

## *Uppföljning av KY-utbildning*

### **Inledning**

Enheten för statistik om utbildning och arbete vid Statistiska centralbyrån (SCB) genomförde under perioden augusti – oktober 2010 en postenkät på uppdrag av Myndigheten för yrkeshögskolan (Yh). Syftet med undersökningen var bland annat att få svar om sysselsättning efter KY-studierna samt att jämföra examinerade med icke examinerade personer som har studerat en KY-utbildning.

Undersökningen genomfördes som en pappersenkät (med två påminnelser) med möjlighet att svara via webben. Populationen utgjordes av personer som examinerats från en kvalificerad yrkesutbildning 2009. Det gjordes en parallell undersökning med personer icke examinerade från en kvalificerad yrkesutbildning åren 2008 och 2009. Samma enkät användes till båda populationerna.

*Den här tekniska rapporten beskriver den första delen, dvs. personer som examinerats från en kvalificerad yrkesutbildning 2009.*

Totalt var det 9 683 personer som ingick i populationen och 4 732 personer i urvalet. 3 009 personer besvarade frågeblanketten, vilket var 64 procent av urvalet. Frågeblanketterna registrerades med skanning. Från det skannade materialet framställdes tabeller.

Sinisa Sauli var undersökningsledare, Johan Löfgren metodansvarig och Inga-Britt Svalstedt produktionsansvarig.

### **Population och urval**

Populationen, det vill säga de objekt som man vill kunna dra slutsatser om, utgörs i denna undersökning av de personer som examinerats från en kvalificerad yrkesutbildning under 2009. Urvalsramen i denna undersökning skapades utifrån ett register från Yh som skickades till SCB under vecka 25. Antalet individer i ramen var 9 762.

För att få fram aktuella adressuppgifter genomfördes en identifikations-kontroll av personerna i populationen mot registret över totalbefolkningen (RTB, version 2010-03-31). Vid kontrollen visade det sig att 79 personer inte längre tillhörde populationen utan utgjorde s.k. övertäckning. Den vanligaste orsaken till övertäckning var att personerna hade emigrerat. Därmed blev den slutliga populationsstorleken 9 683 individer.

Yh beslutade efter diskussion med SCB att urvalsstorleken skulle vara ca 4 700 individer. Stickprovets storlek bestämdes utifrån uppsatta kostnads-, precisions- och redovisningskrav.

Från urvalsramen drogs ett stratifierat obundet slumpmässigt urval om

4 732 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 vara säker på att urvalet innehöll tillräckligt med individer/objekt i redovisningsgrupper utbildningsområde, kön och födelseland definierades de olika utbildningsområdena som egna stratum.

Tabell 1. Population och urval

Stratum	Population	Urval
Data/IT; män; Sverige	412	200
Data/IT; kvinnor; Sverige*	65	65
Ekonomi, adm och försäljning; män; Sverige	676	230
Ekonomi, adm och försäljning; kvinnor; Sverige	1 859	270
Ekonomi, adm och försäljning; kvinnor; övriga världen	213	160
Friskvård och kroppsvård; män; Sverige*	26	26
Friskvård och kroppsvård; kvinnor; Sverige	249	170
Hotell, restaurang och turism; män; Sverige	133	120
Hotell, restaurang och turism; kvinnor; Sverige	593	230
Hälso- och sjukvård samt socialt arbete; män; Sverige*	100	100
Hälso- och sjukvård samt socialt arbete; kvinnor; Sverige	822	240
Kultur, media och design; män; Sverige	321	190
Kultur, media och design; kvinnor; Sverige	320	190
Lantbruk, djurvård, trädgård, skog och fiske; män; Sverige	166	140
Lantbruk, djurvård, trädgård, skog och fiske; kvinnor; Sverige	279	180
Samhällsbyggnad och byggteknik; män; Sverige	453	210
Samhällsbyggnad och byggteknik; kvinnor; Sverige*	101	101
Säkerhetstjänster; män; Sverige*	94	94
Säkerhetstjänster; kvinnor; Sverige*	36	36
Teknik och tillverkning; män; Sverige	947	250

