

Kalibreringsrapport studiecirkeldeltagare 65+

1 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.

2 Hjälpinformation

Viss hjälpinformation utnyttjas vanligtvis även före estimationen, t.ex. för bildande av stratifierade urvalsdesigner. 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.

2.1 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 bortfallsskevheten 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.

I en undersökning med ett stort antal frågor av skiftande karaktär är det främst kriterierna (i) och (iii) som kan beaktas. Tänkbara hjälpvariabler, det vill säga variabler som tros uppfylla dessa kriterier, hämtades ifrån Registret över totalbefolkningen (RTB), Utbildningsregistret (UREG), Yrkesregistret (YREG) samt Studieförbundens verksamhetsregister (STUV). Hjälpvariablerna är definierade enligt tabell 1.

Tabell 1. Tänkbara hjälpvariabler

Variabel (benämning)	Kategorier (koder)
ÄMNE	1 = Konst och media 2 = Humaniora 3 = Övriga
FÖRBUND	1 = ABF 2 = Bilda 3 = FU 4 = Sfr 5 = SV 6 = NBV 7 = Mbsk 8 = Sensus 9 = Ibn Rushd 10 = KBV
KÖN	1 = Kvinnor, 2 = Män
ÅLDER*	1 = 65 – 79, 2 = 80 –
FÖDELSELAND	1 = Sverige, 2 = Övriga Länder
CIVILSTÅND	1 = Gift eller registrerat partnerskap 2 = Övriga
UTBILDNINGSNIVÅ**	1 = Förgymnasial (inkluderar uppgift saknas) 2 = Gymnasial 3 = Eftergymnasial
KOMMUNGRUPP	1 = Storstäder, förortskommuner till storstäder 2 = Större städer, förortskom. till större städer 3 = Övriga kommuner
SAMMANKOMSTER	1 = 0-5 sammankomster 2 = 6-12 sammankomster 3 = 13-19 sammankomster 4 = 20- sammankomster
TIMMAR	1 = 9-14 timmar 2 = 15-38 timmar 3 = 39-59 timmar 4 = 60- timmar
DELTAGARE	1 = 3-6 deltagare 2 = 7-12 deltagare 3 = 13-18 deltagare 4 = 19- deltagare
YRKE	1 = Arbetaryrken 2 = Tjänstemannayrken 3 = Okänt

Variablerna kopplade till person (kön, ålder, födelse-land, civilstånd, utbildningsnivå, boenderegion) avser den utvalda deltagaren.

* Ålder avser den 31 december 2012.

** Utbildningsnivå avser deltagarens högsta utbildning fram till och med vårterminen 2012.

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

3 Analys av hjälpinformation

3.1.1 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 svarsandel i olika grupper, bestämda av respektive hjälpvariabel. Vid beräkningen används designvikten. Se teknisk rapport för mer detaljerad förklaring av hur svarsandelen beräknas.

Den vägda svarsandelen bland studiecirkel där deltagaren var 65 år och äldre skattades till 49,6 procent.

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

Tabell 2 Skattad svarsandel fördelad på ämne, procent.

	Konst och media	Humaniora	Övriga
Vägd	53,7	52,2	43,3
Ovägd	59,9	50,2	42,1

Tabell 3a Skattad svarsandel fördelad på förbund, procent.

	ABF	Bilda	FU	Sfr	SV
Vägd	47,4	42,9	66,1	55,1	47,3
Ovägd	47,7	31,8	57,8	50,0	42,9

Tabell 3b Skattad svarsandel fördelad på förbund, procent.

	NBV	Mbsk	Sensus	lbn Rushd	KBV
Vägd	46,5	59,2	56,5	19,4	64,3
Ovägd	42,5	60,0	61,1	19,4	65,0

Tabell 4 Skattad svarsandel fördelad på kön, procent.

	Kvinnor	Män
Vägd	48,1	53,5
Ovägd	47,0	48,9

Tabell 5 Skattad svarsandel fördelad på ålder, procent.

