

2003-05-28

Folkhälsoenkät i Stockholms län 2002 – dokumentation av viktberäkningar

1. Population och redovisningsgrupper

Populationen utgörs av folkbokförda personer i åldern 18-84 år i Stockholms län. Viktiga redovisningsgrupper utgörs av kön*åldersklasser (i SU; se avsnitt 3) och kön*områden*åldersklasser (i TS1 och TS2). Klassindelningar för ålder som nämnts är 18-45, 46-65 och 66-84 samt indelningar i 5-års-klasser.

2. Parametrar

I de flesta fallen önskas information om parametrar av typen procentuella andelen personer med viss egenskap, t.ex. procentuella andelen med dålig hälsa. Motsvarande skattningar önskas också för redovisningsgrupper.

3. Urval och svarsmängd

Undersökningen består av tre olika delar, nämligen tvärsnittsstudien (TS), särskilda urvalet (SU) samt uppföljningsstudien (US). Vikterna i uppföljningsstudien ska inte kalibreras och därför beskrivs den inte närmare i detta dokument.

I datainsamlingen har två olika blanketter använts. Huvudblanketten (rött färgband) har använts i SU och i halva urvalet till TS. Blanketten med vitt omslag har använts till den andra halvan av urvalet i TS. Blanketterna är i stort sett lika, skillnaden består endast av att blanketten med vitt omslag innehåller en extra fråga och en fråga har en annan placering. Det innebär att TS består av två mängder som behöver var sitt viktsystem, nämligen

TS1: Urval på 50 067 personer med samtliga frågor utom en

TS2: Urvalet på 21 033 med en enda fråga

I SU drogs ett stratifierat obundet slumpmässigt (STOSU) med kön som stratifieringsvariabel. Från respektive strata drogs 4 000 personer. Det innebär att vi har följande mängd:

SU: Urval på 8000 personer (med extra bortfallsuppföljning)

Urvalet till TS1 gjordes i form av en sammanläggning av två urval med syfte att få ett STOSU där kön*område utgör strata. Områdena består av 25 kommuner + en kommun uppdelad i 18 kommundelar, d.v.s totalt 43

2003-05-28

områden. Vissa personer saknade uppgift om områdestillhörighet och fick därför bilda ett särskilt "områdes-strata". Den första urvalsdelen bestod av urvalet i SU som fördelades på de 88 (2*44) strata. Därefter konstaterades hur många ytterligare som behövde dras från de 88 strata för att uppnå 550 i respektive strata (endast 50 drogs ur de två strata som bildades av personer utan uppgift om områdestillhörighet). Denna andra urvalsdragning gjordes med negativ urvalssamordning, d.v.s. sedan den första urvalsdelen plockats bort från den ursprungliga urvalsramen.

Kalibreringen av vikter ska göras för SU och TS1.

Datainsamlingen i SU har resulterat i svar från 5579 personer, d.v.s. 69.7% av de utvalda (0.5% av de utvalda vet vi säkert tillhör övertäckningen). För TS1 har 62.9 % svarat. Svarsandelarna varierar kraftigt mellan olika grupper av utvalda, vilket tabellerna 2-8 resp. 17-23 visar.

4. Olika brister i undersökningen

I en urvalsundersökning är alltid skattningarna behäftade med *urvalsfel* beroende på att endast en delmängd (urval) av populationen studeras. Ett annat fel, *bortfallsfel*, uppkommer p.g.a. att vi inte lyckas få svar från vissa personer (bortfall) och att de avviker från de svarande beträffande det ämne undersökningen handlar om.

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 erhållas på ett indirekt och approximativt sätt genom att utnyttja registervariabler.

Både urvalsfelet och bortfallsfelet kan reduceras genom att använda ett effektivt uppräkningsförfarande. I följande avsnitt visas hur det görs i denna undersökning.

2003-05-28

5. Uppräkningsförfarande (estimation)

5.1. Allmänt om kalibrering av vikter

I undersökningen använder vi tekniken med kalibrering av vikter, vilket är en effektiv och flexibel metod (se Lundström och Särndal, 1999, och Lundström och Särndal, 2001) för att reducera urvalsfelet och bortfallsfelet. I korthet består tekniken i att utnyttja ”starka” hjälpvariabler (registervariabler). De bildade kalibreringsvikterna är sådana att uppräkningsgenomgången ger exakta skattningar av de hjälptotaler som bildas av hjälpvariablerna.

