

## Extra bortfallsanalys

### 1 Inledning

Enligt teorin har varje objekt i urvalet en sannolikhet (större än 0) att på förhand svara i den specifika undersökningen. Om denna sannolikhet vore känd kunde bortfallet justeras helt. Sannolikheten är dock okänd, men den går att uppskatta. Uppskattningen av svarsbenägenheten kan sedan användas i syfte att förbättra skattningarna.

För att skevheter i statistiken orsakade av bortfallet ska uppstå krävs att olika objekt har olika svarssannolikheter samt att de objekt som ingår i bortfallsmängden skiljer sig från de objekt som svarar med avseende på undersökningsvariablerna. Man kan intuitivt tänka sig att om objekten i bortfallsmängden har exakt samma fördelning på en undersökningsvariabel som objekten i svarsmängden så har bortfallet ingen effekt. Således behöver bortfall inte nödvändigtvis vara skadligt utan det beror på flera faktorer och hur dessa samvarierar med varandra. Effekten av bortfallet får utvärderas från undersökning till undersökning.

I följande avsnitt studerar vi bortfallet närmare och vilka effekter det kan ha på skattningar av relevanta målvariabler i undersökningen. Studien ska se som en fortsättning på bortfallsanalysen som gjordes i *Kalibreringsrapport Nationellt urval*.

### 2 Kvarstående bortfallskevheter

#### 2.1 Jämförelse av kalibrerade och ej kalibrerade skattningar

I avseende att utvärdera effekten av bortfallsjusteringen jämförs skattningar som beräknats utifrån kalibrerade vikter, med skattningar gjord med "konventionell" teknik där ingen hjälpinformation används. Skattningarna görs för målvariablerna *Hälsa*, *Värk* och *Kontant* beskrivna i tabell 1.

Tabell 1. Konstruerade målvariabler

Målvariabel	Förklaring (se även frågeformuläret)
Hälsa	Fråga 1: 1=Dåligt och mycket dåligt allmänt hälsotillstånd.
Värk	Fråga 15a-c: 1=Svåra besvär av värk.
Kontant	Fråga 41: 1=Ja, kan betala oväntad utgift (har kontantmarginal).

I tabell 2-4 visas skattad andel personer samt ett 95-procentigt konfidensintervall för skattningen. De redovisas totalt och efter

hjälpvariablerna använda i kalibreringen enligt *Kalibreringsrapport Nationellt urval*.

Tabell 2. Andelsskattningar för målvariabeln Hälsa

Redovisningsgrupp	HT-skattning	Kalibrerad skattning	Antal svar totalt
Totalt	6,2 ± 0,4	6,3 ± 0,4	16 518
Män, 16-29 år	3,4 ± 1,2	3,6 ± 1,2	871
Män, 30-44 år	3,4 ± 1,0	3,5 ± 1,0	1 330
Män, 45-64 år	5,6 ± 0,9	6,6 ± 1,1	2 610
Män, 65-84 år	7,0 ± 0,9	7,4 ± 1,1	2 781
Kvinnor, 16-29 år	6,5 ± 1,4	6,9 ± 1,5	1 236
Kvinnor, 30-44 år	5,7 ± 1,1	5,9 ± 1,2	1 674
Kvinnor, 45-64 år	7,2 ± 0,9	8,2 ± 1,1	3 051
Kvinnor, 65-84 år	7,0 ± 0,9	7,8 ± 1,1	2 965
Låg utbildning	8,3 ± 0,7	8,4 ± 0,7	6 452
Mellannivå	5,5 ± 0,6	5,4 ± 0,6	5 504
Hög utbildning	4,0 ± 0,6	3,8 ± 0,6	4 562
Födda i Sverige	5,8 ± 0,4	5,9 ± 0,4	14 278
Övriga Norden	8,7 ± 2,4	8,8 ± 2,5	530
Övriga Europa	8,1 ± 2,0	7,6 ± 1,9	741
Övriga världen	8,2 ± 1,7	7,6 ± 1,7	969
Gift/RP	5,5 ± 0,5	5,6 ± 0,6	8 300
Ej gifta	6,8 ± 0,5	6,8 ± 0,6	8 218
Boende i storstad	5,2 ± 0,5	5,5 ± 0,6	6 321
Boende utanför storstad	6,7 ± 0,5	6,8 ± 0,5	10 197

