

Enkätundersökning till elever i årskurs 9 för UGU-projektet

Teknisk rapport - En beskrivning av
genomförande och metoder

2021 04 23

Beteckning: 249676/8948143



Innehåll

| | | |
|----------|--|-----------|
| 1 | Sammanfattning..... | 2 |
| 2 | Genomförande och metoder..... | 3 |
| 2.1 | Population och urval..... | 3 |
| 2.2 | Blankettfrågor och registeruppgifter | 3 |
| 2.3 | Datainsamling..... | 4 |
| 2.3.1 | Kontroller..... | 5 |
| 2.4 | Bortfall | 5 |
| 2.5 | Viktberäkning och estimation..... | 6 |
| 2.6 | Statistikens/Datats tillförlitlighet..... | 7 |
| 2.6.1 | Tillförlitlighet totalt..... | 7 |
| 2.6.2 | Osäkerhetskällor | 7 |
| 2.6.3 | Urval..... | 8 |
| 2.6.4 | Ramtäckning | 8 |
| 2.6.5 | Mätning..... | 8 |
| 2.6.6 | Bortfall | 9 |
| 2.6.7 | Bearbetning..... | 9 |
| 2.6.8 | Modellantagande | 9 |
| 2.7 | Beskrivning av datafil..... | 9 |
| 2.7.2 | Jämförbarhet och användbarhet | 10 |
| 2.7.3 | Sekretess och utlämnande | 10 |
| 2.7.4 | Övrigt | 10 |
| 3 | Bilagor..... | 11 |

1 Sammanfattning

Den här rapporten beskriver hur SCB har genomfört uppdraget **UGU årskurs 9** och vilka metoder som har använts.

Om du har frågor är du välkommen att höra av dig till den ansvarig på SCB: Karolina Bock, telefon 010 - 479 65 72 eller e-post karolina.bock@scb.se.

Uppdragets syfte: Syftet med UGU-projektet är att bidra till utvecklingen av kunskap om skolan. Framförallt är avsikten att undersöka hur skolan påverkar eleverna vad gäller vilja och lust att lära, självkänsla, intressen, trivsel, stress, hälsa och välbefinnande, och fortsatta studier.

Uppdragets parter: Enheten för statistik om utbildning och arbete vid SCB genomförde under perioden april – juni 2020 en enkätundersökning på uppdrag av Göteborgs universitet.

Uppdragets bakgrund: Under våren 2020, när undersökningen genomfördes, var det 964 personer som inte samtyckte till att delta i undersökningen. Dessa svarande kodades därför till bortfall. I samband med Coronaenkät UGU fick urvalspersonerna nya frågor om samtycke. Av 964 svarande personer som tidigare inte lämnat sitt samtycke var det 510 personer som valde att lämna in sitt samtycke i efterhand. Denna rapport är baserad på alla svarande som lämnat samtycke, totalt 3 033 personer.

Kontaktpersoner: Undersökningsledare på SCB var Karolina Bock. Johan Löfgren var metodstatistiker. Göteborgs universitets kontaktperson gentemot SCB var Jan-Eric Gustafsson.

Grundläggande kvalitetskrav: Detta uppdrag har genomförts i enlighet med ISO 20252:2012 Marknads-, opinions- och samhällsundersökningar vilket innebär att grundläggande kvalitetskrav har uppfyllts.

Leverans: Resultatet i form av datafil och teknisk rapport levererades vecka 16 enligt överenskommelse.

Urval: Urvalet till denna uppföljning bestod av 9 665 personer varav 94 personer utgjorde övertäckning. Det innebär att urvalet, där övertäckningen är borträknad, var 9 571 personer. Det var totalt 3 033 personer som besvarade frågeblanketten, vilket var 31,7 procent av urvalet (övertäckning borträknad). Den designvägda svarsandelen, vilken tar hänsyn till om olika personer haft olika stor sannolikhet att komma med i urvalet, var 31,7 procent. Denna svarsandel kan ses som en skattning av den svarsandel som en totalundersökning skulle ha resulterat i.

2 Genomförande och metoder

2.1 Population och urval

Populationen, det vill säga de objekt som man vill kunna dra slutsatser om, utgjordes av elever i årskurs 3 då elevpanelen startade våren 2014. Vårterminen 2020 gick majoriteten av dessa elever i årskurs 9. Urvalet består av ca 9 600 personer, vilka utgörs av samtliga elever som kvarstår i den insamling som genomförs för den givna elevpanelen.

