

## Longitudinell individdatabas (LINDA) 2006

LE1900

*I denna beskrivning redovisas först allmänna och legala uppgifter om undersökningen samt dess syfte och historik. Därefter redovisas undersökningens innehåll och tillförlitlighet samt hur den genomförs och hur man kan ta del av resultaten. Genom att klicka på en rubrik i innehållsförteckningen kommer man direkt till aktuellt avsnitt.*

### Innehållsförteckning

<b>A</b>	<b>Administrativa uppgifter</b> .....	<b>2</b>
A.1	Ämnesområde.....	2
A.2	Statistikområde.....	2
A.3	SOS-klassificering.....	2
A.4	Statistikansvarig.....	2
A.5	Statistikproducent.....	3
A.6	Uppgiftsskyldighet.....	3
A.7	Sekretess och regler för behandling av personuppgifter.....	3
A.8	Gallringsföreskrifter.....	3
A.9	EU-reglering.....	3
A.10	Syfte och historik.....	3
A.11	Statistikanvändning.....	4
A.12	Uppläggning och genomförande.....	4
A.13	Internationell rapportering.....	4
A.14	Planerade förändringar i kommande undersökningar.....	5
<b>B</b>	<b>Kvalitetsdeklaration</b> .....	<b>5</b>
B.0	Inledning.....	5
B.1	Innehåll.....	5
1.1	Statistiska målstorheter.....	5
1.1.1	Objekt och population.....	5
1.1.2	Variabler.....	5
1.1.3	Statistiska mått.....	7
1.1.4	Redovisningsgrupper.....	7
1.1.5	Referenstider.....	7
1.2.	Fullständighet.....	7
B.2	Tillförlitlighet.....	7
2.1	Tillförlitlighet totalt.....	7
2.2	Osäkerhetskällor.....	7
2.2.1	Urval.....	7
2.2.2	Ramtäckning.....	8
2.2.3	Mätning.....	8
2.2.4	Svarsbortfall.....	8

2.2.5	Bearbetning .....	8
2.2.6	Modellantaganden .....	9
2.3	Redovisning av osäkerhetsmått .....	9
B.3	<i>Aktualitet</i> .....	9
3.1	Frekvens .....	9
3.2	Framställningstid .....	9
3.3	Punktlighet .....	9
B.4	<i>Jämförbarhet och sammanvändbarhet</i> .....	9
4.1	Jämförbarhet över tiden.....	9
4.2	Jämförbarhet mellan grupper .....	9
4.3	Sammanvändbarhet med annan statistik .....	9
B.5	<i>Tillgänglighet och förståelighet</i> .....	10
5.1	Spridningsformer.....	10
5.2	Presentation .....	10
5.3	Dokumentation.....	10
5.4	Tillgång till primärmaterial .....	10
5.5	Upplysningstjänster .....	10

## A Administrativa uppgifter

### A.1 Ämnesområde

*Ämnesområde:* Inkomst och inkomstfördelning

### A.2 Statistikområde

*Statistikområde:* Hushållens ekonomi

### A.3 SOS-klassificering

*Tillhör (SOS)* Ja



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)

### A.4 Statistikansvarig

*Myndighet/organisation:* Statistiska centralbyrån (SCB)  
*Postadress:* 701 89 Örebro  
*Besöksadress:* Klostergatan 23  
*Kontaktperson:* Marcus Vingren  
*Telefon:* 019-17 63 12  
*Telefax:* 019-17 70 83  
*E-post:* Fornamn.efternamn@scb.se

## **A.5 Statistikproducent**

<i>Myndighet/organisation:</i>	Statistiska centralbyrån (SCB)
<i>Postadress:</i>	701 89 Örebro
<i>Besöksadress:</i>	Klostergatan 23
<i>Kontaktperson:</i>	Marcus Vingren
<i>Telefon:</i>	019-17 63 12
<i>Telefax:</i>	019-17 70 83
<i>E-post:</i>	Fornamn.efternamn@scb.se

## **A.6 Uppgiftsskyldighet**

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

## **A.7 Sekretess och regler för behandling av personuppgifter**

*I myndigheternas särskilda verksamhet för framställning av statistik gäller sekretess enligt 9 kap. 4 § sekretesslagen (1980:100). 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.*

Sekretess gäller för personuppgifterna i LINDA. Uppgifter som behövs för forsknings- eller statistikändamål får dock efter särskild prövning lämnas ut, om det står klart att uppgiften kan röjas utan att den som uppgiften rör eller någon honom närstående lider skada eller men.

## **A.8 Gallringsföreskrifter**

Undersökningen berörs ej av gallringsföreskrifter.

## **A.9 EU-reglering**

Inga tvingande EU-direktiv finns på inkomststatistiken.