Med ”stark” hjälpvariabel menas en variabel som uppfyller en eller flera av följande egenskaper:

(i) variabeln samvarierar väl med svars-benägenheten (-sannolikheten),

eller/och

(ii) den samvarierar väl med (viktiga) målvariabler,

eller/och

(iii) den avgränsar (viktiga) redovisningsgrupper.

I följande avsnitt redovisas de presumtiva hjälpvariablerna och den analys som leder till det slutliga valet.

5.2. Presumtiva hjälpvariabler

Urvalet har dragits från Registret över totalbefolkningen (RTB) och från detta register hämtas också hjälpinformation. För att ytterligare reducera felet hämtas också hjälpinformation från Utbildningsregistret (UTB).

De RTB-variabler som vi i en första gallring funnit intressanta är *kön, ålder, födelseland, inkomst, civilstånd* och av UTB-variablerna *utbildningsnivå*. Även tillhörighet till *område* är intressant.

Vi gör redan i ett första steg hopslagningar av kategorier. Resultatet presenteras i tabell 1 (med undantag för områdestillhörighet).

2003-05-28

Tabell 1. Presumptiva hjälpvariabler

Variabel	Koder
Kön	Man; kvinna
Ålder	18-20; 21-25; 26-30; 31-35; 36-40; 41-45; 46-50; 51-55; 56-60; 61-65; 66-70; 71-75; 76-80; 81-84 år
Födelseland	Födda i norden; övriga
Inkomst	0-149; 150-299; 300- (tusental kr)
Civilstånd	Gift eller registrerat partnerskap; övriga
Utbildningsnivå	Kod 1-7; se bilaga 1
Område	De 43 områdena som används i stratifieringen i TS1

5.3. Analys av presumtiva hjälpvariabler i SU

5.3.1. Variabeln samvarierar med svarsbenägenheten

För att uppfylla egenskap (i) studeras sambandet mellan den dikotoma variabeln svarande/bortfall och hjälpvariablerna. Skattad andel svarande i olika grupper bestämda av respektive hjälpvariabel har därvid beräknats. Vid skattningen används en ”konventionell” teknik utan utnyttjande av hjälpinformation (se anmärkning 6.1).

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

Tabell 2. Skattad procentuell andel svarande fördelat på kön.

Kön	Man	Kvinna
Svarsandel (%)	64.6	74.9

2003-05-28

Tabell 3. Skattad procentuell andel svarande fördelat på åldersklasser.

Ålder	18-20	21-25	26-30	31-35	36-40	41-45	46-50
Svarsandel (%)	64.9	61.3	62.3	63.8	69.6	66.3	72.1

Ålder	51-55	56-60	61-65	66-69	71-75	76-80	81-84
Svarsandel (%)	71.2	75.0	79.9	76.3	79.4	80.8	72.1

Tabell 4. Skattad procentuell andel svarande fördelat på födelseland.

Födelseland	Norden	Övriga
Svarsandel (%)	73.6	50.9

Tabell 5. Skattad procentuell andel svarande fördelat på inkomst.

Inkomst	0-149	150-299	300-
Svarsandel (%)	60.4	74.1	77.3

Tabell 6. Skattad procentuell andel svarande fördelat på civilstånd.

Civilstånd	Gift eller registrerat partnerskap	Övriga
Svarsandel (%)	75.4	65.9

Tabell 7. Skattad procentuell andel svarande fördelat på utbildningsnivå.

Utbildningsnivå	1	2	3	4	5	6	7
Svarsandel (%)	73.0	22.1	62.3	70.2	74.3	77.3	66.2

Vi ser av ovanstående tabeller att skattade svarsandelar varierar kraftigt mellan olika kategorier för alla presumtiva hjälpvariabler. Tabell 3 visar dock att vissa åldersklasser kan slås samman. I en sådan sammanslagning beaktar vi också vilka redovisningsgrupper som ska studeras, d.v.s. vi tar hänsyn till egenskap (iii). Vi har i vissa undersökningar sett stora skillnader i svarsbenägenhet mellan män och kvinnor inom olika åldrar. Därför har vi tagit fram tabell 8 där också nämnda sammanslagning av åldersklasser är gjorda.

2003-05-28

Tabell 8. Skattad procentuell andel svarande fördelat på kön*åldersklasser.

Kön/ålder	Män					Kvinnor				
	18-35	36-45	46-65	66-80	81-84	18-35	36-45	46-65	66-80	81-84
Svarsandel (%)	55.1	60.6	70.1	81.1	77.6	70.9	75.8	77.9	76.8	69.0

Tabell 8 visar att svarsbenägenheten över åldersklasserna är olika för män och kvinnor och därför är det motiverat att ta med information från denna korstabell i kalibreringen av vikter.

