

STATISTIKENS FRAMSTÄLLNING

Undersökningen om vuxnas deltagande i utbildning

Ämnesområde

Utbildning och forskning

Statistikområde

Befolkningens utbildning

Produktkod

UF0538

Referenstid

Aktuella förhållanden september 2016 - februari 2017

Genomförda utbildningar 12 månader bakåt från mättillfället som var någon gång mellan september 2016 - februari 2017

Kontaktuppgifter

Statistikansvarig myndighet	Statistiska Centralbyrån
Kontaktinformation	Mattias Fritz
E-post	mattias.fritz@scb.se
Telefon	010 - 479 62 67

Innehåll

1	Statistikens sammanhang.....	3
2	Undersökningsdesign	3
2.1	Målstorheter	3
2.2	Ramförfarande	3
2.3	Förfaranden för urval och uteslutning	3
2.3.1	Urvalsförfarande.....	3
2.3.2	Uteslutning från insamling (cut-off)	4
2.4	Insamlingsförfarande.....	4
2.4.1	Datainsamlingsmetoder	4
2.4.2	Mätning	5
2.4.3	Bortfallsuppföljning.....	5
2.5	Bearbetningar.....	6
2.6	Granskning.....	6
2.6.1	Granskning under insamlingen	6
2.6.2	Granskning av mikrodata	6
2.6.3	Granskning av makrodata.....	7
2.6.4	Granskning av redovisning	7
2.7	Skattningsförfarande.....	7
2.7.1	Principer och antaganden	7
2.7.2	Skattningsförfarande för målstorheter.....	8
2.7.3	Skattningsförfarande för tillförlitlighet.....	9
2.7.4	Röjandekontroll	10
3	Genomförande	10
3.1	Kvantitativ information.....	10
3.2	Avvikelser från undersökningsdesignen	11

1 Statistikens sammanhang

Statistiken bygger på en intervjuundersökning riktad mot vuxna personer i åldern 25-64 år. Undersökningen är EU-reglerad och visar bland annat andel vuxna befolkningens deltagande i utbildning efter utbildningsnivå, hushållstyp, födelseland, kön och ålder. Statistiken visar hur den vuxna befolkningens deltagande i utbildning under de senaste 12 månaderna varit samt vilka grupper som deltar mer i utbildning och vilka som deltar mindre. Det handlar också om skälen eller hinder till deltagande i utbildning samt nyttan med utbildning och information om utbildningsmöjligheter. Statistikens kvalitet beskrivs i Kvalitetsdeklarationen som finns på www.scb.se/UF0538.

2 Undersökningsdesign

2.1 Målstorheter

De viktigaste målstorheterna är antal och andel vuxna personer i åldern 25-64 år som deltagit i formell eller icke-formell utbildning under de senaste 12 månaderna från mättillfället. Den enskilt viktigaste målstorheten är andelen deltagande personer i icke-formell utbildning.

2.2 Ramförfarande

Ramen som används för att dra urvalet från är Registret över totalbefolkningen (RTB) där alla folkbokförda personer finns med. Rampopulationen är avgränsad med åldersgränser så att ingen person i rampopulationen var under 25 år eller över 64 år under mätperioden september 2016 – februari 2017. Ramen framställdes från RTB med aktualitet 2016-03-31. Aktualiteten för ramen är viktigt för att minska täckningsfelen. Ramen har skapats sex månader före datainsamlingens start vilket behövdes för att hinna hämta registeruppgifter om arbetsställe, utbildningsnivå och programmera och testa dessa variabler i det elektroniska frågeformuläret.

Observationsobjekten är personerna i ovan nämnda population. Observationsobjekten stämmer överens med uppgiftskällorna. Uppgifter om adress har hämtats från RTB och telefonnummer har hämtats från Eniro och Försäkringskassan.

