



Statistics Sweden

Statistiska centralbyrån

Kalibreringsrapport

Föräldraundersökningen 2012, 1 – 5 år

SCB, Stockholm
08-506 940 00

SCB, Örebro
019-17 60 00

www.scb.se



Inledning

I en urvalsundersökning är alltid skattningarna behäftade med *urvalsfel* beroende på att endast en delmängd (urval) av populationen studeras. Ett annat fel uppkommer om vi inte lyckas få svar från alla personer (bortfall) och om de avviker från de svarande med avseende på undersökningsvariablerna. Detta fel kallas för *bortfallsfel*.

För att underlätta användningen av statistiken är det värdefullt om storleken på felet kan uppskattas. Av nämnda feltyper är det endast storleken på urvalsfelet som kan skattas med hjälp av urvalsinformation. Kunskap om bortfallsfelet kan i regel bara fås på ett indirekt och approximativt sätt genom att utnyttja registervariabler.

Både urvalsfel och bortfallsfel kan reduceras genom att använda ett effektivt uppräkningsförfarande. I följande avsnitt redovisas hur det görs i denna undersökning.

Parametrar

I de flesta fallen önskas information om parametrar av typen procentuella andelen barn med viss egenskap, t.ex. procentuella andelen vars vårdnadshavare är nöjda med omsorgsformen. Men även genomsnittsvärden är intressanta, t.ex. genomsnittligt antal timmar barnen vistas i olika omsorgsformer. Motsvarande skattningar önskas också för redovisningsgrupper.

Hjälpinformation

Viss hjälpinformation utnyttjas vanligtvis även före estimationen, t.ex. för bildande av stratifierade urvalsdesigner. I studerade undersökning används, som tidigare påpekats, stratifieringsvariablerna åldersgrupp, utlänsk eller svensk bakgrund och kommun eller kommungrupp. Det kan dock finnas ytterligare hjälpinformation som är effektiv i estimationen.

Det centrala arbetet för att få god kvalitet på skattningarna, då kalibreringsestimatoern används, är att använda "stark" hjälpinformation. I följande avsnitt beskrivs detta arbete för denna undersökning.

Tänkbara hjälpvariabler

Vid val av hjälpvariabler är det tre kriterier som ska beaktas (se Lundström och Särndal 2001):

- Det första kriteriet är att variabeln samvarierar väl med svarsbenägenheten (-sannolikheten). Det är det viktigaste kriteriet eftersom det leder till en minskning av bortfallskevheter för alla skattningar.
- Det andra kriteriet är att variabeln samvarierar väl med (viktiga) målvariabler. Om så är fallet minskar bortfallsbiasen för de skattningar som

byggs upp av dessa målvariabler. Även variansen minskar för dessa skattningar.

- Det tredje kriteriet är att variabeln avgränsar (viktiga) redovisningsgrupper. Det leder framförallt till minskad varians i skattningar för dessa redovisningsgrupper.

Tänkbara hjälpvariabler, det vill säga variabler som tros uppfylla de ovan uppsatta kriterierna, hämtades ifrån olika register. Urvalet har dragits från Registret över totalbefolkningen (RTB), varifrån vi också hämtar hjälpinformation. För att ytterligare reducera felen hämtas också hjälpinformation från Utbildningsregistret (UTB) och Inkomst och taxeringsregistret (IoT). Eftersom någon förälder eller vårdnadshavare antas besvarar enkäten använder vi deras registerinformation. Stratumindelningen bestäms dels av åldersklasser och beroende på åldersklass dessutom av kommun eller utländsk/svensk bakgrund *kommungrupp. Eftersom vi inte vet vilken förälder som har svarat eller som skulle kunnat svara har vi använt hjälpvariabler som är oberoende av vilken förälder som svarat. Det vill säga att vi har använt hushålls gemensamma variabler. En genomgång av variablerna i dessa tre register resulterade i att **5** variabler valdes ut.

