



Metodrapport - reviderad hjälpvektor

Framtagande av data för utvärdering av de integrationspolitiska delmålen

2026-06-18

Thomas Önskog
Metodstatistiker
Avdelningen för social statistik och analys



Bakgrund

Uppföljningen av de integrationspolitiska delmålen om demokratisk, social och kulturell integration baseras delvis på uppgifter från Undersökningarna av levnadsförhållanden (ULF) 2023–2025. Uppgifterna kommer att gälla en annan population (18 och äldre) än den som normalt redovisas i ULF (16 år och äldre). Dessutom kommer andra redovisningsgrupper att användas.

Förändringar i hjälpvektorn

För att uppnå tillräcklig precision i skattningarna i tabellpaketet är det lämpligt att justera den hjälpvektor som normalt används i kalibreringen av skattningar i ULF. Ett antal olika justeringar behöver göras:

Den första justeringen består i att ändra hjälptotalerna så att de gäller den nya populationen 18 år och äldre. För hjälpvariabeln ålder * kön används normalt en åldersindelning i åldersgrupperna 16–19 år respektive 20–24 år. Eftersom personer i åldrarna 16–17 år här inte tillhör populationen och endast två årskullar därför kvarstår i gruppen 16–19 år är det rimligt att slå ihop de båda yngsta åldersgrupperna till en åldersgrupp, 18–24 år.

En andra justering är att använda den hjälpvariabel som togs fram för ULF 2025 även för ULF 2023 och ULF 2024. Vid produktionen av ULF 2025 justerades hjälpvariabeln för valdeltagande så att personer som endast varit röstberättigade vid kommunal- eller landstingsvalet 2022 och valt att inte rösta i något av de valen nu räknas som icke-röstande. Enligt den tidigare definitionen av hjälpvariabeln var det bara personer som varit röstberättigade vid riksdagsvalet 2022 och valt att inte rösta vid det valet som räknades som icke-röstande.

För det tredje ska här statistik framställas för ett antal redovisningsgrupper som inte ingår i ordinarie produktion av ULF eller i den ordinarie hjälpvektorn i ULF. De viktigaste av dessa gäller tid i Sverige, födelseregion/födelsevärldsdel och områdestyp samt utanförskapsområden. Dessa indelningar fångas inte av den ordinarie hjälpvektorn. Ett sätt att förbättra kvaliteten hos skattningarna för dessa grupper är att addera en eller flera av dessa variabler till hjälpvektorn. För de grupper som adderas till hjälpvektorn kommer summan av de kalibrerade vikterna att överensstämma exakt med gruppens andel av populationen.

Revision av den befintliga hjälpvektorn för ULF

De variabler som undersökts för hjälpvektorn är

1. Födelseregion (Norden utom Sverige, Europa utom Norden, världen utom Europa)
2. Födelsevärldsdel (Norden utom Sverige, EU utom Norden, Europa utom EU och Norden, Afrika, Asien, Övriga världen)
3. Tid i Sverige (0–9 år, 10+ år)
4. Tid i Sverige (0–4 år, 5–9 år, 10–19 år, 20–29 år, 30–39 år, 40+ år)
5. Områdestyp (Typ 1+2, Typ 3, Typ 4+5)
6. Utanförskapsområde (Ja, Nej)
7. Valdeltagande* tid i Sverige
8. Studiestöd * tid i Sverige
9. Utbildning * tid i Sverige

Nya variabler till hjälpvektorn

Tid i Sverige

Ur ämnessynpunkt är tid i Sverige den viktigaste redovisningsgruppen av de ovanstående och undersöktes därför först för hur väl fördelningen efter tid i Sverige i populationen (med indelningen i hjälpvariabel 4) återskapas med den ordinarie hjälpvektorn respektive med den ordinarie hjälpvektorn kompletterad med hjälpvariabel 3.

Tabell 1. Fördelning efter vistelsetid i Sverige för populationen 18+ för två olika val av hjälpvektorer i ULF 2023. Procent

Hjälpvektor	Inrikes född	0–4 år	5–9 år	10–19 år	20–29 år	30–39 år	≥40 år
Original	76,8	4,2	6,3	4,6	2,4	2,5	3,1
Original + variabel 3	76,8	3,4	5,0	5,5	2,9	2,8	3,7
Sann fördelning	76,8	3,5	4,8	5,6	3,4	2,5	3,4

Med originalvektorn avviker andelen personer i de olika grupperna relativt kraftigt från den sanna andelen; som mest 1,5 procentenheter för gruppen med vistelsetid i Sverige 5–9 år. När originalvektorn kompletteras med en hjälpvariabel med en grov indelningen efter vistelsetid i Sverige (hjälpvariabel 3), så minskar avvikelserna och uppgår till som mest 0,5 procentenhet för gruppen 20–29 år.