Teknik och tillverkning; kvinnor; Sverige	344	190
Transporttjänster; män; Sverige	212	150
Transporttjänster; kvinnor; Sverige	184	140
Andra utbildningsområden; män; Sverige*	50	50
Andra utbildningsområden; kvinnor; Sverige	158	130
Alla; alla; EU 27 inkl. Norge och Schweiz (exkl. Sverige)*	272	272
Alla; alla; övriga världen*	598	598
<b>Totalt</b>	<b>9 683</b>	<b><del>732</del><sup>4</sup></b>

\*Totalt undersökt stratum

För att reducera bortfallsskevheten har vikter beräknats med hjälp av kalibrering (se även kalibreringsrapporten). Efter avslutad insamling gjordes en ny identifikationskontroll av personerna i populationen mot registret över totalbefolkningen (RTB, version 2010-06-30). Skattningarna kalibrerades mot denna version av RTB. Vid denna kontroll var det 9 690 av de ursprungliga 9 762 personerna som matchade mot RTB.

### **Sekretess och utlämnande**

I ett informationsbrev kunde uppgiftslämnarna läsa om undersökningens bakgrund, syfte och att undersökningen genomfördes i samarbete mellan Yh och SCB. Informationsbrevet informerade också om vilka uppgifter som hämtades från Yh-myndighetens register och vilka uppgifter som hämtades från SCB, att svaren skyddas av personuppgiftslagen och sekretesslagen samt att det var frivilligt att medverka i undersökningen.

Behandlingen av personuppgifter i undersökningen har anmälts till SCB:s juridiska kansli. Beslutet därifrån medger endast att registret (datamaterialet) bevaras på SCB till 2011-01-21 då uppgifterna kommer att förstöras.

### **Variabler**

Yh har utformat frågorna. Blanketten bestod av 12 numrerade frågor. En fråga hade delfrågor vilket genererade totalt 13 frågor. Frågorna handlade bl.a. om sysselsättning före KY-studierna, nuvarande sysselsättning, överensstämmelse mellan nuvarande arbete och avslutad KY-utbildning, anställningsform på nuvarande arbete samt hur nöjd eller missnöjd är man med KY-utbildningen.

### **Datainsamling**

Frågeblanketterna sändes till urvalspersonerna med post. I informationsbrevet ombads de besvara frågorna och skicka tillbaka frågeblanketten till SCB. Det första utskicket genomfördes den 20 augusti 2010. Två påminnelser med ny enkät skickades ut, den första påminnelsen den 7 september och den andra påminnelsen den 23 september. Enligt överenskommelse med Yh förlängdes insamlingen med sex dagar och avslutades den 18 oktober 2010 (istället för den 12 oktober).

Totalt besvarades enkäten av 3 009 personer, vilket motsvarar 63,6 procent (oviktat) av urvalet. 985 personer har

valt att svara via webben vilket är  
33 procent av alla inkomna svar.

**Tabell 2. Beskrivning av inflödet. Antal och ovägd andel**

	Antal	Procent
Efter första utskick	1 923	40,6
Efter enkätpåminnelse 1	660	14,0
Efter enkätpåminnelse 2	426	9,0
Totalt antal svar	3 009	63,6
Bortfall	1 723	36,4
Urval	4 732	100,0

