

# Hushållens ekonomi (HEK) 2007

HE0103

## Innehåll

SCBDOK 3.1	
<p><b>0 Administrativa uppgifter</b></p> <p>0.1 Ämnesområde 0.2 Statistikområde 0.3 SOS-klassificering 0.4 Statistikansvarig 0.5 Statistikproducent 0.6 Uppgiftsskyldighet 0.7 Sekretess och regler för behandling av personuppgifter 0.8 Gallringsföreskrifter 0.9 EU-reglering 0.10 Syfte och historik 0.11 Statistik användning 0.12 Uppläggning och genomförande 0.13 Internationell rapportering 0.14 Förändringar i undersökningen</p>	<p><b>1 Innehållsöversikt</b></p> <p>1.1 Observationsstorheter 1.2 Statistiska målstorheter 1.3 Utflöden: statistik och mikrodata 1.4 Dokumentation och metadata</p>
<p><b>2 Uppgiftsinsamling</b></p> <p>2.1 Ram och ramförfarande 2.2 Urvalsförfarande 2.3 Mätinstrument 2.4 Insamlingsförfarande 2.5 Databeredning</p>	<p><b>3 Statistisk bearbetning och redovisning</b></p> <p>3.1 Skattningar: antaganden och beräkningsformler 3.2 Redovisningsförfaranden</p>
<p><b>4 Slutliga Observationsregister</b></p> <p>4.1 Produktionsversioner 4.2 Arkiveringsversioner 4.3 Erfarenheter från senaste undersökningsomgången</p>	

## 0 Administrativa uppgifter

### 0.1 Ämnesområde


*Ämnesområde:* Hushållens ekonomi

### 0.2 Statistikområde

*Statistikområde:* Inkomster och inkomstfördelning  
Boende

### 0.3 SOS-klassificering

*Tillhör (SOS)* Ja

För undersökningar som ingår i Sveriges officiella statistik  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:* 701 89 Örebro  
*Besöksadress:* Klostergatan 23  
*Kontaktperson:* Johan Lindberg  
*Telefon:* 019-17 60 64  
*E-post:* fornamn.efternamn@scb.se

### 0.5 Statistikproducent

*Myndighet/organisation:* Statistiska centralbyrån (SCB)  
*Postadress:* 701 89 Örebro  
*Besöksadress:* Klostergatan 23  
Johan Lindberg  
*Telefon:* 019-17 60 64  
*E-post:* fornamn.efternamn@scb.se

### 0.6 Uppgiftsskyldighet

Uppgiftsskyldighet föreligger inte enligt lagen om den officiella statistiken (SFS 2001: 99), vilket innebär att uppgiftslämnaren medverkar frivilligt.

## 0.7 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.

Vid utlämnande av mikrodata görs först en sekretessprövning. Data lämnas ut via MONA (Microdata ONline Access).

## 0.8 Gallringsföreskrifter

Undersökningen om Hushållens ekonomi (HEK) omfattas av ett bevarandebeslut.

## 0.9 EU-reglering

Inga tvingande EU-direktiv finns för denna statistik.

## 0.10 Syfte och historik

Syftet är att kartlägga den disponibla inkomstens fördelning bland olika hushåll, belysa inkomststrukturen samt beskriva boendet och boendeutgifterna för hushåll i olika upplåtelseformer.

Den officiella boendestatistiken redovisas i HEK fr.o.m. 2003 års undersökning. Det ersätter hushållsdelen i Bostads- och hyresundersökningen (BHU).

Statistiken innefattar variabler som disponibel inkomst, faktorinkomst, inkomst av kapital, skatt, skattepliktiga och skattefria transfereringar. Statistiken redovisas bland annat efter bakgrunds- och klassificeringsvariablerna hushållstyp (kosthushåll och familjeenhet), ålder, kön, sysselsättningsgrad, socioekonomisk indelning, födelse-land, boendeform/upplåtelseform.

Undersökningen har genomförts varje år sedan 1975. Datamaterialet är användbart för åren 1975, 1978 samt för varje år från 1980 till 2007. För vissa år kan förmögenhetsfördelningen bland hushållen beskrivas. Ett tidsseriebrott i statistiken uppkom på grund av skattereformen 1990/1991.

## 0.11 Statistikanvändning

Statistiken används av många olika användare inom skilda delar av samhället;

- Statistiken från HEK används i stor utsträckning som ett viktigt underlag vid politiska ställningstaganden och beslut.
- Allmänheten använder statistiken för att få en översiktlig beskrivning av inkomstfördelningen samt boendet och boendeutgifterna för både hushåll och personer efter olika bakgrundsvariabler som bland annat ålder, kön, sysselsättning.
- Andra statistikgrenar vid SCB använder HEK som underlag, exempelvis Nationalräkenskaperna och mikrosimuleringsmodellen FASIT. I FASIT-modellen är det

möjligt att simulera budget- och fördelningseffekter av ändringar i skatte- och bidragssystem.

- Användarna representeras bland annat av Programrådet för ekonomisk välfärdsstatistik och Programrådet för Mark och bebyggelse.

Deltagande organisationer i Programrådet för ekonomisk välfärdsstatistik är Finansdepartementet, Socialdepartementet, Socialstyrelsen, LO, Institutet för social forskning, Institutet för framtidsstudier, Örebro universitet, Integration och jämställdhetsdepartementet och Boverket.

Deltagande organisationer i programrådet för Mark och bebyggelse är Finansdepartementet, Sveriges Byggindustrier, SABO, Fastighetsägarna Sverige, Boverket, Kungliga Tekniska högskolan (KTH), Hyresgästernas Riksförbund, Länsstyrelsen, Naturvårdsverket, Lantmäteriet, Sveriges kommuner och landsting och miljödepartementet.

Andra användare är bland annat forskare och företag. Internationella användare är bland annat Eurostat och Luxemburg income study (LIS).

## 0.12 Uppläggning och genomförande

HEK är en urvalsundersökning som genomförs varje år. Populationen består av samtliga hushåll och personer som var folkbokförda i Sverige under undersökningsåret (inkomståret) 2007. Registret över totalbefolkningen (RTB) används som urvalsram. Urvalet utgörs av personer som är 18 år eller äldre. Urvalet är ett s.k. nätverksurval, vilket betyder att både urvalspersonen och de personer som tillhör hans/hennes hushåll ingår i undersökningen. Från och med inkomståret 2000 samordnas urvalet med SCB:s Longitudinella Individdatabas (LINDA), se 2.2. Urvalsstorleken år 2007 var ca 17 000 hushåll. Genom samordningen med LINDA ges möjlighet att följa samma urvalsenheter över tid.

Datainsamlingen sker genom en telefonintervju och genom insamling av uppgifter från administrativa register. Eftersom det är ett urval av befolkningen som ingår i undersökningen kan det bli viss osäkerhet när man redovisar variabler som är extremt snedfördelade, t.ex. kapitalvinst. Det påverkar också osäkerheten för variabler där kapitalvinst ingår. För personer som under inkomståret redovisat stora kapitalvinster/-förluster vid försäljning av t.ex. värdepapper och fastigheter görs därför ett tilläggsurval. Syftet med tilläggsurvalet är att öka säkerheten i skattningarna på de olika inkomstparametrarna.

Vid intervjun, som genomförs under januari-maj året efter inkomståret, ställs frågor om hushållets sammansättning, uppgifter om boende och boendeutgifter, ekonomiskt stöd mellan olika hushåll och ett antal bilfrågor. För personerna i hushållet ställs också frågor om sysselsättning, arbetstid, yrke, barnomsorg, långvarigt sjuk eller funktionshinder, färdtjänst, hemtjänst, samt utgifter för tandvård, läkemedel och sjukvård.

Statistiken över bland annat inkomster, ersättningar, bidrag och skatter bygger på delar av Inkomst- och taxeringsregistret (IoT). Dessa registeruppgifter kommer från Skatteverket (taxeringsuppgifter och kontrolluppgifter), Försäkringskassan, Centrala studiestödsnämnden, Statens pensionsverk, Pliktverket och Socialstyrelsen. För beräkning av bland annat inkomst från näringsverksamhet, olika intäkts- och kostnadsposter för näringsidkare har uppgifter från standardiserade

räkenskapsutdrag (SRU) använts. Statistiken över boende och boendeutgifter kompletteras med uppgifter från fastighetstaxeringsregistret.

### **0.13 Internationell rapportering**

Ingen internationell rapportering sker.

### **0.14 Förändringar i undersökningen**

För undersökningsår 2007 skickades ingen enkät om äldreomsorg och hemtjänst till kommunerna.

I HEK2007 har en mer sofistikerad metod för bortfallsuppföljning tillämpats. Tidigare har endast andelen svar i olika grupper använts som beslutsunderlag för vilka som ska gå till bortfallsuppföljning. Från och med HEK2007 studeras även inkomstuppgifter samt hushållssammansättning från register när beslut tas om vilka som ska bearbetas ytterligare. Detta har medfört att färre hushåll har valts ut för bortfallsuppföljning samtidigt som skattningarna har blivit bättre.

Vid tilläggsurvalet av personer med stora kapitalvinster/förluster tittar vi nu även på ränta och utdelning för att få bättre skattningar på makronivå för dessa snedfördelade variabler och samtidigt undvika outliers.

För HEK 2007 publicerades inget statistiskt meddelande för Boende och boendeutgifter.

