



Statistics Sweden

Statistiska centralbyrån

Teknisk Rapport

En beskrivning av genomförande och metoder

EP7 Enkät årskurs 9

2014-06-13

SCB, Stockholm
08-506 940 00

SCB, Örebro
019-17 60 00

www.scb.se



Inledning

Enheten för statistik om utbildning och arbete vid Statistiska centralbyrån (SCB) genomförde under perioden mars – maj 2014 en enkätundersökning på uppdrag av Göteborgs universitet.

Undersökningen riktade sig till elever som ingår i urvalet i elevpanel 7 och genomfördes inom ramen för samarbetet mellan SCB:s longitudinella elevpaneler och *Utvärdering genom uppföljning (UGU)* vid Göteborgs universitet. Resultatet från undersökningen kommer utgöra underlag till forskning inom framförallt det samhällsvetenskapliga området.

Populationen utgjordes av elever som gick i årskurs 3 då elevpanel 7 startades våren 2008, där majoriten av eleverna våren 2014 finns i årskurs 9. Urvalet bestod av 9 549 personer och det var totalt 4 573 personer som besvarade frågeblanketten, vilket var 48 procent av urvalet.

Resultatet i form av en datafil levererades vecka 24 enligt överenskommelse. Resultat på frågor med öppna svar ingick dock inte i leveransen, då hanteringen av dessa behöver utredas ytterligare innan dessa kan levereras. En kompletterande leverans av dessa beräknas därmed att göras inom kort.

Undersökningsledare på SCB var Anna Andersson och Johan Löfgren var metodstatistiker. Göteborgs universitets kontaktperson gentemot SCB var Michael Hansen.

Omfattning

Population och urval

Populationen, d.v.s. de objekt som man vill kunna dra slutsatser om, utgjordes av elever som gick i årskurs 3 då elevpanel 7 startades våren 2008, där majoriten av eleverna våren 2014 finns i årskurs 9.

För att kunna dra ett urval från populationen skapas en urvalsram som avgränsar, identifierar och möjliggör koppling till objekten i populationen. Urvalsramen bestod år 2008 av grundskolor (ej sameskolor) som hade elever i årskurs 2 under läsåret 2006/07. Skolorna hämtades från registret över skolor. För att avgöra vilka skolor som hade elever i årskurs 2 läsåret 2006/07 användes registret över elever i grundskolan. Båda dessa register håller SCB för Skolverkets räkning.

Målstorleken för elevurvalet till elevpanel 7 bestämdes till ca 10 000 elever. Ett stratifierat urval av skolor med elever i årskurs 3 drogs med urvalssannolikhet proportionell mot redovisat antal elever enligt det elevregister SCB håller för Skolverkets räkning, för läsåret 2007/2008.

Stratifiering gjordes efter typ av skolhuvudman:

1. Skolor med kommunen som huvudman
2. Skolor med annan huvudman/friskolor

Urvalssannolikheterna fastställdes så att 15 procent av urvalet skulle avse elever i friskolor. Eftersom endast ca 8 procent av samtliga elever i årskurs 3 gick i en friskola innebar det en väsentligt förhöjd urvalssannolikhet för friskolor.

Uteslutna från undersökningspopulationen var sameskolor, vilket handlade om fem skolor med totalt 19 elever. Skolor med skoldaghem tilläts ingå i urvalsramen med motiveringen att skoldaghem ofta bara är en del av verksamheten vid en skola och att urvalsramen endast innehöll uppgift om förekomst av skoldaghem. Urvalet som drogs år 2008 omfattade 300 av de totalt 3 888 skolor med elever i årskurs 3. För mer information om population, stratifiering samt ursprungligt urval hänvisas till dokumentationen över SCB:s longitudinella elevpaneler, vilken finns [här](#).

Urvalet bestod i denna enkätundersökning av samtliga 9 549 elever som kvarstod från elevpanelens ursprungliga urval. Vid de insamlingar som gjorts av administrativa data mellan åren 2008-2013 har ett antal elever utgått ur undersökningen. Anledning till detta har bl.a. varit överflyttning till särskola och vägran. Dessa elever exkluderades även från denna undersökning.

