

Vejledning til forskere

Opbygning af SILC

Eurostat guidelines	<p>SILC (Statistics on Income and Living Conditions) er en europæisk levevilkårsundersøgelse, der gennemføres i samtlige EU-lande samt nogle samarbejdslande (bl.a. Norge og Schweiz). Hvert år laver Eurostat en manual til SILC med metodologiske retningslinjer og beskrivelse af de enkelte variable. Hele statistikken er bygget op med udgangspunkt i denne manual. Manualerne findes under fanen Dokumentation her: https://www.dst.dk/da/Statistik/emner/levevilkaar/velfaerdsindikatorer/velfaerdsindikatorer.</p>
Oversigt over variable	<p>Under fanen Dokumentation på ovenstående link findes en oversigt over alle variable, der indgår i den danske SILC tilbage til 2011.</p>
Survey- og registerdata	<p>I Danmark er SILC sammensat dels af surveydata fra en landsdækkende spørgeskemaundersøgelse, som gennemføres hvert forår, og dels af registerdata, som hentes fra Danmarks Statistiks registre om indkomster, uddannelse, boligforhold (BBR) mv., og som bearbejdes til efterlevelse af Eurostats guidelines.</p>
Ad hoc moduler	<p>Udover de faste variable i SILC, indgår der hvert år også et eller flere spørgsmålsmoduler, der ændrer sig år for år. Fra 2021 vil modulerne blive gentaget i fastlagte intervaller med enten 3 eller 6 års mellemrum.</p>
Periode	<p>SILC er indsamlet i Danmark siden 2004, men vi har valgt kun at lægge mikrodata ud fra 2011 og frem, da der er lavet en revision, som kun er ført tilbage til 2011.</p>
Fire forskellige filer	<p>SILC-data er opdelt i fire forskellige datasæt. Afgrænset til ét interviewår er populationerne i filerne som følger:</p> <ul style="list-style-type: none">• D-filen: Alle husstande, der er forsøgt kontaktet det pågældende år.• H-filen: Alle husstande, der har deltaget i årets undersøgelse.• P-filen: Alle personer på 16 år og derover, som på interviewtidspunktet er medlemmer i en husstand, der indgår i H-filen.• R-filen: Alle personer uanset alder, som på interviewtidspunktet er medlemmer i en husstand, der indgår i H-filen. Derudover indgår tidligere husstandsmedlemmer, der siden foregående års undersøgelse er døde eller flyttede fra en husstand, der indgår i undersøgelsen både i år og sidste år. Nuværende husstandsmedlemmer har RB110=1-4. <p>Variable, der starter med D og H, omhandler således den samlede husstand, mens variable, der starter med P og R, er personrettede.</p>
Flag	<p>Til hver variabel hører et flag med navnet [variabelnavn]_f (fx HS031_f). Flagget angiver om variablen er udfyldt (= 1), missing (= -1) eller evt. en grund til, at variablen ikke er udfyldt. Se Eurostats guidelines for detaljer. Fra 2021</p>

kommer der ændringer i flagene til nogle variable, herunder indkomstvariable-
ne.

Nøgler og identer

Identer I SILC er der 4 pseudonymiserede identer.

HB030 og DB030 er husstandsidenter
PB030 og RB030 er personidenter

For at koble personer og husstandsfiler benyttes følgende sammenhænge:

$DB030 = \text{int}(PB030/100)$

$HB030 = DB030$

$PB030 = RB030$

SILC husstande Der er personidenter på data, som gør det muligt at koble med andre registre, men man kan ikke forvente nøjagtigt match med familiemønstre i Danmarks Statistiks øvrige registre. I SILC er det respondenterne selv, som definerer deres husstand. Det betyder også, at der ikke altid er fuld overensstemmelse med person-, familie- og adresseidenter i befolkningsregistret.

Vægte

Fire vægte at vælge imellem I SILC er der skævt bortfald, som følge af, at nogle samfundsgrupper deltager mindre end andre i interviewundersøgelsen. For at korrigere herfor er der beregnet nogle vægte, som sikrer, at svarpersonerne tilsammen summer op til antallet i den danske befolkning. I SILC-datasættene findes 4 primære vægte. Benytter man den forkerte vægt eller glemmer at anvende en vægt, så kan man godt risikere at få lidt skæve resultater. Dette er en hurtig vejledning i brugen af vægtene.

DB090 Household cross-sectional weight

RB050 Personal cross-sectional weight (alle personer).