**Tabell 3. Beskrivning av inflödet per stratum. Antal och ovägd andel**

Stratum	Antal svar	Andel svar
Data/IT; män; Sverige	115	58
Data/IT; kvinnor; Sverige	44	68
Ekonomi, adm och försäljning; män; Sverige	136	59
Ekonomi, adm och försäljning; kvinnor; Sverige	180	67
Ekonomi, adm och försäljning; kvinnor; övriga världen	93	58
Friskvård och kroppsvård; män; Sverige	16	62
Friskvård och kroppsvård; kvinnor; Sverige	114	67
Hotell, restaurang och turism; män; Sverige	65	54
Hotell, restaurang och turism; kvinnor; Sverige	145	63
Hälso- och sjukvård samt socialt arbete; män; Sverige	69	69
Hälso- och sjukvård samt socialt arbete; kvinnor; Sverige	180	75
Kultur, media och design; män; Sverige	92	48
Kultur, media och design; kvinnor; Sverige	119	63
Lantbruk, djurvård, trädgård, skog och fiske; män; Sverige	85	61
Lantbruk, djurvård, trädgård, skog och fiske; kvinnor; Sverige	132	73
Samhällsbyggnad och byggteknik; män; Sverige	132	63
Samhällsbyggnad och byggteknik; kvinnor; Sverige	73	72
Säkerhetstjänster; män; Sverige	61	65
Säkerhetstjänster; kvinnor; Sverige	24	67
Teknik och tillverkning; män; Sverige	154	62

Teknik och tillverkning; kvinnor; Sverige	135	71
Transporttjänster; män; Sverige	97	65
Transporttjänster; kvinnor; Sverige	105	75
Andra utbildningsområden; män; Sverige	32	64
Andra utbildningsområden; kvinnor; Sverige	87	67
Alla; alla; EU 27 inkl. Norge och Schweiz (exkl. Sverige)	187	69
Alla; alla; övriga världen	337	56
<b>Totalt</b>	<b>3 009</b>	<b>63,6</b>

### Bortfall

Bortfallet består dels av objektsbortfall 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. För att reducera bortfallsskevheten har vikter beräknats med hjälp av kalibrering (se även kalibreringsrapporten).

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, t ex på grund av sjukdom. Objektsbortfallet i denna undersökning redovisas i tabell 4.

**Tabell 4. Beskrivning av objektsbortfall**

	Antal
Ej avhörda	1 624
Avböjd medverkan	10
Förhindrad medverkan	1
Postreturer	79
Ej anträffad	9
<b>Totalt</b>	<b>1 723</b>

Med "ej avhörda" menas att ingen uppgift om varför frågeblanketten inte är besvarad har lämnats. Med "avböjd medverkan" menas att SCB (eller Yh) meddelats att uppgiftslämnaren inte vill medverka i undersökningen. "Förhindrad medverkan" innehåller språksvårigheter, fysiskt eller psykiskt hinder. "Ej anträffad" innehåller saknad adress i RTB, hemlig adress och tillfälligt bortrest.

Partiellt bortfall kan bero på att frågan är svår att förstå, är känslig, att lämpligt alternativ saknas eller att uppgiftslämnaren inte vill besvara frågan. I denna undersökning

var det partiella bortfallet litet. Det högsta partiella bortfallet finns i fråga 7. En redovisning finns i bilaga 2.

### **Databeredning**

Databeredningen genomfördes av enkätenheten vid SCB. Frågeblanketterna skannades med programvaran ReadSoft DOCUMENTS for FORMS. Kontroller genomfördes under och efter registreringen. Då kontrollerades bland annat att endast valida värden förekom i materialet.

Svarsdatafilen kompletterades med bakgrundsvariabler från register samt med vikter för uppräknig till populationsnivå.

### Viktberäkning och estimation

Vikter har beräknats så att resultat kan redovisas för hela populationen och inte bara för de svarande. Vikterna kompenserar för objektsbortfallet men inte för det partiella bortfallet och därmed kommer antalet i respektive tabell att vara olika. Om vikterna inte används så kan resultaten bli helt missvisande, speciellt som olika individer haft olika sannolikhet att komma med i urvalet.

Vikterna i denna undersökning kan beskrivas med formeln:

$$w_k = d_k \cdot v_k$$

där  $w_k$  = den totala vikten för objekt k  
 $d_k$  = designvikt  
 $v_k$  = kalibreringsvikt

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.

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

För beräkning av skattningen av 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 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  $w_k$  = den totala vikten för objekt k  
 $y_k$  = variabelvärde för objekt k

summering sker av de svarande (r)

Kalibreringen och de tabeller som presenteras i denna rapport har tagits fram enligt de formler som presenterats ovan med hjälp av SAS och variansprogrammet ETOS.