Ursprungligt urval och urvalsram

För att kunna dra ett urval från populationen skapas en urvalsram som avgränsar, identifierar och möjliggör koppling till objekten i populationen.

Urvalet i denna undersökning består av det urval som drogs till Elevpaneler för longitudinella studier, panel 8. Urvalet drogs år 2014. Urvalsramen bestod år 2014 av grundskolor (ej sameskolor) som hade elever i årskurs 3 under läsåret 2013/14.

Skolenheterna hämtades från registret över skolenheter. För att avgöra vilka skolenheter som hade elever i årskurs 3 det aktuella läsåret användes registret över elever i grundskolan, mätdatum 15 oktober. Båda dessa register håller SCB för Skolverkets räkning. Från fem skolenheter där personuppgifter saknades för eleverna samlades klasslistor in.

Urvalet stratifierades efter typ av skolhuvudman för att säkerställa att ett tillräckligt antal elever i fristående skolenheter blev representerade. För mer information om population, stratifiering samt ursprungligt urval hänvisas till dokumentationen över SCB:s longitudinella elevpaneler.

I urvalet som drogs år 2014 ingick totalt 9 794 elever. Elever som vägrat medverka i *Elevpaneler för longitudinella studier* exkluderades från dessa insamlingar, liksom elever som där kodats bort som övertäckning eller på grund av skyddad identitet. I de fall vårdnadshavaren efter att ha fått sin enkät kontaktade SCB och bad SCB att exkludera eleven har så också skett.

2.2 Blankettfrågor

Göteborgs universitet utformade frågorna i frågeblanketten.

Grundläggande för bra kvalitet i en undersökning är kvaliteten på de data som samlas in. För att minska risken för mätfel har blanketten därför genomgått ett mättekniskt test.

Blanketten bestod av 27 numrerade frågor, flera av dem hade dessutom delfrågor.

2.3 Datainsamling

Innan utskick gjordes en kontroll av personerna i urvalet mot de senaste folkbokföringsuppgifterna, för att ta fram aktuella adressuppgifter. Vid kontrollen och under insamlingsfasen framkom det att 94 personer inte längre tillhörde populationen utan utgjorde känd övertäckning, det vill säga den övertäckning som identifierats. Orsakerna till övertäckning var att personerna emigrerat eller att de inte fanns med i SCB:s register.

Totalt 4 utskick gjordes. Samtliga utskick gjordes med post. I ett informationsbrev ombads urvalspersonerna att besvara frågorna och skicka svaret till SCB. I informationsbrevet kunde urvalspersonerna läsa om undersökningens bakgrund, syfte och att undersökningen genomfördes i samarbete mellan Göteborgs universitet och SCB. Informationsbrevet informerade också om att uppgifter hämtades från SCB:s register och att en avidentifierad datafil levereras till Göteborgs universitet. Brevet informerade även om dataskyddslagstiftningen samt offentlighets- och sekretesslagen och att det var frivilligt att medverka i undersökningen.

Frågeblanketten skickades ut den 20 april 2020. Sedan skickades 3 påminnelser till dem som inte besvarat frågeblanketten. Ett påminnelsebrev skickades den 5 maj. En första påminnelse med ny frågeblankett skickades den 19 maj och en andra påminnelse med ny blankett den 2 juni. Insamlingen avslutades den 26 juni 2020.

Vid samtliga utskick fick uppgiftslämnaren information om att frågorna även kunde besvaras på webben samt inloggningsuppgifter.

Totalt svarade 3 033 personer på frågeblanketten, vilket var 31,7 procent av urvalet efter att den kända övertäckningen, det vill säga den övertäckning som identifierats, tagits bort. Den designvägda svarsandelen, vilken tar hänsyn till om olika personer haft olika stor sannolikhet att komma med i urvalet, var 31,7 procent. Denna svarsandel kan ses som en skattning av den svarsandel som en totalundersökning skulle ha resulterat i.

