot andra datakällor, och försök att ge en uppfattning om tillförlitligheten totalt, t.ex. för olika variabelgrupper eller delpopulationer.

K2.2 Osäkerhetskällor

Ge, om möjligt, en bedömning av vilken eller vilka osäkerhetskällor som mest påverkar tillförlitligheten för objekt, populationsavgränsningar och variabelvärden i registret. En presentation av osäkerhetskällorna var för sig kan ge användarna värdefull information om hur respektive osäkerhetskälla i registret påverkar tillförlitligheten i den aktuella användningen, i första hand motsvarande osäkerhetskälla i den statistik som registret används till.

Beskriv i avsnitten K2.2.1–K2.2.5 var och en av de fem osäkerhetskällorna med betoning på respektive källas påverkan på tillförlitligheten. Om en osäkerhetskälla inte är tillämplig, skriv: Denna källa bidrar inte till osäkerhet i registret.

K2.2.1 Täckning

Ange, för varje målpopulation, brister i registrets täckning och åtföljande osäkerhet med uppdelning på över- och undertäckning. Ge om möjligt kvantitativa mått. Beskriv om det finns delpopulationer som har särskilt stora täckningsbrister.

K2.2.2 Mätning

Beskriv, så gott det går, osäkerheten i de mätningar som ligger till grund för registret. Bedöm och beskriv konsekvenserna av mätfel för registrets tillförlitlighet, både systematiska och slumpmässiga effekter. Inkludera kunskaper från uppgiftskällan. Om direktinsamling förekommer, sammanfatta de erfarenheter från insamlingen som har betydelse för bedömningen av mätfel.

K2.2.3 Bortfall

Beskriv troliga konsekvenser för registrets tillförlitlighet av bortfall, utifrån registrets huvudsakliga användningsområden.

Beskriv osäkerhet i registret utifrån objektbortfall, dvs. kända objekt för vilka information saknas. Jämför, om det är motiverat, med avsnitt K2.2.1 *Täckning*. Ange om bortfallet kan hänföras till olika delpopulationer.

Beskriv osäkerhet i registret utifrån partiellt bortfall. Ange vilka variabler och delpopulationer som har högt respektive lågt partiellt bortfall. Beskriv hur det partiella bortfallet har hanterats, om möjligt både hos uppgiftskällan och hos registerproducenten på SCB.

K2.2.4 Bearbetningar

Beskriv osäkerheten utifrån den gjorda bearbetningen med kodning, granskning och ändringar. Utnyttja erfarenheter och kunskap från tidigare produktionsomgångar och de åtgärder som då vidtagits. Använd i förekommande fall också erfarenheter från integrering med andra register.

K2.2.5 Modellantaganden

Beskriv modellerna, modellantagandena och åtföljande osäkerhet här, om det inte beskrivs under någon av osäkerhetskällorna ovan. Redovisa i förekommande fall om modellantagandena inte är helt uppfyllda och orsakerna till detta. Beskriv osäkerheten till följd av de modeller som används i framställningen av registret.

K2.3 Preliminärt register jämfört med slutligt

Beskriv skillnader mellan preliminära register och det slutliga registret ur ett tillförlitlighetsperspektiv. Ange orsaker till skillnader och konsekvenser, t.ex. vad gäller täckning, mätning och bortfall.

Om preliminära register inte tas fram, skriv: Endast slutligt observationsregister tas fram.

Om preliminära register tas fram men inte omfattas av denna information om registrets framställning och kvalitet, skriv: Denna information om registret avser endast det slutliga observationsregistret.

K3 Aktualitet och punktlighet

Ingen text behövs på den här nivån. I avsnitten nedan ska respektive delkomponent beskrivas för alla varianter och versioner av registret.

K3.1 Framställningstid

Beskriv hur lång framställningstiden är från referenstidpunkten eller referensperiodens slut till den tidpunkt då registret är tillgängligt. Om registret tas fram i olika varianter och versioner, beskriv framställningstiden för var och en av dessa. Begränsa beskrivningen till en översikt om det är många referenstider.

