

Teknisk rapport

Akademisk frihet i praktiken

Inledning

Enheten för utbildning och arbetsmarknadsstatistik vid Statistiska centralbyrån (SCB) genomförde under perioden mars-maj 2005 en enkätundersökning på uppdrag av Höskoleverket och Statsvetenskapliga institutionen vid Uppsala universitet.

Bakgrunden till undersökningen är de stora förändringar som högskolor och universitet genomgått under de senaste 15 åren: nya uppgifter har tillkommit och antalet studenter har vuxit kraftigt.

Undersökningens syfte var att kartlägga vad dagens studenter anser bör vara vägledande för den högre utbildningen, samt hur de betraktar förutsättningarna att arbeta på ett sätt som motsvarar detta ideal.

Undersökningen genomfördes som en postenkät med två påminnelser. Målpopulationen utgjordes av personer som studerade på universitet och högskolor i Sverige vårterminen 2005. Totalt var det 1 867 personer som besvarade frågeblanketten, vilket var 60 procent av urvalet. 372 av de svarande var dock personer som inte längre studerade på högskolan. Denna s.k. svarande övertäckning var en följd av att det saknades ett uppdaterat register att dra ett urval från. Frågeblanketterna registrerades med skanning. Från det skannade materialet framställdes en fil.

Harald Theorin var produktansvarig. Anna-Liisa Halunen arbetade med registrering och skanning av frågeblanketterna.

Population och urval

Målpopulationen, d.v.s. de individer som man vill kunna dra slutsatser om, utgjordes av personer som studerade på svenskt universitet eller svensk högskola under vårterminen 2005.

För att kunna dra ett urval från populationen skapades en urvalsram som avgränsade, identifierade och möjliggjorde koppling till objekten i populationen. Urvalsramen i undersökningen var Universitets- och högskoleregistret för höstterminen 2004. Detta var ingen idealisk urvalsram eftersom det är mycket svårt att förutsäga vilka som är kvar i studier från höstterminen till vårterminen, och för att det saknades många som börjat eller återinträtt i studier under vårterminen. Det var dock den bästa urvalsram som gick att få tag i utan att göra en tidskrävande insamling från högskolorna. Eftersom personer som studerar i liten omfattning har lägre sannolikhet att

vara kvar, från en termin till nästa, exkluderades personer som läste mindre än 5 poäng under höstterminen ur urvalet. Med facit i hand skulle gränsen kanske ha lagts än högre för att få en effektivare urvalsram. I den slutliga urvalsramen ingick 317 932 personer.

Högskoleverket beslutade efter diskussion med SCB att urvalsstorleken skulle vara 3000 individer. Från urvalsramen drogs ett stratifierat obundet slumpmässigt urval om 3127 individer med hjälp av ett av SCB egenutvecklat urvalsprogram. Ett stratifierat obundet slumpmässigt urval innebär att alla objekt inom respektive stratum har samma sannolikhet att komma med i urvalet.

För att urvalet skulle omfatta individer med olika studieinriktning definierades dessa som egna strata. Grupperingen av inriktningar var mer aggregerad än i enkätens motsvarande fråga, men eftersom inriktningen under vårterminen 2005 ändå var okänd ska grupperingen endast ses som en hjälp för att inte få ett alltför snedfördelat urval. Inriktningen grupperades enligt svensk utbildningsnomenklatur. I stratifieringen ingick även kön. Sedan kunde stickprovsstorleken fördelas till dessa stratum på ett sätt som gjorde att urvalet innehöll ett lämpligt antal personer från de olika strata. I de förhållandevis homogena grupperna lärarutbildning, juristutbildning och konstnärlig utbildning drogs ett lite mindre urval än i de övriga grupperna.

Aktuella adressuppgifter hämtades från befolkningsstatistiken.

Tabell 1 Rampopulation och urval

	Population	Urval
TOTALT	317932	3127
Läroutbildning, män	12864	139
Läroutbildning, kvinnor	41439	139
Konst och media, män	4326	140
Konst och media, kvinnor	5726	140
Humaniora och språk, män	10576	229
Humaniora och språk, kvinnor	19786	225
Samhällsvetenskap, män	31123	229
Samhällsvetenskap, kvinnor	48007	229
Juridik, män	4481	140
Juridik, kvinnor	6770	140
Naturvetenskap, män	16009	230
Naturvetenskap, kvinnor	12450	229
Teknik, män	37705	230
Teknik, kvinnor	14256	229
Vård och omsorg samt medicin, män	8989	229
Vård och omsorg samt medicin, kvinnor	43425	230

Sekretess och utlämnande

Med frågeblanketten följde ett informationsbrev där uppgiftslämnarna kunde läsa om undersökningens bakgrund, syfte och att undersökningen

genomfördes i samarbete mellan SCB, Högskoleverket och Uppsala universitet. I brevet stod också att samtliga uppgifter skyddas av Personuppgiftslagen och Sekretesslagen, att det var frivilligt att delta och att materialet skulle lämnas aidentifierat till Högskoleverket.