För att få fram aktuella adressuppgifter till eleverna genomfördes en identifikationskontroll av personerna i urvalet mot befolkningsdata. Vid kontrollen visade det sig att 84 elever inte längre tillhörde populationen utan utgjorde s.k. övertäckning. Samtliga elever hade flyttat utomlands.

Frågor/Variabler

Göteborgs universitet utformade frågorna i frågeblanketten. Därefter genomfördes en snabbgranskning av mätteknisk expertis i syfte att minska risken för mätfel. Grundläggande för bra kvalitet i en undersökning är kvaliteten på de data som samlas in. För att säkerställa att frågorna fungerar så bra som möjligt och enligt intentionerna genomgår därför alla blanketter ett mättekniskt test.

Blanketten bestod av 27 numrerade frågor, flera av dem hade delfrågor vilket genererade totalt 188 antal frågor.

Förutom de variabler som samlades in via frågeblanketten hämtades följande variabler från de elevregister för grundskola respektive gymnasieskola som SCB håller för Skolverkets räkning:

- Årskurs den 15 oktober 2013

Datafilen innehåller även uppgift om elevens UGU-kod, vilket innebär att svaren i och med detta kan kopplas ihop med uppgifter i den databas som hålls inom projektet "Utvärdering genom uppföljning".

För referensperiod för enkätfrågor, se bifogad frågeblankett. Med referensperiod menas vilken tidpunkt svaren avser, t.ex. idag, förra veckan, senaste tre månaderna.

Datainsamling

Frågeblanketterna skickades ut med post. I ett informationsbrev ombads personen att besvara frågorna och skicka svaret till SCB. Möjlighet fanns även att besvara enkäten via webb.

Det första utskicket genomfördes den 6 mars 2014. Sedan skickades ett tack- och påminnelsekort (ToP-kort) samt två påminnelser till dem som inte besvarat frågeblanketten. ToP-kortet skickades den 19 mars. En första påminnelse med ny enkät skickades den 2 april och en andra påminnelse med ny enkät den 16 april. Insamlingen avslutades den 12 maj 2014.

Tabell 1 Beskrivning av inflödet. Antal och andel

	Antal	Andel
Efter första utskick	1 994	20,9
Efter ToP-kort	1 047	11,0
Efter enkätpåminnelse 1	814	8,5
Efter enkätpåminnelse 2	718	7,5
Totalt	4 573	47,9
Bortfall	4 884	51,1
Övertäckning	92	1,0
Urval	9 549	100,0

I informationsbrevet kunde uppgiftslämnarna läsa om undersökningens bakgrund, syfte och att undersökningen genomfördes i samarbete mellan Göteborgs universitet och SCB. Informationsbrevet informerade också om att uppgifter kommer att kopplas på från andra register och att en avidentifierad datafil skulle levereras till Göteborgs universitet. Brevet informerade även om personuppgiftslagen samt offentlighets- och sekretesslagen och att det var frivilligt att medverka i undersökningen.

För att SCB ska kunna lämna ut data från en enkätundersökning krävs informerat samtycke av uppgiftslämnarna. Det innebär att de genom att besvara blanketten och skicka in den godkänner att deras svar kompletteras

med de bakgrundsvariabler, och behandlas på det sätt, som beskrivs i informationsbrevet.

Datansamlingen genomfördes av enkätenheten vid SCB.

Datansamlingen genomfördes via webb samt skanning av de frågeblanketter som kommit in via post.

SCB kan inte garantera att den utvalda personen själv besvarat frågeblanketten. Någon kontroll av att rätt person har besvarat frågeblanketten har inte gjorts.

Svarsdatafilen kompletterades med vikter för uppräkningsnivå.

Bortfall

Bortfallet består dels av objektsbortfall, 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, så kan skattningarna som grundar sig på enbart de svarande vara skeva. För att reducera bortfallsskevheten har vikter beräknats med hjälp av kalibrering (se bilaga).