Vi har även skattat den procentuella andelen svarande bland de 43 områdena och funnit att den minsta andelen, 47.5, finns i område 202 och den största, 80.6, finns i område 103.

Låt oss också titta på i vilken mån dessa variabler uppfyller egenskap ii).

5.3.2. Variabeln samvarierar med (viktiga) målvariabler

Det är inte möjligt att hitta hjälpvariabler som samvarierar med alla målvariabler. Därför väljer vi att studera samvariationen med ett antal viktiga målvariabler. Vi bildar av de ursprungliga variablerna dikotoma variabler. I tabell 9 beskrivs de variabler vi valt.

Tabell 9. Viktiga målvariabler

Tilldelat variabelnamn	Förklaring (se även frågeformuläret)
Gott allmäntillstånd	Fr 3; Mycket gott/gott
Dagligrökare	Fr 22
Lita på grannarna	Fr 61_1; Stämmer mycket väl/ganska väl
God ekonomi	Fr 63; Kan alltid skaffa fram 14000 kr
Frisk	Fr 77; Sjukfrånvarande ≤ 7 dagar
Utvecklas inte i arbetet	Fr 89; Inte alls

I tabellerna 10-16 skattas parametrar för redovisningsgrupper bestämda av hjälpvariablerna. Alla parametrar utgör procentuell andel med viss egenskap. Skattningarna görs med hjälp av ”konventionell” teknik; se (6.4).

2003-05-28

Tabell 10. Kön

Variabel	Man	Kvinna
Gott allmäntillstånd	72.3	66.1
Dagligrökare	15.4	18.5
Litar på grannarna	85.4	84.5
God ekonomi	60.1	45.9
Frisk	57.8	48.6
Utvecklas inte i arbetet	5.0	4.0

Tabell 11. Ålder

Variabel	18-20	21-25	26-30	31-35	36-40	41-45	46-50
Gott allmäntillstånd	76.7	77.8	81.5	78.8	75.2	69.2	70.1
Dagligrökare	16.4	13.8	10.3	13.4	16.7	21.4	21.9
Litar på grannarna	78.1	78.7	80.5	86.4	86.6	84.9	83.8
God ekonomi	38.2	43.4	51.6	54.4	53.8	45.4	54.0
Frisk	37.8	42.8	65.3	65.7	68.7	64.1	68.3
Utvecklas inte i arbetet	5.5	7.8	2.9	5.3	6.6	5.2	5.8

Variabel	51-55	56-60	61-65	66-69	71-75	76-80	81-84
Gott allmäntillstånd	70.6	64.0	62.0	61.7	53.4	49.8	39.3
Dagligrökare	23.4	20.7	19.5	14.0	14.1	12.3	7.6
Litar på grannarna	87.0	89.8	87.1	80.9	86.7	87.7	83.8
God ekonomi	55.3	56.9	61.5	55.5	52.1	54.9	57.0
Frisk	70.4	62.1	45.8	14.0	11.1	10.2	6.2
Utvecklas inte i arbetet	4.2	5.0	4.8	0.8	1.1	0.4	0.7

2003-05-28

Tabell 12. Födelseland

Variabel	Norden	Övriga
Gott allmäntillstånd	70.4	59.8
Dagligrökare	16.7	18.8
Litar på grannarna	87.3	67.3
God ekonomi	55.3	35.2
Frisk	53.9	47.0
Utvecklas inte i arbetet	3.6	10.6

Tabell 13. Inkomst

Variabel	<149999	150000-299999	3000000-
Gott allmäntillstånd	62.7	66.9	81.1
Dagligrökare	17.6	19.0	12.5
Litar på grannarna	78.4	84.1	94.5
God ekonomi	38.4	49.1	77.5
Frisk	39.1	52.2	72.1
Utvecklas inte i arbetet	5.4	5.3	1.7

Tabell 14. Civilstånd

Variabel	Gift eller registrerat partnerskap	Övriga
Gott allmäntillstånd	71.1	67.6
Dagligrökare	14.4	19.0
Litar på grannarna	88.5	82.0
God ekonomi	60.5	46.7
Frisk	56.8	50.1
Utvecklas inte i arbetet	3.2	5.5