K3.2 Frekvens

Beskriv med vilken frekvens registret tas fram. Ange periodiciteten för insamlingen av data från olika källor, inklusive eventuell direktinsamling, samt för tillgängliggörande av registret. Förklara eventuella skillnader mellan dessa. Det kan också finnas skillnader i periodicitet mellan olika datakällor, delpopulationer och variabler. Beskriv eventuella sådana skillnader. Begränsa beskrivningen till en översikt om datakällorna eller skillnaderna är många.

Om flera olika registervarianter framställs, beskriv eventuella skillnader i frekvens mellan dem. Beskriv även eventuella skillnader i frekvens mellan olika versioner.

K3.3 Punktlighet

Ange om tillgängliggörandet är punktligt för respektive variant och version av registret. Om inte, beskriv vad som orsakade förseningen.

K4 Tillgänglighet och tydlighet

Ingen text behövs på den här nivån.

K4.1 Tillgång till registret

Beskriv hur interna respektive externa användare kan få tillgång till registret, men utan interna tekniska detaljer. Använd nedanstående standardtext om uppgifter ur registret kan lämnas ut för forskning och statistikproduktion.

Uppgifter ur <Registrets namn> kan lämnas ut för forskning och statistikändamål. Varje förfrågan till SCB om utlämnande av mikrodata sekretessprövas. SCB lyder under statistiksekretessen som återfinns i 24 kap 8 § Offentlighets- och sekretesslagen (2009:400). Vid automatiserad behandling av personuppgifter gäller även reglerna i dataskyddsförordningen.

Efter godkänd prövning kan forskare och utredare få tillgång till avidentifierade mikrodata för egna bearbetningar i t.ex. MONA.

En skriftlig begäran om de uppgifter som önskas skickas till SCB. Mer information om hur man begär uppgifter samt beställningsblankett finns på SCB:s webbplats, www.scb.se. När uppgifterna ska användas i forskning ser SCB gärna att en projektbeskrivning bifogas. Om forskningsprojektet har prövats av en regional etikprövningsnämnd bör även ansökan och beslut bifogas. I de fall som uppgifterna kan lämnas ut lämnas en offert som innehåller en beskrivning av SCB:s åtagande, leveranstidpunkt och pris.

K4.2 Informationsspridning

Beskriv översiktligt hur interna respektive externa användare informeras om registret och dess kvalitet, t.ex. genom e-post eller informationsmöten.

K4.3 Dokumentation

Beskriv vilka dokumentationer som görs för respektive variant och version av registret.

K5 Jämförbarhet och sam användbarhet

Ingen text behövs på den här nivån. Om jämförelser eller sam användning ligger nära till hands och begränsningar i möjligheter inte är uppenbara för användare, förklara det här eller i relevant(a) avsnitt nedan.

K5.1 Jämförbarhet över tid

Beskriv kort i vilken utsträckning definitioner av objekt och variabler legat fasta respektive ändrats över tiden (här eller i avsnitt U5 *Historik*). Beskriv nya, ändrade och borttagna variabler. En ändring kan gälla värdemängden jämfört med föregående produktionsomgång. Om variabelnamnet och definitionen ändras blir det en ny variabel och den ursprungliga tas bort.

Förändringar i samhället kan också ha betydelse, t.ex. ändringar i regelverk. Ange sådana förändringar kort och beskriv betydelsen för jämförbarheten över tid.

Beskriv de åtgärder som vidtagits för att mildra eventuella tidsseriebrott. Beskriv i vilken utsträckning det är möjligt att göra jämförelser över tid. Beakta ändringar i definitioner och information i avsnitt U5 *Historik* nedan. Beskriv jämförelser både på makronivå (tidsserier för registerbaserad statistik) och på mikronivå (longitudinella jämförelser) för berörda variabler.

K5.2 Sam användbarhet med andra register

Bedöm med tanke på registrets ändamål vilka register som det kan finnas intresse att sam använda detta statistiska register med. Ange den valda avgränsningen.