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	4 760
Postreturer	66
Avböjd medverkan	22
Förhindrad medverkan	9
Ej anträffad	8
Insänd blank/oanvändbar	13
Inkommen för sent	6
Totalt	4 884

Med "Ej avhörda" menas att ingen uppgift om varför frågeblanketten/intervjun inte är besvarad har lämnats. Med "Avböjd medverkan" menas att SCB meddelats att uppgiftslämnaren inte vill medverka i undersökningen. "Förhindrad medverkan" innehåller förhindrad p.g.a. sjukdom,

institution samt övrigt. "Ej anträffad" innehåller saknad adress i Registret över totalbefolkningen (RTB) samt skyddad adress.

Partiellt bortfall kan bero på att en fråga är svår att förstå, är känslig, att uppgiftslämnaren glömmer att besvara frågan, att han/hon kryssar på fel rad eller att instruktionerna vid hoppfrågor misstolkas. Det partiella bortfallet varierar mellan 0,5 och 2 procent för de flesta av frågorna. De delfrågor som innehåller alternativet "annat" har alla ett högt partiellt bortfall, vilket kan antas bero på att endast de som har en annan anledning/orsak eller dylikt än det som räknats upp besvarat denna fråga.

Bortsett från dessa frågor finns det högsta partiella bortfallet, mellan 9,5-10,6 procent, i fråga 2. Några kommentarer tyder på att en anledning till detta kan vara att endast de som upplever sig ha gjort ett val av grundskola besvarar dessa frågor. Utöver dessa frågor finns det bara två som överstiger 2 procent i partiellt bortfall. Detta är dels fråga 8h, med 2,3 procent, där en del kommentarer vid frågan tyder på att större arbeten eller projekt i dessa fall inte har förekommit i matematik. Detta var även något som lyftes fram i rapporten från den mättekniska granskningen.

Den andra frågan med ett partiellt bortfall på över 2 procent är fråga 15b, med 2,7 procent och här tyder en del kommentarer på att man i dessa fall ännu inte upplever sig veta hur de nationella proven påverkar betygen.

I den mättekniska rapporten lyftes det även fram att ett sätt att minska risken för felaktiga svar (i den mån det går) och partiella bortfall (på frågor som ul inte kan välja ett befintligt svarsalternativ på) så är en lösning att erbjuda "Vet inte"-alternativet. I flera fall där uppgiftslämnare hoppat över frågor sker detta i kombination med att man skrivit "?", "Vet inte" eller dylikt vid frågan, vilket kan tyda på att man ibland saknar detta alternativ.

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 (CLAN).

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

v_k = kalibreringsvikt 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 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 bygger på antagandet att ramen återspeglar populationen väl och därmed att över- och undertäckningen är försumbar.

För en mer utförligare beskrivning av kalibreringsestimatern 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 skattningen 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)

Och för beräkning av skattningen 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).

Statistikens tillförlitlighet

Ramtäckning

Täckningsfel, under- och övertäckning, innebär att urvalsram och population inte helt stämmer överens. Undertäckning innebär att vissa objekt som ingår i populationen saknas i urvalsramen. Övertäckning innebär att objekt

som inte ingår i populationen ändå finns i urvalsramen. Ett sätt att minska täckningsfelen är att ha bra och uppdaterade register.

Risken för täckningsfel orsakade av brister i ramens aktualitet eliminerades genom att det slutliga elevurvalet i samband med den första uppgiftsinsamlingen 2008 bestämdes utifrån SCB:s elevregister för de utvalda skolorna avseende läsåret 2007/08.

Inför varje insamlingsomgång inom elevpanelerna görs även en kontroll gentemot SCB:s register över totalbefolkningen (RTB) och individer som t.ex. avlidit hanteras som övertäckning. Så har även skett i de fall uppgift under insamlingens gång lämnats om att en individ i urvalet t.ex. gått över till särskola. Fram till och med år 2013 hade 87 st. individer utgått från insamlingarna på grund av sådana skäl.