Tabell 1 Beskrivning av inflödet. Antal och andel svar

| | Antal | Andel |
|--------------------------|--------------|--------------|
| Efter första utskicket | 1424 | 14,9 |
| Efter andra utskicket | 633 | 6,6 |
| Efter tredje utskicket | 606 | 6,3 |
| Efter fjärde utskicket | 370 | 3,9 |
| Totalt antal svar | 3 033 | 31,7 |
| Bortfall | 6 538 | 68,3 |
| Urval exkl. övertäckning | 9 571 | 100,0 |

Det var 1 229 personer som valde att besvara webblancketten, vilket var 40 procent av samtliga svarande.

De frågeblanketter som besvarats på papper registrerades med hjälp av skanning.

2.3.1 Kontroller

Kontroller har genomförts bland annat för att säkerställa att endast valida värden förekommer i materialet.

De inkomna pappersblanketterna granskades i flera olika skeden av bearbetningen.

Handskrivna tecken kontrollerades efter skanningen och tecken som tolkats felaktigt korrigerades.

Vid frågor där svaret skulle anges i form av skrivna siffror, gjordes en rimlighetsbedömning. Orimliga värden togs ut för kontroll i syfte att säkerställa att rätt värde registrerats.

Under datainsamlingen granskades otillåtna värden löpande, till exempel om det fanns frågor där uppgiftslämnare markerat flera svar trots att frågan skulle besvaras med ett alternativ. Hur otillåtna värden hanterats beskrivs i avsnittet 2.7.1.1 Datafil.

Datamaterialet granskades före leverans med hjälp av frekvenstabeller för att upptäcka och korrigera eventuella kvarstående fel.

SCB kan inte garantera att den utvalda personen själv besvarat frågeblanketten. Någon kontroll av att rätt person har besvarat frågeblanketten har inte gjorts. Detta på grund av att frågeblanketten saknar bakgrundsfrågor.

2.4 Bortfall

Bortfallet består dels av objektbortfall, som innebär att frågeblanketten inte är besvarad alls, och dels av partiellt bortfall som innebär att vissa frågor i blanketten inte är besvarade. Om bortfallet skiljer sig från de svarande, med avseende på undersökningsvariablerna, kan

skattningarna som grundar sig på enbart de svarande vara skeva. För att reducera bortfallsskevheten har vikter beräknats med hjälp av kalibrering (se bilaga 1).

Objektsbortfall kan bland annat bero på att uppgiftslämnaren inte är villig att delta i undersökningen, att uppgiftslämnaren inte går att nå eller att uppgiftslämnaren är förhindrad att medverka. Objektsbortfallet i denna undersökning redovisas i tabell 2 nedan.

Tabell 2 Beskrivning av objektsbortfall

| | Antal |
|----------------------------------|--------------|
| Ej avhörda | 5 963 |
| Avböjd medverkan | 37 |
| Ej anträffad | 74 |
| Inkommen blank | 10 |
| Inte kryssat för samtyckesfrågor | 454 |
| Totalt | 6 538 |

Med "Ej avhörda" menas att ingen uppgift om varför frågeblanketten inte är besvarad har lämnats. Med "Avböjd medverkan" menas att SCB meddelats att uppgiftslämnaren inte vill medverka i undersökningen. "Ej anträffad" innehåller personer som saknar adress i Registret över totalbefolkningen (RTB), har hemlig adress samt postreturer. "Inte kryssat för samtyckesrutor" betyder att uppgiftslämnaren inte har samtyckt till att delta i undersökningen.

Partiellt bortfall kan bero på att en fråga är svår att förstå, är känslig, att uppgiftslämnaren missar att besvara frågan eller att instruktionerna vid hoppfrågor misstolkas. Det partiella bortfallet varierar mellan 1 och 4 procent för de flesta av frågorna.

2.5 Viktberäkning och estimation

För varje svarande person (kallas objekt i fortsättningen) har en vikt beräknats. Syftet med detta är att kunna redovisa resultat för hela populationen och inte bara för de svarande. Vikten kallas därför även för uppräkningsstal.

Vikterna har beräknats utifrån urvalsdesignen samt antaganden om objektsbortfall och täckningsfel. Beräkningen gjordes med hjälp av ett av SCB egenutvecklat SAS-makro (ETOS).