2.3 Förfaranden för urval och uteslutning

2.3.1 Urvalsförfarande

Urvalsdesignen är ett sannolikhetsurval av typen stratifierat obundet slumpmässigt urval utan återläggning. Urvalet av observationsobjekt dras direkt från urvalsramen som omfattar rampopulationen. Ramen avseende folkbokförda personer i Sverige i åldern 25-64 år delades in i 18 urvalsstrata, och ett stratifierat obundet slumpmässigt urval

drogs. Stratifieringen gjordes i syfte att uppnå bra precision i viktiga redovisningsgrupper samt att uppnå Eurostats precisionskrav i skattningen för den viktigaste målstorheten, andelen deltagande personer i icke-formell utbildning.

Urvalsstorleken för respektive stratum har bestämts på ett sådant sätt att precisionen förväntas bli densamma för alla stratum.

Planeringsvärden för svarsandelar i stratum hämtade för den senaste undersökningsomgången som genomfördes 2012 och har använts för att optimera allokeringen.

$$m_h = \frac{\tilde{p}_h(1 - \tilde{p}_h)}{V(\hat{p}_h) + \tilde{p}_h(1 - \tilde{p}_h)/N_h}$$

där

m_h = antal svar som krävs för att uppnå precisionskrav, stratum h

\tilde{p}_h = planeringsvärde för andel personer som deltar i icke-formell utbildning, stratum h

$V(\hat{p}_h)$ = önskad precision, stratum h

N_h = antal personer i urvalsramen, stratum h

Därefter beräknas planerad stickprovsstorlek per stratum

$$n_h = m_h / \tilde{\alpha}_h$$

där

$\tilde{\alpha}_h$ = planeringsvärde för svarsandel, stratum h

2.3.2 Uteslutning från insamling (cut-off)

Ingen del av målpopulationen utesluts från direktinsamling

2.4 Insamlingsförfarande

2.4.1 Datainsamlingsmetoder

Datainsamlingsmetoden är en kombination av direktinsamling från observationsobjekten och insamling från register. För att minska uppgiftslämnarbördan har alla variabler som går att samla in från administrativa register hämtats på detta vis. De variabler som inte finns i register har inhämtats via direktinsamling från observationsobjekten. För vissa variabler, exempelvis arbetsställets näringsgren och högsta utbildningsnivå har de hämtade registeruppgifterna bekräftats eller korrigerats vid intervjun.

De administrativa register som använts för datainsamling är

- Registret över totalbefolkningen (RTB)

- Inkomst- och taxeringsregistret (IoT)
- Utbildningsregistret (UREG)
- Registerbaserad arbetsmarknadsstatistik (RAMS)

För direktinsamling har ett informationsbrev skickats ut med information om undersökningen, att det är frivilligt att delta samt att en intervjuare ringer upp inom en vecka. Det fanns också möjlighet att skicka önskemål till SCB om passande tider för intervjun. Urvalet var uppdelat på två grupper och direktinsamlingen pågick i åtta veckor till vardera grupp följt av en fyra veckor lång bortfallsuppföljning till ett urval av de som inte redan deltagit eller avböjt.

2.4.2 Mätning

Undersökningen avser primärt icke-formell utbildning och det ställs därför fördjupande frågor om icke-formella utbildningar (dock högst två) per undersökningssperson. För formell utbildning ställs endast fördjupande frågor om den senaste utbildningen och för informella utbildningsaktiviteter ställs inga fördjupande frågor. De fördjupande frågorna handlar om utbildningsaktiviteternas egenskaper.

För registeruppgifter har registerversioner använts med aktualitet så nära referenstiden hösten 2016 som möjligt.

Direktinsamlingen genomfördes som en datorstödd telefonintervjuundersökning med inbyggda logiska kontroller för att förebygga mätfel. Frågeformuläret är uppbyggt att fånga deltagande i olika utbildningsformer som sedan följs av fördjupningsfrågor om de utbildningsformer urvalspersonen deltagit i. Om en urvalsperson inte deltagit i en viss utbildningsform sker ett hopp förbi följdfrågorna. Det innebär att en urvalsperson som deltagit i mycket utbildning får svara på många frågor medan en urvalsperson som inte deltagit i någon form av utbildning får svara på ett mycket begränsat antal frågor.

