

Metod för estimation vid sammanslagning av urval med olika design i arbetskraftsundersökningarna



I serien Bakgrundsfakta presenteras bakgrundsmaterial till den statistik som SCB producerar inom området arbetsmarknad och utbildning. Det kan röra sig om produktbeskrivningar, metodredovisningar samt olika sammanställningar som kan ge en överblick och underlätta användandet av statistiken.

Utgivna publikationer från 2000 i serien Bakgrundsfakta till arbetsmarknads- och utbildningsstatistiken

- 2000:1 Övergång till yrkeskodning på fyrsiffernivå (SSYK) och införande av jobbstatuskod i SCB:s lönestatistik
- 2000:2 The Information System for Occupational Injuries and the Work-related Health Problems Survey – A comparative study
- 2000:3 Konferens om utbildningsstatistik den 23 mars 2000
- 2001:1 Avvikelser i lönesummestatistiken – en jämförelse mellan LAPS och LSUM
- 2001:2 En longitudinell databas kring utbildning, inkomst och sysselsättning 1990–1998
- 2001:3 Staff training costs 1994–1999
- 2001:4 Studieresultat i högskolan i form av avklarade poäng
- 2001:5 Urvals- och estimationsförfarandet i de svenska arbetskraftsundersökningarna (AKU)
- 2001:6 Svar, bortfall och representativitet i Arbetsmiljöundersökningen 1999
- 2001:7 Individ- och företagsbaserad sysselsättningsstatistik – en jämförelse mellan AKU och KS
- 2002:1 Tidsseriebrott i utbildningsregistret 2001-01-01
- 2002:2 En longitudinell databas kring utbildning, inkomst och sysselsättning (LOUISE) 1990–1999
- 2003:1 Exempel på hur EU:s "Quality Reports" kan skrivas – avser Labour Cost Survey (LSC) 2000
- 2003:2 Förändrad redovisning av högskolans personal
- 2003:3 Individ- och företagsbaserad sysselsättningsstatistik – en fortsatt jämförelse mellan AKU och KS
- 2003:4 Sjukfrånvarande enligt SCB och sjukskrivna enligt RFV
- 2003:5 Informationssystemet om arbetsskador och undersökningen om arbetsorsakade besvär. En jämförande studie
- 2004:1 Samlad statistik från SCB avseende ohälsa
- 2004:2 Översyn av forskarutbildningsstatistiken. Bedömning av kvaliteten
- 2004:3 Sjukfrånvaro och ohälsa i Sverige – en belysning utifrån SCB:s statistik
- 2005:1 En longitudinell databas kring utbildning, inkomst och sysselsättning (LOUISE) 1990–2002
- 2005:2 Nordisk pendlingskarta. Huvudrapport
- 2005:3 Nordisk pendlingskarta. Delrapport 1–4.
- 2005:4 Flödesstatistik från AKU
- 2005:5 Flow statistics from the Swedish Labour Force Survey
- 2006:1 Sysselsättningsavgränsning i RAMS – Metodöversyn 2005

Fortsättning på omslagets tredje sida!

Bakgrundsfakta

**Metod för estimation vid
sammanslagning av urval med
olika design i
arbetskraftsundersökningarna**

Arbetsmarknads- och utbildningsstatistik 2014:1

**Statistiska centralbyrån
2014**

Background Facts

Labour and Education Statistics 2014:1

Method for estimation when combining samples with different designs in the Swedish Labour Force Surveys

Statistics Sweden
2014

Producent <i>Producer</i>	SCB, avdelningen för befolkning och välfärd Statistics Sweden, Population and Welfare Department Box 24300 SE-104 51 STOCKHOLM
Förfrågningar <i>Inquiries</i>	Frida Videll, +46 8 506 947 22 frida.videll@scb.se Elisabet Andersson, +46 8 506 946 45 elisabet.andersson@scb.se

Det är tillåtet att kopiera och på annat sätt mångfaldiga innehållet i denna publikation.

Om du citerar, var god uppge källan på följande sätt:

Källa: SCB, Bakgrundsfakta, Arbetsmarknads- och utbildningsstatistik 2014:1, *Metod för estimation vid sammanslagning av urval med olika design i arbetskraftsundersökningarna.*

It is permitted to copy and reproduce the contents in this publication.

When quoting, please state the source as follows:

Source: Statistics Sweden, Background Facts, Labour and Education Statistics 2014:1, *Method for estimation when combining samples with different designs in the Swedish Labour Force Surveys*

ISSN 1654-465X (Online)

URN:NBN:SE:SCB-2014-AM76BR1401_pdf

Denna publikation finns enbart i elektronisk form på www.scb.se.

This publication is only available in electronic form on www.scb.se.

Förord

Sedan januari 2010 baseras Arbetskraftsundersökningarna (AKU) på två urval, dels ett ordinarie urval som består av ett månadsurval på 21 500 personer och dels ett tilläggsurval som består av ett månadsurval på 8 000 personer. Regeringen fattade vid årsskiftet 2008/2009 beslut om en utökning av urvalet i AKU med syftet att få en utförligare och förbättrad redovisning av bestånd (stockar) och flöden för små grupper i populationen. Urvalsdesignen för det ordinarie urvalet i AKU är oförändrad sedan 2005. Urvalsdesignen för tilläggsurvalet skiljer sig från det ordinarie urvalet genom att det riktas mot grupper av personer med svagare anknytning till arbetsmarknaden.

Med tillgång till data från två olika urval från en och samma population, inställer sig frågan hur dessa data bäst kombineras för att tillgodose det informationsbehov som föreligger? Klassisk statistisk teori föreskriver en ansats som bygger på att man för varje storhet som ska skattas först tar fram en skattning per urval, vilka därefter vägs samman med ett optimalt, storhetspecifikt viktsystem. I praktiken innebär dock ett sådant förfarande att man måste använda lika många viktsystem som de storheter som ska skattas. För AKU, som varje månad skattar hundratals storheter, många definierade på redovisningsgruppsnivå, är detta inte önskvärt. Tvärtom vill man, av såväl praktiska som pedagogiska skäl, att alla skattningar ska baseras på ett gemensamt viktsystem. Vad som eftersträvas är en kompromisslösning, i form av *ett* viktsystem som kan förmodas fungera väl för flertalet av de storheter som ska skattas.