För att SCB ska kunna lämna ut ett enkätmaterial krävs ”informerat samtycke” av uppgiftslämnarna. Det innebär att de genom att besvara blanketten och skicka in den godkänner att deras svar behandlas på det sätt som beskrivs i informationsbrevet.

SCB har gjort en intern sekretessprövning för utlämnandet av enkätfrågorna. Dessutom har en sekretessöverenskommelse upprättats mellan Uppsala universitet och SCB med avseende på hur det aidentifierade datamaterialet får hanteras.

Behandlingen av personuppgifter i undersökningen har anmälts till SCB:s juridiska kansli. Beslutet däriifrån medger endast att registret (datamaterialet) bevaras på SCB till 2005-08-20 då uppgifterna kommer att förstöras.

Variabler

Uppsala universitet har utformat frågorna i frågeblanketten i samarbete med SCB. Blanketten bestod av 37 numrerade frågor. Flera av dem hade delfrågor vilket genererade omkring 100 frågor. Frågorna anknöt till undersökningens syfte.

Datainsamling

Frågeblanketterna sändes till urvalspersonerna med post. I informationsbrevet ombads de besvara frågorna och skicka tillbaka frågeblanketten till SCB. Det första utskicket genomfördes den 31 mars. Två påminnelser skickades ut till dem som inte besvarat frågeblanketten. Båda påminnelserna, den 13 april och den 26 april var enkätpåminnelser. Insamlingen avslutades den 13 maj.

Tabell 2a Beskrivning av inflödet

	Antal	Procent
Inför enkätpåminnelse 1	908	29,0
Inför enkätpåminnelse 2	1565	50,0
Totalt antal svar	1867	59,7
Bortfall	1260	40,3
Urval	3127	100,0

Tabell 2b Beskrivning av inflödet per kön

Kön	Urval	Antal svar	Andel svar*
Män	1566	851	54,5
Kvinnor	1561	1011	65,1

Tabell 2c Beskrivning av inflödet per åldersgrupp

Ålder	Urval	Antal svar	Andel svar*
Upp till 24 år	1155	674	60,3
25-29 år	892	519	58,5
30-39 år	637	376	60,1
40-49 år	287	193	67,6
50 år eller äldre	156	105	68,0

Tabell 2d Beskrivning av inflödet per inriktning

Inriktning	Urval	Antal svar	Andel svar*
Läroutbildning	278	165	62,9
Konst och media	280	147	54,0
Humaniora och språk	454	273	60,9
Samhällsvetenskap	458	283	62,2
Juridik	280	172	62,2
Naturvetenskap	459	275	58,9
Teknik	459	257	53,7
Vård, omsorg, medicin	459	295	66,3

Tabell 2e Beskrivning av inflödet per kurstyp

Kurstyp	Urval	Antal svar	Andel svar*
Yrkesprogram	1173	703	61,5
Övriga program	656	380	58,8
Fristående kurs	784	784	61,6

Tabell 2f Beskrivning av inflödet per poäng ht 2004

Poäng	Urval	Antal svar	Andel svar*
Mindre än 10	333	197	59,1
10-19	631	359	56,9
20 poäng	1612	973	61,9
21-30 poäng	445	283	65,0
Mer än 30 poäng	106	55	54,7

*Observera att andelen svar är vägt med urvalssannolikheten, d.v.s. det är inte den enkla kvoten mellan svar och urval.

Bortfall

Bortfallet består dels av objektbortfall, som innebär att frågeblanketten inte är besvarad alls, och av partiellt bortfall som innebär att vissa frågor på 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. Skevheten reduceras för de variabler som har används i stratifieringen, kön och inriktning. Inriktning är dock inte helt aktuell längre, så den bortfallsreducerande effekten blir mindre än för kön. För bakgrundsvariabler som inte används i stratifieringen kvarstår bortfallsskevheten, där svarsfrekvensen skiljer sig mycket mellan grupper (t.ex. ålder, se ovan). De skattade antalen som fås vid analys bör därför behandlas med försiktighet – inte bara p.g.a. bortfall utan även för att

målpopulationens storlek är okänd. Försiktighet kan vidtas genom att främst baserera analysen på jämförelse av andelar mellan olika grupper.

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 3 nedan. Med "ej avhörda" menas att ingen uppgift om varför frågeblanketten inte är besvarad har lämnats.