Beskriv i vilken utsträckning målobjekt, målpopulation och målvariabler överensstämmer eller inte. Beskriv de skillnader i metoder som bedöms ha konsekvenser för sam användning, så långt det är möjligt och rimligt.

Allmänna uppgifter

U1 Sekretess och personuppgiftsbehandling

Om sekretessbelagda uppgifter behandlas, ange förutsättningarna och de författningar som gäller för behandlingen. Använd nedanstående text:

I myndigheternas särskilda verksamhet för framställning av statistik gäller sekretess enligt 24 kap. 8 § offentlighets- och sekretesslagen ([2009:400](#)).

Om personuppgifter behandlas, ange de författningar som gäller för behandlingen. Använd nedanstående text:

Vid behandling av personuppgifter, dvs. information som direkt eller indirekt kan hänföras till en person som är i livet, gäller lagen ([2001:99](#)) och förordningen ([2001:100](#)) om den officiella statistiken samt EU:s dataskyddsförordning ([2016/679](#)).

Om registret används i uppdragsverksamheten för mikrodatautlämnade, använd nedanstående text och skriv in registrets namn.

Aidentifierade mikrodata från <Registrets namn> kan lämnas ut till forskning och statistikändamål efter sekretessprövning om SCB bedömer att det finns stöd för att behandla uppgifterna.

U2 Bevarande och gallring

Ange hur uppgifterna i det slutliga observationsregistret hanteras avseende bevarande och gallring. Hänvisa till relevanta regelverk och beslut, t.ex. lag, föreskrifter eller beslut från Riksarkivet. Om länkning till dokument eller webbsidor görs, tänk på att länkar kan vara instabila över tid. Viktiga dokument kan behöva bifogas även som bilaga.

Ange ansvarig registerproducent och utförare av aidentifiering, gallring och arkivering. Ange även beslut och beslutsdatum samt tidsfrist för gallring/ arkivering.

I de undantagsfall där beslut ännu inte finns, använd nedanstående text.

Bevarandebehov är under utredning.

U3 Uppgiftsskyldighet

Ange om uppgiftsskyldighet gäller för hela eller delar av registret, samt i så fall vilka delar.

Om uppgiftsskyldighet gäller, ange de författningar uppgiftsskyldigheten grundar sig på. Länka gärna till författningar eller föreskrifter. Om länkning till dokument eller webbsidor görs, tänk på att länkar kan vara instabila över tid. Viktiga dokument kan behöva bifogas även som bilaga.

Använd nedanstående text om data hämtas in för den officiella statistiken från en extern myndighet.

Uppgiftsskyldighet föreligger enligt förordning om den officiella statistiken (2001:100), vilket innebär att myndigheter ska lämna de uppgifter som behövs för framställning av statistiken.

Om uppgiftsskyldighet inte föreligger, ange detta enligt nedan.

Uppgiftsskyldighet till framställning av annan statistik än den officiella föreligger inte.

U4 EU-reglering och internationell rapportering

Ange relevanta EU-regleringar. Hänvisa till Europaparlamentets och rådets förordning. Komplettera med andra EU-regleringar. Om länkning till dokument eller webbsidor görs, tänk på att länkar kan vara instabila över tid. Viktiga dokument kan behöva bifogas även som bilaga.

Om ingen EU-reglering förekommer, använd nedanstående text.

Ingen EU-reglering finns för <Registrets namn>.

Beskriv kort eventuell rapportering till internationella organisationer.

Beskriv vad regleringar och internationell rapportering innebär för registret.

U5 Historik

Ange när registret framställdes första gången. Beskriv kortfattat större förändringar som inte framgår av avsnitt *K5 Jämförbarhet och sam användbarhet*.

U6 Kontaktuppgifter

Ange kontaktuppgifter till registerproducenten, mikrodatautlämnade samt statistikuppdrag. Använd gruppbrevlåda där det finns.

Information om det statistiska registret	
Kontaktperson	
Enhet	
E-post	
Telefon	

Mikrodatautlämnande	
Kontaktperson	
Enhet	
E-post	
Telefon	

Statistikuppdrag	
Kontaktperson	
Enhet	
E-post	
Telefon	

Ordlista

Ordlistan innehåller begrepp och termer som används om och för statistiska register.