Hur en sådan kompromiss åstadkoms är dock inte självklart. I samband med att tilläggsurvalet tillkom användes inledningsvis ett gemensamt viktsystem, baserat på urvalsstorlekarnas relativa förhållande. Även om viktsystemet var relativt lätt att implementera, visade det sig vara problematiskt genom att det för flera storheter resulterade i oacceptabelt osäkra skattningar. Mot bakgrund av detta startades det projekt, vars resultat redovisas i rapporten.

I rapporten diskuteras olika metoder för sammanvägningar utifrån AKU:s behov. Analysen resulterar i ett förfarande med goda egenskaper och som är förhållandevis lätt att införa i produktionen. Resultat för centrala arbetsmarknadsvariabler visas och jämförs med skattningar som enbart bygger på det ordinarie urvalet i AKU. Därmed belyser rapporten de precisionsförbättringar som uppnås genom såväl det utökade urvalet som den nya skattningsmetoden.

Arbetet har genomförts av en projektgrupp bestående av Jan Selén, Frida Videll och Anna-Karin Westöö. Gruppen har haft ett nära samarbete med en referensgrupp där Hassan Mirza, Martin Axelson, Karl-Erik Kristiansson och Elisabet Andersson har ingått. Ett särskilt tack till Claes Andersson som har bidragit med värdefulla synpunkter.

Statistiska centralbyrån februari 2014

Inger Eklund

Hassan Mirza

Innehåll

Förord	3
Sammanfattning	7
1 Bakgrund	9
1.1 Ordinarie AKU	9
1.2 Tilläggsurval	9
1.3 Estimation	10
2 Sammansatt estimation – sammanvägning på estimatornivå	11
3 Sammansatt estimation – sammanvägning på individnivå	13
4 Resultat och jämförelser	15
5 Referenser	17
Resultattabeller, ny estimator i jämförelse med enbart ORD- urvalet	19
Bilaga	23

Sammanfattning

Sedan 2010 baseras Arbetskraftundersökningarna, AKU, på utökade urval¹. Undersökningen omfattar nu cirka 29 500 individer varje månad där cirka 8 000 tillkommit efter urvalsökningen. Syftet med utökningen var att bland annat förbättra redovisningen för grupper med svagare anknytning till arbetsmarknaden. Utökningen gjordes därför genom tilläggsurval med en annan design än de ordinarie urvalen, riktad mot sådana grupper.

Estimation utifrån flera urval, dragna enligt olika design, kräver anpassade metoder för att få goda egenskaper. I detta dokument beskrivs sådan estimation. En förändring av estimationen baserad på så kallade kumulerade sannolikheter specificeras och undersöks. Tester på viktiga tabeller som ingår i AKU:s redovisning på webbplatsen ger med några få undantag små effekter på punktskattningarna, jämfört med skattningar utan utökningen och enligt det tidigare ordinarie skattningsförfarandet.

Effekterna av förändringarna är däremot tydliga vad gäller precisionen. Den nya estimatorn leder till en precisionsförbättring på mellan 9 och 20 procent i genomsnitt för de undersökta tabellerna. Så mycket lägre blir variationskoefficienten (kvadratroten ur skattad varians/punktskattning) jämfört med skattningar utan utökningen och enligt det tidigare ordinarie skattningsförfarandet.

Arbetet med att ta fram en ny estimator till AKU har skett i form av ett projekt. För vidare läsning om resultat därifrån se SCB (2012).

¹ Tilläggsurval finns från och med april 2009, från början av begränsad omfattning. Tilläggsurvalet ingår i AKU från och med 2010.

1 Bakgrund

AKU är en panelundersökning med roterande urval där varje individ deltar en gång per kvartal i två år, totalt åtta gånger. Urvalet roteras på så sätt att en åttondel byts ut mellan två på varandra följande undersökningstillfällen, som består av månader. För varje urval inträffar detta alltså med tre månaders mellanrum. I Videll (2011) ges en utförlig beskrivning av urvals- och estimationsförfarandet från och med 2005.

Urvalet till undersökningen görs årligen för att täcka behovet av nya individer under det kommande året, januari–december. Från och med januari 2010 baseras skattningarna i AKU på två urval i och med en utökning av månadsurvalet med 8 000 personer. Detta urval om 8 000 personer kallas fortsättningsvis för tilläggsurvalet och urvalet om cirka 21 500 personer före utökningen, för ordinarie AKU. Syftet med utökningen var utökad redovisning för små grupper av populationen, främst grupper med svagare anknytning till arbetsmarknaden. Dessa grupper utgörs av sysselsatta personer frånvarande från arbetet under minst en kalendervecka, arbetslösa samt personer ej i arbetskraften (t.ex. pensionärer, heltidsstuderande, sjuka med flera).

De ursprungliga ordinarie urvalen (ORD) har samma urvalsdesign för varje år. Urvalen är stratifierade men väsentligen självvägda, det vill säga urvalssannolikheterna är ungefär desamma. En i huvudsak konstant design över tid gäller också de nya tilläggsurvalen (NY). Dessa senare urval är dock riktade mot vissa delar av befolkningen varför urvalsdesignen skiljer sig från de ordinarie urvalens.

1.1 Ordinarie AKU

Urvalsförfarandet² för ordinarie AKU kan beskrivas som stratifierat obundet systematiskt urval, OSU, där målpopulationen består av folkbokförda i åldern 15–74 år. Urvalsstrata skapas genom olika kombinationer av region (24), ålder (3) och kön (2) till sammanlagt 14 strata, där region utgörs av boendelän samt storstadskommunerna Stockholm, Göteborg och Malmö.

1.2 Tilläggsurval

Urvalsförfarandet³ för tilläggsurvalet kan beskrivas som stratifierat OSU där målpopulationen består av folkbokförda i åldern 16–66. Urvalsstrata skapas genom att använda information från RTB⁴, LISA⁵ och IoT⁶. För inrikes födda bildas det 70 strata genom att kombinera ålder (3), region (7),

² Från och med det urval som roterade in 2013 är urvalsförfarandet stratifierat systematiskt urval med sortering efter födelseland (inrikes/utrikes född) och personnummer inom stratum. Detta medförde att ålder togs bort som stratifieringsvariabel vilket resulterade i sammanlagt 48 strata.

