

Longitudinell individdatabas (LINDA)

2015

LE1900

Innehåll

0	Allmänna uppgifter	2
0.1	Ämnesområde	2
0.2	Statistikområde	2
0.3	SOS-klassificering	2
0.4	Statistikansvarig	2
0.5	Statistikproducent	2
0.6	Uppgiftsskyldighet	2
0.7	Sekretess och regler för behandling av personuppgifter	2
0.8	Gallringsföreskrifter	3
0.9	EU-reglering	3
0.10	Syfte och historik	3
0.11	Statistikanvändning	3
0.12	Uppläggning och genomförande	3
0.13	Internationell rapportering	4
0.14	Planerade förändringar i kommande undersökningar	4
1	Översikt	4
1.1	Observationsstorheter	4
1.2	Statistiska målstorheter	4
1.3	Utfloeden: statistik och mikrodata	7
1.4	Dokumentation och metadata	7
2	Uppgiftsinsamling	7
2.1	Ram och ramförfarande	8
2.2	Urvalsförfarande	10
2.3	Mätinstrument	11
2.4	Insamlingsförfarande	12
2.5	Databeredning	13
3	Statistisk bearbetning och redovisning	16
3.1	Skattningar: antaganden och beräkningsformler	16
3.2	Redovisningsförfaranden	19
4	Slutliga observationsregister	19
4.1	Produktionsversioner	19
4.2	Arkiveringsversioner	19
4.3	Erfarenheter från senaste undersökningsomgången	20

0 Allmänna uppgifter

0.1 Ämnesområde

Ämnesområde: Hushållens ekonomi

0.2 Statistikområde

Statistikområde: Inkomst och inkomstfördelning

0.3 SOS-klassificering

Tillhör (SOS): Statistiken ingår inte i Sveriges officiella statistik (SOS) men ingår i systemet för Sveriges officiella statistik (SOS).

För undersökningar som ingår i Sveriges officiella statistik gäller särskilda regler när det gäller kvalitet och tillgänglighet, se Förordningen om den officiella statistiken (2001:100).

0.4 Statistikansvarig

Myndighet/organisation: Statistiska centralbyrån (SCB)
Postadress: 70189 ÖREBRO
Besöksadress: Klostergatan 23, Örebro
Kontaktperson: Stina Åsling Rönning
Telefon: 010 – 479 62 07
E-post: fornamn.efternamn@scb.se

0.5 Statistikproducent

Myndighet/organisation: Statistiska centralbyrån (SCB)
Postadress: 70189 ÖREBRO
Besöksadress: Klostergatan 23, Örebro
Kontaktperson: Stina Åsling Rönning
Telefon: 010 – 479 62 07
E-post: fornamn.efternamn@scb.se

0.6 Uppgiftsskyldighet

Uppgiftsskyldighet föreligger inte enligt lagen om den officiella statistiken (SFS 2001: 99).

0.7 Sekretess och regler för behandling av personuppgifter

I myndigheternas särskilda verksamhet för framställning av statistik gäller sekretess enligt 24 kap. 8 § offentlighets- och sekretesslagen (2009:400). Vid automatiserad behandling av personuppgifter gäller

reglerna i personuppgiftslagen (1998:204). På statistikområdet finns dessutom särskilda regler för personuppgiftsbehandling i lagen (2001:99) och förordningen (2001:100) om den officiella statistiken.

0.8 Gallringsföreskrifter

Syftet med LINDA är longitudinella studier, därför gallras inte registret.

0.9 EU-reglering

Statistiken är inte EU-reglerad.

0.10 Syfte och historik

LINDA är en urvalsbaserad longitudinell individdatabas som i första hand riktar sig till forskare. Det huvudsakliga syftet med databasen är att ge forskare möjligheten att studera inkomstutveckling för olika grupper under en lång tid. LINDA innehåller huvudsakligen uppgifter om inkomster, som t.ex. lön, inkomst av näringsverksamhet, kapitalinkomst, skattepliktiga och skattefria transfereringar. LINDA uppdateras en gång om året. Databasen skapades ursprungligen på uppdragsbasis i ett samarbete mellan Statistiska centralbyrån, Riksförsäkringsverket (numera Försäkringskassan), Uppsala Universitet och Finansdepartementet under senare delen av 1990-talet. Undersökningen blev senare en anslagsprodukt.

Sfi-bonus upphörde under 2014 och inga bonusar betalas ut under 2015. Denna variabel har ingått i disponibel inkomst, men är numera alltså borttagen.

0.11 Statistikanvändning

Användare utgörs bl.a. av statliga verk, departement, organisationer, högskolor och universitet. Exempel på användningsområden är analys av välfärdssystemet.

LINDA diskuteras i *Användarrådet för välfärdsstatistik*. Deltagande organisationer är Finansdepartementet, Socialdepartementet, Socialstyrelsen, Institutet för social forskning (SOFI), Institutet för bostads- och urbanforskning, Arbetsmarknadsdepartementet, Uppsala universitet, Konsumentverket och Boverket.

0.12 Uppläggning och genomförande

LINDA (Longitudinell individdatabas) är en urvalsbaserad undersökning och uppgifter till LINDA hämtas i första hand från den totalräknade inkomststatistikens årsregister (IoT), Utbildningsregistret samt från Registret över totalbefolkningen (RTB). Från folk- och bostadsräkningarna, FoB, hämtas även uppgifter för respektive FoB-år (1960,-65, -70, -75, -80, -85 och -90). Utbildningsuppgifter från Högskoleregistret (77–95) och Registret över komvuxstuderande (85–95) samt uppgifter om löner från Strukturlönestatistiken (98–). Från Försäkringskassan (FK) hämtas uppgifter om pensioner och pensionspoäng (91–97), sjukfall (97–07) och föräldrapenning (97–07).