Vikterna i denna undersökning kan beskrivas med formeln:

$$w_k = d_k \cdot v_k$$

där w_k = vikt/uppräkningsstal för objekt k

d_k = designvikt för objekt k

v_k = kalibreringsvikt för objekt k baserad på hjälpinformation

Designvikten är den del av vikten som beror på urvalsdesignen. Vid bortfall kan det vara så att vissa grupper av urvalet svarar i större utsträckning än övriga, t.ex. kan kvinnor svara i högre grad än män. Om de grupper som svarat i högre grad har en annan fördelning med avseende på undersökningsvariablerna än övriga kan detta ha en snedvridande effekt på resultatet. För att kompensera för detta har kalibreringsvikter använts (se bilaga för mer detaljer). Vikterna korrigerar för den övertäckning som identifierats. Ingen korrigerande för eventuell undertäckning har gjorts.

För en mer utförlig beskrivning av kalibreringsestimatorens se Lundström och Särndal (2001): *Estimation in the Presence of Nonresponse and Frame Imperfections, Statistics Sweden*.

Vikterna/uppräkningsstalen multipliceras med objektens variabelvärden för att skapa statistikvärden gällande för populationen. Om vikterna inte används så kan resultaten bli helt missvisande. Vikterna kompenserar för objektbortfallet, men inte för det partiella bortfallet.

För beräkning av skattningar av totaler används följande formel:

$$\hat{Y} = \sum_r w_k y_k$$

där w_k = vikt/uppräkningsstal för objekt k
 y_k = variabelvärde för objekt k

Summering sker av de svarande r .

För beräkning av skattningar av medelvärden används följande formel:

$$\hat{\bar{Y}} = \frac{\sum_r w_k y_k}{\sum_r w_k}$$

där w_k = vikt/uppräkningsstal för objekt k
 y_k = variabelvärde för objekt k

Summering sker av de svarande r .

2.6 Datats tillförlitlighet

2.6.1 Tillförlitlighet totalt

Detta avsnitt handlar om tillförlitligheten för den statistik som ska tas fram. I detta uppdrag har SCB enbart samlat in data och kunden tar själv fram statistik och gör analyser utifrån insamlade data.

Statistiken som ska tas fram kommer att vara behäftad med osäkerhet. Vid bedömning av hur olika osäkerhetskällor påverkar statistiken från

en undersökning skiljer man på slumpmässiga och systematiska avvikelser. Slumpmässiga avvikelser orsakar enligt statistisk teori beräkningsbar osäkerhet hos de skattade resultaten. Systematiska avvikelser påverkar resultaten i en viss riktning. Det är svårt att ange hur tillförlitligheten påverkas av eventuella systematiska avvikelser (skevhet), då det ofta kräver resurskrävande utvärderingsinsatser. En sådan utvärdering har inte gjorts för detta uppdrag. Det är därför svårt att bedöma kvaliteten i statistiken.

Med tanke på att bortfallet är stort, bör generella slutsatser som dras från undersökningen göras med viss försiktighet.

Nedan ges några kommentarer om varje osäkerhetskälla.

2.6.2 Osäkerhetskällor

Bortfall bedöms vara den största källan till osäkerhet när statistik ska tas fram. Urval, täckning, bearbetning och mätning bedöms påverka osäkerheten i mindre grad. Om mindre delgrupper ska analyseras kan även urvalet bidra till osäkerheten.

2.6.3 Urval

Denna kvalitetskomponent avser osäkerhet som uppkommer på grund av att endast ett urval av populationen undersöks. Urvalsosäkerheten är således avvikelser mellan ett skattat värde och det faktiska värdet. Urvalsosäkerheten minskar med en ökad urvalsstorlek. Urvalsosäkerheten bör beaktas när man drar slutsatser från undersökningen.

Urvalet till denna undersökning består av ca 9 600 personer.

2.6.4 Ramtäckning

Under- och övertäckning innebär att ram- och målpopulation inte helt stämmer överens. Undertäckning innebär att vissa objekt som ingår i målpopulationen saknas i ramen. Övertäckning innebär att objekt som inte ingår i målpopulationen ändå finns i ramen. Ett sätt att minska täckningsbrister är att använda bra och uppdaterade register för att ta fram ramar.

Vid urvalsdragningen som gjordes 2014 användes Skolverkets elevregister avseende läsåret 2013/14. Elevregistret håller hög kvalitet och täckningsbristerna i registret bedöms som relativt små.