³ Från och med det urval som roterade in 2013 är urvalsförfarandet stratifierat systematiskt urval med sortering efter personnummer inom stratum.

⁴ Registret över totalbefolkningen.

⁵ Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier.

⁶ Registret över inkomster och taxeringar.

information från LISA och IoT (2) och kön (2). För utrikes födda bildas 35 strata genom att skapa kombinationer av ålder (3), region (7) och information från LISA och IoT (2). Sammanlagt bildas alltså 105 olika strata. Länen delas in i sju regioner där varje region består av ett eller flera län. För åldersgruppen 25–66 år dras urvalet på ett sådant sätt att 80 procent av det totala urvalet ska fånga upp personer ej i arbete enligt indikation från LISA och IoT.

1.3 Estimation

Estimationen i AKU utgår från estimation per månad som baseras på individer från två till tre årsurval. Efter utökningen 2010 består varje årsurval av två olika urval. Förenklat kan undersökningens skattningar ses som att separata skattningar för urvalen från de olika åren vägs samman.

Estimationen i AKU handlar om estimation med sammansatta urval. Tidigare ingick endast urval med samma design vilket inte är fallet efter utökningen. Med olika urvalsdesigner krävs anpassade metoder för skattningarna.

I fortsättningen ska först en metod för direkt sammanvägning av estimatorer beskrivas. Estimationen i AKU före 2010 bygger på sådana sammanvägningar. Därefter behandlas en ny metod som utgår från sammanvägningar på individnivå och som fungerar bättre efter utökningen.

2 Sammansatt estimation – sammanvägning på estimatornivå

Antag att X_1 och X_2 är två estimatorer av en och samma storhet. En sammansatt estimator för denna storhet ges av

$$Z = \alpha * X_1 + (1-\alpha) * X_2 \quad (1)$$

Om X_1 och X_2 är väntevärdesriktiga, är också Z väntevärdesriktig.

Om X_1 och X_2 är okorrelerade följer att variansen för estimatoren Z minimeras för

$$\alpha = \frac{1/V_1}{1/V_1 + 1/V_2} \quad (2)$$

där V_1 och V_2 betecknar varianserna för X_1 respektive X_2 .

Antag att X_1 och X_2 baseras på urval av storleken n_1 respektive n_2 , samt att varianserna kan skrivas på $V_1 = c_1/n_1$ och $V_2 = c_2/n_2$. Då kan (2) skrivas

$$\alpha = \frac{n_1/c_1}{n_1/c_1 + n_2/c_2} \quad (3)$$

Om $c_1 = c_2$ förenklas detta till

$$\alpha = \frac{n_1}{n_1 + n_2} \quad (4)$$

Resultaten ovan kan generaliseras till att gälla sammanvägning av fler än två estimatorer. Motsvarigheten till (4) för J urval med parvis okorrelerade estimatorer blir för det j :te urvalet

$$\alpha_j = \frac{n_j}{n_1 + n_2 + \dots + n_j} \quad (5)$$

Optimeringar av det här slaget medför separata viktsystem för varje redovisning, det vill säga för varje variabel och redovisningsgrupp. Att i löpande statistikproduktion hantera flera olika viktsystem är dock något man i möjligaste mån vill undvika. Av konsistensskäl är separata optimeringar inte lämpliga för tabeller, eftersom en skattning för hela befolkningen av antalet sysselsatta – till exempel – bör stämma med summan av antalet sysselsatta i olika åldersgrupper. Vad som i praktiken efterfrågas är ett viktsystem, som kan användas för alla redovisningar och som kan förmodas fungera bra i allmänhet.

Det är långt från givet att det i praktiken går att konstruera *ett* viktsystem för sammanvägning på estimatornivå, som kan förväntas fungera bra för estimation av flertalet av de storheter och redovisningsgrupper som är av intresse. Före utökningen bygger AKU dock på sådan sammansatt estimation, med vikter proportionella mot årsurvalens storlekar enligt (5). Eftersom årsurvalen var dragna enligt i princip samma urvalsdesign, och i allt väsentligt var att betrakta som stratifierade obundna slumpmässiga urval med proportionell allokering, gäller att ett sådant viktsystem kan sägas vara approximativt optimalt för alla storheter och redovisningsgrupper.

3 Sammansatt estimation – sammanvägning på individnivå

Att konstruera *ett*, väl fungerande, viktsystem för skattningar baserade på både ORD och NY är en större utmaning. När data från urval baserade på helt olika urvalsdesigner ska vägas samman är det nämligen långt ifrån säkert att det ens existerar en acceptabel kompromisslösning. I detta fall kan i stället sammansatt estimation baserad på sammanvägning av data på individnivå användas.

Nedan diskuteras kort den generella ansats som beskrivs i bilagan under benämningen implicit sammanvägning. Ansatsen förutsätter att det är möjligt att för varje utvald individ beräkna sannolikheten att inkluderas i urvalen under var och en av de olika urvalsdesignerna. Utifrån denna information bestäms sedan sammanvägningsvikter på individnivå. Dessa vikter används sedan genomgående i estimationen, oavsett variabel eller storhet och redovisningsgrupp. Eftersom sammanvägningsvikterna betraktas som konstanter på individnivå, är det enkelt att använda dem för designbaserad inferens baserad på såväl Horvitz-Thompson-estimation som generell regressionsestimation.

För AKU används ansatsen för att för årsurvalen kombinera ORD och NY. För beräkning av de eftersökta vikterna för individ i krävs att det är möjligt att fastställa stratumtillhörighet enligt var och en av de urvalsdesigner som ligger till grund för ORD respektive NY. Eftersom stratumindelningarna är baserade på registerinformation är detta möjligt. För ett givet årsurval, låt P_i^{ORD} och P_i^{NY} beteckna inklusionssannolikheterna för individ i under de två urvalsdesignerna, samt låt P_i motsvara summan

$$P_i = P_i^{ORD} + P_i^{NY} \quad (6)$$

För årsurvalet ges vikter för sammanvägning av data på individnivå av P_i^{ORD} / P_i för individer i ORD och P_i^{NY} / P_i för individer i NY.