Exempelvis vet vi från tidigare studier att svarsfrekvensen brukar skilja sig åt mellan individer med lång eller kort utbildning, samt beroende på civilstånd, ålder och kön. Inkomst är en annan variabel som brukar samvariera med svarsfrekvensen. Viktiga redovisningsgrupper är förutom åldersklass och kommun är även föräldrarnas högsta utbildning och utländsk bakgrund därför är även de med i analysen.

De sammanslagningar av kategorier som gjorts baseras på kunskaper från tidigare kalibreringar. Dessutom har hänsyn tagits till hur de redovisningsgrupper som sedan ska användas är avgränsade. Andra hänsyn som tagits är att uppräkningsfaktor inte får bli för stor då det kan ge en för stor vikt till vissa svarande. Hjälpvariablerna är definierade enligt tabell 1.

Tabell 1 Tänkbara hjälpvariabler

Variabel (benämning)	Kategorier (koder)
CIVILSTÅND (Vårdnadshavare)	1 = Gift+registrerat partnerskap 2 = Övriga
HÖGSTA UTBILDNING (Vårdnadshavare)	1 = Förgymnasial och gymnasial utbildning 2 = Eftergymnasial utbildning
FÖDELSELAND	1 = Båda föräldrarna födda i Sverige 2 = Övriga
REGION	Varje kommun har fått en egen kod.
INKOMST (Hushåll)	(Inkomst: tkr) 1 = 0 – 400 2 = 401 – 645 3 = 646 –

I följande avsnitt analyserar vi variablerna i tabell 1 för att slutligen bestämma en hjälpvektor.

Analys av hjälpinformation

Kriterium 1: Variabeln samvarierar med svarsbenägenheten

För att se huruvida hjälpvariablerna uppfyller det första kriteriet, studeras sambandet mellan den dikotoma variabeln svarande/bortfall och hjälpvariablerna. Det görs genom att beräkna skattad andel svarande i olika grupper, bestämda av respektive hjälpvariabel. Vid skattningen används designvikten (vid stratifierat OSU: N_h/n_h).

Vid stora skillnader mellan svarsandelarna utgör variabeln en stark kandidat till hjälpvariabel.

Tabell 2 Skattad procentuell andel svarande fördelat på civilstånd

Civilstånd	Gift, registrerat partnerskap	Övriga
Svarsandel (%)	60,0	53,8

Tabell 3 Skattad procentuell andel svarande fördelat på högsta utbildning

Högsta utbildning	Förgymnasial och gymnasial	Eftergymnasial
Svarsandel (%)	41,7	68,4

Tabell 4 Skattad procentuell andel svarande fördelat på födelseland

Födelseland	Båda föräldrarna födda i Sverige	Övriga
Svarsandel (%)	63,5	44,3

Tabell 5 Skattad procentuell andel svarande fördelat på kommun

Kommun	Minsta svarsandel	Största svarsandel
Svarsandel (%)	35,5	72,0

I tabellen ovan redovisas bara den minsta och största svarsandel någon kommun hade.

Tabell 6 Skattad procentuell andel svarande fördelat på hushålls-inkomst

Inkomst	Mindre än 400 tkr	Mellan 400 och 645 tkr	Mer än 645 tkr
Svarsandel (%)	40,4	61,1	72,3

Tabellerna 2, 3, 4 och 6 visar att de starka hjälpvariablerna (beträffande kriterium 1) framförallt är det högsta utbildning, födelseland och inkomst. Vidare kan det noteras att det är stor skillnad mellan den kommun med minst andel svarande mot den kommun med störst andel svarande. Det ger en stark indikation att även kommun skulle kunna vara en stark hjälpvariabel. Variabeln civilstånd är inte lika stark som de övriga hjälpvariablerna men är i sin karaktär skilld från övriga hjälpvariabler och kan därför tillföra ytterligare information.

Innan någon variabel utesluts undersöks i vilken mån det andra kriteriet uppfylls för de tänkbara hjälpvariablerna.

