# Kalibreringsrapport för undersökningen ”Vetenskap i skolans och förskolans vardag”

## 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 man inte lyckas få svar från alla personer (bortfall) och om dessa 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. Parametrar

För samtliga frågor kommer den procentuella andelen personer med en viss egenskap att redovisas. Resultaten redovisas efter skolform, huvudman, inriktning och befattning.

## 3. 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.

### 3.1 Tänkbara hjälpvariabler

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

1. 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.
2. 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.
3. 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 Registret över pedagogisk personal och Registret över Totalbefolkningen (RTB). En genomgång av variablerna i dessa register resulterade i att sju variabler valdes ut.

Erfarenheten säger exempelvis att svarsfrekvensen brukar skilja sig åt mellan män och kvinnor. Dessutom brukar personer i storstadsregioner ha en benägenhet att svara i mindre utsträckning än övriga landet.

De sammanslagningar av kategorier som gjorts baseras på kunskaper från tidigare kalibreringar. I tabell 1 definieras de utnyttjade hjälpvariablerna.

*Tabell 1.* Definition av de utnyttjade hjälpvariablerna

|  |  |
| --- | --- |
| Variabel (benämning) | Förklaring av koderna |
|  |  |
| Kön | 1 = Man  2 = Kvinna |
| Ålder | 1 = -39 år  2 = 40-49 år  3 = 50-59 år  4 = 60 – 66 år |
| Civilstånd | 1 = Gift eller registrerat partnerskap  2 = Annat civilstånd |
| Födelseland | 1 = Sverige  2 = Annat land |
| Boenderegion | 1 = Storstadskommun (Stockholm, Göteborg, Malmö kommun)  2 = Övriga kommuner |
| Huvudman | 1 = Kommunal skola  2 = Fristående skola |
| Behörighet | 1 = Behörig lärare  2 = Obehörig lärare |
| Tjänstgöringsomfattning | 1 = <100 %  2 = 100 % |

Variablerna kön, ålder, civilstånd och födelseland är frekvent använda hjälpvariabler när kalibrering används vid individundersökningar. Det främsta skälet till detta är att de ofta tros kunna förklara delar av variationen i svarsbenägenheten.

För förskollärarna är inte variablerna behörighet och omfattning kända (eftersom dessa variabler hämtas från Registret över pedagogisk personal där förskollärare inte finns med). Svarsandelarna i avsnitt 4 omfattar därför inte förskollärare för dessa variabler.

## 4. 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: /). Vid stora skillnader mellan svarsandelarna utgör variabeln en stark kandidat till hjälpvariabel.

Den totala svarsandelen skattades till 50,6 procent.

*Tabell 2.* Skattade procentuella andelar svarande fördelade på kön

|  |  |  |
| --- | --- | --- |
| Kön | Man | Kvinna |
| Svarsandel (%) | 46,7 | 51,5 |

*Tabell 3.* Skattade procentuella andelar svarande fördelade på ålder

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ålder | -39 år | 40-49 år | 50-59 år | 60-66 år |
| Svarsandel (%) | 43,5 | 48,8 | 56,1 | 60,1 |

*Tabell 4.* Skattade procentuella andelar svarande fördelade på civilstånd

|  |  |  |
| --- | --- | --- |
| Civilstånd | Gift/partnerskap | Annat |
| Svarsandel (%) | 50,4 | 51,0 |

*Tabell 5.* Skattade procentuella andelar svarande fördelade på födelseland

|  |  |  |
| --- | --- | --- |
| Födelseland | Sverige | Annat |
| Svarsandel (%) | 51,7 | 39,2 |

*Tabell 6*. Skattade procentuella andelar svarande fördelade på boenderegion

|  |  |  |
| --- | --- | --- |
| Boenderegion | Storstadskommun | Övriga kommuner |
| Svarsandel (%) | 38,8 | 52,6 |

*Tabell 7.* Skattade procentuella andelar svarande fördelade på huvudman

|  |  |  |
| --- | --- | --- |
| Huvudman | Kommunal skola | Fristående skola |
| Svarsandel (%) | 53,0 | 40,2 |

*Tabell 8.* Skattade procentuella andelar svarande fördelade på behörighet

|  |  |  |
| --- | --- | --- |
| Behörighet | Behörig | Obehörig |
| Svarsandel (%) | 51,0 | 35,4 |

*Tabell 9.* Skattade procentuella andelar svarande fördelade på omfattning

|  |  |  |
| --- | --- | --- |
| Omfattning | <100 % | 100 % |
| Svarsandel (%) | 51,1 | 47,2 |

Variablerna ålder, födelseland, boenderegion, huvudman och behörighet tycks samvariera med svarsbenägenhet. Variablerna kön, civilstånd och tjänstgöringsomfattning verkar däremot inte starka enligt kriterium 1.