O'Muirheartaigh och Pedlow (2002) diskuterar estimation baserad på så kallade kumulerade sannolikheter. Man kan lätt visa att för AKU:s vidkommande motsvarar punkttestimation under implicit sammanvägning med sammanvägningsvikter enligt ovan, punkttestimation baserad på de kumulerade sannolikheter P_i .

4 Resultat och jämförelser

Tabellerna I.1 till I.12.1 sammanfattar resultat för centrala AKU-tabeller för januari 2012 för grundtabeller i undersökningen och för första kvartalet 2012 för en uppsättning tabeller över arbetskraftsindikatorer. Tabellerna jämför skattningar för en estimator baserad på den ansats som beskrivs i avsnitt 3 med skattningar som baseras på ORD-urvalet enligt det tidigare skattningsförfarandet före 2010.

För en liten del av urvalet, cirka 3 procent, var det svårigheter med att ta fram urvalssannolikheten enligt den design som beskrivs i avsnitt 3. För dessa individer har sannolikheten under den alternativa designen satts till 0, med resultat att den kumulerade sannolikheten P_i underskattas. Under strikt Horvitz-Thompson-estimation skulle detta resultera i en positiv bias, men estimationen i AKU baseras på så kallad generell regressions-estimation. Sannolikt kompenserar detta åtminstone delvis för de problem som orsakas av att den kumulerade sannolikheten P_i underskattas för vissa individer.

I tabellerna visas antalet gånger som den standardiserade skillnaden mellan punktskattningarna överstiger 1,96. Indikatorn är beloppet av differensen dividerad med roten ur summan av varianserna, utan hänsyn till kovariansen mellan punktskattningarna. Då skattningarna i stort baseras på samma urval förväntas de vara positivt korrelerade vilket medför att indikatorn ger en underskattning av antalet signifikanta skillnader⁷. Genomsnittet av de procentuella förändringarna för osäkerhetsmättet för punktskattningarna visas, liksom motsvarande genomsnitt för relativa förändringen av variationskoefficienterna (CV). För de relativa CV-förändringarna visas därutöver största ökning respektive största minskning. Andelen variationskoefficienter som blir lägre med den nya estimatorn avslutar redovisningen.

Tabellerna I.1 till I.12.1 visar att det enbart finns en indikation på förändring i punktskattningarna, detta för frånvarande hela veckan i åldern 25–54. Precisionen i skattningarna förbättras överlag med den nya estimatorn jämfört med då enbart ORD-urvalet används. Detta ses i CV som blir lägre, från 8,9 procent till drygt 20 procent i genomsnitt för de olika tabellerna. Genomsnitten är över åldersgrupper och variabler (tabellkolumner i undersökningens redovisning). Andelen skattningar med lägre CV för den nya estimatorn är mellan 66,7 och 100 procent. I de få fall då försämringar förekommer rör det sig vanligtvis om små grupper bland 65–74 år.

⁷ SCB har inför införandet av estimatorn granskat centrala tidsserier i AKU för att säkra jämförbarheten. I denna granskning observerades inga tidsseriebrott för årsskiftet 2009/2010.

5 Referenser

- Andersson, C. (2008): CLAN97. *Supplement to A user's guide to CLAN97*, SCB.
- O'Muircheartaigh C. och Pedlow S. (2002): *Combining samples vs. Cumulating cases: a comparison of two weighting strategies in NLSY97*, Joint Statistical Meeting – Section on Survey Research.
- Sing, A. C. och Mecatti F. (2011): *Generalized Multiplicity-Adjusted Horvitz-Thompson Estimation as a Unified Approach to Multiple Frame Surveys*, Journal of Official Statistics, Vol. 27, No. 4, s 633–650.
- Särndal, C-E., Swensson B. och J. Wretman (1992): *Model assisted survey sampling*, Springer Series in Statistics.
- Videll, Frida (2011): *Urvals- och estimationsförfarandet i de svenska arbetskraftsundersökningarna (AKU) 2005 – "Bakgrundsfakta till arbetsmarknads- och utbildningsstatistiken 2011:6"*.
- SCB (2012): *Estimation för urval med olika design –den utökade AKU. Resultatrapport ökad hjälpinfo_2012_del 1*. Projektet: Ökad hjälpinformation i estimationen i AKU.

Resultattabeller, ny estimator i jämförelse med enbart ORD-urvalet

Sammanfattande resultat för viktiga AKU-tabeller, standardiserade jämförelser av punktskattningar (absolut differens/ $\sqrt{\text{differensvarians}}$), procentuella skillnader för osäkerhetsmått och för variationskoefficienter (CV). Jämförelser mellan den nya estimatorm, då hela urvalet (ORD och NY) används, och då enbart ORD används. Kovarianserna mellan punktskattningarna beaktas inte i de standardiserade jämförelserna.

Tabell I.1: Grundtabell 1.

Befolkningen fördelad efter arbetskraftstillhörighet. Sammanfattningar över kolumner och åldersgrupper, 15–24, 25–34, 35–44, 45–54, 55–64, 65–74 år. Kol. 1–7 sammanfattar skattningar av totaler i AKU-tabellens kolumner 1–7 och kol 9–11 motsvarande skattningar av andelar. AKU januari 2012

Punktskattningar, jämförelsemått, antal >1,96		Osäkerhet, procentuell förändring, genomsnitt		CV, procentuell förändring, genomsnitt		CV, procentuell förändring, extremvärden		CV, andel lägre, procent
kol 1–7	kol 9–11	kol 1–7	kol 9–11	kol 1–7	kol 9–11	min	max	
0	0	-12,7	-14,4	-14,0	-16,1	-27,1	-3,8	100,0

Tabell I.2: Grundtabell 2.