### **Tabellerna**

Tabellerna har räknats upp till populationsnivå, vilket innebär att resultatet avser hela populationen och inte endast de svarande. Antalsuppgifterna är således skattningar av antal personer i populationen med den aktuella egenskapen. Att antalet individer är olika för olika frågor beror på hoppinstruktioner då olika delpopulationer avses.

För de följdfrågor som föregås av hoppinstruktioner har selekteringar genomförts vid tabellframställandet. Det betyder att för fråga 3 - 7 har endast de som svarat "Arbete" eller "Egen företagare" i fråga 2 ingått i tabellframställningen.

Om en följdfråga är besvarad men dess styrfråga inte är besvarad och svaret på följdfrågan gör att ett visst svarsalternativ på styrfrågan är självklar så har hänsyn till detta tagits vid tabellframställningen.

I andelstabellerna är värdena avrundade till närmaste heltal och värden under 0,5 är därmed avrundade till noll.

Tabellerna har sekretessgranskats så att celler som bygger på mindre än tre svarande har prickats. Mycket osäkra skattningar, där konfidensintervallet är större än andelsskattningen, har också prickats.

De tabeller som presenteras i denna rapport har tagits fram enligt de formler som presenterats i avsnitt "Viktberäkning och estimation" med hjälp av SPSS/SAS och variansprogrammet ETOS.

### **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 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).

Redovisningsgrupperna i den här undersökningen avgränsas med hjälp av registervariablerna kön, ålder, utbildningsområde, utbildningsnivå samt födelseland.

### Referenstider

Referensperioden i den här undersökningen varierar mellan olika frågor, t.ex. sysselsättningen under 6-månadersperioden innan KY-utbildning och sysselsättningen idag (augusti - oktober 2010).

För objekten (individerna) är referensperioden år 2009.

### 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 ansvarade KY-myndigheten för urvalsramen, vilket innebär att SCB inte kan uppskatta täckningsfelen.

#### 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. Urvalsfellets 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 respondenten inte minns de faktiska uppgifterna, missförstår frågan eller medvetet svarar felaktigt.

**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 bedöms registreringsfel vara försumbara eftersom blanketten endast hade fasta svarsalternativ.

**Bortfall**

Bortfallsfel inträffar om objekten i bortfallet och de svarande skiljer sig åt (har en annan fördelning) avseende undersökningsvariablerna. Kalibrering har använts för att reducera bortfallsfelet.

**Aktualitet**

SCB:s arbete omfattade enkätkonstruktion, tryckning, utsändning, insamling, databearbetning samt framställning av tabeller.

Materialet levererades under vecka 45, enligt överenskommelse.

**Jämförbarhet och sam användbarhet****Jämförbarhet över tiden**

Utvärderingen av KY-utbildning har genomförts en gång per år sedan 1999.

Antalet och innehållet i frågorna har förändrats över tiden. 2007, 2008 och 2009 års enkät sändes till examinerade.

Yh-myndigheten är beställaren av uppdraget för första gången. Tidigare åren var det Myndigheten för kvalificerad yrkesutbildning. Årets undersökning gjordes med både examinerade (2009) och icke examinerade personer (åren 2008 och 2009). Det är möjligt att göra vissa jämförelser med tidigare år men de ska göras med försiktighet.

**Tillgänglighet och förståelighet**

SCB levererar tabeller i Excel-format tillsammans med denna rapport. Användning och spridningen regleras i SCB: allmänna avtalsvillkor och i sekretessöverenskommelsen.

**Användning av tabellerna**

För att kunna läsa tabellerna måste (oftast) inställningar i Excel ändras.

**1. Excel 2003**

För att tabellerna skall fungera måste säkerhetsnivån vara medel.

Detta kontrolleras genom att gå in under "Verktyg - Makro - Säkerhet - Säkerhetsnivåfliken"

När Excel-filen öppnas skall även makrot aktiveras.

Vissa funktioner stöds inte av äldre versioner av Excel. Knappar "Pdf" och "Pdf Alla" fungerar inte i Excel 2003

**2. Excel 2007**

Välj Excel-alternativ - Säkerhetscenter - Inställningar

för säkerhetscenter – Makroinställningar  
Markera rutorna ”Aktivera alla makron” och  
”Åtkomst till objektmodell för VBA-projekt betrodd”.