# 1 Översikt

Den årliga statistiken över hushållens ekonomi bygger på uppgifter som samlas in genom en telefonintervju till ett urval av befolkningen samt från administrativa register. Undersökningen har genomförts varje år sedan 1975. Statistiken redovisar inkomstfördelning för olika hushåll och personer efter olika bakgrundsvariabler som bland annat ålder, kön, sysselsättning samt boende och boendeutgifter för hushåll i olika upplåtelseformer för inkomståret 2007.

## 1.1 Observationsstorheter

Målpopulationen för undersökningen utgörs av de hushåll som fanns i Sverige den 31 december 2007. Hushållen består av personer som enligt gällande lagar och förordningar skulle ha varit folkbokförda i Sverige 31 december 2007. I redovisningen av inkomststatistiken används helårshushåll. Helårshushåll utgörs av de personer som var folkbokförda vid såväl årets början som dess slut och vars hushåll har en disponibel inkomst skild från noll. Anledningen till att denna avgränsning görs är för att minimera den övertäckningen som finns, se 2.1.

Urvalspopulationen består av personer som enligt gällande lagar och förordningar skulle ha varit folkbokförda i Sverige 31 december 2007. Eftersom endast personer 18 år eller äldre ingår i rampopulationen, innebär det att personer yngre än 18 år som bor i egna hushåll helt saknas i undersökningen, se 2.2.

Undersökningen använder två definitioner av hushåll; familjeenhet och kosthushåll. Kosthushåll är huvudbegrepp.

En familjeenhet omfattar en eller två vuxna som är gifta/sammanboende samt ev. barn under 18 år. Ett hemmavarande barn 18 år eller äldre bildar en egen familjeenhet. Familjeenhet används främst av mikrosimuleringsmodellen FASIT.

Ett kosthushåll utgörs av alla personer som bodde i samma bostad och som hade gemensam "hushållning" den 31 december 2007. Inneboende ingår inte i kosthushållet. I redovisningen av inkomstfördelningen samt för boendestatistiken används kosthushåll. Ett kosthushåll kan bestå av flera familjeenheter.

Personer som är tillfälligt borta, men som förväntas återkomma, ingår i hushållet. Barn som bor lika mycket hos båda föräldrarna räknas med i det hushållet där de är folkbokförda. Institutionshushåll ingår inte i redovisningen. Deras inkomst- och utgiftsförhållanden är inte jämförbara med övriga hushåll. Personer som är 65 år eller äldre och bor i särskilt boende ingår dock.

När hushållens utgifter för boendet ska beskrivas avgränsas populationen till boende med äganderätt, bostadsrätt och hyresrätt, den så kallade "boendeutgiftspopulationen". I boendeutgiftspopulationen ingår inte: Hushåll i andelslägenhet, i egen hyresfastighet, i småhus på ägd jordbruksfastighet, de med andrahandskontrakt, de som är inneboende, de i särskilt boende, de i studentbostad, de som hyr möblerat samt de som inte har bott i bostaden hela året.

## 1.2 Statistiska målstorheter

A. Målpopulationen dvs. individpopulationen, se avsnitt 1.1.

A0. Personer 0-19 år

A1. Personer 18 år eller äldre

A2. Personer 20 år eller äldre

A3. Personer 20-64 år

A4. Personer 65 år eller äldre

A5. Helårs- och heltidsanställda 20-64 år

A6. Helårs- och heltidsanställda utan näringsinkomst 20-64 år

B. Målpopulationen dvs. hushållspopulationen, se avsnitt 1.1.

B1. Kosthushåll

B1a. Kosthushåll där hushållsföreståndaren är 18 år eller äldre

B1b. Kosthushåll där hushållsföreståndaren är 20-64 år

B2. Familjeenhet

B2a Familjeenhet där hushållsföreståndaren är 18 år eller äldre

B2b Familjeenhet enligt registret över totalbefolkningen (RTB)

C. Makropopulationen dvs. individpopulationen inklusive avlidna/utvandrade personer (övertäckningen), se avsnitt 1.1 samt 2.2.

D. Bostadshushåll från folk- och bostadsräkningen 1990 (FoB90) dvs. bostadshushåll och barn i åldern 0-17 år.

E. Boendeutgiftspopulationen dvs. populationen är avgränsar till boende med äganderätt, bostadsrätt och hyresrätt, se avsnitt 1.1.

Objektgrupp		Variabel	Mått
Population	Indelning i redovisningsgrupper		
<b>Inkomstfördelningsundersökningen</b>			
C	- År	- Faktorinkomst arbete och kapital - Skattepliktiga transfereringar - Skattefria transfereringar Skatt och allmänna egenavgifter - Betalt underhåll och återbetalt studielån m.m. - Disponibel inkomst	- Summor
A	- Decilgrupper - År	- Disponibel inkomst inkl. reavinst per konsumtionsenhet	- Medelvärde - Median - Gini-koefficient - Andel

A	- Decilgrupper - År	- Disponibel inkomst exkl. reavinst per konsumtionsenhet	- Medelvärde - Median - Gini-koefficient - Andel
A	- År - Gränsvärden	- Disponibel inkomst inkl. reavinst per konsumtionsenhet	- Median - Percentilgränser - Gini-koefficient - Andel
A	- År - Gränsvärden	- Disponibel inkomst exkl. reavinst per konsumtionsenhet	- Median - Percentilgränser - Gini-koefficient - Andel
A	- År	- Disponibel inkomst per konsumtionsenhet - Faktorinkomst per konsumtionsenhet	- Medelvärde - Median - Gini-koefficient - Topp 10% - Topp 1%
A	- År	- Disponibel inkomst per konsumtionsenhet - Faktorinkomst per konsumtionsenhet	- Medelvärde - Median - Gini-koefficient - Topp 10% - Topp 1%
A	- År - Hushållstyp - Ålder	- Disponibel inkomst per konsumtionsenhet	- Median
A2	- År - Hushållstyp - Ålder - Kön	- Disponibel inkomst per konsumtionsenhet	- Median
A	- År - Ålder	- Disponibel inkomst per konsumtionsenhet	- Median
A3	- År - Socioekonomisk grupp/ Syssetsättning	- Disponibel inkomst per konsumtionsenhet	- Median
A	- År - Födelseland - Vistelsetid i Sverige	- Disponibel inkomst per konsumtionsenhet	- Median
A	- År - Region	- Disponibel inkomst per konsumtionsenhet	- Median



A	- Hushållstyp - Ålder - Procent av medianvärdet för samtliga personer	- Disponibel inkomst per konsumtionsenhet	- Andel (radprocent och kolumnprocent)
A2	- Hushållstyp - Ålder - Kön - Procent av medianvärdet för samtliga personer	- Disponibel inkomst per konsumtionsenhet	- Andel (radprocent och kolumnprocent)
A	- Ålder - Procent av medianvärdet för samtliga personer	- Disponibel inkomst per konsumtionsenhet	- Andel (radprocent och kolumnprocent)
A3	- Socioekonomisk grupp/ Sysselsättning - Procent av medianvärdet för samtliga personer	- Disponibel inkomst per konsumtionsenhet	- Andel (radprocent och kolumnprocent)
A	- Födelseland - Vistelsetid i Sverige - Procent av medianvärdet för samtliga personer	- Disponibel inkomst per konsumtionsenhet	- Andel (radprocent och kolumnprocent)
A	- Region - Ålder - Procent av medianvärdet för samtliga personer	- Disponibel inkomst per konsumtionsenhet	- Andel (radprocent och kolumnprocent)
A	- Hushållstyp - År - Hushåll med en låg respektive hög	- Disponibel inkomst per konsumtionsenhet	- Andel

	<p>disponibel inkomst (50 eller 200 procent av medianen)</p>		
A2	<p>- Hushållstyp - År - Hushåll med en låg respektive hög disponibel inkomst (50 eller 200 procent av medianen) - Kön</p>	- Disponibel inkomst per konsumtionsenhet	- Andel
A	<p>- Hushållstyp - År - Hushåll med en låg respektive hög disponibel inkomst (50 eller 200 procent av medianen) - Ålder</p>	- Disponibel inkomst per konsumtionsenhet	- Andel
A3	<p>- Socioekonomisk grupp/ Syssetsättning - År - Hushåll med en låg respektive hög disponibel inkomst (50 eller 200 procent av medianen)</p>	- Disponibel inkomst per konsumtionsenhet	- Andel
A	<p>- Födelseland - Vistelsetid i Sverige - År - Hushåll med en låg respektive hög</p>	- Disponibel inkomst per konsumtionsenhet	- Andel

	disponibel inkomst (50 eller 200 procent av medianen)		
A	- Region - År - Ålder - Hushåll med en låg respektive hög disponibel inkomst (50 eller 200 procent av medianen)		
A3	- År - Kön	- Arbetsinkomst - Heltidsarbete	- Andel
A3	- År - Kön	- Arbetsinkomst	- Medelvärde
A5	- År - Kön (kvinnor)	- Arbetsinkomst	- Medelvärde - Percentilgränser - Gini- koefficienten
A5	- År - Kön (män)	- Arbetsinkomst	- Medelvärde - Percentilgränser - Gini- koefficienten
A5	- År	- Arbetsinkomst	- Medelvärde - Percentilgränser - Gini- koefficienten
A5	- År - Kvinnors arbetsinkomst i procent av mäns	- Arbetsinkomst	- Medelvärde - Percentilgränser - Gini- koefficienten
A5	- År - Ålder - Kön - Kvinnors arbetsinkomst i procent av mäns	- Arbetsinkomst	- Median
A5	- År - Socioekonomisk	- Arbetsinkomst	- Median - Andel (kvinnor i