Uppgifter om arbetssökande hämtas från Arbetsförmedlingens (Af) Datalager (91–07) där möjligheten finns att lägga till nya årgångar. En årgång har ca 13 månaders produktionsstid vilket innebär att årgång 2015 är fullständig i februari 2017.

0.13 Internationell rapportering

Statistiken rapporteras inte internationellt.

0.14 Planerade förändringar i kommande undersökningar

Inga planerade förändringar finns för kommande undersökning, förutom de förändringar som eventuella lagändringar vad gäller skatter och bidrag medför.

1 Översikt

Målpopulationen utgörs av samtliga individer som enligt lagar, förordningar och andra regler som gäller för folkbokföringen skulle ha varit folkbokförda i landet någon gång under undersökningsåret, samt deras familjemedlemmar. I vissa fall förekommer även dödsbon och personer som är skattskyldiga i Sverige men som saknar hemortskommun i riket. I LINDA finns för varje individ uppgift om bl.a. skattepliktiga/skattefria transfereringar, skatt, arbetsinkomst, inkomst av näringsverksamhet, faktorinkomst m.m. Det totala antalet variabler sett över åren är drygt 1 000. SCB framställer ingen statistik från LINDA. Allt material ligger som mikrodata för specialbearbetningar efter beställning. LINDA är dokumenterad enligt SCB:s dokumentationssystem.

1.1 Observationsstorheter

Målpopulationen utgörs av samtliga individer som enligt lagar, förordningar och andra regler som gäller för folkbokföringen skulle ha varit folkbokförda i landet någon gång under undersökningsåret, samt deras familjemedlemmar. I vissa fall förekommer även dödsbon och personer som är skattskyldiga i Sverige men som saknar hemortskommun i riket.

Rampopulationen utgörs av de som är folkbokförda någon gång under undersökningsåret, samt deras familjemedlemmar. För att kunna dra några statistiska slutsatser om målpopulationen används ett urval som består av tre procent av hela befolkningen och ett urval som består av 20 procent av de utrikesfödda.

Undersökningen har två objekt, hushåll och individer. Hushåll består av individer från LINDA och alla individer i LINDA tillhör ett hushåll.

Objektet hushåll har följande observationsstorheter

- Antal personer i hushållet
- Hushållets disponibla inkomst
- Folkbokföringsort

Objektet individer har följande observationsstorheter

- Kön
- Ålder
- Civilstånd
- Inkomst
- Skatt
- Högsta utbildning
- Avgifter
- Avdrag

1.2 Statistiska målstorheter

SCB producerar ingen statistik från LINDA.

I LINDA finns för varje individ uppgift om bl.a. skattepliktiga/skattefria transfereringar, skatt, arbetsinkomst, inkomst av näringsverksamhet, faktorinkomst m.m.

Det totala antalet variabler sett över åren är drygt 1 000. Alla variabler som återfinns i undersökningen kan därför inte specificeras här. Alla variabler finns ej i alla årgångar vilket beror på att inkomstsystemet ändras över tid. I nedanstående tablå återfinns däremot en grov beskrivning av några områden inom vilka det finns uppgifter i undersökningen. Detaljerad information om mikrodata finns beskrivet i *Dokumentation av mikrodata (MetaPlus)* på SCB:s webbplats, www.scb.se/LE1900.

Område	Beskrivning
Bakgrundsvariabler	Civilstånd, folkbokföring, mantalsskrivning, samhörighet, medborgarskap, invandringsår, kön, ålder, födelsemånad, utbildning m.m.
Inkomstvariabler	Arbetsinkomst, faktorinkomst, disponibel inkomst, sammanräknad förvärvsinkomst, kapitalinkomst m.m.
Negativa transfereringar	Skatter och övriga negativa transfereringar
Positiva transfereringar skattepliktiga	<p><i>Ersättningar i samband med sjukdom</i> Sjukpenning, rehabiliteringsersättning m.m.</p> <p><i>Pensioner och livräntor</i> Pension, delpension, livränta m.m.</p> <p><i>Föräldrapenning</i> Föräldrapenning tillfällig vård av barn, föräldrapenning vid barns födelse</p> <p><i>Studiestöd</i> Studiepenning, utbildningsbidrag för doktorander m.m.</p> <p><i>Arbetsmarknadsstöd</i> Totalt arbetsmarknadsstöd</p> <p><i>Övriga positiva transfereringar</i> Vårdbidrag, närstående penning, dagpenning vid repetitionsutbildning och civilförsvarsutbildning m.m.</p>
Positiva transfereringar skattefria	<p><i>Ersättningar i samband med sjukdom</i> Sjukpenning (skattefri)</p> <p><i>Pensioner och livräntor</i> Pension, livränta skattefri del, frivillig pension, särskilt pensionstillägg, barnpension skattefri del</p> <p><i>Familjestöd</i> Ekonomiskt bistånd, bostadsbidrag för pensionärer, bostadsbidrag för barnfamiljer samt makar och ensamstående utan barn, barnbidrag och flerbarnstillägg, mottaget bidragsförskott (underhållsstöd)</p> <p><i>Studiestöd</i> Studiemedel (lån och bidrag), vuxenstudiestöd lån m.m.</p>
Variabler som ligger till grund för beskattning och variabler som påverkar beskattningen	Fastighetsskatter, pensionsavgifter, sjukförsäkringsavgifter, grundavdrag, skuldräntor, särskilt grundavdrag, realisationsförluster, allmänna avdrag m.m.