	65-79	80-
Vägd	54,6	35,6
Ovägd	53,0	32,9

Tabell 6 Skattad svarsandel fördelad på födelseland, procent.

	Sverige	Övriga Länder
Vägd	51,5	35,7
Ovägd	50,9	27,8

Tabell 7 Skattad svarsandel fördelad på civilstånd, procent.

	Gift eller registrerat partnerskap	Övriga
Vägd	53,5	45,7
Ovägd	53,6	40,8

Tabell 8 Skattad svarsandel fördelad på utbildningsnivå, procent.

	Förgymnasial	Gymnasial	Eftergymnasial
Vägd	35,9	50,4	65,9
Ovägd	32,4	49,4	61,5

Tabell 9 Skattad svarsandel fördelad på kommungrupp, procent.

	Storstäder, förortskommuner till storstäder	Större städer, förortskom. till större städer	Övriga kommuner
Vägd	51,5	53,0	46,3
Ovägd	47,2	52,8	43,8

Tabell 10 Skattad svarsandel fördelad på sammankomster, procent.

	0-5	6-12	13-19	20
Vägd	50,3	51,7	44,4	48,2
Ovägd	49,7	52,5	39,9	35,4

Tabell 11 Skattad svarsandel fördelad på timmar, procent.

	9-14	15-38	39-59	60-
Vägd	44,5	52,0	46,5	48,9
Ovägd	44,9	52,3	39,5	40,4

Tabell 12 Skattad svarsandel fördelad på deltagare, procent.

	3-6	7-12	13-18	19-
Vägd	52,6	48,5	48,8	57,7
Ovägd	43,9	47,6	53,1	58,0

Tabell 13 Skattad svarsandel fördelad på yrke, procent.

	Arbetaryrken	Tjänstemannayrken	Okänt
Vägd	40,0	60,0	35,4
Ovägd	43,0	58,1	31,6

Tabellerna 2-13 indikerar att samtliga hjälpvariabler utom kön är starka beträffande kriterium 1. Exempelvis är det stora skillnader i svarsbenägenhet beroende på deltagarens utbildning (svarsbenägenheten ökar med högre utbildning).

3.1.2 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.

Kalibreringsestimaten ger konsistenta skattningar i den meningen att estimaten ger exakta skattningar för utnyttjade registertotaler.

Samtliga hjälpvariabler utom civilstånd avgränsar viktiga redovisningsgrupper i föreliggande undersökning. Om det är möjligt bör dessa variabler användas i hjälpvektorn.

Slutligt val av hjälpvektor

Efter en sammanvägning av analysen kring ovanstående kriterier samt efter kontroll av vikternas fördelning används följande hjälpvektor:

$\text{ÄMNE} + \text{FÖRBUND} + \text{KÖN} + \text{ÅLDER} + \text{FÖDELSELAND} + \text{UTBILDNINGSNIVÅ}$

Det finns fler variabler som är tänkbara hjälpvariabler, men ovanstående har bedömts vara de viktigaste variablerna. Att inkludera för många variabler i hjälpvektorn ger problem med stor variation i vikterna, vilket kan leda till dålig precision i statistiken.

4 Teknisk beskrivning av urval och estimation

Vi har en population U bestående av N studiecirkeldeltagare. 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 deltagare k och z_k värdet på en annan variabel för samma deltagare. Vanligtvis är y (och även z) en dikotom variabel, d.v.s.

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

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 deltagare 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

*ÄMNE + FÖRBUND + KÖN + ÅLDER + FÖDELSELAND +
UTBILDNINGSNIVÅ*

som deltagare k tillhör. Vi framställer hjälptotalerna $\sum_{U_d} \mathbf{x}_k$ och utnyttjar denna hjälpinformation i en kalibreringsestimator.

Kalibreringsestimatoren för totalen Y_d har följande utseende:

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

där

$$d_k = N_h / n_h \text{ för } k \in r_h$$

och

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

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

Referenser:

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