2003-05-28

Tabell 15. Utbildningsnivå

Variabel	1	2	3	4	5	6	7
Gott allmäntillstånd	47.3	61.7	59.8	68.8	77.5	78.3	84.7
Dagligrökare	10.7	19.8	27.3	20.6	12.5	9.0	6.6
Litar på grannarna	86.2	58.6	78.8	83.1	87.2	90.5	93.4
God ekonomi	54.6	22.4	38.0	47.9	62.6	65.1	85.0
Frisk	10.9	34.6	42.8	54.4	58.6	66.7	74.6
Utvecklas inte i arbetet	1.0	12.8	8.8	4.7	4.6	2.4	0.0

Tabell 16. Kön*ålder

Kön/ålder	Män					Kvinnor				
	18-35	36-45	46-65	66-80	81-84	18-35	36-45	46-65	66-80	81-84
Gott allmäntillstånd	82.4	74.3	70.5	59.2	40.7	76.2	70.9	63.2	51.3	38.2
Dagligrökare	9.5	18.2	19.7	13.1	10.2	15.9	19.3	23.2	13.8	5.6
Litar på grannarna	82.0	86.7	87.6	85.0	83.1	81.7	85.1	86.4	85.0	84.3
God ekonomi	55.3	56.4	64.5	63.3	62.7	43.5	44.3	48.7	45.2	52.8
Frisk	61.9	73.1	66.3	15.8	8.5	52.9	60.7	59.0	8.0	4.5
Utvecklas inte i arbetet	5.2	6.7	5.7	1.1	1.7	5.0	5.3	4.1	0.5	0.0

Vi ser att de flesta registervariablerna förklarar en viss del av variationen i de viktiga undersökningsvariablerna.

5.3.3. Variabeln avgränsar (viktiga) redovisningsgrupper

Redovisningsgrupperna bestäms av *kön*åldersklasser* och därför är det viktigt att dessa variabler används i kalibreringen.

5.3.4. Slutligt val av hjälpinformation

Den hjälpvektor vi valt att utnyttja i kalibreringen kan skrivas som *kön*åldersklasser+födelseland+inkomst+civilstånd+utbildningsnivå*. Detta val baserar sig på analysen i avsnitten 5.3.1-5.3.3 samt några olika beräkningar av kalibrerade vikter. Vi tar inte med *område* eftersom antalet svarande är litet i vissa områden och därför skulle ge alltför varierande vikter

2003-05-28

(orsakar stor varians) om den hjälvariabeln användes. Dessutom kommer inte skattningar för områden att tas fram i denna del av undersökningen.

De kalibrerade vikterna är sådana att vid skattning av antalet personer i varje grupp bestämda av *kön*åldersklasser* överensstämmer skattningen helt med det kända antalet (från RTB). Vikterna är också sådana att skattningen av antalet personer i olika utbildningsnivåer överensstämmer helt med kända totaler från UTB. Överensstämmelse med RTB-totaler gäller också för variabelerna födelse-land, inkomst och civilstånd.

Vi kan förvänta oss att bortfallsskevheten reduceras kraftigt genom utnyttjandet av denna hjälpinformation. Detta beroende på att variationen i svarsbenägenhet förklaras väl (egenskap (i)). Dessutom samvarierar också hjälpvariablerna till betydande del med de viktiga undersökningsvariablerna (egenskap (ii)), vilket torde något reducera urvalsfelet. Det är också ett ytterligare skäl till att tro att bortfallsskevheten är liten.

5.4. Analys av presumtiva hjälpvariabler i TS1

5.4.1. Inledning

Vi gör i detta avsnitt motsvarande analys av TS1-materialet som för SU-materialet. Därför upprepar vi inte här motiveringen till beräkningarna.

5.4.2. Variabeln samvarierar med svarsbenägenheten

Tabell 17. Skattad procentuell andel svarande fördelat på kön.

Kön	Man	Kvinna
Svarsandel (%)	57.0	68.5

Tabell 18. Skattad procentuell andel svarande fördelat på åldersklasser.

Ålder	18-20	21-25	26-30	31-35	36-40	41-45	46-50
Svarsandel (%)	52.8	51.8	55.8	57.8	57.7	58.9	63.3

Ålder	51-55	56-60	61-65	66-69	71-75	76-80	81-84
Svarsandel (%)	66.8	72.1	73.6	75.8	74.3	72.9	68.6

2003-05-28

Tabell 19. Skattad procentuell andel svarande fördelat på födelseland.