Demografiska uppgifter som härrör från folkbokföringen avser senast kända uppgift för respektive år. Detta innebär att merparten av individerna har referenstid den 31 december aktuellt år. Personer som lämnat landet under året har de uppgifter som gällde vid det datum då de avskrevs från folkbokföringen. Utbildningsvariablerna avser vårterminen. För övriga variabler är referenstiden respektive kalenderår. För longitudinella studier är referensperioden 1968–2015.

1.3 Utflöden: statistik och mikrodata

SCB framställer ingen statistik från LINDA. Allt material ligger som mikrodata för specialbearbetningar efter beställning.

Det slutliga observationsregistret, se kapitel 4, är identifierbart på individnivå. Aidentifierade mikrodata kan göras tillgängligt via SCB:s leveranssystem MONA, efter en sekretessprövning.

1.4 Dokumentation och metadata

LINDA är dokumenterad enligt SCB:s dokumentationssystem. Framställningen av statistikregistret beskrivs i *Dokumentation av statistiken (SCBDOK, innevarande dokument)*. Detaljerad information om mikrodata finns beskrivet i *Dokumentation av mikrodata (MetaPlus)*. Samtliga dokumentationer finns på SCB:s webbplats, www.scb.se/LE1900. Utöver detta finns variabelförteckningar och postbeskrivningar som kan fås efter kontakt med SCB.

2 Uppgiftsinsamling

SCB:s register över totalbefolkningen (RTB) utgör rampopulationen. Personerna måste någon gång under det aktuella året ha varit folkbokförd i Sverige för att ingå i LINDA:s rampopulation. För inkomståret 1994 drogs ett obundet slumpmässigt urval (OSU) om 300 000 personer (ca 3 %) bland samtliga personer som funnits i Sverige under inkomståret. Dessa urvalspersoner bildar en panel som följs både bakåt till 1968, första året som SCB har totalräknad inkomststatistik, och framåt. Urvalet uppdateras årligen så att det vid varje tillfälle motsvarar befolkningen respektive år. Ytterligare ett panelurval har skapats, omfattande ca 20 % av personer som är utrikesfödda.

För LINDA-urvalet finns två önskemål uppsatta. Varje årgång av LINDA ska utgöra ett tvärsnittsurval över alla individer i Sverige under aktuellt år och det ska gå att följa en urvalsindivid under alla år som individen ingått i urvalsramen. LINDA ska alltså även utgöra ett panelurval.

LINDA baseras på ett antal andra SCB-register; Inkomst- och taxeringsregistret (IoT), Utbildningsregistret, Registret över totalbefolkningen (RTB) och Strukturlönestatistik för offentlig sektor och delar av privat sektor. Dessa register har administrativa data som sin källa. Utöver dessa register hämtas uppgifter från enkätundersökningen Longitudinella studier löner (LSL).

2.1 Ram och ramförfarande

SCB:s register över totalbefolkningen (RTB) utgör rampopulationen med försumbara täckningsfel. Personerna måste någon gång under det aktuella året ha varit folkbokförd i Sverige för att ingå i LINDA:s rampopulation.

Följande kvalitetbrister vad gäller täckning av de folkbokförda i RTB:

Undertäckning

Brister i rapporteringen av födslar och invandring resulterar i undertäckning.

Födslar ger i praktiken upphov till mycket små problem, vilket beror på de snabba rapporteringsrutinerna. Skatteverket registrerar 94 procent av födslarna inom 2 dagar. Inom en månad har i princip 100 procent registrerats. För nyfödda uppskattas undertäckningen i observationsregistren vara 0,1 procent av 0-åringarna. Någon uppskattning av undertäckningens storlek på grund av invandring saknas. Invandrare blir registrerade som invandrade vid den tidpunkt då de folkbokförs i Sverige. I vissa fall kan det gå en relativt lång tid mellan ankomsten till Sverige och tidpunkten för folkbokföring. Det beror på att det kan ta lång tid för personer med utomnordiskt medborgarskap innan de får uppehållstillstånd. Undertäckningen uppstår också om en person invandrar till Sverige och personen enligt lagar och förordningar ska folkbokföra sig men inte gör det. Det finns dock starka incitament för en person att faktiskt folkbokföra sig. Den som inte är folkbokförd saknar många rättigheter, t.ex. kan man inte få barnbidrag eller öppna ett bankkonto. Detta talar för att undertäckningen troligen är mycket liten.

Övertäckning

Brister i rapporteringen av dödsfall och utvandring resulterar i övertäckning. Dödsfall ger upphov till mycket små problem beroende på de snabba rapporteringsrutinerna. Inom en månad har i princip 100 procent registrerats. Övertäckningen i observationsregistren med anledning av ej rapporterade dödsfall antas vara mindre 0,1 procent av de döda.

Utvandrare förorsakar övertäckning om utflyttning inte anmäls. Genomförda studier tyder på att folkbokföringen och därmed RTB innehåller ett betydande antal personer som inte längre bor i landet. Detta är förmodligen den allvarligaste kvalitetsbristen i såväl folkbokföringen som i RTB. De flesta av de som flyttar ut ur landet har inget eget intresse av att anmäla utflyttning och en del gör det heller inte. Det kan till och med vara fördelaktigt att behålla sin folkbokföring i Sverige.