Bilaga 1: Kalibreringsrapport

Bilaga 2: Dubbelmarkeringar i fråga 7

## Kalibreringsrapport - Examinerade

### 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 vi inte lyckas få svar från alla individer (bortfall) och om de 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 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 redovisas den procentuella andelen individer med viss egenskap, t ex den procentuella andelen som har ett arbete som den nuvarande huvudsakliga sysselsättningen. Dessutom redovisas antal individer. Resultaten redovisas efter kön, ålder, utbildningsområde, utbildningsnivå och födelseland.

### 3 Hjälpinformation

Viss hjälpinformation utnyttjas vanligtvis även före estimationen, t.ex. för bildande av stratifierade urvalsdesigner. I denna undersökning används stratifieringsvariablerna utbildningsområde, kön och födelseland. 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):

## Bilaga 1

2(12)

2010-11-11

- Det första kriteriet är att variabeln samvarierar väl med svarsbenägenheten (-sannolikheten). Det är det viktigaste kriteriet eftersom det leder till en minskning av bortfallskevheten för alla skattningar.
- 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.
- 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 från det, av Yh-myndigheten, inskickade registret över individer som examinerats från en KY-utbildning 2009, RTB samt Utbildningsregistret. En genomgång av variablerna i dessa register resulterade i att sex variabler valdes ut.

De sammanslagningar av kategorier som gjorts baseras på kunskaper från tidigare kalibreringar. Dessutom har hänsyn tagits till hur de redovisningsgrupper som sedan ska användas är avgränsade.

**Tabell 1. Tänkbara hjälpvariabler**

Variabel (benämning)	Kategorier (koder)
Kön	1 = Kvinna 2 = Man
Ålder	1 = 20-24 år 2 = 25-29 år 3 = 30-39 år 4 = 40 år eller äldre
Region	1 = Storstäder (H1, H8, H9) 2 = Större städer (H3) 3 = Övrigt (H4, H5, H6)
Födelseland	1 = Sverige 2 = EU 27 utom Sverige 3 = Utanför EU
Utbildningsnivå	1 = Eftergymnasial utbildning kortare än 3 år 2 = Eftergymnasial utbildning 3 år eller mer
Utbildningsområde	1 = Data/IT

## Bilaga 1

3(12)

2010-11-11

	2 = Ekonomi, adm. och försäljning 3 = Friskvård och kroppsvård 4 = Hotell, restaurang och turism 5 = Hälsa- och sjukvård samt socialt arbete 6 = Kultur, media och design 7 = Lantbruk, djurvård, trädgård, skog och fiske 8 = Samhällsbyggnad och byggteknik 9 = Säkerhetstjänster 10 = Teknik och tillverkning 11 = Transporttjänster 12 = Övriga
--	---

Utbildningsområdena journalistik och information, juridik, miljövard och miljöskydd, pedagogik och undervisning samt övriga har vid analys av hjälpinformationen slagits samman till en kategori. Detta har gjorts på grund av att det finns få individer i dessa utbildningsområden.

I följande avsnitt analyserar vi variablerna i tabell 1 för att slutligen bestämma en hjälpvektor.

### 4 Analys av hjälpinformation

#### 4.1 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:  $N_h/n_h$ ).

Vid stora skillnader mellan svarsandelarna utgör variabeln en stark kandidat till hjälpvariabel.

