

# Longitudinell individdatabas (LINDA)

## 1968-1991

### LE1900

## Innehåll

<b>0</b>	<b>Allmänna uppgifter .....</b>	<b>2</b>
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
<b>1</b>	<b>Översikt .....</b>	<b>5</b>
1.1	Observationsstorheter .....	5
1.2	Statistiska målstorheter .....	5
1.3	Utfloeden: statistik och mikrodata .....	8
1.4	Dokumentation och metadata .....	8
<b>2</b>	<b>Uppgiftsinsamling .....</b>	<b>9</b>
2.1	Ram och ramförfarande .....	9
2.2	Urvalsförfarande .....	10
2.3	Mätinstrument .....	11
2.4	Insamlingsförfarande .....	11
2.5	Databeredning .....	12
<b>3</b>	<b>Statistisk bearbetning och redovisning .....</b>	<b>16</b>
3.1	Skattningar: antaganden och beräkningsformler .....	16
3.2	Redovisningsförfaranden .....	19
<b>4</b>	<b>Slutliga observationsregister .....</b>	<b>20</b>
4.1	Produktionsversioner .....	20
4.2	Arkiveringsversioner .....	20
4.3	Erfarenheter från senaste undersökningsomgången .....	20

#### Denna dokumentation är upprättad i efterhand

Detta dokument avser referensåren 1968-1991. Det är framtaget 2021 och innehåller den information som var känd 2021. Det innebär att vissa uppgifter kan saknas eller vara bristfälliga, vilket kan vara av betydelse för förståelsen av kvaliteten i statistiken.

Statistikens benämning i detta dokument är den som gällde när statistiken publicerades. Hänvisningar till lagrum görs till dåvarande lagstiftning.

Kontaktpersoner avser de som är kontaktpersoner när detta dokument publiceras.

## 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-systemet).

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:* Klostersgatan 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:* Klostersgatan 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

LINDA är inte EU-reglerad.

## 0.10 Syfte och historik

Longitudinell individdatabas (LINDA) är en urvalsbaserad undersökning som ursprungligen skapades 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.

LINDA innehåller huvudsakligen uppgifter om inkomster, som t.ex. lön, inkomst av näringsverksamhet, kapitalinkomst, skattepliktiga och skattefria transfereringar. Det huvudsakliga syftet med undersökningen är att ge forskare med flera möjligheter att göra longitudinella studier av inkomstutveckling för olika grupper.

## 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

Longitudinell individdatabas (LINDA) är en urvalsbaserad undersökning. Uppgifter till LINDA hämtas i första hand från den totalräknade inkomststatistikens årsregister (IoT), Utbildningsregistret samt från Registret över totalbefolkningen (RTB).

En årgång har ca 13 månaders produktionstid.

### **0.13 Internationell rapportering**

Ingen internationell rapportering sker.

### **0.14 Planerade förändringar i kommande undersökningar**

Arbetet med att förbättra statistikens innehåll, tillförlitlighet, aktualitet, jämförbarhet och sam användbarhet samt tillgänglighet och förståelighet fortlöper genom ett produktnära utvecklingsarbete, huvudsakligen i samarbete användare och producenter av källregistren.

# 1 Översikt

## 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 utrikes födda.

Undersökningen har två objekt, hushåll och individer. Hushåll består av individer från LINDA och alla individer i LINDA som 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](http://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 civilförvarsutbildning 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>

	<i>Studiestöd</i> Studiemedel (lån och bidrag), vuxenstudiestöd lån m.m.
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. Utbildningsvariablerna avser vårterminen. För övriga variabler är referenstiden respektive kalenderår. För longitudinella studier är referensperioden 1968–1991.

### 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](http://www.scb.se/LE1900). Utöver detta finns variabelförteckningar och postbeskrivningar som kan fås efter kontakt med SCB.



## 2 Uppgiftsinsamling

### 2.1 Ram och ramförfarande

Rampopulation är alla i riket folkbokförda personer den 31/12 respektive referensår och avgränsas med hjälp av SCB:s register över totalbefolkningen (RTB). RTB innehåller samtliga folkbokförda personer i riket och källan är Skatteverkets folkbokföring. Det ideala innehållet i folkbokföringens register definieras av de författningar som reglerar folkbokföringen. Det finns personer som borde vara folkbokförda men som inte är det (undertäckning) samt personer som är folkbokförda trots att de inte borde vara det (övertäckning), vilket leder till täckningsfel. Det är brister i rapportering av födselar, dödsfall, invandring och utvandring som resulterar i dessa täckningsfel.

#### *Undertäckning*

Brister i rapporteringen av födselar och invandring resulterar i undertäckning. Födselar ger upphov till mycket små problem beroende på de snabba rapporteringsrutinerna. För nyfödda uppskattas undertäckningen i observationsregistren vara 0,1 procent av 0-åringarna.

Om en person invandrar till Sverige och personen enligt lagar och förordningar borde folkbokföra sig men inte gör det uppstår undertäckning. Uppskattningar som har gjorts tyder på en undertäckning på cirka 10 000 personer.