Flera försök har gjorts för att skatta övertäckningen, bl.a. genom att jämföra dödsrisker för personer med svensk respektive utländsk bakgrund samt genom analys av bortfall och postreturer i olika undersökningar. I en studie av Nilsson (1994:09), framförs hypotesen att övertäckningen bland utrikes födda kan vara så stor som 10 procent. Andra analyser pekar på att omfattningen av övertäckningen kan ligga på 25 000–50 000 personer, vilket utgör 0,25–0,5 procent av hela befolkningen.

Ett försök till indirekt skattning av övertäckningen har gjorts i en studie av Qvist (1999:06). I studien analyseras dödligheten hos utrikes födda personer för att indirekt skatta övertäckningen. Om dessa personer har alltför låg dödlighet i förhållande till inrikes födda personer för jämförbara grupper, kan det tyda på fel i folkmängden i grupperna. Den lägre dödligheten kan alltså tolkas som att det finns övertäckning i RTB. Beräkningar visar att övertäckningen bland utomnordiska invandrare ligger i intervallet 25 000–50 000 personer, vilket motsvarar 4–8 procent av de utomnordiska invandrarna.

Under 2004 har ett arbete gjorts där man försökt att skatta storleken på övertäckningen genom att skapa en övertäckningsvariabel. Idén bakom övertäckningsvariabeln är att personer som är bosatta i Sverige lämnar avtryck i olika administrativa register. De kan t.ex. ha inkomst, studera eller byta bostad. Personer som saknar avtryck i registren, trots att de är folkbokförda, kan då antas vara utflyttade från landet utan att rapportera det till Skatteverket. Dessa personer får en markering i övertäckningsvariabeln. Övertäckningsvariabeln använder årsvis information från tre interna register, baserade på administrativa register från andra myndigheter: Registret över totalbefolkningen (RTB), Inkomst- och taxeringsregistret (IoT) samt STATIV. Beräkningen då gav en övertäckning på 35 000 personer 2002.

Skatteverket har uppmärksammat problemet med folkbokföringsfelet i en rapport av Johansson (2006:11). Bl.a. tittar man på hur många personer som är maskinellt nolltaxerade (person där Skatteverket inte har någon indikation om ekonomisk aktivitet i form av kontrolluppgifter eller bedriven näringsverksamhet) som en möjlig indikation på att man inte längre finns i landet. Uppskattningen av övertäckningen blev då 25 000–35 000 i yrkesverksam ålder. I rapporten ger Skatteverket förslag på åtgärder som ska minska folkbokföringsfelet, t.ex. lagändring som möjliggör ett utökat samarbete mellan berörda myndigheter. Inom Skatteverket skulle kontrollen bli effektivare om folkbokföringen fick tillgång till skatteregistren.

I en rapport av Bengtsson (2010:5) konstateras att övertäckningen i RTB har ökat under 2000-talet. Rapporten kan dock inte ge underlag för att mer precist bedöma storleken av övertäckningen.

Resultatet från den senaste studien finns i en rapport av Åsling Rönning m.fl. (2015:1). I det arbetet skapas en modell för att skatta övertäckningens storlek. Grunden i arbetet har varit att vidareutveckla den övertäckningsvariabel som tidigare projekt från 2004 har utvecklat för att mäta övertäckningen registervägen. Projektet har sedan tagit fram en skattningsmodell för att beräkna övertäckningen. Först sker en avgränsning av de personer som skulle kunna vara tänkbara att ingå i övertäckningen. Gruppen består av personer som ingår i den justerade övertäckningsvariabeln, dessutom tillkommer personer som har mindre än ett prisbasbelopp i inkomst i Inkomst- och taxeringsregistret. Även information från andra register används, t.ex. gränspendling. För att därefter avgöra om en person ingår i övertäckningen eller i RTB-populationen används ett antal olika indikationer som kan peka på om en

person troligen tillhör övertäckningen eller inte. Indikationerna skapas från ett regelverk som kan bygga på registerdata från ett enskilt år eller från flera på varandra följande år.

Modellskattad övertäckning i RTB

År	Modellskattad övertäckning
2000	(35 241)
2001	39 096
2002	39 322
2003	39 973
2004	39 700
2005	42 179
2006	40 816
2007	40 864
2008	45 761
2009	55 591
2010	60 640
2011	(64 755)
2012	(73 355)

Skattningarna för 2000, 2011 och 2012 är osäkrare än övriga år eftersom indikatorer baserade på observationer före eller efter referenstidpunkten inte är tillgängliga. Ytterligare en nackdel med den modellen är att resultatet inte blir tillgängligt förrän 1,5 år efter referenstidpunkten.

2.2 Urvalsförfarande

För inkomståret 1994 drogs ett obundet slumpmässigt urval (OSU) om 300 000 personer (ca 3 %) bland samtliga personer som funnits i Sverige under inkomståret. Dessa urvalspersoner bildar en panel som följs både bakåt till 1968, första året som SCB har totalräknad inkomststatistik, och framåt. Urvalet uppdateras årligen så att det vid varje tillfälle motsvarar befolkningen respektive år. Ytterligare ett panelurval har skapats, omfattande ca 20 % av personer som är utrikesfödda.