Information om de slutliga observationsvariablerna och statistikens detaljerade innehåll finns på SCB:s webbplats. Där beskrivs alla observationsvariabler och värdemängder m.m. Dokumentationen finns på www.metadata.scb.se under *Vuxnas deltagande i utbildning (AES)*.

2.4.3 Bortfallsuppföljning

Efter ordinarie datainsamlingsperiod med grindar för att stänga stratum med "hög" svarsandel genomfördes en bortfallsuppföljning. Där prioriterades observationsobjekt i bortfallet som tillhörde de nio stratum med störst bortfall. Observationsobjekt där det hittades nya telefonnummer prioriterades också samt de observationsobjekt som det gjorts minst kontaktförsök till i ordinarie datainsamlingsperiod.

Uppföljningen vände sig endast till observationsobjekt i bortfallet som inte tidigare avböjt medverkan.

2.5 Bearbetningar

Kodning har skett för variablerna inriktning på utbildningsaktivitet, yrke och näringsgren. När det gäller näringsgren har administrativa data hämtats för att bekräfta eller korrigera uppgift om arbetsställe. Vid korigering har kodning av näringsgren skett i efterhand. För kodning av yrke har en förprogrammerad lista med 7 000 yrken lagts in i frågeformuläret med sökfunktion som genererar förslag. I de fall yrket inte finns i listan blir det fråga om kodning av det öppna svaret i efterhand. För kodning av utbildningsinriktning har en lista med 13 000 kursnamn som erhållits vid tidigare undersökningsomgångar använts vid kodning av det öppna svaret i efterhand.

All kodning som skett i efterhand har kodats av kodningsexperter på SCB i ett kodningsverktyg med relevant hjälpinformation. Detta kodningsarbete har kvalitetskontrollerats via kontrollkodning. Ungefär 10 procent av alla poster har kontrollkodats och resultatet indikerar att det är svårt att koda dessa variabler. Mellan 15 och 30 procent av det kontrollkodade materialet har gett olika koder vid första och andra kodningen.

En stor del av variablerna i det slutliga observationsregistret är härledda utifrån en eller oftast flera frågor i frågeformuläret. Ett exempel är antal lärarledda timmar för en utbildningsaktivitet. Där har urvalspersonen först valt att rapportera i timmar, dagar eller veckor. Sedan en fråga om antalet timmar/dagar/veckor och till sist en fråga om hur många lärarledda utbildningstimmar i genomsnitt per dag/vecka. Variabeln i det slutliga observationsregistret avser antalet timmar för utbildningsaktiviteten.

Inga imputeringar av partiellt bortfall eller objektsbortfall har gjorts.

2.6 Granskning

2.6.1 Granskning under insamlingen

En stor mängd inbygga kontroller fanns i det elektroniska frågeformuläret för att undvika ologiska svar eller dubbelmarkeringar där så inte var tillåtet.

2.6.2 Granskning av mikrodata

Efter en vecka av datainsamlingsperioden levererades den första testfilen. Där kunde materialet granskas så att frågeformuläret fungerade som förväntat och svar fanns där det skulle finnas svar. Datafiler granskades kontinuerligt under datainsamlingsperioden då leveranser av mikrodata skedde i flera omgångar. Poster för kodning kunde också laddas så att kodningsmaterialet kunde granskas. I

veckovisa informationsbrev till intervjuarna kunde information ges för att förbättra insamlingen och minska mätfel.

2.6.3 Granskning av makrodata

Löpande under datainsamlingen granskades centrala skattningar för att bedöma rimligheten i dem. Eftersom antalet variabler är stort var det främst selektiv makrogranskning som genomfördes av de viktigaste variablerna, andel deltagare i formell och icke-formell utbildning.