Kriterium 3: Variabeln avgränsar (viktiga) redovisningsgrupper

Om hjälpvariabeln avgränsar viktiga redovisningsgrupper kan kvaliteten bli bättre i dessa grupper. Framförallt blir skattningarna säkrare om hjälpvariabeln väl avgränsar redovisningsgruppen.

Kalibreringsestimatorn ger konsistenta skattningar i den meningen att estimatorn ger exakta skattningar för utnyttjade registertotaler. Det är därför viktigt att vid val av hjälpvariabler fundera på vilka skattningar man vill få konsistenta.

Variablerna kommun och högsta utbildning avgränsar redovisningsgrupper och bör därför vara med i hjälpvektorn.

Slutligt val av hjälpvektor

Civilstånd, föräldrarnas högsta utbildning och födelseland samt hushålets inkomst är alla starka hjälpvariabler. För att inte få för stora vikter har föräldrarnas högsta utbildning endast delats upp i två kategorier, ur ett redovisnings perspektiv hade det varit bra om samma kategorier kunnat användas både för redovisning och hjälpvariabler. Det hade givet konsistenta skattningar för alla redovisningskategorier. Även kommuner som är en strak hjälpvariabel har fått anpassas på så sätt att de 200 största kommunerna har fått en egen kategori med de 90 minsta har fått en egen kategori på länsnivå. Det innebär att de största kommunerna kommer att ge konsistenta skattningar men inte de minsta kommunerna.

Efter en sammanvägning av analysen kring de tre kriterierna samt efter kontroll av vikternas fördelning används följande hjälpvektor:

Kommun+inkomst+födelseland+högsta utbildning+ civilstånd

Teknisk beskrivning av urval och estimation

Vi har en population U bestående av N personer. De parametrar vi är intresserade av är vanligtvis funktioner av två totaler $Y = \sum_U y_k$ och

$Z = \sum_U z_k$, där y_k är värdet på variabel y för person k och z_k värdet på en annan variabel för samma person. Vanligtvis är y (och även z) en dikotom variabel, d.v.s.

$$y_k = \begin{cases} 1 & \text{om person } k \text{ har studerade egenskap} \\ 0 & \text{för övrigt} \end{cases} \quad (4.1)$$

Vanligtvis är vi också intresserade av parametrar för redovisningsgrupper. Låt oss benämna dessa $U_1, \dots, U_d, \dots, U_D$, där $U = \bigcup_{d=1}^D U_d$. Totalen för redovisningsgrupp d kan skrivas

$$Y_d = \sum_U y_{dk} \quad (4.2)$$

$$\text{där } y_{dk} = \begin{cases} y_k & \text{för } k \in U_d \\ 0 & \text{för övrigt.} \end{cases}.$$

Z_d bildas på likartat sätt.

En generell parameter för redovisningsgrupp d (d kan också avse hela populationen) kan skrivas $\theta_d = C \frac{Y_d}{Z_d}$, där C är en konstant.

Den vanligaste parametern är en procentuell andel, som erhålles när $C = 100$ och $z_k = 1$ för alla k , och y är definierad enligt (4.1). Om vi låter N_d vara antalet personer i redovisningsgrupp d , då kan parametern skrivas

$$P_d = 100 \frac{\sum_U y_{dk}}{N_d} \quad (4.3)$$

Vi drar ett obundet slumpmässigt urval s_h av storleken n_h från stratum h ($h = 1, \dots, H$), men p.g.a. övertäckning och bortfall har vi endast svarsmängden r_h av storleken m_h att utföra beräkningarna på. Storleken på stratum h ger vi beteckningen N_h .

Den "konventionella" estimatorm (för Y_d), har följande form:

$$\hat{Y}_d = \sum_{h=1}^H \frac{N_h}{m_h} \sum_{r_h} y_{dk} \quad (4.4)$$

I estimator (4.4) används ingen ytterligare hjälpinformation än stratifieringsinformationen.