Från och med årgång 1991 ingår i urvalsramarna alla som någon gång under året bott i Sverige, inklusive dödsbon. För att kunna särskilja dödsbon från fysiska personer kan variabeln Objekttyp användas. Denna har värdet 03 för dödsbon. Om man endast är intresserad av individer som fanns i Sverige vid årsskiftet så finns det från och med 1993 en variabel som markerar om individen fanns i RTB vid årsslutet eller inte.

Detta innebär att urvalet innehåller alla som bott i Sverige under hela året, alla som någon gång har bott i landet under året, alla som avlidit, alla som utvandrat, alla som invandrat under året samt alla dödsbon. I urvalsramarna har dödsbon ingått men alla inkomststoppgifter som avsåg dessa har ej funnits bevarade på registren vilket innebär att alla dödsbon

som ingår i urvalet saknar värden på variablerna från SCB:s taxeringsregister.

För år 1990 och tidigare saknas uppgifter om barn 0–15 år i Inkomst- och taxeringsregistret. För att få en heltäckande urvalsram även för dessa år kompletteras urvalsramen med alla 0–15 åringar från SCB:s register över totalbefolkningen (RTB). Kompletteringen har gjorts genom att RTB för den 31/12 det aktuella året har använts. Detta innebär att individer mellan 0–15 år som avlidit eller utvandrat under det aktuella året ej har kommit med i urvalsramen under året.

För LINDA-urvalet finns två önskemål uppsatta. Varje årgång av LINDA ska utgöra ett tvärsnittsurval över alla individer i Sverige under aktuellt år och det ska gå att följa en urvalsindivid under alla år som individen ingått i urvalsramen. LINDA ska alltså även utgöra ett panelurval.

En urvalsmetod för att uppnå dessa båda syften är Poisson-urval med permanenta slumpstal. Grundidén i metoden är att alla individer i urvalsramen tilldelas ett slumpstal. Detta slumpstal behåller individen under alla år som den ingår i urvalsramen. Nyttillkomna individer i urvalsramen (invandrade eller nyfödda) tilldelas också ett slumpstal. Detta slumpstal används sedan för att göra urvalet. Urvalet görs genom att man bestämmer ett intervall $0 \leq a < b \leq 1$ och alla individer i intervallet a – b kommer sedan att ingå i urvalet. För att få samordning mellan åren kommer intervallgränserna att vara samma mellan åren.

Som första urvalsram har 1994 års skattedatabas från skatteverket används. Alla individer i denna databas har tilldelats ett slumpstal rektangulärt fördelat mellan 0 och 1. Databasen har sedan sorterats i storleksordning efter slumpstalen och de 300 000 individerna med lägst slumpstal har sedan valts ut. Värdet på slumpstalet för den individ bland de utvalda med högst värde på slumpstalet har sparats. Detta urval på 300 000 personer har sedan kompletterats med alla personer som ingår i urvalsindividens hushåll.

2.3 Mätinstrument

LINDA baseras på ett antal andra SCB-register; Inkomst- och taxeringsregistret (IoT), Utbildningsregistret, Registret över totalbefolkningen (RTB) och Strukturlönestatistik för offentlig sektor och delar av privat sektor. Dessa register har administrativa data som sin källa. Utöver dessa register hämtas uppgifter från enkätundersökningen Longitudinella studier löner (LSL).

I LSL-undersökningen används två olika typer av enkäter. Företag med 30 eller fler anställda skickar vanligen in uppgifterna via en Excel-fil. De två olika pappersenkäterna innehåller samma frågor; den ena innehåller frågor för tre personer medan den andra vanligare bara innehåller frågor för en person.

2.4 Insamlingsförfarande

I början av varje år hämtas uppgifter från olika register inom SCB för att sedan bilda LINDA, se avsnitt 2.3 *Mätinstrument* för vilka register som ingår.

Undersökningen *Longitudinella studier löner* samlas in av SCB via internet med hjälp av SCB:s elektroniska insamlingsverktyg SIV. Insamlingen kombineras med en pappersenkät.

Löneenkätundersökningens genomförande

SCB har av Medlingsinstitutet fått till uppgift att göra ett longitudinellt register byggt på löneuppgifter. Regeringen beslöt därför att komplettera dessa uppgifter på LINDA:s urvalspersoner. Detta sker dels via befintliga register och dels via enkätundersökningen *Longitudinella studier löner* (LSL).

I början av mars 2016 fick LSL-undersökningen personnummer på de 333 786 individer som ingår i 2015 års urvalspanel. Personer med skyddad identitet samt avlidna personer är exkluderade i insamlingen. När kontrolluppgiftsregistret (KU 2015) blev klart kunde bearbetningarna av urvalspersonerna påbörjas.

Personer utanför intervallet 18 – 66 år exkluderas (204 485 personer) från urvalet. Vidare exkluderas de som inte arbetat i Sverige under 2015 eller tjänat mindre än 10 000 SEK hos en arbetsgivare under året (159 586 personer). Även de personer som inte arbetat under november eller december exkluderas (150 465). Personer med arbetsställesnummer 99990 (personer utan normalt anställningsförhållande), 99993 (utlandsstationerad personal), 99994 (anställda som har den egna bostaden som arbetsplats) och 99998 (person med garantilön) exkluderas också. Därmed återstår 149 695 personer för vilka uppgifter ska samlas in. Varje person kopplas till en arbetsgivare. Om det finns uppgift om flera arbetsgivare i KU-registret väljs den anställning med högst lön. En av anställningarna väljs ut med hjälp av tidsmarkeringen på kontrolluppgifterna genom att ta den anställning/kontrolluppgift som har längst antal månader under året, dvs. den kontrolluppgift med högst lön per månad. En kontroll görs om det möjligen är samma företag som bytt organisationsnummer under året och därför lämnat två kontrolluppgifter.