Födelseland	Norden	Övriga
Svarsandel (%)	66.7	42.9

Tabell 20. Skattad procentuell andel svarande fördelat på inkomst.

Inkomst	0-149	150-299	300-
Svarsandel (%)	52.4	67.3	71.7

Tabell 21. Skattad procentuell andel svarande fördelat på civilstånd.

Civilstånd	Gift eller registrerat partnerskap	Övriga
Svarsandel (%)	69.5	58.2

Tabell 22. Skattad procentuell andel svarande fördelat på utbildningsnivå.

Utbildningsnivå	1	2	3	4	5	6	7
Svarsandel (%)	68.3	22.3	53.7	62.1	66.5	71.6	63.3

Tabell 23. Skattad procentuell andel svarande fördelat på kön*åldersklasser.

Kön/ålder	Män					Kvinnor				
	18-35	36-45	46-65	66-80	81-84	18-35	36-45	46-65	66-80	81-84
Svarsandel (%)	46.9	50.8	64.1	75.5	75.6	63.4	66.0	73.3	73.6	64.9

Vi har även skattat den procentuella andelen svarande bland de 43 områdena. Det klart minsta värdet, 29.8, finns i område 202. Låga värden finns också i områdena 201 och 203, nämligen 50.7 resp. 52.4. De största värdena finns i område 106 och 114, nämligen 69.9 resp. 69.2. Det är alltså rätt stor variation i svarsbenägenhet mellan områdena, även om de flesta ligger nära populationens värde.

Vi ser av ovanstående tabeller att skattade svarsandelar varierar kraftigt mellan olika kategorier för alla presumtiva hjälpvariabler. Låt oss också titta på i vilken mån dessa variabler uppfyller egenskap ii).

2003-05-28

5.4.3. Variabeln samvarierar med (viktiga) målvariabler

Vi studerar samma variabler som i SU-materialet. Resultatet presenteras i tabellerna 24-30.

Tabell 24. Kön

Variabel	Man	Kvinna
Gott allmäntillstånd	71.8	67.4
Dagligrökare	15.4	17.3
Litar på grannarna	85.7	84.7
God ekonomi	59.1	45.8
Frisk	59.3	48.9
Utvecklas inte i arbetet	4.7	4.1

Tabell 25. Ålder

Variabel	18-20	21-25	26-30	31-35	36-40	41-45	46-50
Gott allmäntillstånd	76.3	77.5	81.2	79.4	76.4	71.3	71.3
Dagligrökare	14.3	16.4	10.7	12.3	15.5	20.1	21.9
Litar på grannarna	77.6	75.8	81.1	84.8	87.3	84.7	86.9
God ekonomi	31.2	39.5	50.6	51.7	49.4	45.8	52.5
Frisk	38.6	46.5	66.2	68.8	69.4	68.7	69.8
Utvecklas inte i arbetet	7.8	6.0	3.9	4.7	5.3	5.6	6.1

Variabel	51-55	56-60	61-65	66-69	71-75	76-80	81-84
Gott allmäntillstånd	69.0	64.2	63.6	65.3	52.5	48.3	40.3
Dagligrökare	20.7	21.9	17.2	14.2	14.0	10.8	7.8
Litar på grannarna	88.1	88.5	87.6	85.5	87.3	86.3	85.0
God ekonomi	55.7	59.7	61.8	58.9	54.1	56.2	56.9
Frisk	68.6	63.6	43.5	15.0	10.4	8.0	8.9
Utvecklas inte i arbetet	5.2	5.2	4.2	0.7	0.7	0.2	0.8

2003-05-28

Tabell 26. Födelseland

Variabel	Norden	Övriga
Gott allmäntillstånd	70.8	60.4
Dagligrökare	15.8	20.7
Litar på grannarna	87.6	66.9
God ekonomi	54.7	34.0
Frisk	54.6	49.8
Utvecklas inte i arbetet	3.7	9.8

Tabell 27. Inkomst

Variabel	<149999	150000-299999	3000000-
Gott allmäntillstånd	62.6	67.8	80.8
Dagligrökare	18.2	18.2	11.1
Litar på grannarna	77.8	85.4	93.6
God ekonomi	36.2	48.9	77.2
Frisk	39.9	53.1	72.0
Utvecklas inte i arbetet	5.3	5.4	1.6

Tabell 28. Civilstånd

Variabel	Gift eller registrerat partnerskap	Övriga
Gott allmäntillstånd	70.2	69.0
Dagligrökare	13.8	18.6
Litar på grannarna	88.6	82.3
God ekonomi	59.8	45.9
Frisk	55.7	52.6
Utvecklas inte i arbetet	3.6	5.1