Sysselsatta fördelade efter anknyningsgrad till arbetsmarknaden och överenskommen arbetstid per vecka. Sammanfattande resultat över 9 kolumner och 6 åldersgrupper. AKU januari 2012

Punktskattningar, jämförelsemått, antal >1,96		Osäkerhet, procentuell förändring, genomsnitt		CV, procentuell förändring, genomsnitt		CV, procentuell förändring, extremvärden		CV, andel lägre, procent
kol 1–9	kol 1–9	kol 1–9	kol 1–9	min	max			
0		-11,5		-12,9		-20,7	-6,2	100,0

Tabell I.7: Grundtabell 7.

Totalt antal arbetstimmar per vecka för personer i arbete efter näringsgren och anknyningsgrad till arbetsmarknaden. Sammanfattande resultat över 3 kolumner och 14 näringsgrenar. AKU januari 2012

Punktskattningar, jämförelsemått, antal >1,96		Osäkerhet, procentuell förändring, genomsnitt		CV, procentuell förändring, genomsnitt		CV, procentuell förändring, extremvärden		CV, andel lägre, procent
kol 1–3	kol 1–3	kol 1–3	kol 1–3	min	max			
0		-8,4		-8,9		-28,4	4,4	79,5

Tabell I.12: Grundtabell 12.

Frånvarande från arbetet (huvudsysslan) under referensveckan fördelade efter huvudsaklig frånvaroorsak. Sammanfattning över 8 kolumner och 3 åldersgrupper. AKU januari 2012

Punktskattningar, jämförelsemått, antal >1,96	Osäkerhet, procentuell förändring, genomsnitt	CV, procentuell förändring, genomsnitt	CV, procentuell förändring, extremvärden		CV, andel lägre, procent
			min	max	
kol 1–8	kol 1–8	kol 1–8			
0	-9,4	-10,9	-15,6	-6,8	100,0

Tabell I.15: Grundtabell 15.

Arbetslösa fördelade efter arbetslöshetstidens längd samt efter ålder. Sammanfattning över 9 kolumner och 3 åldersgrupper. AKU januari 2012

Punktskattningar, jämförelsemått, antal >1,96	Osäkerhet, procentuell förändring, genomsnitt	CV, procentuell förändring, genomsnitt	CV, procentuell förändring, extremvärden		CV, andel lägre, procent
			min	max	
kol 1–9	kol 1–9	kol 1–9			
0	-15,3	-20,1	-43,3	-6,3	100,0

Tabell I.16: Grundtabell 16.

Personer ej i arbetskraften fördelade efter huvudsaklig verksamhet. Sammanfattning över 8 kolumner och 2 åldersgrupper. AKU januari 2012

Punktskattningar, jämförelsemått, antal >1,96	Osäkerhet, procentuell förändring, genomsnitt	CV, procentuell förändring, genomsnitt	CV, procentuell förändring, extremvärden		CV, andel lägre, procent
			min	max	
kol 1–8	kol 1–8	kol 1–8			
0	-12,1	-11,3	-16,2	-4,5	100,0

Tabell I.18: Grundtabell 18.

Arbetslösa, undersysselsatta och latent arbetssökande. Sammanfattning över 7 kolumner och 6 åldersgrupper. AKU januari 2012

Punktskattningar, jämförelsemått, antal >1,96	Osäkerhet, procentuell förändring, genomsnitt	CV, procentuell förändring, genomsnitt	CV, procentuell förändring, extremvärden		CV, andel lägre, procent
			min	max	
kol 1–7	kol 1–7	kol 1–7			
0	-15,3	-12,7	-50,0	33,3*	66,7

*) 65–74 år, antal timmar i arbetskraftsutbud.

Tabell I.1.2: Arbetskraftsindikatorer, tabell 1.2.
Personer i arbetskraften fördelade efter arbetskraftstillhörighet samt efter kön och in- och utrikesfödd. Sammanfattning över 10 kolumner och in- och utrikes födda. AKU kvartal 1, 2012

Punktskattningar, jämförelsemått, antal >1,96		Osäkerhet, procentuell förändring, genomsnitt		CV, procentuell förändring, genomsnitt	CV, procentuell förändring, extremvärden		CV, andel lägre, procent
kol 1-5	kol 6-9	kol 1-5	kol 6-9	kol 1-9	min	max	
0	0	-12,1	-13,5	-12,6	-17,5	-7,1	100,0

Tabell I.3.1: Arbetskraftsindikatorer, tabell 3.1.
Personer som är frånvarande hela veckan fördelade efter frånvarotid. Sammanfattning över 5 kolumner och 4 åldersgrupper. AKU kvartal 1, 2012

Punktskattningar, jämförelsemått, antal >1,96		Osäkerhet, procentuell förändring, genomsnitt		CV, procentuell förändring, genomsnitt	CV, procentuell förändring, extremvärden		CV, andel lägre, procent
kol 1-6		kol 1-6		kol 1-6	min	max	
1*		-13,7		-9,5	-16,2	2,2	93,8

*) 25-54 år, frånvarande hela veckan.

Tabell I.4.1: Arbetskraftsindikatorer, tabell 4.1.
Personer som är frånvarande hela veckan pga. sjukdom fördelade efter frånvarotid. Sammanfattning över 5 kolumner och 4 åldersgrupper. AKU kvartal 1, 2012

Punktskattningar, jämförelsemått, antal >1,96		Osäkerhet, procentuell förändring, genomsnitt		CV, procentuell förändring, genomsnitt	CV, procentuell förändring, extremvärden		CV, andel lägre, procent
kol 1-5		kol 1-5		kol 1-5	min	max	
0		-15,0		-13,2	-53,5	24,9*	85,7

*) 65-74 år, sjukfrånvarande hela veckan, 1-2 veckor.

Tabell I.12.1: Arbetskraftsindikatorer, tabell 12.1.
Personer ej i arbetskraften som är sjuka fördelade efter tidigare arbetslivserfarenhet samt efter kön och ålder. Sammanfattning över 4 kolumner och 4 åldersgrupper. AKU kvartal 1, 2012

Punktskattningar, jämförelsemått, antal >1,96		Osäkerhet, procentuell förändring, genomsnitt		CV, procentuell förändring, genomsnitt	CV, procentuell förändring, extremvärden		CV, andel lägre, procent
kol 1-4		kol 1-4		kol 1-4	min	max	
0		-8,7		-9,9	-24,7	12,8*	81,3

*) 15-24 år, sjuka ej i arbetskraften, 0-11 månader sedan senaste arbete.