Tabell 3 Beskrivning av objektsbortfall

	Antal
Ej avhörda	1232
Postreturer	17
Avböjd medverkan	11
Totalt	1260

Partiellt bortfall kan bero på att en fråga är svår att förstå, är känslig, att uppgiftslämnaren glömmer att besvara frågan eller att instruktionerna vid hoppfrågor misstolkas. Det partiella bortfallet varierade mellan olika frågor, men var i de flesta fall ganska lågt. I fråga 13 ("tidigare högskolestudier") var dock det partiella bortfallet mycket högt. Majoriteten av dem som inte fyllde i något alternativ på frågan läste på program, och många hade inte studerat så länge. Sannolikt har en stor del av dem inte läst något annat än de gjorde vid insamlingen av undersökningen (som mäts i fråga 4a), och därför inte fyllt i.

Databeredning

Databeredningen gjordes av Enheten för utbildning och arbetsmarknadsstatistik vid SCB. Samtliga inkomna frågeblanketter prickades av i avregistreringssystemet RODE. Frågeblanketterna skannades med programvaran EYES&HANDS.

Kontroller genomfördes under och efter registreringen. Då kontrollerades bland annat att endast valida värden förekom i materialet.

Svarsdatafilen kompletterades med vikter för uppräkningsnivå.

Viktberäkning och estimation

Vikter har tagits så att resultat kan redovisas för hela populationen och inte bara för de svarande. Om vikterna inte används så kan resultaten bli helt missvisande. Vikterna kompenserar för objektsbortfallet men inte för det partiella bortfallet. Eftersom urvalet är stratifierat så har 16 vikter beräknats,

en för respektive stratum. Vikterna kompenserar därmed för att urvals-
personerna i olika strata har olika stor sannolikhet att komma med i urvalet.

Vikterna för respektive stratum har beräknats med formeln:

$$w_k = \frac{N_h}{n_h} \frac{n_h}{m_h} = \frac{N_h}{m_h}$$

där w_k = Vikten för varje objekt i stratum h ($h=1, 2, 3, \dots$)
 N_h = Antal individer i stratum h
 n_h = Antal individer i bruttourvalet i stratum h
 m_h = Antal individer som svarat samt övertäckning i stratum h

Vikterna bygger på antagandet att inom respektive stratum är övertäckningen
lika stor bland de svarande som i bortfallet och att bortfallet inte skiljer sig
från de svarande med avseende på undersökningsvariablerna.

För beräkning av skattning för totaler används följande formel:

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

där w_k = den totala vikten för objekt k
 y_k = variabelvärde för objekt k
summering sker över de svarande (r)

Och för beräkning av skattning för medelvärden används följande formel:

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

där w_k = den totala vikten för objekt k
 y_k = variabelvärde för objekt k
summering sker över de svarande (r)

Vikterna som finns på filen ska användas vid framställning av tabeller och
diagram samt vid beräkning av statistiska mått.

Kvalitet

SCB tillämpar en bred definition på kvalitet. För att hålla en hög kvalitet och
täcka kvalitetsdimensionerna används kontroller i undersökningens olika
faser, allt från kundkontakter till dokumentation och uppföljning. I SCB MIS
2001:1 "Kvalitetsbegrepp och riktlinjer för kvalitetsdeklaration av officiell
statistik" ges en ingående beskrivning av SCB:s kvalitetsbegrepp och de olika
kvalitetskomponenterna. Nedan beskrivs de begrepp som har betydelse för
denna undersökning.

Innehåll

Statistiska mått

Materialet levereras som en fil. Inga mått har tagits fram av SCB. Antal och andel brukar vara de vanligaste måtten.

Redovisningsgrupper

Ofta redovisas statistik inte bara för hela populationen utan också för delgrupper, redovisningsgrupper. Redovisningsgrupperna i den här undersökningen kan avgränsas utifrån svaren på frågor.

Referenstider

Referensperioden i den här undersökningen är vårterminen 2005 för objekten och frågorna. Ramuppgiften avseende högskolestudier är höstterminen 2004.

Tillförlitlighet

Ramtäckning

Täckningsfel, under- och övertäckning, innebär att urvalsram och population inte helt stämmer överens. Undertäckning innebär att vissa enheter som ingår i populationen saknas i urvalsramen. Övertäckning innebär att enheter som inte ingår i populationen ändå finns i urvalsramen. Ett sätt att minska täckningsfelen är att ha bra och uppdaterade register. I denna undersökning användes en version av Universitets- och högskoleregistret som var en termin gammal, vilket orsakade täckningsbrister.

I undersökningen visade sig 372 av de svarande vara övertäckningsobjekt. Det var personer som inte längre studerade på universitet eller högskola. Mycket troligt finns det ytterligare övertäckning i bortfallet.