I syfte att erhålla en estimator med mindre urvalsfel och bortfallsskevhet än estimator (4.4) utnyttjar vi hjälpinformation också i estimationen. Vi bildar en hjälpvektor \mathbf{x}_k , som anger till vilka kategorier av

Kommun+inkomst+födelseland+högsta utbildning+ civilstånd

som person k hör. Från Inkomst och taxeringsregistret, RTB och Utbildningsregistret framställer vi hjälptotalerna $\sum_{U_d} \mathbf{x}_k$. Vi utnyttjar denna hjälpinformation i en kalibreringsestimator.

Kalibreringsestimatoren för totalen Y_d har följande utseende:

$$\hat{Y}_{wd} = \sum_r d_k^* g_k y_{dk} \quad (4.5)$$

där

$$d_k^* = d_k \cdot f_k = 1 / \left(\frac{\pi_k}{\hat{\theta}_k} \right) \text{ för } k \in r,$$

så att

w_k = den totala vikten för objekt k

π_k = inklusionssannolikhet för objekt k

$\hat{\theta}_k$ = den skattade svarssannolikheten där det antas att personer svarar med samma sannolikhet och oberoende av varandra

$d_k = \left(\frac{\pi_k}{w_k} \right)$ designvikt

$f_k = \left(\frac{\hat{\theta}_k}{\pi_k} \right)$ bortfallsvikten

g_k = justeringsfaktor som baseras på hjälpinformationen

och

$$g_k = 1 + \left(\frac{\sum_U \mathbf{x}_k - \sum_r d_k^* \mathbf{x}_k}{\sum_r d_k^* \mathbf{x}_k \mathbf{x}_k'} \right) \mathbf{x}_k \quad (4.6)$$

Vid skattning av en parameter av typen $\theta_d = C \frac{Y_d}{Z_d}$ skattas respektive total med hjälp av kalibreringsvikterna $d_k^* g_k$.

Referenser:

Lundström S. och Särndal C.-E. (2001). *Estimation in the Presence of Nonresponse and Frame Imperfection*. Stockholm: Statistics Sweden



Statistics Sweden

Statistiska centralbyrån

Kalibreringsrapport

Föräldraundersökningen 2012, 6 – 12 år

SCB, Stockholm
08-506 940 00

SCB, Örebro
019-17 60 00

www.scb.se



Inledning

I en urvalsundersökning är alltid skattningarna behäftade med *urvalsfel* beroende på att endast en delmängd (urval) av populationen studeras. Ett annat fel uppkommer om vi inte lyckas få svar från alla personer (bortfall) och om de avviker från de svarande med avseende på undersökningsvariablerna. Detta fel kallas för *bortfallsfel*.

För att underlätta användningen av statistiken är det värdefullt om storleken på felen kan uppskattas. Av nämnda feltyper är det endast storleken på urvalsfelet som kan skattas med hjälp av urvalsinformation. Kunskap om bortfallsfelet kan i regel bara fås på ett indirekt och approximativt sätt genom att utnyttja registervariabler.

Både urvalsfel och bortfallsfel kan reduceras genom att använda ett effektivt uppräkningsförfarande. I följande avsnitt redovisas hur det görs i denna undersökning.

Parametrar

I de flesta fallen önskas information om parametrar av typen procentuella andelen barn med viss egenskap, t.ex. procentuella andelen vars vårdnadshavare är nöjda med omsorgsformen. Men även genomsnittsvärden är intressanta, t.ex. genomsnittligt antal timmar barnen vistas i olika omsorgsformer. Motsvarande skattningar önskas också för redovisningsgrupper.

Hjälpinformation

Viss hjälpinformation utnyttjas vanligtvis även före estimationen, t.ex. för bildande av stratifierade urvalsdesigner. I studerade undersökning används, som tidigare påpekats, stratifieringsvariablerna åldersgrupp, utlänsk eller svensk bakgrund och kommun eller kommungrupp. Det kan dock finnas ytterligare hjälpinformation som är effektiv i estimationen.

Det centrala arbetet för att få god kvalitet på skattningarna, då kalibreringsestimatorn används, är att använda "stark" hjälpinformation. I följande avsnitt beskrivs detta arbete för denna undersökning.