Bilaga

Några kommentarer rörande sammanvägning av data från flera urval från samma population
av Martin Axelson

Utgångspunkten är att man har J stycken urval, vilka dras från en och samma population $U = \{1, \dots, k, \dots, N\}$. Låt s_j beteckna det j :te urvalet, vilket drags enligt urvalsdesignen p_j , $j = 1, \dots, J$. Låt

$$I_{j,k} = \begin{cases} 1 & \text{om } k \in s_j \\ 0 & \text{annars} \end{cases}$$

samt låt $\pi_{j,k}$ och $\pi_{j,kl}$ beteckna första respektive andra ordningens inklusionssannolikheter under urvalsdesignen p_j , $j = 1, \dots, J$.

I praktiken är det vanligt att man använder sig av negativ urvalssamordning, i form av en urvalsdragningsmetod som garanterar att $\Pr(s_j \cap s_{j'} = \emptyset) = 1$ för alla par $j \neq j'$. När urvalssamordning används gäller att urvalen s_j och $s_{j'}$ i statistik bemärkelse inte är oberoende, men i många fall är beroendet så svagt att man kan bortse från detta. Framställningen nedan utgår därför från att urvalen s_j , $j = 1, \dots, J$, dras oberoende av varandra.

Horvitz-Thompson-estimation, explicit sammanvägning

Värdet på y_k erhålls för alla $k \in s_j$, $j = 1, \dots, J$. Då gäller att HT-estimatorn

$$\hat{t}_{\pi,j} = \sum_{s_j} \frac{y_k}{\pi_{j,k}}$$

är en väntevärdesriktig estimator för populationstotalen

$$t = \sum_U y_k$$

Låt nu a_j , $j = 1, \dots, J$, beteckna vikter sådana att $\sum_{j=1}^J a_j = 1$. Då gäller att en explicit sammanvägd estimator för $t = \sum_U y_k$ ges av

$$\hat{t}_{e\pi} = \sum_{j=1}^J a_j \hat{t}_{\pi,j}$$

Eftersom $E_{p_j}(\hat{t}_{\pi,j}) = t$, $j = 1, \dots, J$, följer att

$$E(\hat{t}_{e\pi}) = E\left(\sum_{j=1}^J a_j \hat{t}_{\pi,j}\right) = \sum_{j=1}^J a_j E_{p_j}(\hat{t}_{\pi,j}) = t \sum_{j=1}^J a_j = t$$

Vidare följer, eftersom urvalen s_j , $j = 1, \dots, J$, är statistiskt oberoende, att variansen för $\hat{t}_{e\pi}$ ges av

$$V(\hat{t}_{e\pi}) = V\left(\sum_{j=1}^J a_j \hat{t}_{\pi,j}\right) = \sum_{j=1}^J a_j^2 V_{p_j}(\hat{t}_{\pi,j})$$

där

$$V_{p_j}(\hat{t}_{\pi,j}) = \sum \sum_U (\pi_{j,kl} - \pi_{j,k} \pi_{j,l}) \frac{y_k}{\pi_{j,k}} \frac{y_l}{\pi_{j,l}}$$

En väntevärdesriktig estimator för $V(\hat{t}_{e\pi})$ ges av

$$\hat{V}(\hat{t}_{e\pi}) = \sum_{j=1}^J a_j^2 \hat{V}_{p_j}(\hat{t}_{\pi,j})$$

där

$$\hat{V}_{p_j}(\hat{t}_{\pi,j}) = \sum \sum_s \frac{\pi_{j,kl} - \pi_{j,k} \pi_{j,l}}{\pi_{j,kl}} \frac{y_k}{\pi_{j,k}} \frac{y_l}{\pi_{j,l}}$$

Horvitz-Thompson-estimation, implicit sammanvägning

Låt $b_{j,k}$, $j = 1, \dots, J$, vara en icke-stokastisk variabel sådan att $\sum_{j=1}^J b_{j,k} = 1$ för varje $k \in U$. Då gäller att populationstotalen $t = \sum_U y_k$ kan uttryckas som

$$t = \sum_U y_k = \sum_U y_k \sum_{j=1}^J b_{j,k} = \sum_{j=1}^J \sum_U b_{j,k} y_k = \sum_{j=1}^J t_j$$

där $t_j = \sum_U b_{j,k} y_k$, $j = 1, \dots, J$. Eftersom

$$\hat{t}_{j\pi} = \sum_{s_j} \frac{b_{j,k} y_k}{\pi_{j,k}}$$

är en väntevärdesriktig estimator för t_j , gäller att en väntevärdesriktig estimator för $t = \sum_U y_k$ ges av

$$\hat{t}_{i\pi} = \sum_{j=1}^J \hat{t}_{j\pi} = \sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} y_k}{\pi_{j,k}}$$

Vidare följer, eftersom urvalen s_j , $j = 1, \dots, J$, är statistiskt oberoende, att variansen för $\hat{t}_{i\pi}$ ges av

$$V(\hat{t}_{i\pi}) = V\left(\sum_{j=1}^J \hat{t}_{j\pi}\right) = \sum_{j=1}^J V_{p_j}(\hat{t}_{j\pi})$$

där

$$V_{p_j}(\hat{t}_{j\pi}) = \sum \sum_U (\pi_{j,kl} - \pi_{j,k}\pi_{j,l}) \frac{b_{j,k}y_k}{\pi_{j,k}} \frac{b_{j,l}y_l}{\pi_{j,l}}$$

En väntevärdesriktig estimator för $V(\hat{t}_{i\pi})$ ges av

$$\hat{V}(\hat{t}_{i\pi}) = \sum_{j=1}^J \hat{V}_{p_j}(\hat{t}_{j\pi})$$

där

$$\hat{V}_{p_j}(\hat{t}_{j\pi}) = \sum \sum_s \frac{\pi_{j,kl} - \pi_{j,k}\pi_{j,l}}{\pi_{j,kl}} \frac{b_{j,k}y_k}{\pi_{j,k}} \frac{b_{j,l}y_l}{\pi_{j,l}}$$

Kommentar: Om $b_{j,k} = a_j$, $j = 1, \dots, J$ och $k \in U$, så gäller att

$$\hat{t}_{i\pi} = \sum_{j=1}^J \hat{t}_{j\pi} = \sum_{j=1}^J \sum_{s_j} \frac{b_{j,k}y_k}{\pi_{j,k}} = \sum_{j=1}^J a_j \sum_{s_j} \frac{y_k}{\pi_{j,k}} = \sum_{j=1}^J a_j \hat{t}_{\pi,j} = \hat{t}_{e\pi}$$

dvs. explicit sammanvägning är inget annat än ett specialfall av implicit sammanvägning.