Inför varje insamlingsomgång inom elevpanelerna görs sedan en kontroll gentemot SCB:s register över totalbefolkningen (RTB) och individer som t.ex. avlidit eller flyttat utomlands hanteras som övertäckning. Så sker även i de fall uppgift under insamlingens gång lämnats om att en individ i urvalet t.ex. gått över till särskola.

Från det ursprungliga urvalet finns det totalt 163 personer som har kodats bort som övertäckningsobjekt. 69 av dessa ingick inte i urvalet till denna uppföljning, då de tidigare kodats bort (vanligaste orsaken är att de övergått till särskola sedan urvalet drogs 2014). Ytterligare 94 personer som ingick i urvalet till denna uppföljning, har identifierats som övertäckning, den vanligaste orsaken till detta var att eleven inte bodde i Sverige vid undersökningstillfället.

Sammanfattningsvis är SCB:s bedömning att bristerna i ramen, både med avseende på under- och övertäckning, inte snedvrider statistiken i någon nämnvärd omfattning.

2.6.5 Mätning

Ett fel som kan uppstå vid mätning är att lämnade uppgifter skiljer sig från faktiska uppgifter. Felet kallas mätfel och kan uppkomma då uppgiftslämnaren inte minns de faktiska uppgifterna, missförstår frågan eller medvetet svarar felaktigt.

2.6.6 Bortfall

Skevheter i skattningar på grund av bortfall inträffar om objekten i bortfallet och de svarande skiljer sig åt avseende frågorna/variablerna i undersökningen.

Eventuella bortfallsfel kan reduceras genom att på ett effektivt sätt justera undersökningens vikter. En beskrivning av förfarandet finns i den bortfallsanalys som gjorts i samband med valet av hjälpinformation till kalibreringen (se bilaga 1).

I bortfallsanalysen redovisas även svarsandelar för viktiga delgrupper. Någon utvärdering av kvarstående skevheter efter kalibrering har inte gjorts, men med tanke på att bortfallet är relativt stort finns det risk för att skevheter kvarstår även efter kalibrering. Svarsandelen i undersökningen var 31,7 procent (designviktad svarsandel).

2.6.7 Bearbetning

Vid den manuella och maskinella bearbetningen av data kan osäkerhet uppstå. Exempel på bearbetningsbrister är registreringsfel och kodningsfel. Dessa kan förhindras och upptäckas i de kontroller som genomförs vid dataregistreringen. I den här undersökningen bedöms osäkerheten på grund av registrering vara litet eftersom frågeblanketten endast hade fasta svarsalternativ.

2.6.8 Modellantagande

Inga modellantaganden har gjorts utöver de som redovisats ovan.

2.7 Beskrivning av datafil

2.7.1.1 Datafil

Svarsdatafilen kompletterades med vikter. De bifogade vikterna är avsedda att användas vid skattning av målstorheter i populationen, exempelvis totaler och medelvärden.

Vid analyser av samband med hjälp av exempelvis regressionsanalys bör det uppmärksammas att det inte är givet hur vikterna ska hanteras. Urvalets design samt bortfall är komplikationer som kan störa analyser av samband.

Om det är partiellt bortfall (uppgift saknas) i en fråga är värdet blankt.

2.7.1.2 Statistiska mått

Materialet lämpar sig att användas för sambandsanalyser eller redovisas som totaler eller procentuella andelar för hela målpopulationen eller fördelat på olika redovisningsgrupper.

2.7.1.3 Redovisningsgrupper

Ofta redovisas statistik inte bara för hela populationen utan också för delgrupper (redovisningsgrupper). Redovisningsgrupper i den här undersökningen kan avgränsas med hjälp av registervariabler eller bakgrundsfrågor i enkäterna.

2.7.2 Jämförbarhet och användbarhet

2.7.2.1 Jämförbarhet över tid

Vid flera tillfällen har SCB på uppdrag av Göteborgs universitet genomfört liknande undersökningar.

2.7.3 Sekretess och utlämnande

SCB har gjort en sekretessprövning för utlämnandet av data på mikronivå.

2.7.4 Övrigt

En nyckel med koppling mellan löpnummer och personnummer kommer att sparas på SCB till 2023-09-01, med möjlighet till förlängning. Omprövning av nyckeln sker vart tredje år.