Tänkbara hjälpvariabler

Vid val av hjälpvariabler är det tre kriterier som ska beaktas (se Lundström och Särndal 2001):

- Det första kriteriet är att variabeln samvarierar väl med svarsbenägenheten (-sannolikheten). Det är det viktigaste kriteriet eftersom det leder till en minskning av bortfallskevheter för alla skattningar.
- Det andra kriteriet är att variabeln samvarierar väl med (viktiga) målvariabler. Om så är fallet minskar bortfallsbiasen för de skattningar som

byggs upp av dessa målvariabler. Även variansen minskar för dessa skattningar.

- Det tredje kriteriet är att variabeln avgränsar (viktiga) redovisningsgrupper. Det leder framförallt till minskad varians i skattningar för dessa redovisningsgrupper.

Tänkbara hjälpvariabler, det vill säga variabler som tros uppfylla de ovan uppsatta kriterierna, hämtades ifrån olika register. Urvalet har dragits från Registret över totalbefolkningen (RTB), varifrån vi också hämtar hjälpinformation. För att ytterligare reducera felen hämtas också hjälpinformation från Utbildningsregistret (UTB) och Inkomst och taxeringsregistret (IoT). Eftersom någon förälder eller vårdnadshavare antas besvarar enkäten använder vi deras registerinformation. Stratumindelningen bestäms dels av åldersklasser och beroende på åldersklass dessutom av kommun eller utländsk/svensk bakgrund *kommungrupp. Eftersom vi inte vet vilken förälder som har svarat eller som skulle kunnat svara har vi använt hjälpvariabler som är oberoende av vilken förälder som svarat. Det vill säga att vi har använt hushålls gemensamma variabler. En genomgång av variablerna i dessa tre register resulterade i att 5 variabler valdes ut.

Exempelvis vet vi från tidigare studier att svarsfrekvensen brukar skilja sig åt mellan individer med lång eller kort utbildning, samt beroende på civilstånd, ålder och kön. Inkomst är en annan variabel som brukar samvariera med svarsfrekvensen. Viktiga redovisningsgrupper är förutom åldersklass och kommun är även föräldrarnas högsta utbildning och utländsk bakgrund därför är även de med i analysen.

De sammanslagningar av kategorier som gjorts baseras på kunskaper från tidigare kalibreringar. Dessutom har hänsyn tagits till hur de redovisningsgrupper som sedan ska användas är avgränsade. Andra hänsyn som tagits är att uppräkningsfaktor inte får bli för stor då det kan ge en för stor vikt till vissa svarande. Hjälpvariablerna är definierade enligt tabell 1.

Tabell 1 Tänkbara hjälpvariabler

Variabel (benämning)	Kategorier (koder)
CIVILSTÅND (Vårdnadshavare)	1 = Gift+registrerat partnerskap 2 = Övriga
HÖGSTA UTBILDNING (Vårdnadshavare)	1 = Förgymnasial och gymnasial utbildning 2 = Eftergymnasial utbildning
FÖDELSELAND	1 = Båda föräldrarna födda i Sverige 2 = Övriga
REGION	Varje kommun har fått en egen kod.
INKOMST (Hushåll)	(Inkomst: tkr) 1 = 0 – 400 2 = 401 – 645 3 = 646 –

I följande avsnitt analyserar vi variablerna i tabell 1 för att slutligen bestämma en hjälpvektor.

Analys av hjälpinformation

Kriterium 1: Variabeln samvarierar med svarsbenägenheten

För att se huruvida hjälpvariablerna uppfyller det första kriteriet, studeras sambandet mellan den dikotoma variabeln svarande/bortfall och hjälpvariablerna. Det görs genom att beräkna skattad andel svarande i olika grupper, bestämda av respektive hjälpvariabel. Vid skattningen används designvikten (vid stratifierat OSU: N_h/n_h).

Vid stora skillnader mellan svarsandelarna utgör variabeln en stark kandidat till hjälpvariabel.