Av de 149 695 personerna är 136 016 med i föregående års undersökning. Personerna matchas sedan mot Strukturlöneundersökningen för att föra på löneuppgifter.

42 665 personer förväntas återfinnas i Strukturlöneundersökningen för den offentliga sektorn.

37 379 personer förväntas återfinnas i Strukturlöneundersökningen för den privata sektorn.

17 312 personer förväntas återfinnas i företag som tillhör arbetsgivarorganisationer som SCB samarbetar med.

2 438 personer återfinns i form av överskottsuppgifter ifrån ordinarie lönestatistiska undersökningar från Strukturlöneundersökningen. Det kan

röra sig om uppgifter som har kommit in från andra arbetsgivare än förväntat.

Företag knutet till försvaret, ambassadpersonal, företag utan adress i Företagsdatabasen, utländska företag, samt semesterkassor exkluderas. Det återstår då 48 218 personer på 33 400 företag till vilka LSL-enkäten skickas.

I LSL-enkäten har uppgifter för 40 829 personer inkommit och godkänts i granskningen. Det *kända* bortfallet utgjordes av 328 personer. För dessa personer saknas uppgifter, men hänvisning som t.ex. ”vägrare”, ”företaget nedlagt, konkurs” etc. finns. Det kan även vara så att uppgifter med bristande kvalitet inkommit, utan att återkontakt varit möjlig i samband med granskning. Det *okända* bortfallet utgjordes av 7 061 personer. För dessa personer saknas uppgifter helt.

Ytterligare bortfall utgörs av de personer som förväntats återfinnas hos de arbetsgivarorganisationer för vilka en separat komplettering görs. Dessa uppgår till 1 419 personer.

2.5 Databeredning

Granskning av LINDA sker genom jämförelser med tidigare årgångar av LINDA. Källregistren samt enkätundersökningen granskas hos respektive producent på SCB före leverans till LINDA.

I LINDA skapas också variabler som inte finns i något annat register. Dessa variabler är bearbetningar av redan befintliga variabler men anpassas för att kunna göra mer longitudinella jämförelser. Det handlar om variablerna:

- arbetsinkomst,
- faktorinkomst,
- negativa transfereringar,
- slutgiltig skatt,
- övriga negativa transfereringar,
- positiva transfereringar,
- disponibel inkomst,
- studiestöd,
- familjestöd,
- pension,
- arbetsmarknadsstöd,
- föräldrapenning,
- sjuk och aktivitetsersättning,
- sjukpenning,
- övriga positiva transfereringar.

Variablerna finns i två olika varianter i LINDA. Dels variablerna från Inkomst- och taxeringsregistret och dels de skapade LINDA-variablerna. För att kunna skilja dem åt avslutas varje skapad variabel med "LINDA" i variabelnamnet. I tabblån nedan beskrivs de bearbetningar som görs för inkomstvariablerna

Variabel i LINDA	Bearbetning	Grundvariabel
Slutgiltig skatt LINDA	Kopia av	Slutgiltig skatt
Arbetsinkomst LINDA	Kopia av	Arbetsinkomst
Faktorinkomst LINDA	Kopia av	Faktorinkomst
Övriga negativa transfereringar LINDA	Summering av	Återbetalt belopp studielån Givet underhållsbidrag
Negativa transfereringar LINDA	Summering av	Slutgiltig skatt LINDA, Övriga negativa transfereringar LINDA
Disponibel inkomst LINDA	Summering av	Faktorinkomst LINDA Skattepliktiga positiva transfereringar LINDA Skattefria positiva transfereringar LINDA
	minus	Negativa transfereringar LINDA
Övriga positiva transfereringar, Skattepliktiga LINDA	Summering av	Vårdbidrag, Närståendepenning, Dagpenning för frivilliga vid krigsförbandsövning, Bidrag från Sveriges författarfond
Studiestöd, skattepliktig LINDA	Summering av	Utbildningsbidrag för doktorander, Teckenspråksundervisning för föräldrar
Familjestöd LINDA	Summering av	Socialbidrag, bostadstillägg för pensionärer och personer som har sjuk- och aktivitetsersättning, särskilt bostadstillägg för pensionärer och personer som har sjuk- och aktivitetsersättning, Bostadsbidrag, barnbidrag, äldreförsörjningsstöd, underhållstöd,

		Jämställdhetsbonus, vårdnadsbidrag, bostadsersättning, etableringsersättning, etableringstillägg, boendetillägg
Skattepliktig pension LINDA	Summering av	Arbetskadelivränta, Tjänstepension, Summa allmän pension/tjänstepension, privat pensionsförsäkring,
Arbetsmarknadsstöd LINDA	Kopia av	Arbetsmarknadsstöd
Föräldrapenning LINDA	Summering av	Tillfällig vård av barn, föräldrapenning vid barnets födelse
Sjuk och aktivitetsersättning LINDA	Kopia av	Sjuk och aktivitetsersättning
Sjukpenning LINDA	Summering av	Sjukpenning, Smittbärrpenning, Graviditetspenning, Rehabiliteringsersättning, Arbetsskadeersättning, Ersättning som utgör inkomst av anställning
Skattefria positiva transfereringar LINDA	Summering av	Utvecklingsersättning, Skattefri pension LINDA, Familjestöd LINDA, Skattefri studiestöd, Ersättning i samband med frivillig militär grundutbildning, Ersättningar för hemvärn
Skattefri pension LINDA	Summering av	Skattefri del av livränta och pension, Skattefri frivillig pension, Särskilt pensionstillägg, skattefri barnpension, Efterlevandestöd till barn
Skattefri studiestöd LINDA	Kopia av	Studiemedel/studiehjälp, lån och bidrag