**Tabell 2. Skattad svarsandel fördelat på kön**

Kön	Svarsandel (%)
Kvinna	68,0
Man	59,2
<b>Totalt</b>	<b>64,4</b>

**Tabell 3. Skattad svarsandel fördelat på ålder**

Ålder	Svarsandel (%)
-------	----------------

## Bilaga 1

4(12)

2010-11-11

20-24 år	57,9
25-29 år	58,2
30-39 år	68,6
40 år eller äldre	78,8
<b>Totalt</b>	<b>64,4</b>

**Tabell 4. Skattad svarsandel fördelat på region**

Region	Svarsandel (%)
Storstäder	61,1
Större städer	66,0
Övrigt	68,9
<b>Totalt</b>	<b>64,4</b>

## Bilaga 1

5(12)

2010-11-11

**Tabell 5. Skattad svarsandel fördelat på födelse land**

Födelse land	Svarsandel (%)
Sverige	64,9
EU 27 utom Sverige	69,0
Utanför EU	57,0
<b>Totalt</b>	<b>64,4</b>

**Tabell 6. Skattad svarsandel fördelat på utbildningsnivå**

Utbildningsnivå	Svarsandel (%)
Eftergymnasial utbildning kortare än 3 år	64,7
Eftergymnasial utbildning 3 år eller mer	64,1
<b>Totalt</b>	<b>64,4</b>

**Tabell 7. Skattad svarsandel fördelat på utbildningsområde**

Utbildningsområde	Svarsandel (%)
Data/IT	57,9
Ekonomi, adm. och försäljning	64,1
Friskvård och kroppsvård	66,9
Hotell, restaurang och turism	60,9
Hälso- och sjukvård samt socialt arbete	73,6
Kultur, media och design	55,6
Lantbruk, djurvård, trädgård, skog och fiske	67,8
Samhällsbyggnad och byggteknik	64,0
Säkerhetstjänster	66,7
Teknik och tillverkning	62,8
Transporttjänster	69,9
Övriga	66,7
<b>Totalt</b>	<b>64,4</b>

Tabellerna 2-7 indikerar att hjälpvariablerna kön, ålder, födelse land och utbildningsområde kan anses som "starka" (beträffande kriterium 1).

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

Det är inte möjligt att hitta hjälpvariabler som samvarierar med alla målvariabler. Därför studeras samvariationen med en målvariabel, som bedömts som viktig i den här undersökningen. Den målvariabel som valts är andelen individer som hade ett arbete/eget företag som helt eller till största delen var inom samma område som KY-utbildningen.

## Bilaga 1

7(12)

2010-11-11

Skattningarna görs med hjälp av "konventionell" teknik (vikten sätts till  $N_h/m_h$ ).

där

$N_h$  = antalet individer i urvalsramen i stratum  $h$

och

$m_h$  = antalet svarande individer i stratum  $h$

**Tabell 8.** Andel som hade arbete inom "rätt område" fördelat på kön

Kön	Andel (%)
Kvinna	61
Man	61
<b>Totalt</b>	<b>61</b>

**Tabell 9.** Andel som hade arbete inom "rätt område" fördelat på ålder

Ålder	Andel (%)
20-24 år	59
25-29 år	60
30-39 år	63
40 år eller äldre	61
<b>Totalt</b>	<b>61</b>

**Tabell 10.** Andel som hade arbete inom "rätt område" fördelat på region

Region	Andel (%)
Storstäder	61
Större städer	62
Övrigt	58
<b>Totalt</b>	<b>61</b>

**Tabell 11.** Andel som hade arbete inom "rätt område" fördelat på födelse land

8(12)

**Bilaga 1**

2010-11-11

Födelse land	Andel (%)
Sverige	61
EU27 exkl. Sverige	66
Utanför EU	65
<b>Totalt</b>	<b>61</b>

**Tabell 12. Andel som hade arbete inom "rätt område" fördelat på utbildningsnivå**

Utbildningsnivå	Andel (%)
Eftergymnasial utbildning kortare än 3 år	60
Eftergymnasial utbildning 3 år eller mer	61
<b>Totalt</b>	<b>61</b>

**Tabell 13. Andel som hade arbete inom "rätt område" fördelat på utbildningsområde**

Utbildningsområde	Andel (%)
Data/IT	66
Ekonomi, adm. och försäljning	61
Friskvård och kroppsvård	43
Hotell, restaurang och turism	56
Hälso- och sjukvård samt socialt arbete	70
Kultur, media och design	71
Lantbruk, djurvård, trädgård, skog och fiske	71
Samhällsbyggnad och byggteknik	72
Säkerhetstjänster	47
Teknik och tillverkning	46
Transporttjänster	54
Övriga	67
<b>Totalt</b>	<b>61</b>

Tabell 8-13 indikerar utbildningsområde bör vara en stark hjälpvariabel enligt kriterium 2.

