

# Teknisk Rapport

En beskrivning av genomförande och metoder

Gymnasieenkät till UGU 2017

2018-01-22



Statistics Sweden

Statistiska centralbyrån

010-479 40 00

[www.scb.se](http://www.scb.se)



## Inledning

Enheten för statistik om utbildning och arbete vid Statistiska centralbyrån (SCB) genomförde under perioden mars – augusti 2017 en enkätundersökning på uppdrag av Göteborgs universitet.

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

Enkätundersökningen ingår i forskningsprojektet "Utvärdering genom uppföljning" (UGU) vid Göteborgs universitet. Undersökningen är ett led i den nationella utvärderingen av skolan och utgör underlag till forskning inom framförallt det samhällsvetenskapliga området.

Populationen bestod av individer som gick i grundskolans årskurs 3 läsåret 2007/08. Våren 2017 gick majoriteten av dem i gymnasieskolans årskurs 3.

Utskicket gjordes till 9 532 personer och det var totalt 3 552 personer som besvarade frågeblanketten, vilket var 37 procent.

Resultatet har delats upp i två olika leveranser. Den första leveransen av resultat i form av datafil levererades vecka 4. I den leveransen ingick alla variabler som ej bedömts som känsliga, vilket innebär att fråga 12L, 19A och alla delfrågor inom fråga 20 har exkluderats. Leveransen av de exkluderade variablerna sker sedan när beslut från etikprövningsnämnden finns under förutsättning att Göteborgs universitet ges stöd för behandlingen.

Undersökningsledare på SCB var Anna Andersson och Johan Löfgren var metodstatistiker. Göteborgs universitet 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 personer som gick i grundskolans årskurs 3 läsåret 2007/08, som inte varit inskrivna i särskola någon gång i årskurs 3-9 och som bodde kvar i Sverige vid undersökningstillfället. Våren 2017 gick majoriteten av dem i gymnasieskolan.

För att kunna dra ett urval från populationen skapades 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 7. Urvalet drogs 2008. 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. Det slutliga elevurvalet bestämdes utifrån klasslistor för de utvalda skolorna i samband med den första uppgiftsinsamlingen. Ett stratifierat urval av skolor med elever i årskurs 3 drogs med urvalssannolikhet proportionell mot redovisat antal elever. För mer information om population, stratifiering samt ursprungligt urval hänvisas till dokumentationen över SCB:s longitudinella elevpaneler, vilken finns [här](#).

I urvalet som drogs år 2008 ingick totalt 9 671 elever. Elever som vägrat medverka i Elevpaneler för longitudinella studier exkluderades från denna insamling, liksom elever som där kodats bort som övertäckning eller på grund av skyddad identitet.

Aktuella adresser hämtas från Registret över totalbefolkningen (RTB). Individer som ej är folkbokförda i Sverige alternativt inte har ett giltigt svenskt personnummer ingår därmed ej i insamlingen.

### **Frågor/Variabler**

Göteborgs universitet utformade frågorna i frågeblanketten i samarbete med SCB.

Därefter genomfördes en mätteknisk standardgranskning 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 har därför blanketten genomgått ett mättekniskt test.

Blanketten bestod av 27 numrerade frågor, flera av dem hade delfrågor vilket genererade totalt 138 frågor.

Förutom de variabler som samlades in via frågeblanketten hämtades följande variabler från Skolverkets elevregister:

- Gymnasieprogram (sökta/registrerade)
- Skolenhetskod löpnummer (sökta/ registrerade)
- Huvudman (sökta/registrerade)

Utöver det hämtas Studiedeltagande HT avseende år 2015 från LISA och UGU-kod från registret över *Elevpaneler för longitudinella studier*.

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

## Datainsamling

Frågeblanketterna skickades ut med post. I ett informationsbrev ombads personen att besvara frågorna och skicka svaret till SCB. Det första utskicket genomfördes den 7 mars 2017. Sedan skickades tre påminnelser till dem som inte besvarat frågeblanketten. Ett tack- och påminnelsekort (ToP-kort) skickades den 22 mars. En påminnelse med ny enkät skickades den 10 april och ett andra ToP-kort den 17 maj. Insamlingen avslutades den 24 augusti 2017.

**Tabell 1 Beskrivning av inflödet. Antal och andel**

	Antal	Andel
Urval	9 671	
<i>varav utskick till</i>	9 532	
Efter första utskick	1 875	
Efter ToP nr 1	797	
Efter påminnelse	664	
Efter ToP nr 2	216	
Totalt svarande	3 552	37,4
Bortfall	5 895	62,6
Övertäckning	224	

I informationsbrevet beskrevs bakgrund, syfte och att undersökningen genomfördes i samarbete mellan Göteborgs universitet och SCB. Informationsbrevet informerade också om att uppgifter hämtades från register och att en aidentifierad datafil levererades 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.