Nedan följer en lista på andra variabler som skapas i LINDA;

BANT	Antal personer i hushållet
BURVKODF	Urvalskod för panelurvalet 0=ingår ej i urvalet 1=urvalsperson 2=familjemedlem
BLILLAN	Lilla n för panelurvalet
BSTORAN	Stora N för panelurvalet
BANTF	Antal personer i panelurvalet inom hushållet
BURVKODI	Urvalskod för utrikesfödda urvalet 0=ingår ej i urvalet 1=urvalsperson 2=familjemedlem
BLILLANI	Lilla n för utrikesfödda urvalet
BSTORANI	Stora N för utrikesfödda urvalet
BANTI	Antal personer i utrikesfödda urvalet inom hushållet
BANTUI	Antal utrikesfödda personer i utrikesfödda urvalet inom hushållet

3 Statistisk bearbetning och redovisning

3.1 Skattningar: antaganden och beräkningsformler

För skattningarna kan urvalet betraktas som ett OSU urval.

Eftersom urvalet kompletteras med alla familjemedlemmar, innebär det att ett nätverksurval har skapats. Detta kan beskrivas som att varje hushåll utgör ett nätverk med en eller flera medlemmar. När urvalet dras väljs det bland medlemmarna i nätverken och för den individ som valts hämtas övriga medlemmar i nätverket som kan knytas till honom. I det här fallet innebär det att övriga hushållsmedlemmar hämtas in.

På de filer som levereras finns följande variabler som ska användas för skattningarna: BSTORAN, BLILLAN, BANT, BANTF och BURVKODF, se avsnitt 2.5 *Databeredning* för förklaring av variablerna. Genom det urvalsförfarande som använts kan samma familj ha valts ut på mer än ett sätt, dvs. både t.ex. ett barn och en vuxen medlem av familjen kan ingå i urvalet. Detta problem kan hanteras på olika sätt men i LINDA-urvalet har följande hantering valts. För varje hushåll dels antalet individer i hushållet angivits (BANT) dels har antalet urvalsindivider i hushållet angivits (BANTF). Det innebär att varje nätverk (hushåll) i

urvalet endast kommer att representeras en gång. För att för ett urvalsår göra skattningar av totaler och medelvärde bör följande vikter bildas

$BVIKT = BSTORAN/BLILLAN$ denna vikt utgör urvalssannolikheten för urvalsindividerna.

$BSMEDV = BANTF/BANT (R(d)/a(d))$ denna vikt används för att korrigera för att en individ i urvalet kan komma med dels genom att själv bli utvald och dels genom att någon/några andra i hushållet väljs ut.

Den vikt som ska användas för skattningar av totaler eller medelvärden för urvalet ett givet år blir då:

$$VIKT = BVIKT * BSMEDV$$

Följande formler gäller för beräkningar av totaler för skattningar från urvalet.

För att göra skattningar på *hushållsnivå* gäller nedanstående formler. Observera att här förutsätts att det finns ett hushållsregister som bearbetningarna görs på.

Låt

$a(d)$ = antal primärobject i nätverk d som ingår i urvalsramen V ,
 $d = 1, 2, \dots, D$, (antal medlemmar i hushåll d som ingår i urvalsramen)

$R(d)$ = antal gånger nätverk d utvaldes i det replikerade nätverksurvalet
(antal medlemmar i hushåll d som ingår i urvalet)

Väntevärdesriktig estimation av $\tau(\mathbf{y})$ ges av :

$$\hat{\tau}(\mathbf{y}) = \frac{N}{n} \cdot \sum_{\substack{d \in \text{distinkta} \\ \text{nätverksurvalet}}} y_d \cdot \frac{R(d)}{a(d)}$$

Observera att $BVIKT = N/n$ och att $BSMEDV = R(d)/a(d)$.

Estimatorns varians skattas väntevärdesriktigt av:

$$V[\hat{\tau}(\mathbf{y})] = \frac{N^2}{n} \cdot \frac{1}{n-1} \cdot \left\{ \sum_{\substack{d \in \text{distinkta} \\ \text{nätverksurvalet}}} y_d^2 \cdot \frac{R(d)}{a(d)^2} - \frac{n \cdot \hat{\tau}(\mathbf{y})^2}{N^2} \right\} \cdot \left(1 - \frac{n}{N}\right)$$

För att göra motsvarande beräkningar baserade på *individnivå* användes nedanstående formler. Dessa bygger på att det finns tillgång till ett individregister omfattande alla individer som kommit att ingå i urvalet (dvs. alla urvalsindivider och alla familjemedlemmar till dessa).