Tabell 3. Andelsskattningar för målvariabeln Värk

Redovisningsgrupp	HT-skattning	Kalibrerad skattning	Antal svar totalt
Totalt	15,4 ± 0,5	14,9 ± 0,6	16 710
Män, 16-29 år	5,1 ± 1,5	5,0 ± 1,5	870
Män, 30-44 år	7,7 ± 1,4	8,3 ± 1,5	1 341
Män, 45-64 år	13,7 ± 1,3	15,3 ± 1,5	2 642
Män, 65-84 år	16,2 ± 1,4	17,6 ± 1,5	2 826
Kvinnor, 16-29 år	10,4 ± 1,7	10,8 ± 1,8	1 247
Kvinnor, 30-44 år	12,4 ± 1,6	13,4 ± 1,7	1 687
Kvinnor, 45-64 år	19,7 ± 1,4	22,0 ± 1,6	3 089
Kvinnor, 65-84 år	22,0 ± 1,5	23,9 ± 1,7	3 008
Låg utbildning	20,9 ± 1,0	19,9 ± 1,0	6 555
Mellannivå	13,2 ± 0,9	12,3 ± 0,9	5 557
Hög utbildning	10,3 ± 0,9	9,6 ± 1,0	4 598
Födda i Sverige	14,8 ± 0,6	13,9 ± 0,6	14 440
Övriga Norden	21,3 ± 3,4	21,4 ± 3,6	540
Övriga Europa	18,5 ± 2,8	17,7 ± 2,8	752
Övriga världen	20,0 ± 2,5	19,3 ± 2,5	978

Gift/RP	15,8 ± 0,8	16,3 ± 0,9	8 386
Ej gifta	15,1 ± 0,8	14,0 ± 0,8	8 324
Boende i storstad	13,6 ± 0,8	13,5 ± 0,9	6 380
Boende utanför storstad	16,6 ± 0,7	15,9 ± 0,7	10 330

Tabell 4. Andelsskattningar för målvariabeln Kontant

Redovisningsgrupp	HT- skattning	Kalibrerad skattning	Antal svar totalt
Totalt	83,9 ± 0,6	79,9 ± 0,7	16 620
Män, 16-29 år	76,9 ± 2,8	74,4 ± 3,0	866
Män, 30-44 år	82,8 ± 2,0	80,5 ± 2,2	1 332
Män, 45-64 år	86,3 ± 1,3	83,7 ± 1,6	2 632
Män, 65-84 år	88,5 ± 1,2	85,9 ± 1,5	2 807
Kvinnor, 16-29 år	75,4 ± 2,4	73,8 ± 2,5	1 234
Kvinnor, 30-44 år	79,6 ± 1,9	76,4 ± 2,2	1 685
Kvinnor, 45-64 år	83,5 ± 1,3	79,8 ± 1,6	3 081
Kvinnor, 65-84 år	86,1 ± 1,2	82,8 ± 1,6	2 983
Låg utbildning	78,5 ± 1,0	74,8 ± 1,2	6 494
Mellannivå	84,1 ± 1,0	81,0 ± 1,2	5 542
Hög utbildning	91,2 ± 0,8	88,2 ± 1,2	4 584
Födda i Sverige	86,3 ± 0,6	84,0 ± 0,7	14 367
Övriga Norden	83,3 ± 3,2	81,8 ± 3,5	533
Övriga Europa	73,6 ± 3,2	71,8 ± 3,4	746
Övriga världen	55,5 ± 3,1	53,0 ± 3,2	974
Gift/RP	89,5 ± 0,7	85,4 ± 0,9	8 355
Ej gifta	78,2 ± 0,9	76,1 ± 1,0	8 265
Boende i storstad	85,2 ± 0,9	81,2 ± 1,1	6 358
Boende utanför storstad	83,0 ± 0,7	79,1 ± 0,9	10 262

Vid en jämförelse mellan kalibrerade och ej kalibrerade skattningar ser man att kalibrering har viss effekt. Skillnaden mellan punktskattningarna är olika stor för olika redovisningsgrupper. För *Hälsa* verkar andelen öka något i de flesta redovisningsgrupper vid användning av kalibrerade vikter. Det motsatta gäller dock för personer födda utanför Norden. För *Värk* varierar motsvarande skillnader både uppåt och nedåt.

För variabeln *Kontant* ligger den kalibrerade skattningen i flera av redovisningsgrupperna lägre än HT-skattningen. Det tyder på att personer med lägre inkomst (och mindre kontantmarginal) har lägre svarsbenägenhet än de med högre inkomst. Det är samstämmigt med resultaten i *Kalibreringsrapport Nationellt urval*. Det tyder på att kalibreringen har önskvärd effekt.

