

Kalibreringsrapport

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 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.

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) samt registret över folkhögskollärare. Hjälpvariablerna är definierade enligt tabell 1.

Tabell 1. Tänkbara hjälpvariabler

Variabel (benämning)	Kategorier (koder)
HUVUDMAN	1 = Offentlig (kommunal, landsting eller statlig) 2 = Enskild/fristående
H-REGION*	1 = Stockholm/Södertälje 2 = Göteborg/Malmö/Lund/Trelleborg 3 = Större städer 4 = Mellanbygden 5 = Tätbygden 6 = Glesbygden
STORSTAD**	1 = Stockholm/Göteborg/Malmö region 2 = Övriga
KÖN	1 = Män 2 = Kvinnor
ÅLDER***	1 = 25 – 34 2 = 35 – 45 3 = 46 – 60 4 = 61 –
FÖDELSELAND	1 = Sverige 2 = Övriga Världen
UTLÄNDSK/SVENSK BAKGRUND	1 = Inrikes född med två inrikes födda föräldrar 2 = Övriga
CIVILSTÅND	1 = Gift eller registrerat partnerskap 2 = Övriga
UTBILDNINGSNIVÅ****	1 = Förgymnasial, Gymnasial (inkluderar uppgift saknas) 2 = Eftergymnasial kortare än tre år 3 = Eftergymnasial tre år och längre
PEDAGOGISK UTBILDNING	1 = Folkhögskolläraryxamen 2 = Annan lärarexamen 3 = Annan högskoleexamen 4 = Annan utbildning inkl. uppgift saknas

Tabell 1. Tänkbara hjälpvariabler forts.

Variabel (benämning)	Kategorier (koder)
ANSTÄLLNINGSSLAG	1 = Tillsvidareanställd 2 = Visstidsanställd inkl. uppgift saknas
BARN6	1 = Inga hemmavarande barn i åldern 0 – 6 år 2 = Har hemmavarande barn i åldern 0 – 6 år
BARN18	1 = Inga hemmavarande barn i åldern 0 – 18 2 = Har hemmavarande barn i åldern 0 – 18
ANSTÄLLD ANTAL ÅR I FOLKHÖGSKOLA	1=0-5 år, inkl. uppgift saknas 2=6-10 år 3=11 år eller mer
TJÄNSTGÖRINGS-OMFATTNING PEDAGOGIK	1 = 0 – 40 2 = 41 – 80 3= 81 –
INRIKTNING TJÄNSTGÖRINGSOMFATTNING	1=Huvudsakligen övrigt (UU, AD, SY) 2=Huvudsakligen allmän kurs (UA) 3=Huvudsakligen särskild och kort kurs (US) 4=Blandat
LÖNEINKOMST*****	1= -- 232032 2 = 232033 – 314152 3 = 314153 -- 355206 4 = 355207 --

Variablerna kopplade till person (kön, ålder, födelseland, civilstånd, antal hemmavarande barn, utbildningsnivå, pedagogisk utbildning, anställningsår, tjänstgöringsomfattning, anställningsslag och löneinkomst) avser folkhögskolläraren.

*** Ålder avser den 31 december 2012.

**** Utbildningsnivå avser folkhögskollärarens hösta utbildning fram till och med vårterminen 2012.

* H-region avser den region där folkhögskolan finns. För en mer detaljerad beskrivning av H-regioner se variabelbeskrivningen.

** Storstadskommuner avser också den kommun där folkhögskolan finns. Till storstad räknas Stockholm, Göteborg och Malmö region.

***** Årlig löneinkomst avser år 2011. Intervallgränserna för årlig löneinkomst är beräknade utifrån kvartiler och median för samtliga 2 599 folkhögskollärare som ingår i urvalsramen för undersökningen.

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 svarsandel i olika grupper, bestämda av respektive hjälpvariabel. I svarsandelen inkluderas inte övertäckningen dvs. personer som inte arbetar som folkhögskollärare längre.

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

Tabell 2 Svarsandel fördelad på Huvudman.

	Offentlig	Enskild
Svarsandel (%)	55,4	50,0

Tabell 3 Svarsandel fördelad på H-region.

	Stockholm	Göteborg/Malmö	Större städer	Mellanbygden	Tät- och glesbygd
Svarsandel (%)	49,5	56,1	47,9	52,5	56,4

Tabell 4 Svarsandel fördelad på storstad.

	Storstadsregioner	Övriga regioner
Svarsandel (%)	51,7	51,7

Tabell 5 Svarsandel fördelad på kön.

	Kvinnor	Män
Svarsandel (%)	52,6	50,6

Tabell 6 Svarsandel fördelad på ålder.

	25-34	35-45	46-60	61 -
Svarsandel (%)	45,5	45,3	56,2	51,9

Tabell 7 Svarsandel fördelad på födelseland.