## **A.10 Syfte och historik**

LINDA är en urvalsbaserad longitudinell individdatabas som ursprungligen skapats på uppdrag. Uppbyggnaden av LINDA gjordes i ett samarbete mellan Statistiska centralbyrån, Riksförsäkringsverket (numera Försäkringskassan), Uppsala Universitet och Finansdepartementet. Undersökningen är idag en anslagsprodukt.

Det huvudsakliga syftet med undersökningen är att ge forskare med flera möjligheter att göra longitudinella studier av inkomstutveckling för olika grupper. Data finns för åren 1968-2006.

Möjligheter finns att göra longitudinella studier över åren men även att studera enskilda år tack vare urvalets uppbyggnad.

## A.11 Statistikanvändning

Viktiga användare utgörs av statliga verk, departement, organisationer, högskolor och universitet. Exempel på användningsområden är analys av välfärdssystemet.

En referensgrupp är kopplad till LINDA. Referensgruppen består av representanter från följande organisationer: Försäkringskassan, Uppsala Universitet, Finansdepartementet, Socialdepartementet och SCB.

LINDA diskuteras även i ett programråd. Det är programrådet för ekonomisk välfärdsstatistik, som inkluderar representanter från Finansdepartementet, Socialdepartementet, Boverket, Konsumentverket, Försäkringskassan, Skatteverket och Uppsala universitet.

## A.12 Uppläggning och genomförande

Inkomståret 1994 drogs ett obundet slumpmässigt urval 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. Dessutom kompletteras urvalspersonerna med familjemedlemmar för varje inkomstår. 1968-1998 utgörs familjebegreppet av totala inkomststatistikens taxeringshushåll. Från 1999 utgörs familjebegreppet av RTB:s familjebegrepp. Årgångar 1991-1998 har även kompletterats med RTB:s familjebegrepp.

Ytterligare ett panelurval har skapats omfattande ca 20 % av personer som är utrikesfödda. Även panelurvalet över utrikesfödda har kompletterats med familjemedlemmar för respektive inkomstår.

Det två urvalen omfattar tillsammans ca 1,1 miljoner personer.

Uppgifter till LINDA hämtas i första hand från den totalräknade inkomststatistikens årsregister samt från registret över totalbefolkningen, RTB. Från folk- och bostadsräkningarna, FoB, hämtas även uppgifter för respektive FoB-år (60, 65, 70, 75, 80, 85 och 90). Utbildningsuppgifter från högskoleregistret (77 - 95) och registret över komvuxstuderande (85 - 95) samt löneregistret (98 - ). Från Riksförsäkringsverket/Försäkringskassan hämtas uppgifter om pensioner och pensionspoäng (91 - ) sjukfall (97 - ) och föräldrapenning (97 - ). Uppgifter om arbetssökande hämtas från Arbetsförmedlingens (AF) händelsedatabas, HÄNDEL (91 - ), och databasen över utbetalda A-kasse- och KAS-ersättningar, ASTAT (93 - ), hämtas från Inspektionen för arbetslöshetsförsäringen (IAF).

## A.13 Internationell rapportering

Ingen internationell rapportering sker.

## A.14 Planerade förändringar i kommande undersökningar

Inga förändringar är planerade.

## B Kvalitetsdeklaration

### B.0 Inledning

Det huvudsakliga syftet med undersökningen är att ge forskare med flera möjligheter att genomföra longitudinella studier av inkomstutvecklingen för olika grupper.

### B.1 Innehåll

#### 1.1 Statistiska målstorheter

##### 1.1.1 Objekt och population

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.

##### 1.1.2 Variabler

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 är drygt 1000. Alla variabler som återfinns i undersökningen kan därför inte specificeras här. I nedanstående tablå återfinns däremot en grov beskrivning av några områden inom vilka det finns uppgifter i undersökningen.

Område	Beskrivning
Bruttoinkomstbegrepp	Arbetsinkomst, faktorinkomst, disponibel inkomst, sammanräknad förvärvsinkomst, kapitalinkomst m.m.
Negativa transfereringar	Skatter och övriga negativa transfereringar
Positiva transfereringar skattepliktiga	<i>Ersättningar i samband med sjukdom</i> Sjukpenning, rehabiliteringsersättning m.m. <i>Pensioner och livräntor</i> Pension, delpension, livränta m.m. <i>Föräldrapenning</i> Föräldrapenning tillfällig vård av barn, föräldrapenning vid barns födelse <i>Studiestöd</i> Studiepenning, utbildningsbidrag för doktorander m.m.