Låt

$a(d)$ = antal primärobject i nätverk d som ingår i urvalsramen V ,
 $d = 1, 2, \dots, D$, (antal medlemmar i hushåll d som ingår i urvalsramen)

$R(d)$ = antal gånger nätverk d utvaldes i det replikerade nätverksurvalet
(antal medlemmar i hushåll d som ingår i urvalet)

Väntevärdesriktig estimation av $\tau(\mathbf{y})$ ges av :

$$\hat{\tau}(\mathbf{y}) = \frac{N}{n} \cdot \sum_{\substack{d \in \text{distinkta} \\ \text{nätverksurvalet}}} \frac{R(d)}{a(d)} \sum_{\substack{\text{allamedlemmar} \\ \text{ihushåll } d}} y_{di}$$

Estimatorns varians skattas väntevärdesriktigt av:

$$V[\hat{\tau}(\mathbf{y})] = \frac{N^2}{n} \cdot \frac{1}{n-1} \cdot \left\{ \left(\sum_{\substack{d \in \text{distinkta} \\ \text{nätverksurvalet}}} \frac{R(d)}{a(d)^2} \left(\sum_{i \in \text{hushåll } d} y_{di} \right)^2 \right) - \frac{n \hat{\tau}(\mathbf{y})^2}{N^2} \right\} \cdot \left(1 - \frac{n}{N} \right)$$

Med det urvalsförfarande som har använts för urvalet är det endast urvalsindividerna (BURVKODF=1) som utgör själva panelurvalet. Familjemedlemmarna kommer endast med om familjens utseende är detsamma under flera år.

För att göra **longitudinella studier** kan två olika ansatser användas.

1) Endast urvalsindividerna följs mellan åren. Detta innebär att det estimationsförfarande som ska användas när undersökningen används som en panel blir:

BVIKT=BSTORANF/BLILLANF Denna vikt utgör urvalssannolikheten för urvalsindividerna.

Väntevärdesriktig estimation av $\tau(\mathbf{y})$ ges av ;

$$\hat{\tau}(\mathbf{y}) = \frac{N}{n} \cdot \sum y_d \cdot$$

Estimatorns varians skattas väntevärdesriktigt av.

$$V[\hat{\tau}(\mathbf{y})] = \frac{N^2}{n} \cdot \frac{1}{n-1} \cdot \left\{ \sum y_d - \frac{n \cdot \hat{\tau}(\mathbf{y})^2}{N^2} \right\} \cdot \left(1 - \frac{n}{N} \right).$$

Vid panelansatsen utgår det från något startår för studien och då används startårets vikter för att göra studien. Det vill säga, utgår urvalet år t och låter uppgifterna för åren $t+1, \dots, t+n$ utgöra egenskaper för individen år t .

2) Alternativ två är att urvalsindividerna följs över åren och för varje nytt år används uppgifterna för den familj som bildas detta år. Detta innebär att familjens sammansättning kan variera mellan åren, dvs. år t är individen ensamstående, år $t+1$ är individen sammanboende med gemensamma barn, år $t+j$ kan individen åter vara ensamstående alternativt ensamstående med barn.

Om denna ansats används så bildas vikterna enligt formeln nedan.

$$\text{VIKT} = \text{BVIKT} * \text{BSMEDV}.$$

För detta fall måste ett medvetande finnas om att ett hushåll kan väljas på flera sätt, dvs. mer än en individ i hushållet utgör urvalsindivid. Men varje hushåll representeras endast av ett hushåll i registret. Detta innebär att om hushållet splittras kommer endast ett av hushållen att följas i fortsättningen. Detta är viktigt att tänka på om hushållsidentiteten används för att följa hushållet över tiden. Det är då hushållet splittras kommer det gamla hushållet att representeras av två hushåll ett med hushållsidentiteten lika med tidigare och ett nytt hushåll vars hushållsidentitet ej fanns tidigare.

Motsvarande problem uppstår om två urvalspersoner som tidigare har utgjort varsitt hushåll bildar ett hushåll tillsammans. Då kommer de tidigare två hushållen att representeras av endast ett hushåll vars hushållsidentitet kommer finnas både aktuellt år och tidigare år.

3.2 Redovisningsförfaranden

SCB publicerar ingen statistik från LINDA. När en ny årgång är klar får potentiella användare kontakta SCB för att eventuellt få tillgång till mikrodata. Utlämnande sker alltid efter en sekretessprövning. Varje ny årgång är klar 13 månader efter referenstiden.

4 Slutliga observationsregister

4.1 Produktionsversioner

I det här dokumentet (SCBDOK) har framtagningen av nedanstående slutliga observationsregister beskrivits.

Register	Longitudinell individdatabas (LINDA)
Registervariant	Panelurval av befolkningen
Registerversion	2015
Register	Longitudinell individdatabas (LINDA)
Registervariant	Tilläggsurval av utrikesfödda
Registerversion	2015
Register	Longitudinell individdatabas (LINDA)
Registervariant	Panelurval - yrke och lön
Registerversion	2015
Register	Longitudinell individdatabas (LINDA)
Registervariant	Panelurval inkl. familjemedlem
Registerversion	2015

Fortsatt dokumentation, av registrens detaljerade innehåll, finns på SCB:s webbplats. Där beskrivs alla variabler och värdemängder m.m. Dokumentationen hittar du här: <https://www.h2.scb.se/metadata>. Klicka dig fram med hjälp av namnen på Register, Registervariant och Registerversion som är angivna i ovanstående tabell.

4.2 Arkiveringsversioner

Ej aktuellt att arkivera eftersom syftet med LINDA är att kunna göra longitudinella studier.

4.3 Erfarenheter från senaste undersökningsomgången

Inga nya erfarenheter har framkommit under denna undersökningsomgång.