	grupp/ Sysstättning - Kön		procent av män)
B1a	- Hushållstyp - Ålder - År	- Disponibel inkomst	- Median
B1a	- Hushållstyp - Ålder - År - Kön	- Disponibel inkomst	- Median
B1a	- Deciler - Inkomststruktur	- Disponibel inkomst per konsumtionsenhet	- Medelvärde
B1a	- Deciler - Inkomststruktur	- Disponibel inkomst per konsumtionsenhet efter justering för skatt, betalt underhållsbidrag och återbetalda studielån	- Andel
B1a	- Hushållstyp - Inkomststruktur - Ålder	- Disponibel inkomst per konsumtionsenhet	- Medelvärde
B1a	- Hushållstyp - Inkomststruktur - Ålder	- Disponibel inkomst per konsumtionsenhet efter justering för skatt, betalt underhållsbidrag och återbetalda studielån	- Andel
B1a	- Hushållstyp - Inkomststruktur - Ålder - Kön	- Disponibel inkomst per konsumtionsenhet	- Medelvärde
B1a	- Hushållstyp - Inkomststruktur - Ålder - Kön	- Disponibel inkomst per konsumtionsenhet efter justering för skatt, betalt underhållsbidrag och återbetalda studielån	- Andel
B1a	- Inkomststruktur - Ålder	- Disponibel inkomst per konsumtionsenhet	- Medelvärde
B1a	- Inkomststruktur	- Disponibel inkomst per	- Andel

	- Ålder	konsumtionsenhet efter justering för skatt, betalt underhållsbidrag och återbetalda studielån	
B1b	- Inkomststruktur - Socioekonomisk grupp/ Syssetsättning	- Disponibel inkomst per konsumtionsenhet	- Medelvärde
B1b	- Inkomststruktur - Socioekonomisk grupp/ Syssetsättning	- Disponibel inkomst per konsumtionsenhet efter justering för skatt, betalt underhållsbidrag och återbetalda studielån	- Andel
B1a	- Inkomststruktur - Födelseland - Vistelsetid i Sverige	- Disponibel inkomst per konsumtionsenhet	- Medelvärde
B1a	- Inkomststruktur - Födelseland - Vistelsetid i Sverige	- Disponibel inkomst per konsumtionsenhet efter justering för skatt, betalt underhållsbidrag och återbetalda studielån	- Andel
B1A	- Inkomststruktur - Region	- Disponibel inkomst per konsumtionsenhet	- Medelvärde
B1A	- Inkomststruktur - Region	- Disponibel inkomst per konsumtionsenhet efter justering för skatt, betalt underhållsbidrag och återbetalda studielån	- Medelvärde
	- Länder	- Disponibel inkomst per konsumtionsenhet	- Median - Ginikoefficient - Topp 20 % inkomst av botten 20 % inkomst - Andel med inkomst < 60 procent av medianen

### 1.3 Utflöden: statistik och mikrodata

**Statistik:** Undersökningen har två redovisningsområden; Inkomstfördelningen och Boende och boendeutgifter. Statistiken över inkomstfördelningen sprids i form av pressmeddelande, statistiska meddelanden (serie HE21) och tabeller på SCB:s webbplats. Även statistiken över Boende och boendeutgifter sprids i form av pressmeddelande och tabeller på SCB:s webbplats. På webbplatsen finns också resultaten av statistiken inlagt i Sveriges statistiska databaser (SSD), där tabelluttag kan göras för ett antal variabler. Statistiken sprids även i form av tabeller i *Statistisk årsbok för Sverige* samt i *Bostads- och byggnadsstatistisk årsbok*.

**Mikrodata:** Slutregister från undersökningen lagras i SAS-miljö. Undersökningen har genomförts varje år sedan 1975. Datamaterialet är användbart för åren 1975, 1978 samt för varje år från 1980 till 2007.

HEK sprids även genom SCB:s mikrosimuleringsmodell FASIT. FASIT är ett verktyg för att simulera budget- och fördelningseffekter av tänkta regelförändringar i t.ex. skatte- och bidragssystemen. HEK utgör basmaterial i denna modell. Tillgång till FASIT får man genom ett abonnemang.

### 1.4 Dokumentation och metadata

Utöver dokumentationen SCBDOK skapas varje år en variabelbeskrivning som innehåller information om population och redovisningsgrupper, bakgrunds- och klassificeringsvariabler och inkomst- och utgiftsvariabler. Dessutom skapas två variabelförteckningar, över hushållsvariabler och individvariabler, vilka är förteckningar över alla intervjuvariabler samt alla variabler som kommer från externa register.

På SCB:s webbplats finns dokumentationen *Beskrivning av statistiken* publicerad. ([www.scb.se](http://www.scb.se) – Hitta statistiken - Statistik efter ämne – Hushållens ekonomi – Hushållens ekonomi (HEK)).

## 2 Uppgiftsinsamling

### 2.1 Ram och ramförfarande

Målpopulationen för hushåll utgörs av de hushåll som den 31 december 2007 fanns i Sverige. Hushållen består av personer som enligt gällande lagar och förordningar skulle ha varit folkbokförda i Sverige 31 december 2007. Undersökningens rampopulation bestod av personer som är 18 år eller äldre och som var folkbokförda i Sverige 31 december 2007.

De brister som finns när det gäller rampopulationens täckning av målpopulationen är små.

#### **Övertäckning:**

Övertäckningen kan delas upp i två grupper. Dessa grupper utgör drygt 1 procent i undersökningens rampopulation:

- Emigranter förorsakar övertäckning om utflyttning inte anmäls. Genomförda studier tyder på att folkbokföringen innehåller ett betydande antal personer som inte längre bor i landet. Detta är förmodligen den allvarligaste kvalitetsbristen i folkbokföringen. Många som flyttar ut ur landet har inget eget intresse av att anmäla utflyttningen och en del gör det heller inte. Vid den kontroll som gjordes i samband med folk- och bostadsräkningen 1985 utgjorde den oanmälda utflyttningen ca 0,1 procent av hela befolkningen. Räknat på delpopulationen utländska medborgare utgjorde den oanmälda utflyttningen 1,0 procent för nordiska medborgare och 2,8 procent för övriga medborgarskap tillsammans.
- Personer som har mycket låga eller ofullständiga inkomst-, bidrags- eller taxeringsuppgifter.

#### **Undertäckning**

Brister i rapporteringen av födselar och invandringen resulterar i undertäckningen i befolkningsstatistiken. Den undertäckning som förekommer förklaras främst av det glapp i tid som finns mellan urvalsdragning och referenstidpunkt, vilket betyder att vi saknar information om de personer som inte fanns i Registret över totalbefolkningen (RTB) vid urvalstillfället men som tillkommit fram till den 31/12 inkomståret.

Undertäckningen i undersökningen bedöms dock vara liten.

### 2.2 Urvalsförfarande

Undersökningens rampopulation bestod av personer som är 18 år eller äldre och som under 2007 var folkbokförda i Sverige. Rampopulationen hämtas från Registret över totalbefolkning (RTB). Urvalet dras i november samma år som mätåret och det är ett stratifierat obundet slumpmässigt urval (STOSU) utan återläggning med fem strata. Urvalet är ett s.k. nätverksurval, vilket innebär att urvalspersonens hushållsmedlemmar också ingår i urvalet. Urvalsstorleken har under åren varierat mellan ca 10 000 och 19 300 hushåll. År 2007 drogs ca 17 000 personer med OSU (obundet slumpmässigt urval) inom varje stratum. För varje utvald person skapas ett förslag på hushållssammansättning med hjälp av RTB. Dessa personer bildade ca 17 000 hushåll och det totala antalet undersökta personer uppgick till ca 38 000.

Rampopulationen är stratifierad och stratifieringen har varierat över åren. För HEK 2007 var stratifieringen enligt följande;

- 18-75 år (urval 10167 personer)
- 75 år eller äldre (urval 2287 personer)
- Bostadsrätt (urval 4290 personer)
  
- Avlidna/emigranter (urval 200 personer)
- Personer med stora kapitalvinster/förluster (urval 40 personer)

För att möjliggöra särredovisning av dessa grupper med tillräcklig god kvalitet bildar de separata strata.

Ett extra stratum görs där gruppen avlidna/emigranter, dvs. de personer som räknas som övertäckning, dras för att få makropopulationen att stämma överens med verkligheten. Denna makropopulation används i huvudsak i FASIT-modellen för att kunna simulera budget- och fördelningseffekter av ändringar i skatte- och bidragssystem. Makropopulation används även i redovisningen av makrosummor.

Eftersom det är ett urval av befolkningen som ingår i undersökningen kan det bli viss osäkerhet när variabler som är extremt snedfördelade, t.ex. kapitalvinst redovisas. Det påverkar också osäkerheten för variabler där kapitalvinst ingår. För personer som under inkomståret redovisat stora kapitalvinster/-förluster vid försäljning av t.ex. värdepapper och fastigheter görs ett tilläggsurval. Syftet med tilläggsurvalet är att öka säkerheten i skattningsarna på inkomstparametrarna. Detta stratum ingår inte i intervjun.

### **Samordnat urval**

Från och med inkomståret 2000 samordnas urvalet med SCB:s Longitudinella Individdatabas (LINDA). Genom samordningen med LINDA ges möjlighet att följa urvalsenheterna över flera år.