Datainsamlingen genomfördes av enkätenheten vid SCB.

De enkäter som kommit in via post registrerades med hjälp av skanning.

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

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 bortfallsrisken 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 bland de individer som utskick skedde till i denna undersökning redovisas i tabell 2 nedan.

**Tabell 2 Beskrivning av objektsbortfall**

	An- tal
Ej avhörda	5769
Postreturer, ej anträffad	70
Avböjd medverkan, frivillighet	11
Insänd blank, oanvändbar	3
Förhindrad medverkan	1
Exkluderade från utskick i undersökningen p.g.a. uppgift från Elevpaneler om skyddad identitet, vägran, ogiltigt personnr etc.	41
<b>Totalt</b>	<b>5895</b>

Med "Ej avhörda" menas att ingen uppgift om varför frågeblanketten inte är besvarad har lämnats. Med "Avböjd medverkan" menas att SCB meddelats att uppgiftslämnaren av någon anledning inte vill medverka i undersökningen.

Partiellt bortfall kan bero på att en fråga är svår att förstå, är känslig eller att uppgiftslämnaren glömmer att besvara frågan. Det partiella bortfallet ligger under 2 procent för de flesta av frågorna. Det högsta partiella bortfallet, omkring 5 procent, finns i fråga 23 som handlar om varför man är intresserad av att läsa på universitet eller högskola.

## 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 korrigerar för den övertäckning som identifierats. Ingen korrigering för eventuell undertäckning har gjorts.

För en mer utförligare beskrivning av kalibreringsestimatoren 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 objektsbortfallet 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 reducerades 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 gjordes sedan en kontroll gentemot SCB:s register över totalbefolkningen (RTB) och individer som t.ex. avlidit hanterades som övertäckning. Så skedde även i de fall uppgift under insamlingens gång lämnats om att en individ i urvalet t.ex. gått över till särskola.

I denna undersökning har totalt 224 personer kodats bort som övertäckningsobjekt. Den vanligaste orsaken till detta var att eleven inte befann sig i Sverige vid undersökningstillfället. Den näst vanligaste orsaken var att eleven redan under insamlingen för *Elevpaneler för longitudinella studier* kodats bort då denne varit inskriven i särskola.

## 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. Urvalet till denna undersökning hämtades från registret *Elevpaneler för longitudinella studier*.

## 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.

## 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 registreringsfelet vara litet eftersom frågeblanketten endast 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.



Totalt 655 st av dem som besvarat fråga 1 med svarsalternativ 1-3 hade dock ändå besvarat fråga 2. Av dem hade alla utom två besvarat frågan med "Ja" och dessa har fått värde 0 på fråga 2. De övriga 2 st hade dock besvarat fråga 2 med Nej och för dem har det svaret fått kvarstå.

Fyra av dem som svarade nej på fråga 2 har ändå besvarat frågor mellan 3-16. Dessa svar har fått kvarstå.

Totalt 11 st som besvarat fråga 6 med nej har ändå besvarat fråga 7. Av dem hade två enbart skrivit en kommentar om att de inte bytt. Dessa har fått 0 på fråga 7 men har kvar sitt svar på F7\_txt. För de övriga 9 st kvarstår svaret på fråga 7.

De som lämnat kommentarer i fråga 7 eller fråga 16 utan att kryssa i den aktuella rutan för detta har fått en 1:a som värde i den fråga som avser kommentaren.

Totalt 6 st hade besvarat fråga 22 med nej men ändå besvarat fråga 23 eller 24. Deras svar har fått kvarstå.

Totalt 139 st hade besvarat fråga 22 med alternativ 1-4 men ändå besvarat fråga 25. Deras svar har fått kvarstå. Av dessa hade majoriteten, 118 st, besvarat fråga 22 med alternativ 4, att de inte hade bestämt sig.

Ett par saker att ta med sig till framtida enkäter skulle kunna vara att det i fråga 3 bland det partiella bortfallet finns dem som skrivit att de läser på en skola som har landstinget som huvudman. I fråga 6 finns det de som kommenterat att man bytt skola, men inte program.

### **Statistiska mått**

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

### **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 enkäten.

## **Jämförbarhet**

### **Jämförbarhet över tid**

Vid flera tillfällen har SCB på uppdrag av Göteborgs universitet genomfört liknande undersökningar. Senast var år 2012.

Undersökningen år 2012 genomfördes på liknande sätt med avseende på urvalsmetod och urvalsstorlek. Då var dock insamlingsmetoden enbart webb. Frågorna skiljer sig också åt en del och 2012 genomfördes undersökningen ett år efter att de flesta elever gått ut gymnasieskolan medan det denna gång var i årskurs 3.

## **Sekretess och utlämnande**

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

## **Bilagor**

Frågeblankett och informationsbrev

Kalibreringsrapport