Födelsevärldsdel/födelseregion

Efter det undersöktes hur väl fördelningen efter födelsevärldsdel återskapas med originalvektorn samt densamma kompletterad med olika kombinationer av hjälpvariabler för tid i Sverige och/eller födelseregion.

Tabell 2. Fördelning efter födelsevärldsdel för populationen 18+ för sex olika val av hjälpvektorer i ULF 2023. Procent

Hjälpvektor	Inrikes född	Norden utom Sverige	EU utom Norden	Europa utom Norden	Afrika	Asien	Övriga världen
Original	76,8	2,2	3,7	3,2	2,8	10,0	1,4
Original + variabel 3	76,8	2,5	3,7	3,3	2,7	9,6	1,4
Original + variabel 4	76,8	2,4	3,7	3,4	2,7	9,6	1,4
Original + variabel 1	76,8	2,5	4,1	3,6	2,6	9,2	1,2
Original + variabel 1 + 3	76,8	2,5	4,1	3,6	2,6	9,1	1,3
Original + variabel 1 + 4	76,8	2,5	4,0	3,7	2,6	9,2	1,3
Sann fördelning	76,8	2,5	4,1	3,6	2,6	8,9	1,5

Originalvektorn och de båda vektorerna som är kompletterade med en hjälpvariabel för tid i Sverige (variabel 3 eller 4) uppvisar relativt stora avvikelser för några av grupperna, främst Asien där skillnaden är runt en procentenhet. När en hjälpvariabel med en geografisk indelning adderas till hjälpvektorn blir avvikelserna mycket mindre. För hjälpvektorn där en grov indelning efter tid i Sverige (variabel 3) och en grov indelning efter födelseregion (variabel 1) adderats till originalvektorn är avvikelserna som mest 0,2 procentenheter.

Områdestyp/Utanförskapsområde¹

Härnäst undersöker vi hur väl originalvektorn respektive en hjälpvektor där hjälpvariabel 1 och 3 adderats till originalvektorn återskapar fördelningarna efter områdestyper respektive utanförskapsområden.

Tabell 3. Fördelning efter områdestyp respektive utanförskapsområden för populationen 18+ för två olika val av hjälpvektorer i ULF 2023. Procent

Hjälpvektor	Områdes-typ 1+2	Områdes-typ 3	Områdes-typ 4+5	Utanförskap	Ej utanförskap
Original	13,0	22,8	64,3	6,5	93,5
Original + variabel 1 + 3	12,7	22,7	64,6	6,3	93,7
Sann fördelning	13,1	22,8	64,1	6,5	93,5

Originalvektorn återskapar fördelningarna efter områdestyp och utanförskap mycket väl och bättre än hjälpvektorn där vi lagt till variabler för tid i Sverige och födelseregion. För att avvikelserna för områdestyper inte ska bli alltför stora läggs även en hjälpvariabel för områdestyp till hjälpvektorn.

Fördelningen av vikter

Fördelningen av vikter undersöks för några av de hjälpvektorer som vi analyserat ovan. En sådan analys kan antingen baseras på de kalibrerade vikterna i sig eller

¹ I enlighet med [Fördjupad analys om utanförskap. Slutrapport regeringsuppdrag 2024:2](#)

på de så kallade g-vikterna, vilka definieras som kvoten mellan den kalibrerade vikten och den bortfallsjusterade designvikten. Medelvärdet av g-vikterna kommer att ligga nära ett. Vi använder g-vikter som ett mått på hur mycket vikterna justeras vid kalibreringen och använder 0,1 som minsta tillåtna värde på g-vikten.

Tabell 4. Nyckeltal för viktfordelningen för sex olika hjälpvektorer för ULF 2023.

Hjälpvektor	Minsta kalibrerad vikt	Största kalibrerad vikt	Minsta g-vikt	Antal g-vikter \leq 0.1	Största g-vikt	Standard avvikelse g-vikter	Skevhet g-vikter	Kurtosis g-vikter
Original	83	3615	0,24	0	3,10	0,38	1,52	2,84
Original + variabel 3	34	4109	0,10	5	3,37	0,40	1,51	3,13
Original + variabel 1	45	3686	0,13	0	3,28	0,38	1,50	2,96
Original + variabel 1 + 3	34	4005	0,10	5	3,49	0,40	1,50	3,16
Original + variabel 1 + 3 + 5	34	3964	0,10	5	3,52	0,40	1,50	3,16
Original + variabel 2 + 4 + 5	34	3940	0,10	7	3,60	0,40	1,51	3,19

Så fort som en hjälpvariabel för den viktigaste redovisningsgruppen tid i Sverige adderas till hjälpvektorn, så får minst fem individer i svarsmängden den lägsta tillåtna g-vikten 0,1. Även det största värdet på den kalibrerade vikten blir avsevärt större med en sådan variabel i hjälpvektorn. I övrigt påverkas inte fördelningen för g-vikten nämnvärt utan både standardavvikelse och skevhet är mer eller mindre desamma för alla hjälpvektorerna. För kurtosisen är förändringen större, men avvikelser från en normalfördelning (kurtosis lika med 3) är i stort sett oförändrat mellan originalvektorn och de vektorer som innehåller en hjälpvariabel för tid i Sverige.