Urvalet har en positiv samordning med LINDA-undersökningen för att möjliggöra longitudinella studier av vissa variabler. Samordningen görs genom unika permanenta slumpstal för varje person i populationen. (*Teorin om samordnade urval finns beskrivet inom företagsundersökningar och SAMU-systemet*<sup>1</sup>). Nyckelvariabel i RTB är personnummer. När ett nytt personnummer dyker upp i RTB får det ett slumpstal om det inte är frågan om återinvandring. Slumptalen är likformigt fördelade mellan noll och ett. Så länge personen/personnumret existerar i RTB bibehåller den sitt unika slumpstal. Det innebär att ett personnummer får ett unikt slumpstal när den dyker upp för första gången i RTB och så länge personnumret existerar kommer detta att vara kopplad med det unika permanenta slumpstalet. Om personerna återinvandrar återfår de sina ”gamla” slumpstal.

### **Hur kan samordningen se ut?**

Samordning av urval görs då undersökningar har samma ram, liknande eller olika ämne samt önskemål om att fördela uppgiftslämnarbelastning jämnare över ramobjekten. Nedan kommer att presenteras en kort sammanfattning av ”*Bakgrundsfakta till Ekonomisk statistik 2003:3*”.

För att samordna två, eller fler, urval med givet stickprovsstorlek  $n_1$  och  $n_2$  väljs två godtyckliga konstanter  $k_1$  och  $k_2$  i intervallet (0, 1). Element med  $n_1$  antal slumpstal med början från  $k_1$  i en riktning (höger eller vänster) ingår i första urvalet. Det andra urvalet omfattar element med  $n_2$  antal slumpstal närmast  $k_2$ , i samma eller den motsatta riktningen som  $n_1$ .

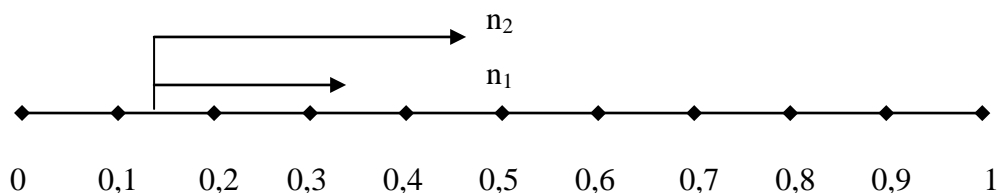
---

<sup>1</sup> Lindblom, A. (2003). Bakgrundsfakta till Ekonomisk statistik. 2003:3 SAMU The system for co-ordination of frame populations and samples from the Business Register at Statistics Sweden. SCB

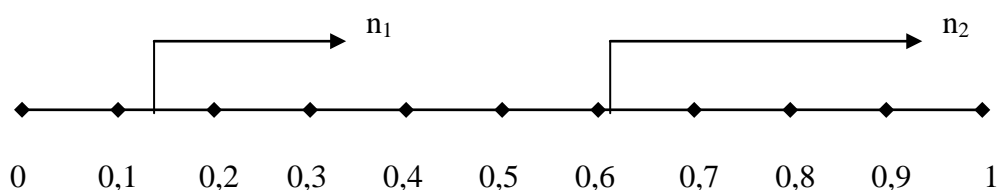


En samordning mellan två, eller fler, urval kan vara positiv eller negativ. Den positiva samordningen resulterar optimalt i flest gemensamma element och den negativa samordningen resulterar optimalt i minst eller inga gemensamma element. Den maximala positiva samordningen mellan två undersökningar uppnås då det väljs samma startpunkt och samma riktning. Den negativa samordningen kan uppnås genom att välja åtskilda start punkter men i samma riktning. Ett alternativ är att börja från samma startpunkt (eller två punkter ganska nära varandra) och dra urval i olika riktningar. Se bild 1, 2 och 3. Det finns inte alltid tillräckligt med element för att uppnå det maximala negativa samordningen, dvs. inga gemensamma element, men på det här sättet minimeras antalet gemensamma element.

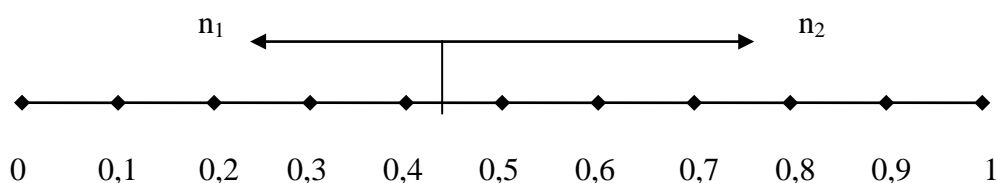
Figur 1. Positiv samordning



Figur 2. Negativ samordning alternativ 1



Figur 3. Negativ samordning alternativ 2



Genom den här tekniken kan ramen delas i flera block allt detta för att sprida uppgiftslämnarbördan. På det här sättet kan undersökningar inom närbesläktade områden väljas inom samma block.

## 2.3 Mätinstrument

Mätinstrumentet är ett intervjuformulär som används i en datorstött telefonintervju.

Exempel på felkällor vid intervjutillfället är att uppgiftslämnaren kan missförstå frågorna intervjuaren ställer eller svaret uppgiftslämnaren ger kan registreras felaktigt. För att minimera dessa felkällor finns i datorstödet ett inbyggt kontrollsystem. Minnesfel kan också förekomma eftersom en del av frågorna avser förhållanden minst 12 månader bakåt i tiden.

## 2.4 Insamlingsförfarande

Uppgifter samlas in genom telefonintervjuer och från administrativa register.

### *Intervju*

Intervjuerna genomförs med datorstöd (WinDATI). Intervjuerna genomförs under januari-maj året efter inkomståret. Frågor om hushållets sammansättning och boendeform avser förhållandet den 31/12 inkomståret, medan övriga frågor gäller hela inkomståret. Frågor kring boendeutgifter ställs till dem som bott i sin bostad hela året. Vissa frågor ställs om alla personer i hushållet. Det innebär att en stor del av frågorna besvaras i en s.k. indirekt intervju.

Intervjuarbetet är uppdelat i fyra omgångar. För att göra en indelning av urvalet i fyra undersökningsomgångar hämtas uppgifter om boendeform (hyreslägenhet, bostadsrätt eller småhus) från Fastighetstaxeringsregistret. Fastighetstaxeringsregistret innehåller uppgifter om bland annat fastighet. De tillgängliga uppgifterna är minst 1 år gamla. Det kan vara svårt att koppla ihop en person med en särskild fastighet. Det gäller främst för bostadsrätter och hyreslägenheter eftersom en fastighet kan bestå av många lägenheter.

Felkällor i intervjun är exempelvis att uppgiftslämnaren kan missförstå frågorna eller att svaret registreras felaktigt. För att minimera dessa felkällor finns kontroller inbyggda i intervjun. Den största felkällan är minnesfel eftersom de flesta frågor avser förhållanden minst 12 månader bakåt i tiden. Frågor om utgifter för tandvård, sjukvård och läkemedel kan vara svåra att besvara eftersom bland annat högkostnadsskydd och rabattsystem påverkar utgifterna. För att försöka minska risken för minnesfel finns ett utdrag av frågorna i informationsbrevet. Hushållet har då möjlighet att förbereda sig genom att ta fram uppgifter innan intervjun genomförs.

### *Administrativa register*

De administrativa registren är i första hand till för administrativa ändamål och svarar därmed inte alltid mot statistikens behov. Ett exempel på detta är hushållets sammansättning. I telefonintervjun frågar vi efter hushållets medlemmar och kan på så vis skapa ett hushåll kring urvalspersonen. För bortfallet används uppgifter från RTB. RTB-familjen kan bestå av en ensamstående eller personer som är gifta med varandra (även registrerat partnerskap), eventuella barn till dessa personer som är folkbokförda på samma adress ingår också, detta oavsett barnets ålder. RTB-familjen kan även bestå av sammanboende som har gemensamma barn. RTB-familjen kan bestå av maximalt två generationer. För sammanboende par utan gemensamma barn finns inga uppgifter för att skapa korrekta hushåll. Därmed sker en överskattning av andelen ensamstående och ensamstående hushåll.

För att kontrollera kvaliteten på andra viktiga variabler görs en avstämning mellan olika register. Kvaliteten på uppgifterna i de administrativa registren bedöms vara bra.

## 2.5 Databeredning

Ett omfattande gransknings- och rättningsarbete utförs.

### *Intervju*

För intervjun genomförs kontroller i samband med intervjun och efter avslutad intervju, logiska kontroller och värdekontroller. Trots detta förekommer fel i materialet. Exempelvis är sysselsättningsnivån under året svår att mäta. Det görs en automatisk kodning av socioekonomisk

grupp under intervjun/i WinDATI. Det är ungefär 90 procent som kodas på detta sätt, övriga 10 procent kodas manuellt.

#### *Administrativa register*

De variabler som hämtas från olika administrativa register bedöms hålla god kvalitet. När uppgifterna aviserats till SCB granskas de i olika skeden av bearbetningarna innan de godkänns för inläggning i respektive register.

#### *Granskning av data*

Vid skapande av variabler kan fel uppstå. Omfattande kontroller görs där de skapade variablerna bl.a. stäms av mot uppgifterna från de administrativa registren.