Begrepp	Förklaring
Administrativa data, administrativt register	Data respektive register som används i ett administrativt informationssystem, t.ex. ett företags lönesystem, Skatteverkets folkbokföringssystem och olika skattesystem.
Framställningstid	Tidsavståndet från referenstidpunkten, eller referensperiodens slut, till den tidpunkt när registret tillgängliggörs.
Identifierande variabel	En variabel som används för att entydigt identifiera objekt. Motsvarande IT-term är primärnyckel.
Imputering	Ett förfarande där ett saknat, eller orimligt, värde ersätts med ett beräknat värde.
Intresseobjekt, intressepopulation, intressevariabel	En objekttyp, en population respektive en variabel som följer av användarnas behov och önskemål.
Integrering av register	En sammanvändning av två eller flera register med ett eller flera nya register som resultat.
Kodning	Regler används för att kategoriindela variabelvärden som t.ex. inhämtats med öppna svar. Siffervärden eller andra koder används.
Konsistenskontroll	Samgranskning av mikrodata från flera register eller datakällor i syfte att upptäcka inkonsistenser avseende objekt och variabler.
Kopplingsvariabel	En variabel som används för entydig identifiering av enskilda objekt eller grupper av objekt.
Mikrodata	En uppsättning enskilda objekt med variabelvärden.

Modellbaserad beräkning	En beräkning som bygger på ett eller flera antaganden, t.ex. genom en modell som beskriver samband mellan variabler.
Målobjekt, målpopulation, målvariabel	En objekttyp, en population respektive en variabel som registret är avsett att innehålla
Objektbortfall	Ett objekt som det saknas observationer för, men objektet är känt som objekt.
Objekttyp	Ett slag av objekt, t.ex. person, hushåll, arbetsställe, fastighet.
Observationsobjekt	Ett objekt som det samlas in uppgifter om.
Observationsvariabel	En variabel som det samlas in värden på.
Observationsregister	Ett observationsregister kan ses som en eller flera tabeller där rader och kolumner innehåller objekt respektive variabler.
Partiellt bortfall	Värden saknas för vissa observationsvariabler för ett observationsobjekt.
Population	Avgränsning i tid och rum av en mängd objekt av samma typ och med vissa egenskaper.
Preliminärt register	En registerversion i preliminär form.
Processdata	Information om en process, t.ex. om en delprocess i granskningen.
Referenstid	Objekt, populationer och variabler har, var och en, referenstider i form av tidsperioder eller tidpunkter.
Register	En förteckning över objekt; dessa bör vara identifierbara och tillhöra en uttalad objektmängd. Ett register innehåller normalt även variabler.
Registervariant	En delmängd av ett register vad gäller population eller variabelinnehåll för att tillgodose ett visst behov.

Registerversion	Ett fysiskt register som kan betraktas som en realisering av en register-variant till en viss tidpunkt eller tidsperiod.
Samgranskning	En gemensam granskning av mikro-data från flera register eller data-källor, med syfte att få ett bättre resultat än vid separata granskningar, t.ex. att upptäcka inkonsistenser avseende objekt och variabler.
Slutligt observationsregister	En slutlig version av ett observationsregister. (Ett slutligt observationsregister för en <i>statistikprodukt</i> innehåller det datamaterial som behövs för att framställa den redovisade statistiken.)
Standardiserad population, standardiserad variabel	En population respektive en variabel för vilken det finns ett utpekat ansvar för definition, namngivning och dokumentation.
Statistiskt register	Ett register som är utformat för statistiska ändamål, dvs. för att användas i statistikproduktion. Det utgår t.ex. från ett eller flera administrativa register.
Täckningsbrister	Skillnader mellan registrets objekt-mängd och dess målpopulation.
Undertäckning	Registret saknar objekt som tillhör målpopulationen.
Variabel	En mätbar egenskap hos ett objekt.
Övertäckning	Registret innehåller objekt som inte tillhör målpopulationen.