3 Bilaga

Bilaga 1. Kalibreringsrapport

Johan Löfgren
PMU/MIÖ

Bilaga 1

Kalibreringsrapport

UGU åk 9 våren 2020 – vikter utifrån nytt samtycke

2021-04-22



Inledning

I en urvalsundersökning är skattningarna behäftade med *urvalsfel* beroende på att endast en delmängd av populationen studeras. Ett annat fel uppkommer om vi inte lyckas få svar från alla och om personerna i bortfallet 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 urvalsfelen 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 urvals- och bortfallsfel kan reduceras genom att använda ett effektivt uppräkningsförfarande. I undersökningen används tekniken med kalibrering av vikter, vilket är en effektiv och flexibel metod för att reducera dessa fel. I korthet går tekniken ut på att vid uppräkningsutnyttja en vektor bestående av variabler som är kända för både de svarande och populationen. Dessa variabler ska helst samvariera med svarsbenägenheten och undersökningsvariabler (målvariabler) samt avgränsa viktiga redovisningsgrupper.

Bortfallsanalys

Syftet med bortfallsanalysen är att utreda vilka variabler som bör användas vid bortfallsjusteringen. Tre kriterier bör beaktas (Särndal & Lundström 2005):

- 1) Variabeln samvarierar väl med svarsbenägenheten. Det är det viktigaste kriteriet eftersom det leder till en minskning av bortfallskevheter för alla skattningar.
- 2) Variabeln samvarierar väl med målvariabler. Om så är fallet minskar bortfallsbiasen för de skattningar som byggs upp av dessa målvariabler. Även variansen minskar för dessa skattningar.
- 3) Variabeln avgränsar viktiga redovisningsgrupper. Det leder framförallt till minskad varians i skattningar för dessa redovisningsgrupper.

Tänkbara hjälpvariabler

I tabell 1 presenteras de variabler som ingår i analysen och som är kandidater till att användas som sk hjälpvariabler vid bortfallsjusteringen. Variablerna är hämtade från Registret över totalbefolkningen (RTB), Skolenhetsregistret, Betygsregister årskurs sex och Utbildningsregistret (UREG) och har antagits samvariera med svarsbenägenheten och/eller målvariabler.

Tabell 1 Tänkbara hjälpvariabler

| Variabel (benämning) | Kategorier |
|--------------------------------|--|
| KÖN | 1 = Pojke 2 = Flicka |
| FÖDELSELAND | 1 = Sverige 2 = Övriga Världen |
| HÖGSTA UTBILDNING FÖRÄLDRAR | 1 = Förgymnasial (eller uppgift saknas) 2 = Gymnasial utbildning 3 = Eftergymnasial utbildning |
| HUVUDMAN | 1 = Kommunal skola 2 = Fristående skola |
| BETYGSPÖÄNG ÅRSKURS SEX | 1 = 0 - 35 2 = 37.5 - 42.5 3 = 45 - 50 4 = 52.5 - 60 5 = Uppgift saknas |
| MERITVÄRDE ÅRSKURS NIO | 1 = Meritvärde saknas 2 = 0 - 197.5 3 = 200 - 240 4 = 242.5 - 275 5 = 277.5 - 320 |

Med högsta utbildning föräldrar avses högsta avslutade utbildning till och med vårterminen 2019. Utbildningen avser den förälder som har högst utbildning.

Betygspoängen har räknats ut genom att poängsätta och summera betygen i årskurs sex i svenska eller svenska som andraspråk, matematik och engelska. Poängsättningen i respektive ämne har gjorts på samma sätt som det görs för meritvärdet i årskurs nio.

Meritvärde årskurs nio är en ny hjälpvariabel, då den inte fanns tillgänglig när undersökningen genomfördes under våren/sommaren 2020.

Svarsandelar

Utifrån de nya medgivanden som samlas in har svarsandelar uppdaterats. Sambandet mellan de svarande och bortfallet studeras genom att beräkna svarsandelen per kategori för de tänkta hjälpvariablerna (se tabell 2-7).

Stora skillnader mellan svarsandelarna tyder på en skev svarsfördelning för hjälpvariabeln och ett starkt samband med svarsbenägenheten, vilket efterfrågas enligt kriterium 1.

I tabellerna redovisas svarsandelar för alla svar inklusive de som inte lämnat sitt medgivande i enkäten, samt svarsandelar baserat på de som lämnat medgivande.