#### *Framställning av resultat*

Punktskattningar och konfidensintervall har beräknats med hjälp av ett av SCB:s egenutvecklade estimationsprogram, EXTRACT. I bakgrunden av EXTRACT ligger CLAN97 och EOS för beräkningen av estimatens medelfel. Ytterligare bearbetningar görs i EXCEL för att skapa de tabeller som sedan publiceras på SCB:s webbplats, [www.scb.se](http://www.scb.se), se avsnitt 1.3.

### 3 Statistisk bearbetning och redovisning

#### 3.1 Skattningar: antaganden och beräkningsformler

Undersökningens målpopulation är hushåll. I avsnitt 1.2 (Statistiska målstorheter) redogörs för vilka målstorheter som man utifrån HEK vill skatta för respektive objektstyp. Flertalet av de variabler som anges i tabellen finns direkt tillgängliga i, eller kan härleds utifrån, SCB:s register. Med undantag för de indelningar som baseras på socioekonomisk gruppstillhörighet, gäller dessutom att samtliga indelningar i redovisningsgrupper delvis eller helt baseras på registerinformation.

För målstorheter definierade utifrån variabler för vilka SCB saknar registerinformation baseras skattningarna på den variabelinformation som finns tillgänglig för de svarande. För målstorheter definierade utifrån variabler för vilka information finns tillgänglig i SCB:s register finns två möjligheter. I de fall då registerinformation kan användas, ev. som komplement till insamlad svarsinformation, för indelning i redovisningsgrupper, utnyttjas variabelinformation för samtliga objekt i urvalet i skattningsförfarandet. I övriga fall baseras skattningarna på den variabelinformation som finns tillgänglig för de svarande.

Nedan redogörs i detalj för hur skattningsförfarandet går till för målstorheter definierade som totalsummor. För målstorheter definierade utifrån de statistiska måtten medelvärde, percentil samt Gini-koefficient är framställningen inte så detaljerad, utan istället hänvisas till lämplig litteratur.

##### 3.1.1 Storheter definierade som totalsummor

Låt  $U_h$  beteckna mängden av de  $N_h$  personer som tillhör urvalsramen i stratum  $h$  ( $h = 1, \dots, 5$ ). Låt vidare  $U_{HH}$  beteckna mängden av de  $N_{HH}$  hushåll som är identifierbara utifrån de sammanlagt

$N = \sum_{h=1}^5 N_h$  personer som ingår i urvalsramen  $U = \bigcup_{h=1}^5 U_h$ . Slutligen, för hushåll  $i \in U_{HH}$ , låt  $U'_i$

beteckna mängden av de  $N'_i$  personer som utgör hushållet och låt  $U_i$  beteckna mängden av de  $N_i \leq N'_i$  personer i hushållet som är valbara utifrån urvalsramen  $U$ .

Låt  $y$  beteckna den variabel som är av intresse och låt  $y_k$  beteckna variabelns värde för person  $k$ . Då målobjekten utgörs av hushåll gäller att den totala för variabeln  $y$  som kan skattas ges av:

$$t_{y_{HH_d}} = \sum_{i \in U_{HH}} I_{di} t_{yi}$$

där:

$$t_{yi} = \sum_{k \in U'_i} y_k = \text{Totalsumman för variabeln } y \text{ i hushåll } i$$

och:

$$I_{di} = \begin{cases} 1 & \text{om hushåll } i \text{ tillhör redovisningsgrupp } d \\ 0 & \text{annars} \end{cases}$$

Den total som kan skattas då målobjekten utgörs av personer ges av:

$$t_{yIND_d} = \sum_{i \in U_{HH}} t_{y_d^i}$$

där:

$$t_{y_d^i} = \sum_{k \in U_i'} I_{dk} y_k = \text{Totalsumman för variabeln } y \text{ för de medlemmar i hushåll } i \\ \text{som tillhör redovisningsgrupp } d$$

med:

$$I_{dk} = \begin{cases} 1 & \text{om person } k \text{ tillhör redovisningsgrupp } d \\ 0 & \text{annars} \end{cases}$$

### 3.1.2 Beräkningsformler vid skattning av totalsummor

För att skatta storheterna  $t_{yHH_d}$  respektive  $t_{yIND_d}$  dras, vilket också beskrivs i avsnitt 2.2, enligt urvalsdesignen OSU-UÅ ett urval  $s_h$  om  $n_h$  personer från de  $N_h$  personer som tillhör i urvalsramen i stratum  $h$  ( $h = 1, \dots, 5$ ). I stratum  $h$  erhålls slutligen svar från  $m_h$  av de  $n_h$  personer som tillhör urvalet  $s_h$  ( $h = 1, \dots, 5$ ). Mängden svarande i stratum  $h$  betecknas  $r_h$  ( $h = 1, \dots, 5$ ).

Av avsnitt 2.2 framgår även att urvalet i HEK är ett nätverksurval. Under den urvalsdesign som används har inte alla hushåll som tillhör mängden  $U_{HH}$  samma sannolikhet att inkluderas i ett urval. Denna sannolikhet är nämligen en funktion av hur många personer i ett hushåll som är valbara utifrån urvalsramen. Något förenklat gäller att ju fler valbara personer som finns i ett hushåll, desto större chans är det att hushållet inkluderas i ett urval. Genom att använda en s.k. nätverksvikt vid skattningsförfarandet beaktas detta faktum. I HEK ges nätverksvikten för person  $k \in U_i$  av:

$$a_k = \frac{n_{h(k)} / N_{h(k)}}{\sum_{l \in U_{i(k)}} n_{h(l)} / N_{h(l)}}$$

där  $h(k)$  och  $i(k)$  betecknar det stratum respektive det hushåll som person  $k$  tillhör.

*Beräkningsformler då variabelinformation på urvalsnivå utnyttjas i skattningsförfarandet*

Då information för samtliga objekt i urvalet utnyttjas, används Horvitz-Thompson-estimation för att skatta totalerna ovan. Låt  $w_k = N_h a_k / n_h$  för  $k \in s_h$  ( $h = 1, \dots, 5$ ). Storheten  $t_{yHH_d}$  skattas med:

$$\hat{t}_{yHH_d} = \sum_{h=1}^5 \sum_{k \in s_h} w_k I_{di(k)} t_{yi(k)}$$

medan storheten  $t_{yIND_d}$  skattas med:

$$\hat{t}_{yIND_d} = \sum_{h=1}^5 \sum_{k \in s_h} w_k t_{y_i(k)}$$

Förutsatt att inga mätfel existerar är ovanstående estimatorer väntevärdesriktiga.

Estimatorernas varians skattas med:

$$\hat{V}(\hat{t}_{yHH_d}) = \sum_{h=1}^5 \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \frac{\sum_{k \in s_h} \left( a_k I_{di(k)} t_{yi(k)} - \sum_{k \in s_h} a_k I_{di(k)} t_{yi(k)} / n_h \right)^2}{n_h - 1}$$

respektive:

$$\hat{V}(\hat{t}_{yIND_d}) = \sum_{h=1}^5 \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \frac{\sum_{k \in s_h} \left( a_k t_{y_i(k)} - \sum_{k \in s_h} a_k t_{y_i(k)} / n_h \right)^2}{n_h - 1}$$

*Beräkningsformler då variabelinformation på svarsmängdsnivå utnyttjas i skattningsförfarandet*

Bortfall uppstår om man inte lyckas få svar från samtliga utvalda objekt. Förekomsten av bortfall medför alltid en ökning av urvalsfelet. Om de icke-svarande objekten dessutom systematiskt avviker från de svarande beträffande de variabler som undersöks tillkommer ytterligare ett fel. Storleken av detta fel, som är av systematisk natur och benämns bortfallsskevheter eller bortfallsbias, kan i praktiken inte kvantifieras. Ofta kan man dock reducera de fel som bortfall ger upphov till genom att utnyttja s.k. hjälpinformation i form av registervariabler. Kalibrering av vikter är en i detta sammanhang både effektiv och flexibel metod. (För mer information, se t.ex. Lundström 1997; Lundström och Särndal 1999, 2001).<sup>2</sup>

Om den hjälpinformation som finns tillgänglig samvarierar väl med de utvalda objektens svarsbenägenhet, kan man förvänta sig att bortfallsbiasen reduceras genom utnyttjandet av hjälpinformationen. Om hjälpvariablerna dessutom till betydande del samvarierar med viktiga

<sup>2</sup> Lundström, S. (1997). Calibration as a standard method for treatment of nonresponse. Doktorsavhandling, Stockholms universitet.

Lundström, S. och Särndal, C.E. (1999). Calibration as a standard method for treatment of nonresponse. Journal of Official Statistics, 15, 305-327.

Lundström, S. och Särndal, C.E. (2001). Estimation in the Presence of Nonresponse and Frame Imperfections. SCB.

undersökningsvariabler, bör utnyttjandet av hjälpinformationen ytterligare reducera bortfallsbiasen samtidigt som även urvalsfelet reduceras.

Då endast variabelinformation för  $k \in r_h$  utnyttjas i skattningsförfarandet, skattas totalerna  $t_{yHH_d}$  och  $t_{yIND_d}$  utifrån estimatorer som utnyttjar kalibrerade vikter. Vid val av hjälpinformation bör man, om möjligt, välja den på så sätt

- (i) att den samvarierar med svarsbenägenheten
- (ii) att den samvarierar med (viktiga) undersökningsvariabler
- (iii) att den avgränsar (viktiga) redovisningsgrupper.