2003-05-28

Tabell 29. Utbildningsnivå

Variabel	1	2	3	4	5	6	7
Gott allmäntillstånd	47.6	62.8	59.6	68.9	77.4	79.5	80.8
Dagligrökare	10.3	22.6	25.5	19.5	12.7	9.6	10.6
Litar på grannarna	85.7	71.2	79.3	83.8	88.2	89.7	91.7
God ekonomi	55.6	29.0	37.8	46.7	58.9	64.9	79.6
Frisk	11.3	40.9	40.5	56.2	60.6	67.2	79.8
Utvecklas inte i arbetet	0.7	7.1	8.0	5.6	3.3	2.1	1.4

Tabell 30. Kön*ålder

Kön/ålder	Män					Kvinnor				
	18-35	36-45	46-65	66-80	81-84	18-35	36-45	46-65	66-80	81-84
Gott allmäntillstånd	81.2	75.8	69.7	58.9	42.5	77.5	72.5	64.3	53.4	38.7
Dagligrökare	10.7	15.9	19.9	13.6	8.3	14.7	19.3	21.4	12.7	7.5
Litar på grannarna	80.9	86.3	88.5	87.5	84.9	80.8	85.9	87.2	85.2	85.1
God ekonomi	53.4	54.6	63.5	63.9	62.8	40.5	41.2	51.0	49.5	52.4
Frisk	65.6	74.6	67.1	15.2	13.9	54.9	63.7	57.3	7.8	5.2
Utvecklas inte i arbetet	4.8	6.1	5.9	0.6	1.4	5.3	4.8	4.5	0.5	0.3

Vi ser att de flesta registervariablerna förklarar en viss del av variationen i de viktiga undersökningsvariablerna.

5.4.4. Variabeln avgränsar (viktiga) redovisningsgrupper

Redovisningsgrupperna bestäms av *kön*åldersklasser* och därför är det viktigt att dessa variabler används i kalibreringen. Dessutom kommer resultaten att redovisas för de 43 delområdena och därför är det viktigt att variabeln *område* är med i hjälvektorn.

5.4.5. Slutligt val av hjälpinformation

Den hjälpvektor vi valt att utnyttja i kalibreringen kan skrivas som *kön*åldersklasser+födelseland+inkomst+civilstånd+utbildningsnivå+område*. Detta val baserar sig på analysen i avsnitten 5.4.2-5.4.4 samt några olika beräkningar av kalibrerade vikter.

2003-05-28

De kalibrerade vikterna är sådana att vid skattning av antalet personer i varje grupp bestämda av *kön*åldersgrupper* överensstämmer skattningen helt med det kända antalet (från RTB). Vikterna är också sådana att vid skattningen av antalet personer i olika utbildningsnivåer överensstämmer helt med kända totaler (från UTB). Överensstämmelse med RTB-uppgifter gäller också för variabelerna födelseland, inkomst, civilstånd och område.

Vi kan förvänta oss att bortfallsskevheten reduceras kraftigt genom utnyttjandet av denna hjälpinformation. Detta beroende på att variationen i svarsbenägenhet förklaras väl (egenskap (i)). Dessutom samvarierar också hjälpvariablerna till betydande del med de viktiga undersökningsvariablerna (egenskap (ii)), vilket torde något reducera urvalsfelet. Det är också ett ytterligare skäl till att tro att bortfallsskevheten är liten. Även egenskap (iii) beaktar vi och därmed bör särskilt urvalsfelet minska för skattningar av domänstorheter.

6. Teknisk beskrivning av urval och estimation

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

$$y_k = \begin{cases} 1 & \text{om person } k \text{ har studerade egenskap} \\ 0 & \text{för övrigt} \end{cases} \quad (6.1)$$

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

$$Y_d = \sum_U y_{dk} \quad (6.2)$$

$$\text{där } y_{dk} = \begin{cases} y_k & \text{för } k \in U_d \\ 0 & \text{för övrigt.} \end{cases}$$

Z_d bildas på likartat sätt.

2003-05-28

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 (6.1). Om vi låter N_d vara antalet personer i redovisningsgrupp d , då kan parametern skrivas

$$P_d = 100 \frac{\sum_U y_{dk}}{N_d} \quad (6.3)$$

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 svarmängden r_h av storleken m_h att utföra beräkningarna på. Storleken på stratum h ger vi beteckningen N_h .