Tabell 2. Designvägd svarsandel efter kön

| Grupp | Kategori | Antal svar | Svarsandel, % |
|---------------------|----------|------------|---------------|
| Tidigare svarsandel | Pojke | 1 053 | 21,3 |
| Tidigare svarsandel | Flicka | 1 470 | 31,2 |
| Tidigare svarsandel | Totalt | 2 523 | 26,2 |
| Nytt medgivande | Pojke | 1 274 | 25,8 |
| Nytt medgivande | Flicka | 1 759 | 37,8 |
| Nytt medgivande | Totalt | 3 033 | 31,7 |

Tabell 3. Designvägd svarsandel efter födelseland

| Grupp | Kategori | Antal svar | Svarsandel, % |
|---------------------|----------------|------------|---------------|
| Tidigare svarsandel | Sverige | 2 380 | 26,6 |
| Tidigare svarsandel | Övriga Världen | 143 | 20,5 |
| Nytt medgivande | Sverige | 2 851 | 32,2 |
| Nytt medgivande | Övriga Världen | 182 | 26,8 |

Tabell 4. Designvägd svarsandel efter föräldrars utbildningsnivå

| Grupp | Kategori | Svar | Svarsandel, % |
|---------------------|----------------|-------|---------------|
| Tidigare svarsandel | Förgymnasial | 80 | 13,9 |
| Tidigare svarsandel | Gymnasial | 577 | 18,3 |
| Tidigare svarsandel | Eftergymnasial | 1 866 | 32,2 |
| Nytt medgivande | Förgymnasial | 103 | 18,3 |
| Nytt medgivande | Gymnasial | 694 | 22,5 |
| Nytt medgivande | Eftergymnasial | 2 236 | 38,8 |

Tabell 5. Svarsandel efter huvudman

| Grupp | Kategori | Antal svar | Svarsandel, % |
|---------------------|------------|------------|---------------|
| Tidigare svarsandel | Kommunal | 2 113 | 25,7 |
| Tidigare svarsandel | Fristående | 410 | 30,2 |
| Nytt medgivande | Kommunal | 2 546 | 31,3 |
| Nytt medgivande | Fristående | 487 | 35,8 |

Tabell 6. Designvägd svarsandel efter betygspoäng (årskurs 6)

| Grupp | Kategori | Antal svar | Svarsandel, % |
|---------------------|----------------|------------|---------------|
| Tidigare svarsandel | 0 - 35 | 327 | 13,0 |
| Tidigare svarsandel | 37.5 - 42.5 | 540 | 21,9 |
| Tidigare svarsandel | 45 - 50 | 828 | 32,8 |
| Tidigare svarsandel | 52.5 - 60 | 771 | 44,0 |
| Tidigare svarsandel | Uppgift saknas | 57 | 24,7 |
| Nytt medgivande | 0 - 35 | 446 | 18,1 |
| Nytt medgivande | 37.5 - 42.5 | 676 | 27,6 |
| Nytt medgivande | 45 - 50 | 966 | 38,4 |
| Nytt medgivande | 52.5 - 60 | 876 | 50,3 |
| Nytt medgivande | Uppgift saknas | 69 | 29,1 |

Tabell 7. Designvägd svarsandel efter meritvärde (årskurs 9)

| Grupp | Kategori | Antal svar | Svarsandel, % |
|-----------------|-------------------|------------|---------------|
| Nytt medgivande | Meritvärde saknas | 43 | 19,9 |
| Nytt medgivande | 0 - 197.5 | 355 | 15,3 |
| Nytt medgivande | 200 - 240 | 568 | 25,8 |
| Nytt medgivande | 242.5 - 275 | 844 | 38,6 |
| Nytt medgivande | 277.5 - 320 | 1 223 | 52,7 |

Svarsmönster från tidigare kvarstår efter insamlingen av medgivanden. Det är framför allt variablerna kön, föräldrarnas utbildningsnivå betygspoäng årskurs sex och meritvärde årskurs 9 som verkar ha ett samband med svarsbenägenheten. Exempelvis har pojkar svarat i lägre grad än flickor och elever med lägre betyg har svarat i lägre grad än elever med höga betyg. Detta talar för att dessa variabler borde användas som hjälpvariabler vid viktberäkningar.