GREG^s-estimation, explicit sammanvägning

Låt \mathbf{x} beteckna en vektorvärd variabel för vilken populationstotalen

$\mathbf{t}_x = \sum_U \mathbf{x}_k$ är känd. Värdet på y_k och \mathbf{x}_k erhålls för alla $k \in s_j$, $j = 1, \dots, J$.

Låt

$$g_{j,k} = 1 + \left(\sum_U \mathbf{x}_k - \sum_{s_j} \frac{\mathbf{x}_k}{\pi_{j,k}}\right) \left(\sum_{s_j} \frac{q_k \mathbf{x}_k \mathbf{x}_k'}{\pi_{j,k}}\right)^{-1} q_k \mathbf{x}_k$$

^s Generell Regression

där q_k är en på förhand bestämd vikt. Då gäller att GREG-estimatorn

$$\begin{aligned}\hat{t}_{G,j} &= \sum_{s_j} \frac{g_{j,k} y_k}{\pi_{j,k}} \\ &= \sum_{s_j} \frac{y_k}{\pi_{j,k}} + \left(\sum_U \mathbf{x}_k - \sum_{s_j} \frac{\mathbf{x}_k}{\pi_{j,k}} \right)' \hat{\mathbf{B}}_j \\ &\approx \sum_{s_j} \frac{y_k}{\pi_{j,k}} + \left(\sum_U \mathbf{x}_k - \sum_{s_j} \frac{\mathbf{x}_k}{\pi_{j,k}} \right)' \mathbf{B}\end{aligned}$$

där

$$\hat{\mathbf{B}}_j = \left(\sum_{s_j} \frac{q_k \mathbf{x}_k \mathbf{x}_k'}{\pi_{j,k}} \right)^{-1} \sum_{s_j} \frac{q_k \mathbf{x}_k y_k}{\pi_{j,k}}$$

och

$$\mathbf{B} = \left(\sum_U q_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \sum_U q_k \mathbf{x}_k y_k,$$

är approximativt väntevärdesriktig för populationstotalen $t = \sum_U y_k$. Låt

nu a_j , $j = 1, \dots, J$, beteckna vikter sådana att $\sum_{j=1}^J a_j = 1$. Då gäller att en explicit sammanvägd GREG-estimator för $t = \sum_U y_k$ ges av

$$\hat{t}_{eG} = \sum_{j=1}^J a_j \hat{t}_{G,j}$$

Eftersom $E_{p_j}(\hat{t}_{G,j}) \approx t$, $j = 1, \dots, J$, följer att

$$E(\hat{t}_{eG}) = E\left(\sum_{j=1}^J a_j \hat{t}_{G,j}\right) = \sum_{j=1}^J a_j E_{p_j}(\hat{t}_{G,j}) \approx t \sum_{j=1}^J a_j = t$$

Eftersom urvalen s_j , $j = 1, \dots, J$, är statistiskt oberoende, följer vidare att variansen för \hat{t}_{eG} ges av

$$V(\hat{t}_{eG}) = V\left(\sum_{j=1}^J a_j \hat{t}_{G,j}\right) = \sum_{j=1}^J a_j^2 V_{p_j}(\hat{t}_{G,j}) \approx \sum_{j=1}^J a_j^2 AV_{p_j}(\hat{t}_{G,j})$$

där

$$AV_{p_j}(\hat{t}_{G,j}) = \sum \sum_U (\pi_{j,kl} - \pi_{j,k} \pi_{j,l}) \frac{E_k}{\pi_{j,k}} \frac{E_l}{\pi_{j,l}}$$

med $E_k = y_k - \mathbf{x}_k' \mathbf{B}$. En estimator för $V(\hat{t}_{eG})$ ges av

$$\hat{V}(\hat{t}_{eG}) = \sum_{j=1}^J a_j^2 \hat{V}_{p_j}(\hat{t}_{G,j})$$

där

$$\hat{V}_{p_j}(\hat{t}_{G,j}) = \sum \sum_{s_j} \frac{\pi_{j,kl} - \pi_{j,k}\pi_{j,l}}{\pi_{j,kl}} \frac{g_{j,k}e_{j,k}}{\pi_{j,k}} \frac{g_{j,l}e_{j,l}}{\pi_{j,l}}$$

med $e_{j,k} = y_k - \mathbf{x}'_k \hat{\mathbf{B}}_j$.

GREG-estimation, implicit sammanvägning

Låt $b_{j,k}$, $j = 1, \dots, J$, vara en icke-stokastisk variabel sådan att $\sum_{j=1}^J b_{j,k} = 1$ för

varje $k \in U$. Låt vidare

$$h_k = 1 + \left(\sum_U \mathbf{x}_k - \sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} \mathbf{x}_k}{\pi_{j,k}} \right) \left(\sum_{j=1}^J \sum_{s_j} \frac{q_k b_{j,k} \mathbf{x}_k \mathbf{x}'_k}{\pi_{j,k}} \right)^{-1} q_k \mathbf{x}_k$$