Tabell 2 Skattad procentuell andel svarande fördelat på civilstånd

Civilstånd	Gift, registrerat partnerskap	Övriga
Svarsandel (%)	58,7	48,5

Tabell 3 Skattad procentuell andel svarande fördelat på högsta utbildning

Högsta utbildning	Förgymnasial och gymnasial	Eftergymnasial
Svarsandel (%)	42,4	65,4

Tabell 4 Skattad procentuell andel svarande fördelat på födelseland

Födelseland	Båda föräldrarna födda i Sverige	Övriga
Svarsandel (%)	60,4	41,6

Tabell 5 Skattad procentuell andel svarande fördelat på kommun

Kommun	Minsta svarsandel	Största svarsandel
Svarsandel (%)	32,7	71,6

I tabellen ovan redovisas bara den minsta och största svarsandel någon kommun hade.

Tabell 6 Skattad procentuell andel svarande fördelat på hushålls-inkomst

Inkomst	Mindre än 400 tkr	Mellan 400 och 645 tkr	Mer än 645 tkr
Svarsandel (%)	39,0	55,5	67,4

Tabellerna 2, 3, 4 och 6 visar att de starka hjälpvariablerna (beträffande kriterium 1) framförallt är det högsta utbildning, födelseland och inkomst. Vidare kan det noteras att det är stor skillnad mellan den kommun med minst andel svarande mot den kommun med störst andel svarande. Det ger en stark indikation att även kommun skulle kunna vara en stark hjälpvariabel. Variabeln civilstånd är inte lika stark som de övriga hjälpvariablerna men är i sin karaktär skilld från övriga hjälpvariabler och kan därför tillföra ytterligare information.

Innan någon variabel utesluts undersöks i vilken mån det andra kriteriet uppfylls för de tänkbara hjälpvariablerna.

Kriterium 3: Variabeln avgränsar (viktiga) redovisningsgrupper

Om hjälpvariabeln avgränsar viktiga redovisningsgrupper kan kvaliteten bli bättre i dessa grupper. Framförallt blir skattningarna säkrare om hjälpvariabeln väl avgränsar redovisningsgruppen.

Kalibreringsestimaton ger konsistenta skattningar i den meningen att estimaton ger exakta skattningar för utnyttjade registertotaler. Det är därför viktigt att vid val av hjälpvariabler fundera på vilka skattningar man vill få konsistenta.

Variablerna kommun och högsta utbildning avgränsar redovisningsgrupper och bör därför vara med i hjälpvektorn. Barn mellan 10 och 12 år redovisas för kommungrupp och utländskbakgrund i stället för kommuner.

Slutligt val av hjälpvektor

Civilstånd, föräldrarnas högsta utbildning och födelse-land samt hushålets inkomst är alla starka hjälpvariabler. För att inte få för stora vikter har föräldrarnas högsta utbildning endast delats upp i två kategorier, ur ett redovisnings perspektiv hade det varit bra om samma kategorier kunnat användas både för redovisning och hjälpvariabler. Det hade givet konsistenta skattningar för alla redovisningskategorier. Istället för kommuner som är en strak hjälpvariabel och kommungrupper med och utan utländskbakgrund har stratum används. Stratum har sedan fått anpassas på så sätt att de 200 största kommunerna har behållt sin stratum kod med de 90 minsta har fått en ny stratumkod på länsnivå. Det innebär att de största kommunerna kommer att ge konsistenta skattningar men inte de minsta kommunerna.