## 2.2 Skevheter i registervariabler

En metod för att bedöma hur stora fel bortfallet kan leda till är att använda sig av registerinformation. Här används registervariablerna enligt tabell 5. Variablerna finns för hela urvalet och därmed är det möjligt att jämföra skattningar från de svarande med skattningar från hela urvalet (med hjälpinformation). Dessa skattningar med 95-procentiga konfidensintervall kan studeras i tabell 6-8. Även populationsvärdet i hela urvalsramen är känd och presenteras.

Tabell 5. Registervariabler

Variabel (benämning)	Värden/Kategorier (koder)
Disponibel inkomst <sup>1</sup> (CDISP04)	$\geq 0$
Sysselsättningsstatus <sup>1</sup> (SyssStat)	1 = Förvärvsarbetande 2 = Övriga
Familjeställning <sup>2</sup> (FamStall)	1 = Ensamstående (FamStall=400) 2 = Övriga

Tabell 6. Skattad medianinkomst, kkr

Redovisningsgrupp	Svarsmängd HT- skattning	Svarsmängd Kalibrerad skattning	Urval Kalibrerad skattning	Urvalsram
Totalt	235,7 ± 2,7	221,5 ± 2,8	216,6 ± 1,6	218,3
Män, 16-29 år	86,3 ± 14,4	75,3 ± 15,6	99,9 ± 5,3	101,6
Män, 30-44 år	302,6 ± 6,9	294,5 ± 6,9	278,7 ± 3,8	279,7
Män, 45-64 år	329,1 ± 5,6	316,5 ± 5,6	304,1 ± 3,5	305,6
Män, 65-84 år	212,9 ± 5,7	196,0 ± 4,4	197,0 ± 3,6	200,3
Kvinnor, 16-29 år	86,1 ± 12,9	64,3 ± 12,3	103,6 ± 5,7	107,0
Kvinnor, 30-44 år	244,9 ± 5,2	237,8 ± 5,8	235,4 ± 3,0	236,1
Kvinnor, 45-64 år	278,6 ± 4,4	265,1 ± 4,1	258,5 ± 2,5	259,0
Kvinnor, 65-84 år	165,7 ± 3,1	158,5 ± 2,3	156,4 ± 1,6	155,9
Låg utbildning	179,2 ± 3,4	171,3 ± 3,7	164,9 ± 2,0	167,4
Mellannivå	243,4 ± 4,2	234,3 ± 5,0	228,9 ± 2,5	229,0
Hög utbildning	302,3 ± 4,7	289,6 ± 5,1	289,6 ± 3,4	291,0
Födda i Sverige	242,4 ± 2,8	233,0 ± 2,9	230,6 ± 1,7	230,8
Övriga Norden	185,6 ± 12,5	179,1 ± 12,5	177,7 ± 8,8	183,5
Övriga Europa	210,0 ± 14,3	205,9 ± 14,2	180,3 ± 7,2	185,3
Övriga världen	151,8 ± 12,5	136,3 ± 11,8	137,6 ± 4,8	146,1
Gift/RP	257,4 ± 3,9	255,7 ± 4,0	250,5 ± 2,4	252,4

<sup>1</sup> Aktualitet 2016

<sup>2</sup> Aktualitet 2017-12-31

Ej gifta	210,8 ± 4,1	193,9 ± 3,7	191,8 ± 2,1	192,9
Boende i storstad	256,6 ± 4,6	235,0 ± 5,0	229,3 ± 2,7	232,8
Boende utanför storstad	224,5 ± 3,5	214,0 ± 3,5	209,4 ± 2,0	210,2

Medianinkomsten totalt bland de svarande är 235 700 kronor.

Medianinkomsten i urvalsskattningen är 216 600 kronor. Detta tyder på personer med hög disponibel inkomst är något överrepresenterade bland de svarande och att inkomsten överskattas om ingen bortfallsjustering görs.

Medianinkomsten skattad med kalibrering är 221 500 kronor.

Bortfallsjustering med hjälp av kalibrering reducerar därmed överskattningen av inkomst. Skattningens felmarginal täcker också in populationsvärdet i urvalsramen.

På liknande sätt reduceras bortfallsskevheter i de flesta redovisningsgrupper utom för 16-29-åringar, där de kalibrerade skattningarna ger ett värde som ligger längre ifrån motsvarande urvalsskattning än HT-skattningarna.

Det finns vissa kvarstående bortfallsskevheter i variabeln inkomst även om den är liten i de flesta redovisningsgrupper. För unga personer ökar dock de kalibrerade skattningarna bortfallsskevheter.