Då gäller att en GREG-estimator för $t = \sum_U y_k$ ges av

$$\begin{aligned} \hat{t}_{iG} &= \sum_{j=1}^J \sum_{s_j} \frac{h_k b_{j,k} y_k}{\pi_{j,k}} \\ &= \sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} y_k}{\pi_{j,k}} + \left(\sum_U \mathbf{x}_k - \sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} \mathbf{x}_k}{\pi_{j,k}} \right)' \hat{\mathbf{B}} \\ &\approx \sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} y_k}{\pi_{j,k}} + \left(\sum_U \mathbf{x}_k - \sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} \mathbf{x}_k}{\pi_{j,k}} \right)' \mathbf{B} \end{aligned}$$

där

$$\hat{\mathbf{B}} = \left(\sum_{j=1}^J \sum_{s_j} \frac{q_k b_{j,k} \mathbf{x}_k \mathbf{x}'_k}{\pi_{j,k}} \right)^{-1} \sum_{j=1}^J \sum_{s_j} \frac{q_k b_{j,k} \mathbf{x}_k y_k}{\pi_{j,k}}$$

och

$$\mathbf{B} = \left(\sum_{j=1}^J \sum_U q_k b_{j,k} \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \sum_{j=1}^J \sum_U q_k b_{j,k} \mathbf{x}_k y_k = \left(\sum_U q_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \sum_U q_k \mathbf{x}_k y_k$$

Eftersom

$$E \left(\sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} \mathbf{x}_k}{\pi_{j,k}} \right) = \sum_{j=1}^J \sum_U b_{j,k} \mathbf{x}_k = \sum_U \mathbf{x}_k$$

gäller att

$$E(\hat{t}_{iG}) \approx E \left(\sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} y_k}{\pi_{j,k}} \right) = E(\hat{t}_{i\pi}) = \sum_U y_k$$

dvs. \hat{t}_{iG} är approximativt väntevärdesriktig för $t = \sum_U y_k$.

Vidare gäller att

$$\hat{t}_{iG} \approx \sum_U \mathbf{x}'_k \mathbf{B} + \sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} (y_k - \mathbf{x}'_k \mathbf{B})}{\pi_{j,k}} = \sum_U \mathbf{x}'_k \mathbf{B} + \sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} E_k}{\pi_{j,k}}$$

vilket implicerar att

$$V(\hat{t}_{iG}) \approx V\left(\sum_{j=1}^J \sum_{s_j} \frac{b_{j,k} E_k}{\pi_{j,k}}\right) = \sum_{j=1}^J \sum_U \sum_U (\pi_{j,kl} - \pi_{j,k} \pi_{j,l}) \frac{b_{j,k} E_k}{\pi_{j,k}} \frac{b_{j,l} E_l}{\pi_{j,l}}$$

där det sista steget följer av att urvalen s_j , $j = 1, \dots, J$, är statistiskt oberoende. En varians estimator ges av

$$\hat{V}(\hat{t}_{iG}) = \sum_{j=1}^J \sum_{s_j} \sum_{s_j} \frac{\pi_{j,kl} - \pi_{j,k} \pi_{j,l}}{\pi_{j,kl}} \frac{h_k b_{j,k} e_k}{\pi_{j,k}} \frac{h_l b_{j,l} e_l}{\pi_{j,l}}$$

där $e_k = y_k - \mathbf{x}'_k \hat{\mathbf{B}}$.

Avslutande kommentarer

Kommentar 1: Den estimationsansats som föreslås svarar i princip mot \hat{t}_{iG} enligt ovan, med valet

$$b_{j,k} = \pi_{j,k} / \sum_{j=1}^J \pi_{j,k}, \quad j = 1, \dots, J$$

för varje $k \in U$. Således gäller att $V(\hat{t}_{iG})$ kan skattas med $\hat{V}(\hat{t}_{iG})$, som utgår från den design enligt vilken varje urval s_j , $j = 1, \dots, J$, dragits.

Kommentar 2: Resultaten ovan bygger på att det är rimligt att anta att urvalen s_j , $j = 1, \dots, J$, är statistiskt oberoende. När så ej är fallet måste man fundera på vilka konsekvenser avsaknaden av oberoende får. Om den negativa urvalssamordningen skapar ett icke-ignorerbart beroende, bör det vara möjligt att utnyttja flerfasteori. För fallet då man har $J = 2$ perfekt negativt samordnade urval kan man t.ex. utgå från innehållet i avsnitt 9.9 i Särndal, Swensson, Wretman (1992).

- 2007:1 Några väsentliga sysselsättningsbegrepp i den officiella statistiken
- 2007:2 Registerbaserad aktivitetsstatistik
- 2008:1 Ungdomar utan fullföljd gymnasieutbildning – en undersökning med många utmaningar
- 2009:1 Longitudinell Integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA) 1990–2007
- 2010:1 Principiella grunder för Arbetskraftsundersökningarna (AKU) och arbetsmarknadsstatistiken
- 2010:2 Rekryteringsstatistik från AKU
- 2011:1 Basic principles for Labour Force Surveys (LFS) and labour market statistics
- 2011:2 Recruitment Statistics for the Swedish Labour Force Survey
- 2011:3 Arbetskraftsundersökningarna (AKU) 50 år. Fyra forskarperspektiv på arbetsmarknaden
- 2011:4 Longitudinell integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA) 1990–2009
- 2011:5 Yrkesregistret med yrkesstatistik. En beskrivning av innehåll och kvalitet
- 2011:6 Urvals- och estimationsförfarandet i de svenska arbetskraftsundersökningarna (AKU)
- 2011:7 Konjunkturberoende i inflödet till och utflödet från högre studier
- 2012:1 Actual Hours Worked in the Swedish LFS. Four articles
- 2013:1 Ungdomsarbetslöshet – jämförbarhet i statistiken mellan ett antal europeiska länder
- 2013:2 Revising Surveys – Linking Old and New Data
- 2013:3 Consistent Seasonal Adjustment and Trend-cycle Estimation
- 2013:4 Youth unemployment – comparability in statistics between a number of European countries
- 2013:5 Länkning av centrala serier i Arbetskraftsundersökningarna (AKU) 1970–1986
- 2013:6 Deltidssysselsatta arbetssökande
- 2013:7 Part-time employed job seekers
- 2013:8 Arbetskraftsundersökningarnas (AKU) och Arbetsförmedlingens (Af) arbetslöshetsstatistik – En jämförande studie

ISSN 1654-465X (Online)

All officiell statistik finns på: **www.scb.se**

Statistikservice: tfn 08-506 948 01

All official statistics can be found at: **www.scb.se**

Statistics Service service, phone +46 8 506 948 01