Invandrare blir registrerade som invandrade vid den tidpunkt då de folkbokförs. I vissa fall kan det gå en relativt lång tid mellan ankomsten till Sverige och folkbokföringstidpunkten. Vid inflyttning till Sverige genomgår personer med utomnordiskt medborgarskap i vissa fall en omfattande administrativ process innan beslut om uppehållstillstånd kan fattas. Detta gör att tiden mellan inresan till Sverige och registrering av inflyttningen kan bli lång. Under perioder med stor invandring tenderar de administrativa handläggningstiderna i varje enskilt fall att öka. Detta gäller särskilt personer som inte fått sin ansökan om uppehållstillstånd beviljad och som överklagar beslutet, vilket medför att det blir en avsevärd tidsskillnad mellan inresan och det slutliga beslutet.

För de invandrare som fick sina beslut om uppehållstillstånd under 1996 var mediantiden från det att ansökan registrerats fram till folkbokföring 21 veckor. Före ansökan finns polisens handläggningstid, om vars längd vi inte har någon uppgift. Uppgift om tidsskillnaden mellan invandring och registreringstidpunkt för nordiska invandrare saknas men den är kortare eftersom nordbor kan bosätta sig i Sverige utan uppehållstillstånd.

#### *Övertäckning*

Brister i rapporteringen av dödsfall och utvandring resulterar i övertäckning. Dödsfall ger upphov till mycket små problem beroende på de snabba rapporteringsrutinerna. 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 dem 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. Vid den kontroll som gjordes i samband med folk- och bostadsräkningen 1985 var övertäckningen ca 0,1 procent av hela befolkningen. Räknat på delpopulationen utländska medborgare var övertäckningen 1,0 procent för nordiska medborgare och 2,8 procent för övriga medborgarskap tillsammans.

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 från 1994 (Nilsson, Bristar i folkbokföringen, PM, 1994) framförs hypotesen att övertäckningen bland utrikes födda kan vara så stor som 10 procent. Andra analyser från den tiden pekade på att omfattningen av övertäckningen kan ligga mellan 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 publicerad 2002 (Qvist, Täckningsproblem i Registret över totalbefolkningen RTB. SCB, R&D Report 1999:1). I studien analyseras invandras dödlighet under perioden 1993–97 med intensitetsregression. Analysen visar att det finns en relativ underdödlighet bland personer födda utomlands jämfört med personer födda i Sverige. Denna relativa underdödlighet tolkas som en relativ övertäckning i RTB. Beräkningar visar att övertäckningen bland utomnordiska invandrare ligger i intervallet 25 000–50 000 personer, vilket motsvarar 4–8 procent av den utomnordiska befolkningen.

Under år 2004 har studier gjorts där man försökt att skatta storleken på övertäckningen med hjälp av registerstatistik. En preliminär beräkning pekade på en övertäckning på minst 35 000 personer.

Personerna måste någon gång under det aktuella året ha varit folkbokförd i Sverige för att ingå i LINDA:s rampopulation.

## 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.

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 inkomstuppgifter 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). Dessa register har administrativa data som sin källa.

## 2.4 Insamlingsförfarande

I början av varje år inhämtas individuppgifter från befintliga register på SCB för att sedan bilda LINDA. Demografiska variabler hämtas från Registret över totalbefolkningen (RTB) och inkomstvariabler från

Inkomst- och taxeringsregistret (IoT) samt utbildningsvariabler från Utbildningsregistret (UREG).

## 2.5 Databeredning

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

Från och med 1991 skapas det i LINDA också inkomstvariabler 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 (under åren kan antalet variabler dock variera):

- 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 kolumnnamnet. I tabblån nedan beskrivs de bearbetningar som görs för inkomstvariablerna:

<b>Variabel i LINDA</b>	<b>Bearbetning</b>	<b>Grundvariabel</b>
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  minus	Faktorinkomst LINDA Skattepliktiga positiva transfereringar LINDA Skattefria positiva transfereringar LINDA  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,

Variabel i LINDA	Bearbetning	Grundvariabel
		vårdnadsbidrag, bostadsersättning, etableringsersättning, etableringstillägg, boend till.
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, Smittbärapenning, 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 urvalsindividen.

$BSMEDV=BANTF/BANT (R(d)/a(d))$  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 primärobjekt 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 urvalsindivider (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 urvalsindividen följs mellan åren. Detta innebär att det estimationsförfarande som ska användas när undersökningen används som en panel blir:

$BVIKT = \frac{BSTRORANF}{BLILLANF}$  Denna vikt utgör urvalssannolikheten för urvalsindividen.

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 utgås 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,...,t+n utgöra egenskaper för individen år t.

2) Alternativ två är att urvalsindividen 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.

$VIKT = BVIKT \cdot BSMEDEV$ .

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.

## 4 Slutliga observationsregister

### 4.1 Produktionsversioner

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

<b>Register</b>	Longitudinell individdatabas (LINDA)
<b>Registervariant</b>	Panelurval av bef inkl fam
<b>Registerversion</b>	1991
<b>Register</b>	Longitudinell individdatabas (LINDA)
<b>Registervariant</b>	Panelurval av bef inkl taxeringshushåll
<b>Registerversion</b>	1968-1991
<b>Register</b>	Longitudinell individdatabas (LINDA)
<b>Registervariant</b>	Urval av utrikesfödda inkl taxeringsh
<b>Registerversion</b>	1968-1991

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.