Om personerna i bortfallet skiljer sig från de svarande med avseende på undersökningsvariablerna så kan skattningarna som grundar sig på enbart de svarande bli skeva. Om exempelvis de svarande till högre andel bor i hyresrätt än de som inte svarat finns risk för överskattning av hushåll boende i hyresrätt. För att försöka reducera eventuella bortfallsskevheter har vikter beräknats med hjälp av kalibrering. Storleken på kvarstående skevheter orsakade av bortfallet är svåra att mäta.

Efter en sammanvägning av analysen kring de tre kriterierna (se avsnittet om konfidensintervall sid 38) samt efter kontroll av vikternas fördelning används hjälpvektor beskriven i tabellen. Uppgifterna i hjälpvektor för en viss variabel kan vara antal personer med värde på den variabel och/eller totalen för den variabel. I tabellen markeras vilken typ av uppgiften som används.

Hjälpvariabel (klasser)	Antal personer	Antal månader	Totalen
Åldersklass (10)	X		
Kön	X		
H-region (9)	X		
Ekonomisk bistånd	X	X	X
Barnbidrag	X		X
Underhållsstöd	X		X
Föräldrapenning vid födelse	X		X
Föräldrapenning vid tillfällig vård av barn	X		X
Bostadsbidrag	X		X
Bostadstillägg till pensionärer	X		X
Sjukpenning	X		X
Arbetsmarknadsstöd (A-kassa m.m.)	X		X
Studiestöd	X		X
Löneinkomst (5)	X		X
Födelseland (6)	X		
Invandringsår (3)	X		
Hushållstyp/familjetyp (10)	X		

För variabeln *arbets tid* och de beskrivande *boende*variablerna används imputeringsmetod för bortfallskompensation (se avsnitt 4.1.4). Metoden tas fram med hjälp av statistiskt förutbestämda regler tillsammans med registervariabler samt inkomna värden från intervjun.

Variabeln *barnomsorgsutgift* tas fram i FASIT-modellen.

Låt  $\mathbf{x}$  beteckna den hjälpvektor som skapas utifrån hjälpvariablerna. Storheten  $t_{yHH_d}$  skattas med:

$$\hat{t}_{yHH_d} = \sum_{h=1}^5 \sum_{k \in r_h} w_k I_{di(k)} t_{yi(k)}$$

medan storheten  $t_{yIND_d}$  skattas med:

$$\hat{t}_{yIND_d} = \sum_{h=1}^5 \sum_{k \in s_h} w_k t_{yi(k)}$$

där  $w_k = N_h g_k a_k / m_h$  för  $k \in r_h$  ( $h=1, \dots, 5$ ). Vikten  $g_k$ , som är den kalibreringsvikt som erhålls då hjälpinformationen används för kalibrering, ges av:

$$g_k = 1 + \left( \mathbf{t}_x - \sum_{h=1}^5 \frac{N_h}{m_h} \sum_{k \in r_h} a_k \mathbf{t}_{xi(k)} \right)' \left( \sum_{h=1}^5 \frac{N_h}{m_h} \sum_{k \in r_h} a_k \mathbf{t}_{xi(k)} \mathbf{t}_{xi(k)}' \right)^{-1} \mathbf{t}_{xi(k)}$$

där  $\mathbf{t}_{xi} = \sum_{k \in U_i'} \mathbf{x}_k$  och  $\mathbf{t}_x$  avser samtliga kosthushåll i Sverige och därför hämtas från Registret över totalbefolkningen (RTB) eller från Inkomst- och taxeringsregistret (IoT).

Ovanstående estimatorers varians skattas med:

$$\hat{V}(\hat{t}_{yHH_d}) = \sum_{h=1}^5 \frac{N_h^2}{m_h} \left( 1 - \frac{m_h}{N_h} \right) \frac{\sum_{k \in r_h} \left( u_{HH_dk} - \sum_{k \in r_h} u_{HH_dk} / m_h \right)^2}{m_h - 1}$$

respektive:

$$\hat{V}(\hat{t}_{yIND_d}) = \sum_{h=1}^5 \frac{N_h^2}{m_h} \left( 1 - \frac{m_h}{N_h} \right) \frac{\sum_{k \in r_h} \left( u_{IND_dk} - \sum_{k \in r_h} u_{IND_dk} / m_h \right)^2}{m_h - 1}$$



där:

$$u_{HH_dk} = g_k a_k (I_{di(k)} t_{yi(k)} - \mathbf{t}'_{xi(k)} \hat{\mathbf{B}}_{HH_d})$$

med:

$$\hat{\mathbf{B}}_{HH_d} = \left( \sum_{h=1}^5 \frac{N_h}{m_h} \sum_{k \in r_h} a_k \mathbf{t}_{xi(k)} \mathbf{t}'_{xi(k)} \right)^{-1} \sum_{h=1}^5 \frac{N_h}{m_h} \sum_{k \in r_h} a_k \mathbf{t}_{xi(k)} I_{di(k)} t_{yi(k)}$$

och:

$$u_{IND_dk} = g_k a_k (t_{ydi(k)} - \mathbf{t}'_{xi(k)} \hat{\mathbf{B}}_{IND_d})$$

med:

$$\hat{\mathbf{B}}_{IND_d} = \left( \sum_{h=1}^5 \frac{N_h}{m_h} \sum_{k \in r_h} a_k \mathbf{t}_{xi(k)} \mathbf{t}'_{xi(k)} \right)^{-1} \sum_{h=1}^5 \frac{N_h}{m_h} \sum_{k \in r_h} a_k \mathbf{t}_{xi(k)} t_{ydi(k)}$$

### 3.1.3 Skattningsförfarande för storheter definierade utifrån övriga statistiska mått

#### Medelvärden

Då målobjekten utgörs av hushåll gäller att hushållsmedelvärdet för variabeln  $y$  i redovisningsgrupp  $d$  ges av:

$$\bar{y}_{HH_d} = \frac{\sum_{i \in U_{HH}} I_{di} \sum_{k \in U'_i} y_k}{\sum_{i \in U_{HH}} I_{di}} = \frac{t_{yHH_d}}{N_{HH_d}}$$

Notera att om  $z_k = 1/N'_i$  för  $k \in U'_i$  så gäller att  $\sum_{k \in U'_i} z_k = 1$  för alla  $i \in U_{HH}$ , vilket i sin tur innebär

att  $t_{zHH_d} = N_{HH_d}$ . Således gäller att nämnaren i uttrycket för  $\bar{y}_{HH_d}$  kan betraktas som en storhet definierad som en totalsumma.

Då målobjekten utgörs av personer gäller att personmedelvärdet för variabeln  $y$  i redovisningsgrupp  $d$  ges av:

$$\bar{y}_{IND_d} = \frac{\sum_{i \in U_{HH}} \sum_{k \in U'_i} I_{dk} y_k}{\sum_{i \in U_{HH}} \sum_{k \in U'_i} I_{dk}} = \frac{t_{yIND_d}}{N'_{IND_d}}$$

Notera att om  $z_k = 1$  för alla  $k$  så gäller att  $t_{zIND_d} = N'_{IND_d}$ , dvs. även nämnaren i uttrycket för  $\bar{y}_{IND_d}$  kan betraktas som en storhet definierad som en totalsumma.

I HEK skattas ovanstående medelvärden med:

$$\hat{y}_{HH_d} = \frac{\hat{t}_{yHH_d}}{\hat{N}_{HH_d}}$$

respektive:

$$\hat{y}_{IND_d} = \frac{\hat{t}_{yIND_d}}{\hat{N}'_{IND_d}}$$

där formeluttryck för  $\hat{t}_{yHH_d}$ ,  $\hat{N}_{HH_d}$ ,  $\hat{t}_{yIND_d}$  och  $\hat{N}'_{IND_d}$  ges i avsnitt 4.1.2.

Varianserna för estimatorerna  $\hat{y}_{HH_d}$  och  $\hat{y}_{IND_d}$  skattas utifrån uttryck som erhålls då formeluttrycken för  $\hat{y}_{HH_d}$  och  $\hat{y}_{IND_d}$  approximeras genom Taylorutveckling av första ordningen. (För mer information om hur variansestimatorer erhålls med hjälp av Taylorutveckling av första ordningen, se t.ex. kap 5.5 i Särndal, Swensson, Wretman 1992.)<sup>3</sup>

#### *Percentiler och percentilgruppstotaler*

Då målobjekten utgörs av hushåll, låt:

$$F_{HH_d}(t) = \frac{\sum_{i \in U_{HH}} I_{di} I\{t_{yi} \leq t\}}{N_{HH_d}}$$

där  $I\{\cdot\}$  är en indikatorfunktion sådan att:

$$I\{A\} = \begin{cases} 1 & \text{om villkoret } A \text{ är uppfyllt} \\ 0 & \text{annars} \end{cases}$$

Funktionen  $F_{HH_d}(t)$  anger andelen hushåll i redovisningsgrupp  $d$  vars hushållstotal inte överstiger det numeriska värdet  $t$ . Låt  $t_{dp}$  beteckna det värde för vilket  $F_{HH_d}(t_{dp}) = p$ , dvs. för 100 $p$  % av hushållen i redovisningsgrupp  $d$  gäller att hushållstotalen inte överstiger det numeriska värdet  $t_{dp}$ . Talet  $t_{dp}$  benämns den  $p$ :te percentilen i redovisningsgrupp  $d$ .