I början av datainsamlingen såg nivån på skattningen för andel deltagare i icke-formell utbildning ut som förväntat men ju längre datainsamlingen gick desto större blev skillnaderna mellan förväntat och faktiskt värde. Förklaringen tros vara att personer som deltagit i utbildning svarade först och de som inte deltagit inte var lika motiverade att delta i undersökningen men till slut medverkade allt fler av de som inte deltagit och då sjönk den löpande skattningen allt eftersom. När datainsamlingen var klar hade skillanden mellan skattningen 2016 och 2011 blivit stor och en utredning visade att modeeffekter troligen föreligger. Med modeeffekter menas att övergången från blandad insamling av besöks-/telefonintervjuer till endast telefonintervjuer påverkade skattningarna vilket ledde till tidsseriebrott. Denna indikation syntes inte i början av datainsamlingens granskning.

2.6.4 Granskning av redovisning

Inför publicering laddades samtliga delar av materialet i SCB:s webbpubliceringsverktyg. Det innebär att det slutgiltiga materialet som ska publiceras kunde granskas precis i den form det publiceras.

2.7 Skattningsförfarande

Undersökningen är en urvalsundersökning med bortfall. Kalibrering används för att kompensera för bortfallet.

2.7.1 Principer och antaganden

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. Om vikterna inte används så kan resultaten bli helt missvisande, speciellt som olika urvalspersoner haft olika sannolikhet att komma med i urvalet.

De kalibrerade vikterna har den egenskapen att de vid skattning av populationstotalen, för var och en av de hjälpvariabler som används för att fastställa de kalibrerade vikterna, resulterar i en skattning som överensstämmer exakt med den registertotal som används för att fastställa de kalibrerade vikterna. Detta innebär bland annat att alla antalsskattningar för hjälpvariablerna som baseras på de kalibrerade

vikterna överensstämmer med de storheter som finns i undersökningens rampopulation.

Den estimator som används i undersökningen är en kalibreringsestimator. För en allmän beskrivning av kalibreringsestimatorn, se "Estimation in the presence of Nonresponse and Frame Imperfections" (2001).

Syftet med kalibreringsestimatorn är att reducera urvalsfel och bortfallskevheter med hjälp av extern information, så kallade hjälpvariabler. Kalibreringen (punktskattningar och konfidensintervall) och de tabeller som presenteras i temarapporten har tagits fram med hjälp av SAS och variansprogrammet ETOS.

2.7.2 Skattningsförfarande för målstorheter

I en undersökning uppkommer fel om vi inte lyckas få svar från alla personer och om de avviker från de svarande med avseende på undersökningsvariablerna. Detta fel kallas för bortfallsfel. Kunskap om bortfallsfelet kan i regel bara fås på ett indirekt och approximativt sätt genom att utnyttja registervariabler.

Bortfallsfel kan reduceras genom att använda ett effektivt uppräkningsförfarande (vägning som kompenserar för bortfallet). I följande avsnitt redovisas hur det görs i denna undersökning.

Vägning görs genom att beräkna s.k. kalibrerade uppräkningsvikter. Detta förfarande utnyttjar hjälpinformation i form av registervariabler. I kalibreringen används följande registervariabler (hjälpvariabler):

Kön, ålder, utbildningsnivå, födelseland, civilstånd, förvärvsinkomst, sysselsättningsstatus och urbaniseringsgrad för folkbokföringskommun.

Var och en av de hjälpvariabler som används i kalibreringen har valts utifrån minst ett av följande tre kriterier:

- (i) de samvarierar med svarsbenägenheten
- (ii) samvarierar med (viktiga) undersökningsvariabler
- (iii) avgränsar (viktiga) redovisningsgrupper

Vikterna i denna undersökning kan beskrivas med formeln

$$w_k = 1 + (\sum_s \mathbf{x}_k - \sum_r \mathbf{x}_k)' (\sum_r \mathbf{x}_k \mathbf{x}_k')^{-1} \mathbf{x}_k$$

där w_k = vikt/uppräkningsstal för objekt k

$$\sum_s \mathbf{x}_k = \text{hjälptotaler för hela urvalet}$$

$$\sum_r \mathbf{x}_k = \text{hjälpotaler för de svarande}$$

Vikterna/uppräkningsstalen multipliceras med objektens variabelvärden för att skapa statistikvärden gällande för populationen. 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

Och för beräkning av skattningen av medelvärden används följande formel:

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

summering sker av de svarande (r).