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

$$\hat{Y}_d = \sum_{h=1}^H \frac{N_h}{m_h} \sum_{r_h} y_{dk} \quad (6.4)$$

I estimator (6.4) används ingen ytterligare hjälpinformation än stratifieringsinformationen.

I syfte att erhålla en estimator med mindre varians och bortfallsskevhet än estimator (6.4) utnyttjar vi hjälpinformation också i estimationen. Vi bildar en hjälpvektor \mathbf{x}_k , som anger till vilka kategorier av *kön*åldersklasser+födelseland+inkomst+civilstånd+utbildningsnivå+område* (för TS1) som person k hör. Från RTB och UTB framställer vi hjälptotalerna $\sum_U \mathbf{x}_k$. Vi utnyttjar denna hjälpinformation i en kalibreringsestimator.

Kalibreringsestimatorn för totalen Y_d har följande utseende:

$$\hat{Y}_{wd} = \sum_r w_k y_{dk} \quad (6.5)$$

där $w_k = d_k v_k$ med

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

och

2003-05-28

$$v_k = 1 + (\sum_U \mathbf{x}_k - \sum_r d_k \mathbf{x}_k)' (\sum_r d_k \mathbf{x}_k \mathbf{x}_k')^{-1} \mathbf{x}_k \quad (6.6)$$

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

Anmärkning 6.1. I tabellerna 2-8 resp. 17-23 beräknas viktat procentuell andel objektsbortfall utifrån det ursprungliga urvalet s_h ($h = 1, \dots, H$) med hjälp av vikterna N_h / n_h .

Anmärkning 6.2. I den slutliga filen står VIKT1 för w_k .

7. Åldersstandardisering

7.1. Allmän beskrivning

I syfte att underlätta jämförelsen mellan olika redovisningsgrupper kommer skattningarna att i ett senare skede ålderstandardiseras. Vilka redovisningsgrupper som ska jämföras är inte bestämt i nuläget och därför kan inte korrigeringsvikten framställas, men för att underlätta beräkningen har vi tagit fram en delkomponent. Korrigeringsvikten erhålles sedan genom att delkomponenten divideras med (det skattade) antalet personer i redovisningsgruppen som tillhör åldersgruppen.

Åldersstandardisering påverkas av vilken standardpopulation och vilka åldersklasser som väljs. I denna undersökning kommer två olika standardpopulationer att användas, nämligen rikets befolkning resp. länets den 31/12, 2002. Fördelningen på åldersklasser framgår av tabell 31 och 32.

Tabell 31. Procentuell andel personer i olika åldersklasser i riket den 31/12 2002.

Ålder	18-20	21-25	26-30	31-35	36-40	41-45	46-50
Procentuell andel	4.5	7.7	8.6	9.2	9.6	8.6	8.6

Ålder	51-55	56-60	61-65	66-69	71-75	76-80	81-84
Procentuell andel	9.2	9.1	6.6	5.5	5.1	4.6	3.0

2003-05-28

Tabell 32. Procentuell andel personer i olika åldersklasser i länet den 31/12 2002.

Ålder	18-20	21-25	26-30	31-35	36-40	41-45	46-50
Procentuell andel	4.1	7.8	10.6	10.9	10.9	9.1	8.4

Ålder	51-55	56-60	61-65	66-69	71-75	76-80	81-84
Procentuell andel	8.7	8.9	5.9	4.5	4.0	3.8	2.5

7.2. Teknisk beskrivning

Antag att vi vill jämföra redovisningsgrupperna $U_1, \dots, U_d, \dots, U_D$ avseende förekomsten av en viss egenskap. Resultattabellerna ger skattningar av $P_1, \dots, P_d, \dots, P_D$ (se 6.5). Låt oss anta att vi valt åldersgrupperna $U_1, \dots, U_q, \dots, U_Q$ som har storleken $N_1, \dots, N_q, \dots, N_Q$. Då kan P_d skrivas

$$P_d = 100 \frac{\sum_U y_{dk}}{N_d} = 100 \sum_{q=1}^Q \frac{N_{dq}}{N_d} \frac{1}{N_{dq}} \sum_{U_q} y_{dk}, \text{ där } N_{dq} \text{ är antalet personer i}$$