Samvariation med målvariabler

Vi bedömer att alla hjälpvariabler som enligt analysen är relevanta också kommer att samvariera med många målvariabler. Därför gör vi ingen särskild analys av denna samvariation.

Avgränsning av redovisningsgrupper

Om hjälpvariabeln avgränsar viktiga redovisningsgrupper kan kvaliteten bli bättre i dessa grupper. Framförallt blir skattningarna säkrare om hjälpvariabeln väl avgränsar redovisningsgruppen enligt kriterium 3. Kalibreringsestimatoern ger konsistenta skattningar i den meningen att estimatoern ger exakta skattningar för utnyttjade registertotaler. Det är därför viktigt att vid val av hjälpvariabler fundera på vilka skattningar man vill få konsistenta.

Kön och födelseland har bedömts vara viktiga redovisningsgrupper och bör om möjligt användas vid viktberäkningar.

Slutligt val av hjälpvektor

Efter en sammanvägning av ovanstående analys samt efter kontroll av vikternas fördelning används följande hjälpvektor i kalibreringen:

*KÖN + FÖDELSELAND + FÖRÄLDRARS UTBILDNINGSNIVÅ +
MERITVÄRDE ÅRSKURS NIO*

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. Ofta är y (och även z) en dikotom variabel, d.v.s.

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

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

$$Y_d = \sum_U y_{dk} \quad (2)$$

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

Z_d bildas på likartat sätt.

En generell parameter för redovisningsgrupp d (d kan också avse hela populationen) kan skrivas $P_d = C \frac{Y_d}{Z_d}$, 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 (1). Om vi låter N_d vara antalet personer i redovisningsgrupp d , då kan parametern skrivas

$$P_d = 100 \frac{\sum_U y_{dk}}{N_d} \quad (3)$$

I denna undersökning drogs ett stratifierat pps-urval av skolenheter. Ett pps-urval innebär i detta fall att urvalssannolikheten sätts proportionell mot antal elever på skolenheten. På varje utvald skolenhet ingick sedan alla elever i aktuell årskurs i urvalet.

Idealt skulle hela urvalet av elever (urvalet av elever betecknas med s och har storleken n) användas vid skattningar. Men på grund av bortfall och övertäckning av elever har vi endast svarsmängden r av storleken m att utföra beräkningarna på.

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

$$\hat{Y}_d = \sum_r w_{1k} w_{2k} y_{dk} \quad (4)$$

där

$w_{1k} =$ designvikt för utvald skola (baserad på pps-urvalet)

$w_{2k} = \frac{\text{totalt antal elever i respektive skola/}}{\text{antal elever som svarat på enkäten i respektive skola}}$

I estimator (4) används ingen ytterligare hjälpinformation än stratifieringsinformationen. Bortfallskompensation görs inom varje skola.

I syfte att erhålla en estimator med mindre urvalsfel och bortfallsskevheter än estimator (4) utnyttjar vi hjälpinformation i estimationen. Vi bildar en hjälpvektor \mathbf{x}_k , som anger till vilka kategorier av

KÖN + FÖDELSELAND + FÖRÄLDRARS UTBILDNINGSNIVÅ + MERITVÄRDE ÅRSKURS NIO

som elev k tillhör. Från register framställer vi

hjälptotalerna $\sum_S \mathbf{x}_k$. Vi utnyttjar denna hjälpinformation i en kalibreringsestimator.

Kalibreringsestimatoren för totalen Y_d har följande utseende:

$$\hat{Y}_{wd} = \sum_r w_{1k} w_{2k} v_{2k} y_{dk} = \sum_r w_k y_{dk} \quad (5)$$

där

$$v_{2k} = 1 + (\sum_S \mathbf{x}_k - \sum_r w_{2k} \mathbf{x}_k)' (\sum_r w_{2k} \mathbf{x}_k \mathbf{x}_k')^{-1} \mathbf{x}_k \quad (6)$$

är en justeringsvikt som kalibrerar skattningarna mot kända antal elever i urvalet.

Vid skattning av en parameter av typen $P_d = C \frac{Y_d}{Z_d}$ skattas respektive total med hjälp av vikterna $w_k = w_{1k} w_{2k} v_{2k}$.

Referenser

Särndal C.-E. & Lundström S. (2005). *Estimation in Surveys with Nonresponse*. New York: Wiley