Tabell 4a Beskrivning av täckningsbrist per kurstyp

Kurstyp	Urval	Antal svarande övertäckning	Andel svarande övertäckning
Yrkesprogram	1173	86	7,3
Övriga program	656	33	5,0
Fristående kurs	784	253	19,5

Tabell 4b Beskrivning av täckningsbrist per poäng ht 2004

Poäng	Urval	Antal svarande övertäckning	Andel svarande övertäckning
Mindre än 10	333	94	28,2
10-19	631	110	17,4
20 poäng	1612	143	8,9
21-30 poäng	445	23	5,2
Mer än 30 poäng	106	2	1,9

Av tabellerna 4a-4b framgår att det framförallt var bland dem som läste på fristående kurser och/eller läste få poäng under höstterminen 2004 som övertäckningen var stor.

Eftersom övertäckningssvaren inte används vid framställning av tabeller bör man tänka på att resultaten underskattar alla antalsuppgifter, jämfört med det faktiska antalet som studerar på högskola under vt 2005. Framförallt är det antalet på fristående kurser och antalet som läser få poäng som underskattas.

Undertäckningen i undersökningen torde vara av samma storleksordning som övertäckningen. Den består av studenter vårterminen 2005 som inte var registrerade under höstterminen 2004.

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 (estimat) och det faktiska värdet (parametern) som beror på att man inte undersöker alla objekt i populationen. Urvalsfelets storlek minskar med en ökad urvalsstorlek. Urvalsfelets storlek kan redovisas med osäkerhetsintervall i form av konfidensintervall.

Mätning

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

För att försöka minimera mätfelen har frågeblanketten granskats av SCB:s mättekniska laboratorium och kontroller har genomförts, t ex kontroll av konsistens mellan olika frågor.

Det finns viss inkonsistens mellan svaren i fråga 12 i förhållande till fråga 9 och 11. I klartext innebär detta att det finns personer som, enligt egen uppgift, läst ett antal poäng som är stort i förhållande till både startår och antalet lästa terminer. Det handlar inte om oerhört många personer och i de flesta fall kan "avvikelsen" säkert vara korrekt. Det finns även några avvikelser mellan fråga 9 och 11 (fler lästa terminer än vad startterminen möjliggör).

Bearbetning

Vid den manuella och maskinella bearbetningen av datamaterialet kan så kallade 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 var det frågorna 9, 11-12 och 28, om starttermin, lästa terminer, godkända poäng och födelseår, som var svårast att registrera, då skannern skulle tolka handskrivna text. Eftersom SCB har tillgång till födelseår kunde en direkt jämförelse med fråga 28 göras. Antalet födelseår som skannern tolkade fel var 22, vilket

motsvarande 1,5 procent. För fråga 9, 11 och 12 gjordes rimlighetskontroller mellan frågorna. Extrema och avvikande svar kontrollerades. Kvarstående avvikelser beror inte på skanningfel utan på att personen har lämnat motsägande uppgifter.

Kodning förekom i mycket liten utsträckning.

Bortfall

Bortfallsfel inträffar om objekten i bortfallet och de svarande skiljer sig åt (har en annan fördelning) avseende undersökningsvariablerna.

En beskrivning av bortfallet finns i avsnittet "Bortfall" och i avsnittet "Datainsamling" finns svarsandelar fördelat på olika grupper.

Aktualitet

Med statistikens aktualitet menas hur väl statistiken beskriver nuläget.

Faktorer som påverkar denna komponent är exempelvis framställningstid, punktlighet och hur snabbt verkligheten förändras.

Undersökningen var en engångsundersökning.

SCB:s arbete vilket omfattade enkätkonstruktion, tryckning, insamling, tabellframställning m.m. pågick i cirka 2 månader – från första kontakten till leveransen.

Fil levererades vecka 20, 2005, enligt överenskommelsen.

Jämförbarhet och sammanvändbarhet

Är främst aktuell vid återkommande undersökningar och vid jämförelse mellan olika undersökningar.

Sammanvändbarhet med annan statistik

Undersökningens urvalsram bygger på uppgifter från Universitets- och högskoleregistret för höstterminen 2004. Antalsuppgifter kommer inte att vara jämförbara med de kommande registeruppgifterna för vårterminen 2005, eftersom det förekom både över- och undertäckning i undersökningen.

Tillgänglighet och förståelighet

SCB levererar en avidentifierad fil tillsammans med den här rapporten.

Att tänka på vid användandet av datamaterialet

I SPSS-filen finns variabler med variabelnamn och en kortfattad förklaring om variabelns innehåll samt värdemängd. Om ett svar är utelämnat har koden "9" (om värdet 9 finns: '99', '999' eller '9999') registrerats. Kryssfrågor som inte ska besvaras p.g.a. hoppinstruktioner har värdet "0".

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.