åldersklass q som tillhör redovisningsgrupp d . Kvoten $\frac{N_{dq}}{N_d}$ anger för

redovisningsgrupp d andelen av individerna som tillhör åldersklass q . Vid åldersstandardisering ersätts dessa storheter med motsvarande storheter i standardpopulationen. Antag att de senare andelarna är A_q (i tabeller 31 och 32 anges $100 \times A_q$). Då är den åldersstandardiserade andelen med studerade egenskap i område d :

$$P_{dSTAND} = 100 \sum_{q=1}^Q \frac{1}{N_{dq}} \sum_{U_q} A_q y_{dk} \quad (7.1)$$

I undersökningen kan P_{dSTAND} skattas med

$$\hat{P}_{dSTAND} = 100 \sum_{q=1}^Q \frac{1}{\hat{N}_{dq}} \sum_{r_q} w_k A_q y_{dk}$$

där r_q är svarmängden i åldersklass q , $\hat{N}_{dq} = \sum_{r_q} w_k I_k$ och

$$I_k = \begin{cases} 1 & \text{om } k \in U_d \\ 0 & \text{för övrigt} \end{cases}.$$

2003-05-28

En annat alternativ (vanligtvis något sämre) vid skattning av P_{dSTAND} är att sätta $\hat{N}_{dq} = N_{dq}$.

Anmärkning 7.1. I den slutliga filen står VIKT2 för A_q med riket som standardpopulation och VIKT3 för A_q med länet som standardpopulation.

Referenser

Lundström, S. and Särndal, C.E. (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, 15, 305-327.

Lundström, S. and Särndal, C.E. (2001). Estimation in the Presence of Nonresponse and Frame Imperfections. *Statistics Sweden*.

2003-05-28

Bilaga 1

Utbklass	SUN2000Niva	SUN2000NivaText
1	Saknas (blank)	
2	999	Okänt
3	100 102 106 200 204 206	Övrig/ospec förgymnasial utbildning kortare än 9 år Grundskoleutbildning, årskurs 1-6 Folkskoleutbildning Övrig/ospec förgymnasial utbildning, 9 (10) år Realskoleutbildning Grundskoleutbildning, årskurs 7-9
4	310 312 313 316 317 320 322 323 326 327 330 332 333 336 337	Ospecc gymnasial utbildning, kortare än två år Gymn. utb. kortare än två år, teoretisk/studieförb, ej slutbetyg Gymnasial utb. kortare än två år, yrkesinriktad, ej slutbetyg Gymnasial utbildning kortare än två år, teoretisk/studieförb. Gymnasial utbildning kortare än två år, yrkesinriktad Ospecc gymnasial utbildning, två år Gymnasial utb. två år, teoretisk/studieförb. - ej slutbetyg Gymnasial utbildning två år, yrkesinriktad - ej slutbetyg Gymnasial utbildning två år, teoretisk/studieförb. Gymnasial utbildning två år, yrkesinriktad Ospecc gymnasial utbildning, tre år Gymnasial utb. tre år, teoretisk/studieförb, ej slutbetyg Gymnasial utbildning tre år, yrkesinriktad, ej slutbetyg Gymnasial utbildning tre år, teoretisk/studieförb. Gymnasial utbildning tre år, yrkesinriktad
5	410 412 413 415 417	Ospecc eftergymnasial utbildning kortare än två år Högskoleutbildning, 20 högskolepoäng Gymnasial påbyggnadsutbildning Eftergymn. utb. kortare än två år - ej universitet/högskola Högskoleutbildning kortare än två år, yrkesinriktad
6	520 522 525 526 527 530 532 535 536 537 540 545 546 547 550 555 556 557	Ospecc eftergymnasial utbildning två år Högskoleutbildning, 80 högskolepoäng - ej examen Eftergymn. utb. två år - ej universitet/högskola Högskoleutbildning två år, generell Högskoleutbildning två år, yrkesinriktad Ospecc eftergymnasial utbildning tre år Högskoleutbildning, 120 högskolepoäng - ej examen Eftergymn. utb. tre år - ej universitet/högskola Högskoleutbildning tre år, generell Högskoleutbildning tre år, yrkesinriktad Ospecc eftergymnasial utbildning fyra år Eftergymn. utb. fyra år - ej universitet/högskola Högskoleutbildning fyra år, generell Högskoleutbildning fyra år, yrkesinriktad Ospecc eftergymnasial utbildning minst fem år Eftergymn. utb. minst fem år - ej universitet/högskola Högskoleutbildning fem år eller längre, generell Högskoleutbildning fem år eller längre, yrkesinriktad

2003-05-28

7	600	Övrig/ospec forskarutbildning
	620	Licentiatutbildning
	640	Doktorsutbildning