Efter en sammanvägning av analysen kring de tre kriterierna samt efter kontroll av vikternas fördelning används följande hjälpvektor:

Stratum+inkomst+födelse-land+högsta utbildning+ civilstånd

Teknisk beskrivning av urval och estimation

Vi har en population U bestående av N personer. De parametrar vi är intresserade av är vanligtvis funktioner av två totaler $Y = \sum_U y_k$ och

$Z = \sum_U z_k$, där y_k är värdet på variabel y för person k och z_k värdet på en annan variabel för samma person. Vanligtvis är y (och även z) en dikotom variabel, d.v.s.

$$y_k = \begin{cases} 1 & \text{om person } k \text{ har studerade egenskap} \\ 0 & \text{för övrigt} \end{cases} \quad (4.1)$$

Vanligtvis är vi också intresserade av parametrar för redovisningsgrupper. Låt oss benämna dessa $U_1, \dots, U_d, \dots, U_D$, där $U = \bigcup_{d=1}^D U_d$. Totalen för redovisningsgrupp d kan skrivas

$$Y_d = \sum_U y_{dk} \quad (4.2)$$

$$\text{där } y_{dk} = \begin{cases} y_k & \text{för } k \in U_d \\ 0 & \text{för övrigt.} \end{cases}.$$

Z_d bildas på likartat sätt.

En generell parameter för redovisningsgrupp d (d kan också avse hela populationen) kan skrivas $\theta_d = C \frac{Y_d}{Z_d}$, där C är en konstant.

Den vanligaste parametern är en procentuell andel, som erhålles när $C = 100$ och $z_k = 1$ för alla k , och y är definierad enligt (4.1). Om vi låter N_d vara antalet personer i redovisningsgrupp d , då kan parametern skrivas

$$P_d = 100 \frac{\sum_U y_{dk}}{N_d} \quad (4.3)$$

Vi drar ett obundet slumpmässigt urval s_h av storleken n_h från stratum h ($h = 1, \dots, H$), men p.g.a. övertäckning och bortfall har vi endast svarsmängden r_h av storleken m_h att utföra beräkningarna på. Storleken på stratum h ger vi beteckningen N_h .

Den "konventionella" estimatorn (för Y_d), har följande form:

$$\hat{Y}_d = \sum_{h=1}^H \frac{N_h}{m_h} \sum_{r_h} y_{dk} \quad (4.4)$$

I estimator (4.4) används ingen ytterligare hjälpinformation än stratifieringsinformationen.

I syfte att erhålla en estimator med mindre urvalsfel och bortfallsskevhet än estimator (4.4) utnyttjar vi hjälpinformation också i estimationen. Vi bildar en hjälpvektor \mathbf{x}_k , som anger till vilka kategorier av

Kommun+inkomst+födelseland+högsta utbildning+ civilstånd

som person k hör. Från Inkomst och taxeringsregistret, RTB och Utbildningsregistret framställer vi hjälptotalerna $\sum_{U_d} \mathbf{x}_k$. Vi utnyttjar denna hjälpinformation i en kalibreringsestimator.

Kalibreringsestimatoren för totalen Y_d har följande utseende:

$$\hat{Y}_{wd} = \sum_r d_k^* g_k y_{dk} \quad (4.5)$$

där

$$d_k^* = d_k \cdot f_k = 1 / \left(\frac{\pi_k \hat{\theta}_k}{w_k} \right) \text{ för } k \in r,$$

så att

w_k = den totala vikten för objekt k

π_k = inklusionssannolikhet för objekt k

$\hat{\theta}_k$ = den skattade svarssannolikheten där det antas att personer svarar med samma sannolikhet och oberoende av varandra

$d_k = \left(\frac{\pi_k}{w_k} \right)$ designvikt

$f_k = \left(\frac{\hat{\theta}_k}{w_k} \right)$ bortfallsvikten

g_k = justeringsfaktor som baseras på hjälpinformationen

och

$$g_k = 1 + \left(\sum_U \mathbf{x}_k - \sum_r d_k^* \mathbf{x}_k \right) \left(\sum_r d_k^* \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k \quad (4.6)$$

Vid skattning av en parameter av typen $\theta_d = C \frac{Y_d}{Z_d}$ skattas respektive total med hjälp av kalibreringsvikterna $d_k^* g_k$.

Referenser:

Lundström S. och Särndal C.-E. (2001). *Estimation in the Presence of Nonresponse and Frame Imperfection*. Stockholm: Statistics Sweden