Svarsandelar för olika grupper har beräknats enligt nedanstående formel

$$SA = \frac{T_S}{T_S + (1-a)T_B}$$

där T_S är antal svarande personer, T_S utgörs av antal personer bortfallet och a är en skattning av hur stor andel av bortfallet som inte tillhör målpopulationen. Skattning av a görs enligt

$$a = \frac{T_O}{T_O + T_S}$$

2.7.3 Skattningsförfarande för tillförlitlighet

Konfidensintervall är beräknade enligt en metod som i 95 procent täcker in det sanna värdet och beräknas enligt formeln:
punktskattning + 1,96 * skattning av estimatorns medelfel.

Konfidensintervallet är korrekt under förutsättning att bortfallsfel inte finns. I undersökningar med bortfall finns i princip alltid bortfallsfel. Kalibrerade vikter reducerar bortfallsfel om hjälpinformationen är stark, men en del av bortfallsfelet finns kvar.

2018-02-26

2.7.4 Röjandekontroll

I publicerade tabeller och diagram redovisas inga resultat för variabler med färre än 30 svar samt för svarsalternativ med färre än 3 svar. I det första fallet sekretessprickas svar för samtliga svarsalternativ. I det andra fallet sekretessprickas enbart svarsalternativet med färre än 3 svar.

3 Genomförande

3.1 Kvantitativ information

Totalt ingick 4 977 900 personer i populationen varav 6 695 personer drog till urvalet där 93 personer utgjorde övertäckning identifierad via kontroll precis före datainsamlingenperioden samt via anhöriga som hörde av sig under datainsamlingsperioden. Den korrigerade urvalsstorleken blev

<i>Stratifiering, indelning efter kön, ålder, högsta utbildningsnivå</i>	<i>Populationsstorlek</i>	<i>Urvalsstorlek</i>	<i>Antal svar</i>
Man, 25 - 34 år, högst gymnasial 1 år	138846	311	79
Man, 25 - 34 år, gymnasial 2-3 år och eftergymnasial < 2 år	322081	493	204
Man, 25 - 34 år, eftergymnasial minst 2 år	198569	192	106
Kvinna, 25 - 34 år, högst gymnasial 1 år	99616	243	69
Kvinna, 25 - 34 år, gymnasial 2-3 år och eftergymnasial < 2 år	252604	384	158
Kvinna, 25 - 34 år, eftergymnasial minst 2 år	273483	269	132
Man, 35 - 49 år, högst gymnasial 1 år	185112	329	100
Man, 35 - 49 år, gymnasial 2-3 år och eftergymnasial < 2 år	466185	607	259
Man, 35 - 49 år, eftergymnasial minst 2 år	319958	289	169
Kvinna, 35 - 49 år, högst gymnasial 1 år	137935	254	69
Kvinna, 35 - 49 år, gymnasial 2-3 år och eftergymnasial < 2 år	368821	482	227
Kvinna, 35 - 49 år, eftergymnasial minst 2 år	431227	390	223
Man, 50 - 64 år, högst gymnasial 1 år	207991	334	125
Man, 50 - 64 år, gymnasial 2-3 år och eftergymnasial < 2 år	485486	609	301

Man, 50 - 64 år, eftergymnasial minst 2 år	206470	250	140
Kvinna, 50 - 64 år, högst gymnasial 1 år	191060	318	113
Kvinna, 50 - 64 år, gymnasial 2-3 år och eftergymnasial < 2 år	390135	487	266
Kvinna, 50 - 64 år, eftergymnasial minst 2 år	302321	361	236
Totalt	4 977 900	6602	2 976

Svar inkom från 2 976 personer. Den ovägda svarsandelen var 45,1 procent och den vägda svarsandelen som tar hänsyn till urvalsdesignen var 48,1 procent.

3.2 Avvikelser från undersökningsdesignen

Inga avvikelser har gjorts.