### 4.2 Kriterium 2: Variabeln samvarierar med (viktiga) målvariabler

Eftersom undersökningen innehåller målvariabler av skiftande karaktär är detta kriterium av underordnad betydelse vid analysen av hjälpinformationen.

### 4.3 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.

Om täckningsfel inte föreligger ger kalibreringsestimatorn konsistenta skattningar i den meningen att esti­matorn ger exakta skattningar för utnyttjade registertotaler.

Variablerna stratum (som konstruerats utifrån skolform, huvudman, inriktning och befattning) avgränsar redovisningsgrupper och bör därför vara med i hjälpvektorn enligt det tredje kriteriet.

### 4.4 Slutligt val av hjälpvektor

De variabler som enligt analysen bör ingå i hjälpvektorn är ålder, födelseland, boenderegion, behörighet och huvudman. Huvudman ingår i variabeln stratum som väljs utifrån kriterium 3.

Efter kontroll av vikternas fördelning används följande hjälpvektor:

*Stratum+ Ålder + Födelseland +Boenderegion + Behörighet*

För förskollärarna används dock bara variabeln stratum, då information om övriga variabler inte är kända på populationsnivå.

## 5. Teknisk beskrivning av urval och estimation

Vi har en population *U* bestående av *N* personer. De parametrar vi är intresserade av är främst funktioner av två totaler  och , där  är värdet på variabel *y* för person *k* och  värdet på en annan variabel för samma person. Vi kan definiera *y* (och även *z*) som en dikotom variabel, dvs.

 (5.1)

Det finns givetvis också intresse av parametrar för olika redovisningsgrupper. Låt oss benämna dessa , där . Totalen för redovisningsgrupp *d* kan skrivas

 (5.2)

där 

 bildas på likartat sätt.

En generell parameter för redovisningsgrupp *d* (*d* kan också avse hela populationen) kan skrivas , där *C* är en konstant.

Den vanligaste parametern är en procentuell andel, som erhålles när  och  för alla k, och *y* är definierad enligt (5.1). Om vi låter  vara antalet personer i redovisningsgrupp *d*, då kan parametern skrivas

 (5.3)

Vi drar ett obundet slumpmässigt urval  av storleken  från stratum h (), men p.g.a. övertäckning och bortfall har vi endast svars­mängden  av storleken  att utföra beräkningarna på. Storleken på stratum *h* ger vi beteckningen .

Den ”konventionella” estimatorn (för ), har följande form:

 (5.4)

I estimator (5.4) används ingen ytterligare hjälpinformation än stratifierings­informationen. Denna estimationsmetod brukar kallas ”rak uppräkning inom strata”.

I syfte att erhålla en estimator med mindre urvalsfel och bortfallsskevhet än estimator 5.4) utnyttjar vi hjälpinformation också i estimationen. Vi bildar en hjälpvektor , som anger till vilka kategorier av

*Stratum+ Ålder + Födelseland +Boenderegion + Behörighet*

som person *k* tillhör. Från registret över totalbefolkningen och registret över pedagogisk personal framställer vi sedan hjälptotalerna . Vi utnyttjar denna hjälp­information i en kalibreringsestimator.

Kalibreringsestimatorn för totalen  har följande utseende:

 (5.5)

där

 för 

och

 (5.6)

Vid skattning av en parameter av typen  skattas respektive total med hjälp av kalibreringsvikterna .

Denna kalibreringsvikt uppfyller kalibreringsvillkoret: , vilket innebär att om vikterna läggs på variabler som ingår i hjälpvektorn summeras dessa upp till de hjälptotaler vi hämtat från registren.

**Referenser:**

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