	Sverige	Övriga Världen
Svarsandel (%)	52,3	46,9

Tabell 8 Svarsandel fördelad på utländsk/svensk bakgrund

	Inrikes född och två inrikes födda föräldrar	Övriga personer
Svarsandel (%)	52,2	49,6

Tabell 9 Svarsandel fördelad på civilstånd.

	Gift/RP	Övriga
Svarsandel (%)	53,3	49,8

Tabell 10 Svarsandel fördelad på utbildningsnivå.

	Förgymnasial och Gymnasial	Eftergymnasial < 3 år	Eftergymnasial 3 år och längre
Svarsandel (%)	39,2	51,1	53,7

Tabell 11 Svarsandel fördelad på pedagogisk utbildning

	Folkhögskolläraryr- examen	Annan lärar examen	Annan högskoleexamen	Annan utbildning
Svarsandel (%)	60,8	52,2	45,4	43,3

Tabell 12 Svarsandel fördelad på anställningslag

	Tillsvidareanställd	Visstidsanställd
Svarsandel (%)	54,4	35,6

Tabell 13 Svarsandel fördelad på antal hemmavarande barn 0 – 6 år

	Minst ett barn i åldern 0 – 6 år	Inga barn i åldern 0 – 6 år
Svarsandel (%)	47,0	52,6

Tabell 14 Svarsandel fördelad på antal hemmavarande barn 0 – 18 år

	Minst ett barn i åldern 0 – 18 år	Inga barn i åldern 0 – 18 år
Svarsandel (%)	50,8	52,2

Tabell 15 Svarsandel fördelad på antal anställningsår i folkhögskola.

	0 – 5 år	6 – 10 år	11 år eller mer
Svarsandel (%)	43,3	50,4	58,5

Tabell 16 Svarsandel fördelad på tjänstgöringsomfattning inom pedagogik/undervisning.

	0 – 40%	41 – 80%	81% eller mer
Svarsandel (%)	37,4	51,0	56,5

Tabell 17 Svarsandel fördelad på tjänstgöringsomfattningens inriktning

	Huvudsakligen övrigt	Huvudsakligen Allmän kurs	Huvudsakligen särskild eller kort kurs	Blandat
Svarsandel (%)	44,5	54,6	52,1	44,4

Tabell 17 Svarsandel fördelad på årlig löneinkomst

	- 232032	232033 - 314152	314153 - 355206	355207 -
Svarsandel (%)	44,0	49,6	55,4	55,5

Tabellerna 2-17 indikerar att majoriteten av hjälpvariablerna är starka beträffande kriterium 1. Exempelvis är det stora skillnader i svarsbenägenhet efter antal anställningsår i folkhögskola (svarsbenägenheten ökar med stigande anställningsår). Det hjälpvariabler som inte tycks vara starka beträffande kriterium 1 är storstad, kön, utländsk/svensk bakgrund, civilstånd och antal hemmavarande barn 0 – 18 år

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.

Kalibreringsestimaton ger konsistenta skattningar i den meningen att estimaton ger exakta skattningar för utnyttjade registertotaler.

Hjälpvariablerna huvudman, region, antal anställningsår i folkhögskola, kön, ålder, födelse land, pedagogisk utbildning och tjänstgöringsomfattning inom pedagogik avgränsar viktiga redovisningsgrupper i föreliggande undersökning. Om det är möjligt bör dessa variabler användas i hjälpvektorn.

3.2 Slutligt val av hjälpvektor

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

*KÖN*HUVUDMAN + H-REGION + ANSTÄLLNINGSSLAG +
UTBILDNINGSNIVÅ+PEDAGOGISK UTBILDNING + FÖDELSELAND +
BARN 0-6 ÅR +ANTAL ANSTÄLLNINGSSÅR FOLKHÖGSKOLA + ÅLDER
+ TJÄNSTGÖRINGSOMFATTNING PEDAGOGIK + LÖNEINKOMST*

4 Teknisk beskrivning av urval och estimation

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

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

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 folkhögskollärare i redovisningsgrupp d , då kan parametern skrivas

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

Samtliga personer av storleken N i urvalsramen U ingår i undersökningen. På grund av övertäckning och bortfall har vi endast svarsmängden r av storleken m att utföra beräkningarna på.

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

$$\hat{Y}_d = \frac{N}{m} \sum_r y_{dk} \quad (4.4)$$

I estimator (4.4) används ingen hjälpinformation.

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

*KÖN*HUVUDMAN + H-REGION + ANSTÄLLNINGSSLAG +
UTBILDNINGSNIVÅ+PEDAGOGISK UTBILDNING + FÖDELSELAND +
BARN 0-6 ÅR +ANTAL ANSTÄLLNINGÅR FOLKHÖGSKOLA + ÅLDER
+ TJÄNSTGÖRINGSOMFATTNING PEDAGOGIK + LÖNEINKOMST*

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

Kalibreringsestimatorn för totalen Y_d har följande utseende:

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

där

$$d_k = N / N \text{ för } k \in r$$

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