Tabell 7. Skattad procentuell andel förvärvsarbetande

Redovisningsgrupp	Svarsmängd HT-skattning	Svarsmängd Kalibrerad skattning	Urval Kalibrerad skattning	Urvalsram
Totalt	59,0 ± 0,7	60,9 ± 0,7	60,9 ± 0,4	60,8
Män, 16-29 år	43,6 ± 3,3	41,4 ± 3,2	48,8 ± 1,5	48,0
Män, 30-44 år	86,0 ± 1,9	83,7 ± 2,1	81,2 ± 1,1	81,3
Män, 45-64 år	89,2 ± 1,2	87,5 ± 1,4	84,1 ± 0,9	83,7
Män, 65-84 år	24,7 ± 1,6	22,8 ± 1,6	23,9 ± 1,3	24,5
Kvinnor, 16-29 år	45,4 ± 2,8	41,6 ± 2,7	49,7 ± 1,5	49,9
Kvinnor, 30-44 år	84,4 ± 1,7	81,2 ± 2,1	79,1 ± 1,1	78,5
Kvinnor, 45-64 år	88,3 ± 1,1	85,9 ± 1,4	82,2 ± 0,9	82,3
Kvinnor, 65-84 år	18,9 ± 1,4	17,1 ± 1,4	17,1 ± 1,1	16,6
Låg utbildning	41,1 ± 1,2	42,2 ± 1,1	43,3 ± 0,7	43,5
Mellannivå	66,9 ± 1,2	71,1 ± 1,2	72,1 ± 0,7	71,5
Hög utbildning	75,2 ± 1,2	80,2 ± 1,3	76,6 ± 0,9	76,3
Födda i Sverige	60,4 ± 0,8	63,6 ± 0,7	64,2 ± 0,4	64,0
Övriga Norden	41,4 ± 4,1	43,0 ± 4,4	44,2 ± 2,9	43,2
Övriga Europa	59,1 ± 3,5	59,4 ± 3,6	56,4 ± 1,8	55,5
Övriga världen	48,7 ± 3,1	46,2 ± 3,1	44,9 ± 1,4	45,7
Gift/RP	60,7 ± 1,0	65,4 ± 1,0	63,9 ± 0,6	63,8

Ej gifta	57,4 ± 1,1	57,7 ± 1,0	58,8 ± 0,6	58,6
Boende i storstad	62,0 ± 1,2	62,6 ± 1,1	63,3 ± 0,6	62,6
Boende utanför storstad	57,2 ± 1,0	59,7 ± 0,9	59,4 ± 0,5	59,5

Andelen förvärvsarbetande totalt bland de svarande är 59,0 procent. Andelen i urvalsskattningen är 60,9 procent. Det betyder att förvärvsarbetande är något underrepresenterade bland de svarande och att andelen underskattas om ingen bortfallsjustering görs. Bortfallsjustering med hjälp av kalibrering reducerar denna underskattning. Andelen förvärvsarbetande skattad med kalibrering är 60,9 procent vars felmarginal täcker populationsvärdet i urvalsramen.

För de flesta enskilda redovisningsgrupper reducerar den kalibrerade skattningen bortfallsskevheter på liknande sätt. För 16-29 åringar samt personer med hög utbildning ligger dock HT-skattningen närmare urvalsskattningen.

Vissa mindre bortfallsskevheter i andel förvärvsarbetande kvarstår även efter kalibrering. För unga personer och högt utbildade ökar dock de kalibrerade skattningarna bortfallsskevheter.

Tabell 8. Skattad procentuell andel ensamstående

Redovisningsgrupp	Svarsmängd HT-skattning	Svarsmängd Kalibrerad skattning	Urval Kalibrerad skattning	Urvalsram
Totalt	31,6 ± 0,7	33,9 ± 0,7	34,9 ± 0,4	35,1
Män, 16-29 år	39,0 ± 3,2	38,2 ± 3,2	42,8 ± 1,5	42,4
Män, 30-44 år	34,2 ± 2,5	35,3 ± 2,5	40,2 ± 1,3	39,8
Män, 45-64 år	28,6 ± 1,7	31,4 ± 1,8	33,1 ± 1,1	33,1
Män, 65-84 år	30,3 ± 1,7	35,7 ± 2,0	34,8 ± 1,3	34,9
Kvinnor, 16-29 år	37,1 ± 2,7	34,7 ± 2,6	37,8 ± 1,5	39,7
Kvinnor, 30-44 år	21,0 ± 1,9	22,2 ± 2,0	21,9 ± 1,1	23,1
Kvinnor, 45-64 år	24,7 ± 1,5	28,2 ± 1,7	26,4 ± 1,1	26,2
Kvinnor, 65-84 år	42,6 ± 1,8	49,0 ± 2,0	46,3 ± 1,3	46,4
Låg utbildning	32,6 ± 1,1	32,7 ± 1,1	35,6 ± 0,7	35,5
Mellannivå	32,0 ± 1,2	35,0 ± 1,3	35,7 ± 0,7	36,2
Hög utbildning	29,5 ± 1,3	34,5 ± 1,6	32,3 ± 0,9	32,6
Födda i Sverige	31,4 ± 0,8	34,0 ± 0,8	34,6 ± 0,4	34,7
Övriga Norden	40,0 ± 4,1	43,4 ± 4,3	42,9 ± 2,9	43,5
Övriga Europa	33,1 ± 3,4	35,5 ± 3,5	35,9 ± 1,8	36,4
Övriga världen	27,9 ± 2,8	29,4 ± 2,9	34,6 ± 1,4	35,1
Gift/RP	2,0 ± 0,3	2,6 ± 0,4	3,8 ± 0,3	3,7