Urval

Denna kvalitetskomponent avser fel som uppkommer på grund av att endast ett urval av populationen undersöks. Urvalsfel är således den avvikelse mellan ett skattat värde och det faktiska värdet som beror på att man inte undersöker alla objekt i populationen. Urvalsfelets storlek minskar med en ökad urvalsstorlek.

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.

Vissa kommentarer på enkäten vid några av frågorna tyder på att det ibland kanske skulle vara bra att ha med ytterligare förklaringar. I fråga 19 undrar exempelvis en del över ordet "begripligt" i delfråga j) och vid delfråga l) undrar en del vad som menas. Vad gäller delfråga 26d undrar flera vad magont omfattar (om t.ex. mensvärk ska räknas med).

Det finns även vissa frågor som genererat fler dubbelmarkeringar, d.v.s. fall där flera svarsalternativ markerats där endast ett efterfrågas, än övriga. En av dessa är fråga 22, vilken diskuteras mer nedan. Andra frågor är fråga 12 samt fråga 25, där det även i den mättekniska rapporten lyftes fram att det blir glapp mellan skalstegen i dessa frågor. På så vis blir det svårt för uppgiftslämnarna att veta vilket alternativ de ska välja.

Bearbetning

Vid den manuella och maskinella bearbetningen av datamaterialet kan bearbetningsfel uppstå. Exempel på bearbetningsfel är registreringsfel och kodningsfel. Dessa fel kan förhindras och upptäckas i de kontroller som genomförs vid dataregistreringen. I den här undersökningen bedöms regi-

streringsfelet vara litet eftersom frågeblanketten för de flesta frågorna hade fasta svarsalternativ.

Bortfall

Bortfallsfel inträffar om objekten i bortfallet och de svarande skiljer sig åt avseende frågorna/variablerna i undersökningen. En beskrivning av bortfallet finns i den bortfallsanalys som gjorts i samband med valet av hjälpinformation till kalibreringen (se bilaga).

Beskrivning av datafil

Datafil

Observera att de bifogade vikterna ska användas vid analys och resultatframställning. I ett stratifierat urval kan datamaterialet inte ses som ett urval av oberoende lika fördelade observationer, som ofta antas i traditionell statistisk teori. Det är därför viktigt att tänka på att de flesta förprogrammerade dataprogram inte klarar av att analysera datamaterialet ifrån en urvalsundersökning på ett korrekt sätt. Var noga med att kontrollera att de dataprogram som används för analys behandlar vikterna på ett riktigt sätt. Att ta fram frekvenser med beaktande av vikter klarar däremot de flesta statistikprogram.

Om det är partiellt bortfall (uppgift saknas) i en fråga har värdet 88 angetts. Vid hoppfrågor har värdet 0 angetts för dem som inte ska besvara följdfrågorna. Vid fråga 13_2, avseende årskurs man tycker att man ska få betyg ifrån, används dock värde 99 för de som inte svarat ja på frågan om man ska få betyg i grundskolan, detta då värde "0" förekommer som en årskurs.

En excelfil bifogas vilken visar vilka dubbelmarkeringar som förekommit samt hur de har hanterats. I de allra flesta frågor slumpas värdet om de svarsalternativ som markerats är närliggande och sätts till 88 om de ej är det. Om "Spelar ingen roll" eller "Jag har inte bestämt mig än" ingår som ett alternativ prioriteras dock säkra svarsalternativ före det osäkra.

Vissa frågor har dock hanterats på lite annorlunda sätt än majoriteten och en av dessa är fråga 13. Här hade 287 st angett en årskurs i 13_2, men hade inte kryssat i 13_1 (om man anser att man ska få betyg i grundskolan). Dessa har fått 13_1=1. Övriga som saknade värde på 13_1 och som inte angett en årskurs har fått 88 på både 13_1 och 13_2. Även de som kryssat Ja på 13_1 men saknar värde på 13_2 har fått värde 88 på 13_2. De som kryssat "Nej" eller "Spelar ingen roll" på 13_1 har dock fått 99 på 13_2.