Då målobjekten utgörs av individer, låt:

$$F_{IND_d}(y) = \frac{\sum_{i \in U_{HH}} \sum_{k \in U_i} I_{dk} I\{y_k \leq y\}}{N_{IND_d}}$$

<sup>3</sup> Särndal, C.E., Swensson, B. och Wretman, J.H. (1992). Model Assisted Survey Sampling. New York: Springer-Verlag

Funktionen  $F_{IND_d}(y)$  anger andelen personer i redovisningsgrupp  $d$  vars variabelvärde inte överstiger det numeriska värdet  $y$ . Låt  $y_{dp}$  beteckna det värde för vilket  $F_{IND_d}(y_{dp}) = p$ , dvs. för 100  $p$  % av personerna i redovisningsgrupp  $d$  gäller att variabelvärdet inte överstiger det numeriska värdet  $y_{dp}$ . Talet  $y_{dp}$  benämns den  $p$ :te percentilen i redovisningsgrupp  $d$ .

Eftersom såväl  $F_{HH_d}(t)$  som  $F_{IND_d}(y)$  är definierade som kvoter av totalsummor, skattas de genom att totalsummorna ersätts med motsvarande estimatorer enligt avsnitt 4.1.2. Om variabelinformation på urvalsnivå utnyttjas i skattningsförfarandet ges en estimator för  $t_{dp}$  av:

$$\hat{t}_{dp} = \min_t \{t_{i(k)} \mid \hat{F}_{HH_d}(t_{i(k)}) \geq p \text{ och } k \in \bigcup_{h=1}^H s_h\}$$

medan:

$$\hat{t}_{dp} = \min_t \{t_{i(k)} \mid \hat{F}_{HH_d}(t_{i(k)}) \geq p \text{ och } k \in \bigcup_{h=1}^H r_h\}$$

används om enbart variabelinformation på svarsmängdsnivå utnyttjas i skattningsförfarandet. På motsvarande sätt skattas  $y_{dp}$  med:

$$\hat{y}_{dp} = \min_y \{y_k \mid \hat{F}_{IND_d}(y_k) \geq p \text{ och } k \in \bigcup_{h=1}^H \bigcup_{l \in s_h} U'_{i(l)}\}$$

om variabelinformation på urvalsnivå utnyttjas i skattningsförfarandet, medan:

$$\hat{y}_{dp} = \min_y \{y_k \mid \hat{F}_{IND_d}(y_k) \geq p \text{ och } k \in \bigcup_{h=1}^H \bigcup_{l \in r_h} U'_{i(l)}\}$$

används om enbart variabelinformation på svarsmängdsnivå utnyttjas i skattningsförfarandet.

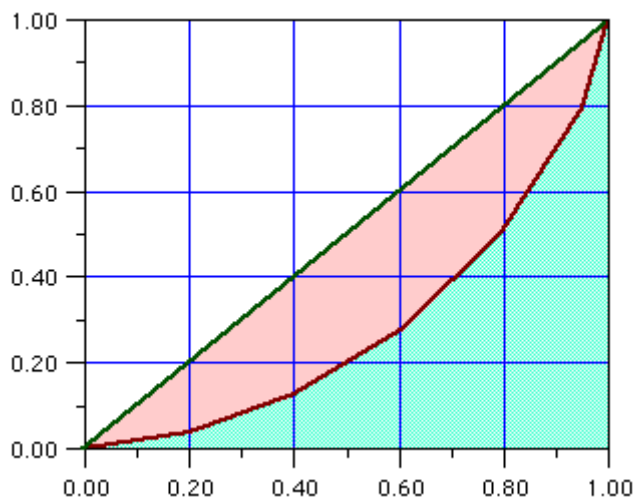
Vid skattning av storheter i HEK som utnyttjar redovisningsgrupper definierade från percentilmått, används skattade percentiler enligt ovan för att skapa redovisningsgruppsindelningen.

För estimatorer av percentiler beräknas inga variansestimater.

### *Gini-koefficienter*

Gini-koefficienten är ett mått som används för att redovisa hur stor spridningen är i en inkomst- eller förmögenhetsfördelning. Koefficienten, som beräknas utifrån den s.k. Lorenz-kurvan, är ett ojämlikhetsmått. Förutsatt att den variabel som studeras endast kan anta icke-negativa värden, gäller att Gini-koefficienten definitionsmässigt antar ett värde mellan 0 och 1, diagram 1. Då gäller att ju högre värde koefficienten antar, desto mer ojämlik är den fördelningen som studeras.

Diagram 1. Lorenz-kurvan



Gini-koefficienten är den rosa ytans andel av hela triangeln.

Då målobjekten utgörs av hushåll definieras storheten Gini-koefficient för redovisningsgrupp  $d$  som:

$$G_{HH_d} = \frac{\sum_{i \in U_{HH}} \sum_{j \in U_{HH}} I_{di} I_{dj} |t_{yi} - t_{yj}|}{2N_{HH_d} t_{yHH_d}}$$

medan definitionen då målobjekten utgörs av individer ges av:

$$G_{IND_d} = \frac{\sum_{i \in U_{HH}} \sum_{j \in U_{HH}} \sum_{k \in U'_i} \sum_{l \in U'_j} I_{dk} I_{dl} |y_k - y_l|}{2N_{IND_d} t_{yIND_d}}$$

Om variabelinformation på urvalsnivå utnyttjas i skattningsförfarandet ges en estimator för  $G_{HH_d}$  av:

$$\hat{G}_{HH_d} = \frac{\sum_{h=1}^5 \sum_{k \in s_h} I_{di(k)} 2\hat{F}_{HH_d}^*(t_{yi(k)}) w_k t_{yi(k)}}{\hat{t}_{yHH_d}} - 1$$

där:

$$\hat{F}_{HH_d}^*(t_{yi(k)}) = \frac{\sum_{h=1}^5 \sum_{l \in s_h} I_{di(l)} w_l [I\{t_{yi(l)} < t_{yi(k)}\} + I\{i(l) = i(k)\} / 2]}{\hat{N}_{HH_d}}$$

medan:

$$\hat{G}_{HH_d} = \frac{\sum_{h=1}^5 \sum_{k \in r_h} I_{di(k)} 2\hat{F}_{HH_d}^*(t_{yi(k)}) w_k t_{yi(k)}}{\hat{t}_{yHH_d}} - 1$$

där:

$$\hat{F}_{HH_d}^*(t_{yi(k)}) = \frac{\sum_{h=1}^5 \sum_{l \in r_h} I_{di(l)} w_l [I\{t_{yi(l)} < t_{yi(k)}\} + I\{i(l) = i(k)\} / 2]}{\hat{N}_{HH_d}}$$

används om enbart variabelinformation på svarsmängdsnivå utnyttjas i skattningsförfarandet.

På motsvarande sätt skattas  $G_{IND_d}$  med:

$$\hat{G}_{IND_d} = \frac{\sum_{h=1}^5 \sum_{l \in s_h} w_l \sum_{k \in U_{i(l)}} 2I_{dk} \hat{F}_{IND_d}^*(y_k) y_k}{\hat{t}_{yIND_d}} - 1$$

där:

$$\hat{F}_{IND_d}^*(y_k) = \frac{\sum_{h=1}^5 \sum_{l \in s_h} w_l \sum_{l' \in U_{i(l)}} I_{dl'} [I\{y_{l'} < y_k\} + I\{l' = k\} / 2]}{\hat{N}_{IND_d}}$$

om variabelinformation på urvalsnivå utnyttjas i skattningsförfarandet, medan:

$$\hat{G}_{IND_d} = \frac{\sum_{h=1}^5 \sum_{l \in r_h} w_l \sum_{k \in U_{i(l)}} 2I_{dk} \hat{F}_{IND_d}^*(y_k) y_k}{\hat{t}_{yIND_d}} - 1$$

där:

$$\hat{F}_{IND_d}^*(y_k) = \frac{\sum_{h=1}^5 \sum_{l \in r_h} w_l \sum_{l' \in U_{i(l)}} I_{dl'} [I\{y_{l'} < y_k\} + I\{l' = k\} / 2]}{\hat{N}_{IND_d}}$$

används om enbart variabelinformation på svarsmängdsnivå utnyttjas i skattningsförfarandet.

Varianserna för ovanstående estimatorer av Gini-koefficienten skattas med hjälp av den varians estimator som ges i Tabell 4.1, s. 13 – 14, i Sandström, Wretman och Waldén (1985).<sup>4</sup>

### Konfidensintervall

---

<sup>4</sup> Sandström, A., Wretman, J.H. och Waldén, B. (1985). Variance estimators of the Gini coefficient - probability sampling. Promemorior från P/STM nr. 17, SCB.

Osäkerhetsintervall beräknas enligt formeln

$$\text{punktskattning} \pm 1,96 * (\text{skattning av estimatorns medelfel})$$

där såväl punkt- som medelfelsskattning beräknas med hjälp av ett av SCB egenutvecklat program, EXTRACT. Under förutsättning att stickprovsstorleken är stor samt

- (i) att den estimator som används är approximativt väntevärdesriktig, dvs. icke-skev
- (ii) att estimatorns varians endast beror av samplingvariansen och bortfallsvariansen

och

- (iii) att de registervariabler som används i kalibreringen uppfyller de tre kriterierna

gäller att sannolikheten är ca 95 procent att man med det använda urvals- och skattningsförfarandet erhåller ett intervall som täcker in den sökta storheten.