	<p><i>Arbetsmarknadsstöd</i>          Totalt arbetsmarknadsstöd  <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)  <i>Pensioner och livräntor</i>          Pension, livränta skattefri del, frivillig pension, särskilt pensionstillägg, barnpension skattefri del  <i>Familjestöd</i>          Socialbidrag, bostadsbidrag för pensionärer, bostadsbidrag för barnfamiljer samt makar och ensamstående utan barn, barnbidrag och flerbarnstillägg, mottaget bidragsförskott (underhållsstöd)  <i>Studiestöd</i>          Studiemedel (lån och bidrag), vuxenstudiestöd lån m.m.  <i>Ersättning i samband med värnplikt</i>          Totalt värnpliktsstöd</p>
Taxeringen	<p><i>Variabler som ligger till grund för beskattning</i>          Underlag för: fastighetsskatter, pensionsavgifter, sjukförsäkringsavgifter m.m.  <i>Variabler som påverkar beskattningen</i>          Grundavdrag, skuldräntor, särskilt grundavdrag, realisationsförluster, allmänna avdrag m.m.</p>
Folkbokföringsuppgifter	<p>Civilstånd, folkbokföring, mantalsskrivning, samhörighet, medborgarskap, invandringsår m.m.</p>
Övriga externa variabler	<p><i>Föräldrapenningsregister</i>          Föräldrapenningdagar  <i>Sjukfallsregister</i>          Ersatta sjukdagar  <i>Pensionspoängregister</i>          Pensionspoäng  <i>FK:s statistikregister</i>          Diverse transfereringsvariabler bl.a. koder för pensionsvariabler, vårdbidrag, BTP m.m.  <i>AF:s register</i>          Händelsedatabas  <i>IAF:s register</i>          Arbetskraftsstatistik</p>

Utöver nämnda variabelgrupper finns även några vanliga bakgrundsvariabler som exempelvis kön, ålder, födelsemånad och utbildning.

### *1.1.3 Statistiska mått*

Ej relevant.

### *1.1.4 Redovisningsgrupper*

Ej relevant.

### *1.1.5 Referenstider*

Data finns från och med 1968. För förmögenhetsuppgifter är referenstiden 31 december kalenderåret. För övriga variabler är referenstiden respektive kalenderår. För longitudinella studier är referensperioden 1968-2006.

## **1.2. Fullständighet**

Registret saknar uppgifter som ej kommit fram i kontrolluppgifter med mera. Det innebär att exempelvis "svarta" inkomster saknas i statistiken. Överföringar mellan hushåll saknas också, t.ex. ekonomiskt stöd som vissa föräldrar ger sina vuxna barn.

## **B.2 Tillförlitlighet**

### **2.1 Tillförlitlighet totalt**

LINDA bygger på uppgifter från inkomstregistret. Eftersom inkomstregistret bygger på administrativa register och är totalräknad får den en mycket hög precision i sina beloppsberäkningar. Variabelkvaliteten på familjesammansättningen är något sämre på grund av att sambo-familjer (ej gifta) utan gemensamma barn blir betraktade som två ensamstående hushåll. Antalet familjer blir därför något högre än det verkliga antalet. I sammanhanget bör betonas att LINDA i huvudsak är en longitudinell *individdatabas*.

Statistiken bygger på uppgifter som redovisas i självdeklarationen, kontrolluppgifter m.m. Det innebär att inkomster och förmögenheter som undanhålls skattemyndigheterna lämnas helt utanför. Storleken på denna dolda sektor i ekonomin är relativt okänd, men torde uppgå till betydande belopp.

### **2.2 Osäkerhetskällor**

#### *2.2.1 Urval*

LINDA är en panelurvalsundersökning. Inkomståret 1994 drogs ett obundet slumpmässigt urval 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.

För att urvalet skall kunna användas för tvärsnittsstudier varje enskilt år kompletteras urvalet med nya urvalspersoner så att urvalet motsvarar 3 % av befolkningen respektive år. De nya urvalspersonerna (nyfödda och invandrare) ersätter de som på något sätt lämnat urvalsramen.

Dessutom kompletteras urvalspersonerna med familjemedlemmar för varje inkomstår. 1968-1998 utgörs familjebegreppet av totala inkomststatistikens taxeringshushåll. Från 1999 utgörs familjebegreppet av RTB:s familjebegrepp. Årgångar 1991-1998 har även kompletterats med RTB:s familjebegrepp.

Ytterligare ett panelurval har skapats omfattande ca 20 % av personer som är utrikesfödda. Även panelurvalet över utrikesfödda har kompletterats med familjemedlemmar för respektive inkomstår.

Det två urvalen omfattar tillsammans ca 1,1 miljoner personer.