Ej gifta	61,4 ± 1,0	55,9 ± 1,1	56,7 ± 0,6	57,1
Boende i storstad	33,0 ± 1,2	35,0 ± 1,1	36,1 ± 0,6	36,3
Boende utanför storstad	30,7 ± 0,9	33,2 ± 0,9	34,1 ± 0,5	34,3

Andelen ensamstående totalt bland de svarande är 31,6 procent. Andelen i urvalsskattningen är 34,9 procent. Det betyder att ensamstående är underrepresenterade bland de svarande och att andelen underskattas om ingen bortfallsjustering görs. Bortfallsjustering med hjälp av kalibrering reducerar denna underskattning, men inte tillräckligt. Andelen ensamstående skattad med kalibrering är 33,9 procent men felmarginalen täcker inte populationsvärdet i urvalsramen.

För de flesta enskilda redovisningsgrupper reduceras bortfallsskevheter vid användning av den kalibrerade skattningen. För 16-29 år, speciellt kvinnor, ökar dock skevheter.

Det finns vissa kvarstående bortfallsskevheter i andelen ensamstående även om den är mindre i de flesta redovisningsgrupper.

### 3 Sammanfattning och slutsatser

Svarsandelen i undersökningen var 42,1 procent. Bortfallet är inte av slumpmässig art utan det finns variabler som påverkar benägenheten att svara, vilket visades i *Kalibreringsrapport Nationellt urval*. Bortfallsanalysen visade på att främst variablerna ålder, kön, utbildning, inkomst, födelseland, civilstånd och invandringsår samvarierar med benägenheten att svara.

I hjälpvektorn vid framtagandet av kalibrerade vikter användes dessa variabler (minus invandringsår) samt storstad. Inkomst togs bort eftersom den gav instabila vikter. Län användes som modellgrupp.

Kalibrerade skattningar utifrån hjälpinformation har jämförts med skattningar där alla svarande ges samma vikt inom stratum. Skattningar har gjorts för några viktiga undersökningsvariabler samt för tre registervariabler.

För undersökningsvariablerna *Hälsa* och *Värk* ger kalibreringen viss effekt. För *Kontant* är effekten större vilken tyder på att personer med lägre inkomst och mindre kontantmarginal har lägre svarsbenägenhet än de med högre inkomst.

För registervariablerna reduceras de skevheter som bortfallet orsakar när kalibrering används för totalnivån. De kalibrerade skattningarna ligger närmare urvalsskattningen (med hjälpinformation). För enskilda redovisningsgrupper reduceras bortfallsskevheter för de flesta men ökar i

några fall, som i fallet med 16-29 åringar. Anledningen kan vara att gruppen är för homogen relativt övrig hjälpinformation inom gruppen.

Det är okänt hur de som inte svarat på undersökningen skulle ha svarat och vilka konsekvenser det får för den statistik som baseras på de svarande. Den analys som genomförts visar att bortfallet orsakar skevheter och kan variera mellan målvariabler och redovisningsgrupper. Situationen avhjälpas med kalibrerade vikter som reducerar bortfallsskevheter, men leder inte alltid till förbättringar i estimaten för alla redovisningsgrupper.

Förutsatt att den valda hjälpinformationen samvarierar med undersökningsvariablerna så bör bortfallsskevheter för dessa inte vara alltför stor. Det finns dock vissa kvarstående bortfallsskevheter. För andra målvariabler och redovisningsgrupper som inte utgörs av hjälpvariablerna är effekten okänd. Vi rekommenderar därför att det kalibrerade vikterna i första hand används för redovisning där grupperna utgörs av hjälpvariablerna.