PB040 Personal cross-sectional weight (personer på 16 år og derover)

PB060 Personal cross-sectional weight for Selected Respondent

DB090 DB090 benyttes til opgørelser, hvor enheden, der tælles over, er antallet af husstande. Det vil typisk kun gælde for opgørelser, som udelukkende er baseret på H-filen. Fx:

- *Antal husstande*, hvor det er svært eller meget svært at få pengene til at slå til (HS120)

RB050 RB050 benyttes til opgørelser, hvor husstandskarakteristika (oplysninger fra H-filen) er koblet på alle husstandsmedlemmer (R-filens population). Fx:

- *Antallet af personer, der indgår i en husstand*, hvor det er svært eller meget svært at få pengene til at slå til (HS120)

Derudover bruges RB050, når der laves opgørelser på variable, der omhandler personer under 16 år. Dvs. spørgsmål om børnepasning eller modulspørgsmål, der omhandler børn.

PB040 PBO40 benyttes til opgørelser, hvor husstandskarakteristika (oplysninger fra H-filen) er koblet på alle husstandsmedlemmer på 16 år og derover (P-filens population). Fx:

- *Antallet af personer på 16 år og derover, der indgår i en husstand, hvor det er svært eller meget svært at få pengene til at slå til (HS120)*

Derudover bruges PBO40, når der laves opgørelser på variable, der er udfyldt for alle personer i P-filen. Det gælder fx en række variable om personernes tilknytning til arbejdsmarkedet/socioøkonomiske status (bl.a. PLO31).

Vægtene PBO40 og RBO50 er ens for alle personer, der indgår i P-filen.

PB060 En række spørgsmål er kun stillet til den person i husstanden, der er udtrukket til undersøgelsen (selected respondent). PBO60 benyttes i opgørelser af disse variable, der bl.a. omfatter helbred (PH) og personlige afsavn (PD). Selected respondent-variablene kan identificeres på, at der kun er ét svar pr. husstand, og at der for de øvrige husstandsmedlemmer er benyttet flaget '-3'. I manualen kan de genkendes ved, at flaget '-3' er nævnt i variabelbeskrivelsen.

Usikkerhed

Det er vigtigt at være opmærksom på, at SILC er en stikprøveundersøgelse.

Det anbefales derfor,

- at man altid gør opmærksom på stikprøveusikkerhed i forbindelse med offentliggørelser af resultater på baggrund af SILC.
- at man aldrig offentliggør resultater på undergrupper med mindre end 100 respondenter. De bør præsenteres, som missing i tabeller og figurer.
- at man læser disse retningslinjer:
<https://www.dst.dk/da/Statistik/dokumentation/metode/stikkvalitet>

To former for usikkerhed: almindelig usikkerhed og skævt bortfald

Som bruger af SILC bør man være meget opmærksom på både den almindelige stikprøveusikkerhed samt risikoen for en skæv stikprøve som følge af skævt bortfald. For eksempel deltager husstande med høj indkomst oftere i undersøgelsen end personer med lav indkomst, midaldrende deltager oftere end de unge, og raske personer må antages at deltage mere end de syge.

Måling af almindelig stikprøveusikkerhed

Som indikation på omfanget af den almindelige stikprøveusikkerhed kan benyttes standardværktøjer og procedurer, som for eksempel Proc Surveysselect i SAS. Husk at basere beregningerne på de korrekte vægte.

Usikkerheden kan være overvurderet

Standardprocedurerne for beregning af stikprøveusikkerhed tager ikke hensyn til vægtningen af SILC-data. SILC-stikprøven kalibreres relativt grundigt på indkomster. Det sikres for eksempel, at antallet med under 60 procent af medianindkomsten og medianen holdes på plads i forhold til registrerdata. Det betyder, at disse standardberegninger måske overvurderer usikkerheden en smule på variable, der er stærkt korreleret med indkomsten.

Stor risiko for skævhed på fx helbred

Hvor SILC korrigeres grundigt på indkomst og alder, er der ingen korrektion for fx helbred. Da der ikke korrigeres for helbred i de almindelige vægte, så bør man være ekstra forsigtig med konklusioner på baggrund af helbredsoplysninger alene, da der her er stor risiko for, at stikprøven er systematisk skæv. Laves

fx analyser på variabelen PHO10 om generelt helbred, så er der en stor risiko for, at SILC systematisk undervurderer andelen af befolkningen, som er angiver at have et meget dårligt helbred.

Gør opmærksom på de to typer usikkerhed

Ved præsentation af resultater fra SILC anbefaler vi derfor, at man i tillæg til den almindelige stikprøveusikkerhed altid gør særligt opmærksom på risikoen for skævt bortfald, når man arbejder med variable, der ikke er meget stærkt korreleret med indkomster og demografi.

Yderligere dokumentation

Man kan læse mere om, hvordan SILC vægtes i [statistikdokumentationen](#).