Den justerade hjälpvektorn i ULF bör inte ge mer extrema kalibrerade vikter än originalvektorn men ingen av de hjälpvektorerna som vi hittills undersökt uppfyller detta.

Variabler i originalvektorn undersöks

För att få kunskap om variablerna i den befintliga hjälpvektorn, studeras vilka värden på hjälpvektorn som karakteriserar de individer som får de allra största respektive minsta g-vikterna när vi lägger till hjälpvariablerna 1, 3 och 5.

Vilka variabler som har störst påverkan

Analysen visar att variablerna födelseland, utbildning, valdeltagande och studiestöd har stora effekter. Därmed undersöks om viktfordelningen blir mindre skev med en hjälpvektor som består av originalvektorn, kompletterad med nya hjälpvariabler korsad med någon eller några av hjälpvariablerna för utbildning eller valdeltagande. För alla dessa varianter av hjälpvektorn finns mellan tre och sju personer med g-vikt lika med 0,1.

Ny reviderad hjälpvektor

Knappt tio procent av utrikes födda saknar information om utbildningsnivå i utbildningsregistret. En utredning som gjordes på SCB 2022 visade att för personer med kort tid i Sverige är lång eftergymnasial utbildning vanligast bland de som saknas i utbildningsregistret. För personer med lång tid i Sverige samt för inrikes födda är gymnasial utbildning vanligast bland de som saknas i utbildningsregistret. Vi utnyttjar detta faktum i konstruktionen av den reviderade hjälpvariabeln:

Utbildning * Vistelsetid i Sverige (Inrikes född med förgymnasial utbildning, inrikes född med gymnasial eller övrig utbildning, inrikes född med kort eftergymnasial utbildning, inrikes född med lång eftergymnasial utbildning, utrikes född med vistelsetid i Sverige 0–9 år och förgymnasial utbildning, utrikes född med vistelsetid i Sverige 0–9 år och gymnasial eller övrig utbildning, utrikes född med vistelsetid i Sverige 0–9 år och kort eftergymnasial utbildning, utrikes född med vistelsetid i Sverige 0–9 år och lång eftergymnasial utbildning, utrikes född med vistelsetid i Sverige minst 10 år och förgymnasial utbildning, utrikes född med vistelsetid i Sverige minst 10 år och gymnasial utbildning, utrikes född med vistelsetid i Sverige minst 10 år och kort eftergymnasial utbildning, utrikes född med vistelsetid i Sverige minst 10 år och lång eftergymnasial eller övrig utbildning)

I tabell 5 nedan (nedersta raden) redovisas att med denna hjälpvektor fås inte g-vikter som är lika med 0,1, samtidigt som maxvärdena på de kalibrerade vikterna och g-vikterna är i samma storleksordning som för originalvektorn. Standardavvikelsen och skevheten för g-viktfördelningen är något större än för originalvektorn men skillnaden är marginell.

Hjälpvektorn består nu av originalvektorn, utan variablerna för utbildning och valdeltagande, utökat med hjälpvariablerna

- Födelseregion
- Områdestyp
- Valdeltagande * tid i Sverige
- Utbildning * tid i Sverige.

Den justerade hjälpvektorn återspeglar viktiga redovisningsgrupper i tabellpaketet bättre än originalvektorn utan att viktfördelningen förändras nämnvärt. Hjälpvektorn har använts vid framtagandet av underlag från såväl ULF 2023–2025 som den nya undersökningen om social sammanhållning (USS).

Tabell 5. Nyckeltal för viktfordelningen för fem olika hjälpvektorer för ULF 2023.

Hjälpvektor	Minsta kalibrerad vikt	Största kalibrerad vikt	Minsta g-vikt	Antal g-vikter ≤ 0.1	Största g-vikt	Standard avvikelse g-vikter	Skevhet g-vikter	Kurtosis g-vikter
Original	83	3615	0,24	0	3,10	0,38	1,52	2,84
Original + variabel 1 + 3 + 5	34	3964	0,10	5	3,52	0,40	1,50	3,16
Original utan val + variabel 1 + 5 + 7	34	3860	0,10	1	3,47	0,41	1,70	3,70
Original utan studiestöd och val + variabel 1 + 5 + 7 + 8	34	3827	0,10	3	3,45	0,41	1,69	3,67
Original utan utbildning och val + variabel 1 + 5 + 7 + 9	88	3592	0,21	0	2,98	0,40	1,56	3,05