#### 2.2.2 *Ramtäckning*

Rampopulation för LINDA är den kopia av Skatteverkets folkbokföringsregister som lagras i Registret över totalbefolkningen (RTB). Detta innebär att personer som enligt gällande regelverk skall vara folkbokförda men inte är det utgör undertäckning. Omvänt gäller att personer som är folkbokförda trots att de enligt gällande regelverk inte skall vara det utgör övertäckning. Det är framför allt brister i rapportering av födselar, dödsfall, invandring och utvandring som resulterar i dessa täckningsfel.

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

#### 2.2.3 *Mätning*

De administrativa registren är i första hand till för administrativa ändamål och svarar därmed inte alltid mot statistikens behov. Statistiken bygger på uppgifter som redovisas i självdeklarationen, kontrolluppgifter m.m. Det innebär att inkomster och förmögenheter som undanhålls skattemyndigheterna lämnas helt utanför. Storleken på denna dolda sektor i ekonomin är relativt okänd, men torde uppgå till betydande belopp.

Inom LINDA tas det fram ett antal härledda variabler.

#### 2.2.4 *Svarsbortfall*

Undersökningen är registerbaserad. Svartsbortfall förekommer ej.

#### 2.2.5 *Bearbetning*

Eftersom data hämtas från IoT, sker ingen "egen" databeredning för LINDA. Ev. inverkan på total kvaliteten beror därför på databeredningen i IoT. Viss egenkon-



troll sker dock, vilket kan leda till justeringar i IoT. Omfattningen av detta är dock relativt liten.

Ett mindre antal härledda variabler skapas vid produktionen av LINDA. Exempelvis skapas variabeln "Skattefritt studiestöd" genom att IoT-variablerna "Rekryteringsbidrag" och "Studiemedel/studiehjälp, lån och bidrag" summeras. Kvaliteten på variabeln beror därför på kvaliteten i IoT.

#### 2.2.6 *Modellantaganden*

-

### **2.3 Redovisning av osäkerhetsmått**

Ej relevant eftersom inga målstorheter skattas.

## **B.3 Aktualitet**

### **3.1 Frekvens**

Undersökningen uppdateras årligen.

### **3.2 Framställningstid**

Cirka 1 år och 2 månader.

### **3.3 Punktlighet**

Enligt plan (februari 2008).

## **B.4 Jämförbarhet och sammanvändbarhet**

### **4.1 Jämförbarhet över tiden**

Ändrade bidrags- och skatteregler medför att jämförbarheten över tiden påverkas. Från och med år 1994 redovisas entals kronor i grundmaterialet. Tidigare avrundades beloppen för varje individ till närmast lägre 100-tal kronor. Ändringen medför att vi får fler inkomsttagare som har låga inkomster eller avdrag.

Ett ändrat familjebegrepp från och med årgång 1991 och framåt påverkar jämförbarheten med tidiga årgångar avseende gifta/sambo och ensamstående.

### **4.2 Jämförbarhet mellan grupper**

Allmänt gäller att jämförbarheten är god för grupper som kan konstrueras utifrån LINDA. Viss försiktighet bör dock iaktas vid jämförelser av olika familjetyper, samt grupper som utgörs av individer med mycket låg inkomst.

### **4.3 Sammanvändbarhet med annan statistik**

LINDA hämtar mycket data från totalräknade inkomststatistikens årsregister. Detta gör även andra undersökningar såsom Hushållens Ekonomi (HEK) och Hushållens Utgifter (HUT).

Urvalet i 1997 års HEK samt HEK-urvalen från och med årgång 2000 är fullständigt samordnade med LINDA. Från och med 2003 är dessutom urvalet i HUT samordnat med LINDA.

## **B.5 Tillgänglighet och förståelighet**

### **5.1 Spridningsformer**

Registret distribueras på CD-Rom-skivor eller via MONA-systemet (Microdata On line Access) till användarna.

MONA-systemet syftar till att öka tillgängligheten till mikrodata samtidigt som säkerhet och sekretess i hanteringen stärks. Genom MONA får användaren tillgång till databaser och kan göra bearbetningar av dem via Internet. Fysiskt lagras dock databaserna hos SCB.

### **5.2 Presentation**

LINDA är en databas som främst är avsedd för forskning. I den mån statistik publiceras svarar användarna för detta.

### **5.3 Dokumentation**

Postbeskrivningar beskriver registerinnehåll i de olika registren från undersökningen. Variabeldokumentation för variabler som är gemensamma med inkomststatistiken finns att tillgå.

### **5.4 Tillgång till primärmaterial**

Registret är primärmaterial. Forskare får efter prövning tillgång till avidentifierat primärmaterial för egna bearbetningar. Specialbearbetningar kan utföras på uppdragsbasis.

### **5.5 Upplysningstjänster**

Vid frågor om statistiken kontakta:

Marcus Vingren, 019-17 63 12, E-post linda@scb.se,  
telefax 019-17 69 22.