#### **4.2 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. Hjälpvariablerna kön, ålder, utbildningsnivå, utbildningsområde, födelseland och region avgränsar viktiga redovisningsgrupper och bör därför om möjligt vara med i hjälpvektorn.

Om täckningsfel inte föreligger, vilket sannolikt gäller för denna undersökning, ger kalibreringsestimatorn konsistenta skattningar i den meningen att estimatorn ger exakta skattningar för utnyttjade registertotaler.

### 4.3 Slutligt val av hjälpvektor

Efter en sammanvägning av analysen kring de tre kriterierna samt efter kontroll av vikternas fördelning används följande hjälpvektor:

*(Kön \* Utbildningsområde) + Ålder + Region + Födelse land + Utbildningsnivå*

## 5 Teknisk beskrivning av urval och estimation

Vi har en population  $U$  bestående av  $N$  individer. 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 individ  $k$  och  $z_k$  värdet på en annan variabel för samma individ. Vanligtvis är  $y$  (och även  $z$ ) en dikotom variabel, d.v.s.

$$y_k = \begin{cases} 1 & \text{om person } k \text{ har studerade egenskap} \\ 0 & \text{f öövrigt} \end{cases} \quad (5.1)$$

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

$$Y_d = \sum_U y_{dk} \quad (5.2)$$

$$\text{där } y_{dk} = \begin{cases} y_k & \text{f } \text{ör } k \in U_d \\ 0 & \text{f öövrigt} \end{cases}.$$

$Z_d$  bildas på likartat sätt.

En generell parameter för redovisningsgrupp  $d$  ( $d$  kan också avse hela populationen) kan skrivas  $\theta_d = C \frac{Y_d}{Z_d}$ , 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 (5.1). Om vi låter  $N_d$  vara antalet individer i redovisningsgrupp  $d$ , då kan parametern skrivas

## Bilaga 1

11(12)

2010-11-11

$$P_d = 100 \frac{\sum_U y_{dk}}{N_d} \quad (5.3)$$

Vi drar ett obundet slumpmässigt urval  $s_h$  av storleken  $n_h$  från stratum  $h$  ( $h = 1, \dots, H$ ), men p.g.a. övertäckning och bortfall har vi endast svars-  
mängden  $r_h$  av storleken  $n_h$  att utföra beräkningarna på. Storleken på  
stratum  $h$  ger vi beteckningen  $N_h$ .

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

$$\hat{Y}_d = \sum_{h=1}^H \frac{N_h}{m_h} \sum_{r_h} y_{dk} \quad (5.4)$$

I estimator (5.4) används ingen ytterligare hjälpinformation än stratifieringsinformationen.

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  $\mathbf{x}_k$ , som anger till vilka kategorier av

*(Kön \* Utbildningsområde) + Ålder + Region + Födelse land + Utbildningsnivå*

som individ  $k$  hör. Från befintliga register framställer vi hjälptotalerna  $\sum_{U_d} \mathbf{x}_k$ . Vi utnyttjar denna hjälpinformation i en kalibreringsestimator.

Kalibreringsestimatorn för totalen  $Y_d$  har följande utseende:

$$\hat{Y}_{wd} = \sum_r d_k v_k Y_{dk} \quad (5.5)$$

där

$$d_k = N_h / n_h \text{ för } k \in r_h$$

Vikten  $v_k$  används i avseende att kompensera för objektsbortfall.

$$v_k = I + (\sum_U \mathbf{x}_k - \sum_r d_k \mathbf{x}_k)' (\sum_r d_k \mathbf{x}_k \mathbf{x}_k')^{-1} \mathbf{x}_k \quad (5.6)$$

Vid skattning av en parameter av typen  $\theta_d = C \frac{Y_d}{Z_d}$  skattas respektive total med hjälp av kalibreringsvikterna  $d_k v_k$ .

## Referenser:

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