I 7 st fall med dubbelmarkeringar på 13_1, där svarsalternativ 1 ingår, har värdet satts till 1 då en årskurs var ifylld. Det fanns även 18 st fall där 13_2 var besvarad med en årskurs trots att man kryssat "Nej" eller "Spelar ingen

roll" på 13_1. I och med att endast de som valde "Ja" på fråga 13_1 skulle besvara fråga 13_2 har för dessa fall den angivna årskursen ersatts med 99.

I åtta fall där svar på fråga 22 saknades och ingen av fråga 23-24 var ifyllda, men däremot någon av delfrågorna inom fråga 25, har svar på fråga 22 satts till svarsalternativ 4 och värde på 23-24 till 0. I övriga fall där svar saknades på fråga 22 har detta och svar på eventuellt efterföljande obesvarade frågor satts till 88. Det fanns även 14 st som valt svarsalternativ 4 på fråga 22 och ändå fyllt i någon fråga av 23 och 24. Dessa har nu fått 0 istället som svar på 23-24. Eventuella kommentarer i textfälten kvarstår dock.

Statistiska mått

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

Jämförbarhet och användbarhet

Jämförbarhet över tid

SCB har på uppdrag av Göteborgs universitet genomfört snarlika undersökningar tidigare. Senast gjordes detta år 2008 och avsåg då elevpanel 6.

Undersökningarna år 2014 och 2008 genomfördes på i stort sett samma sätt med avseende på urvalsmetod, kalibrering och frågeformulär. År 2014 var dock första gången som eleverna även hade möjlighet att besvara enkäten via webben.

Detta har exempelvis fått konsekvenser för fråga 22 där det frågas om gymnasieskolan eleven valt. Om eleven besvarade enkäten via papper kunde flera alternativ väljas, medan endast ett kunde väljas om eleven svarade via webben. Att denna fråga genererat flest dubbelmarkeringar av alla frågor tyder på att det inte är ovanligt att flera skolor valts. Ett förslag inför framtiden kan vara att antingen göra om fråga 22 till en fråga som tillåter att flera svarsalternativ kryssas i och då tydligt instruera om detta, alternativt att fråga alla elever enbart efter förstahandsvalet.

Även i fråga 13 finns en skillnad mellan de som besvarade undersökningen på papper och de som besvarade den på webben. De som besvarade den på webben valde från en lista som innehöll årskurs 1-9. I pappersblanketten var frågan öppen, vilket innebär att vi därifrån fått in svar som inkluderar årskursintervall samt årskurs 0. Även detta bör ses över inför framtida undersökningar.

Både för fråga 13 och fråga 22 kvarstår dubbelmarkeringar i datafilen.

En annan skillnad mellan 2014 års undersökning och 2008 års undersökning är att denna gång inkluderades i påminnelserna information om att alla elever skulle besvara enkäten, inte enbart de som gick i årskurs 9, trots att informationsbrevet adresserades till dessa. Tidigare år verkar ingen så-

dan information ha inkluderats i påminnelserna och elever verkar även ha exkluderats som övertäckning i de fall de meddelat att de ej fanns i årskurs 9. På så vis skiljer sig populationerna mellan undersökningsomgångarna något åt, men skillnaden bedöms vara liten. Utifrån registeruppgifter kan vi se att runt 96-98 procent av eleverna våren 2014 bör finnas i årskurs 9.

Det går även att se att bland de elever som finns i elevregistren för grundskolan respektive gymnasieskolan för hösten 2014 är eleverna i årskurs 9 mer benägna att besvara enkäten än vad elever i övriga årskurser är. Runt 49 procent av eleverna i årskurs 9 har besvarat enkäten jämfört med runt 27 procent av eleverna som finns i övriga årskurser.

Sekretess och utlämnande

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

Dessutom har en sekretessöverenskommelse upprättats mellan Göteborgs universitet och SCB med avseende på hur det avidentifierade datamaterialet får hanteras.

Bilagor

Frågeblankett och informationsbrev

Kalibreringsrapport

Dokumentation över dubbelmarkeringar