Naturligtvis är inte samtliga ovanstående förutsättningar perfekt uppfyllda i praktiken. Även om hjälpvariablerna är klokt valda kommer estimatorn att vara behäftad med viss skevhet, delvis till följd av kvarstående bortfalls- och täckningsfel men också till följd av andra fel av systematisk karaktär. Vidare är det troligt att estimatorns varians underskattas. Utöver samplingvariansen och bortfallsvariansen tillkommer även ett variansbidrag från andra osäkerhetskällor av slumpmässig natur, men vanligtvis kan endast en del av detta variansbidrag ”fångas upp” av den varians estimator som används. Tyvärr är det inte möjligt att avgöra i vilken utsträckning som avvikelser från förutsättningarna ovan föreligger och vilka effekter dessa avvikelser får för tolkningen av de osäkerhetsintervall som presenteras i termer av faktisk konfidensgrad. En rekommendation är dock att man bör vara försiktig med att tolka de presenterade intervallen som regelrätta konfidensintervall.

### 3.1.4 Svartsbortfall

Svartsbortfall, vanligen förkortat till bortfall, uppstår när värden på en eller flera observationsvariabler (ex. registervariabler eller intervjufrågor) i en undersökning inte kan inhämtas.

- Saknas alla värden för en person är det frågan om **objektsbortfall**.
- Saknas enbart vissa värden, handlar det om **partiellt bortfall**.

#### Andelen svartsbortfall i HEK2007

Avböjde medverkan	14,9
Ej anträffade	13,3
Förhindrad medverkan	3,1
Totalt svartsbortfall	31,4

*Svartsbortfallet är oviktat*

I bortfallet är följande grupper överrepresenterade: De som saknar löneinkomst, de som erhållit ekonomiskt bistånd och de som invandrat till Sverige från länder utanför Norden. Bland de svarande är barnfamiljerna något överrepresenterade.

Svartsbortfallet i HEK har ökat under 2000-talet. Bortfallet har dock varierat mellan åren. De år då urvalet varit stort har bortfallet varit högre. I samband med att målpopulationen utökades till att

omfatta äldre personer som bor i s.k. särskilt boende (boende på institution) höjdes bortfallsnivån eftersom det är en svår grupp att nå för intervju.

För att minska snedvridande effekter av bortfallet görs en riktad bortfallsuppföljning. Det primära i HEK:ens bortfallsuppföljningen är att minska effekterna av bortfallet snarare än att minska bortfallsnivån.

Administrativ data används som hjälpinformation vid beslut om vilka som ska gå till uppföljningen. De grupper som bidrar mest till bortfallets snedvridande effekt prioriteras i bortfallsuppföljningen.

Kalibrering och imputering används för bortfallskompensation i urvalet (se sid 30). Kalibrering används för de flesta estimaten som baseras på de svarande i intervjun. Vi räknar med att denna metod kraftigt reducerar den snedvridande effekt som bortfallet har. Imputering används för vissa av boende variabler bl.a. hushållstyp, uppvärmning m.fl. Imputeringen av hushållstyp baseras på registeruppgifter för hushållssammansättningen, vilket bidrar till att antalet ensamboende och ensamstående med barn överskattas (och antalet sammanboende par underskattas). Vid redovisning av variabler som finns i administrativa register används uppräknings utan hjälpinformation baserad på hela urvalet.

### 3.2 Redovisningsförfaranden

De bakgrunds- och klassificeringsvariabler som hämtas från intervjun och administrativa register används för att bilda redovisningsgrupper. Den huvudsakliga redovisningen av inkomstfördelningen görs för personer/hushåll (kosthushåll och familjeenhet). Eftersom statistiken redovisar inkomster och utgifter för inkomståret används helårspopulationen, de personer som enligt gällande lagar och förordningar skulle ha varit folkbokförda i Sverige under hela 2007. Institutionshushåll ingår inte i redovisningen. Deras inkomst- och utgiftsförhållanden är inte jämförbara med övriga hushåll. Personer som är 65 år eller äldre och bor i särskilt boende ingår dock.

Kosthushåll kan redovisas fr.o.m. 1991 års undersökning. Undantag gäller för år 1992 där begreppet kosthushåll inte finns samt för åren 1993-1994 där begreppet endast finns för de svarande i undersökningen. För redovisning över en längre tidsperiod, från 1975 och framåt, används familjeenhet.

För att göra jämförelser av t.ex. disponibel inkomst mellan olika typer av hushåll används ett viktsystem där inkomsten relateras till hushållets sammansättning. Den disponibla inkomsten divideras med den vikt som gäller för hushållet. Viktsystemet redovisas nedan.

**Tabell 1** Ekvivalensskalan/konsumtionsenhetsskalan

	2007
Ensamboende och första vuxna i ett samboförhållande	1,00
Andra vuxen i samboförhållande	0,51
Ytterligare vuxen	0,60
Första barnet 0-19 år	0,52
Andra och påföljande barn 0-19 år	0,42

Ovanstående skala är baserad på ”Svensk konsumtionsenhetsskala” som föreslogs i utredningen om *Förbättrad statistik om hushållens inkomster, SOU 2002:73*. Efter diskussioner med olika användare har SCB beslutat att använda denna reviderade version av ”Svensk konsumtionsenhetsskala”.

Disponibel inkomst per konsumtionsenhet (ke) är huvudredovisningen för personer. Disponibel inkomst per ke kan ses som en indikator på den ekonomiska välfärd som ett hushåll förfogar över. Alla personer i ett hushåll har tillgång denna ekonomiska välfärd eller ekonomiska standard. Genom att redovisa hushållets disponibla inkomst per ke på alla individer i ett hushåll kan man också visa fördelningen av den ekonomiska standarden för hela befolkningen. Detta är en rekommenderad metod<sup>5</sup> som ofta används i forskarvärlden och också i den statistik som tas fram inom EU och andra nationella statistikbyråer.

I den officiella publiceringen av inkomstfördelningen görs ingen redovisning på regional nivå. I den registerbaserade inkomststatistiken (se ”Inkomststatistik – totalräknad”, Beskrivning av statistiken) redovisas däremot individstatistik efter kommuner. I den officiella redovisningen av befolkningens boende redovisas boendeutgifter för regionerna Stor-Stockholm, Stor-Göteborg, övriga kommuner med fler än 75 000 invånare samt övriga kommuner med mindre än 75 000 invånare.

Vid redovisning av **boendeutgifter** används boendeform för att skapa redovisningsgrupperna äganderätt, bostadsrätt och hyresrätt.

Undersökningen har två redovisningsområden; Inkomstfördelningen och Boende och boendeutgifter. Statistiken över inkomstfördelningen sprids i form av pressmeddelande, statistiska meddelanden (serie HE21) och tabeller på SCB:s webbplats. Statistiken över Boende och boendeutgifter sprids i form av pressmeddelande och tabeller på SCB:s webbplats. På webbplatsen finns också resultaten av statistiken inlagt i Sveriges statistiska databaser (SSD), där tabelluttag kan göras för ett antal variabler. Statistiken sprids även i form av tabeller i *Statistisk årsbok för Sverige* samt i *Bostads- och byggnadsstatistisk årsbok*.

HEK sprids även genom SCB:s mikrosimuleringsmodell FASIT. FASIT är ett verktyg för att simulera budget- och fördelningseffekter av tänkta regelförändringar i t.ex. skatte- och bidragssystemen. HEK utgör basmaterial i denna modell. Tillgång till FASIT får man genom ett abonnemang.

---

<sup>5</sup> Expert Group on Household Income Statistics - The Canberra group, *Final Report and Recommendations*, Ottawa 2001, ISBN 0-9688524-0-8, Sid 41



## 4 Slutliga observationsregister

### 4.1 Produktionsversioner

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

<b>Register</b>	Hushållens ekonomi
<b>Registervariant</b>	Kosthushåll
<b>Registerversion</b>	2007
<b>Register</b>	Hushållens ekonomi
<b>Registervariant</b>	Individer
<b>Registerversion</b>	2007

Fortsatt dokumentation, av registrens detaljerade innehåll, finns på SCB:s webbplats. Där beskrivs alla variabler och värдемängder 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 ännu.

### 4.3 Erfarenheter från senaste undersökningsomgången

Fel upptäcktes i variabeln KUTDSF (Skattefri aktieutdelning till näringsidkare) i Inkomst- och taxeringsregistren. KUTDSF ingår bland annat i disponibel inkomst (2004 års definition). Felen innebar att den disponibla inkomsten underskattades för inkomstår 2006 och överskattades för inkomstår 2007.

Skattefri aktieutdelning till näringsidkare (KUTDSF) ingår i följande variabler, som följaktligen också påverkades av felet:

CKAP	Räntor och utdelningar exkl. räntebidrag
CDISP04	Disponibel inkomst, 2004 års definition
CDISP04H	Disponibel inkomst, 2004 års definition
CFAKT04	Faktorinkomst, 2004 års definition
CDISPKEH04	Disponibel inkomst per konsumtionsenhet, 2004 års definition, familj

Felen bedömdes inte som stora på totalen men då det kunde bli större konsekvenser om materialet bröts ned eller användes på individnivå informerades kunder som köpt mikromaterial som innehöll ovanstående variabler.

Efter att felet hade rättats upp uppdaterades efterfrågade tabeller och diagram på hemsidan. Publiceringen av Inkomstfördelningen 2007 försenades några veckor till följd av felet.