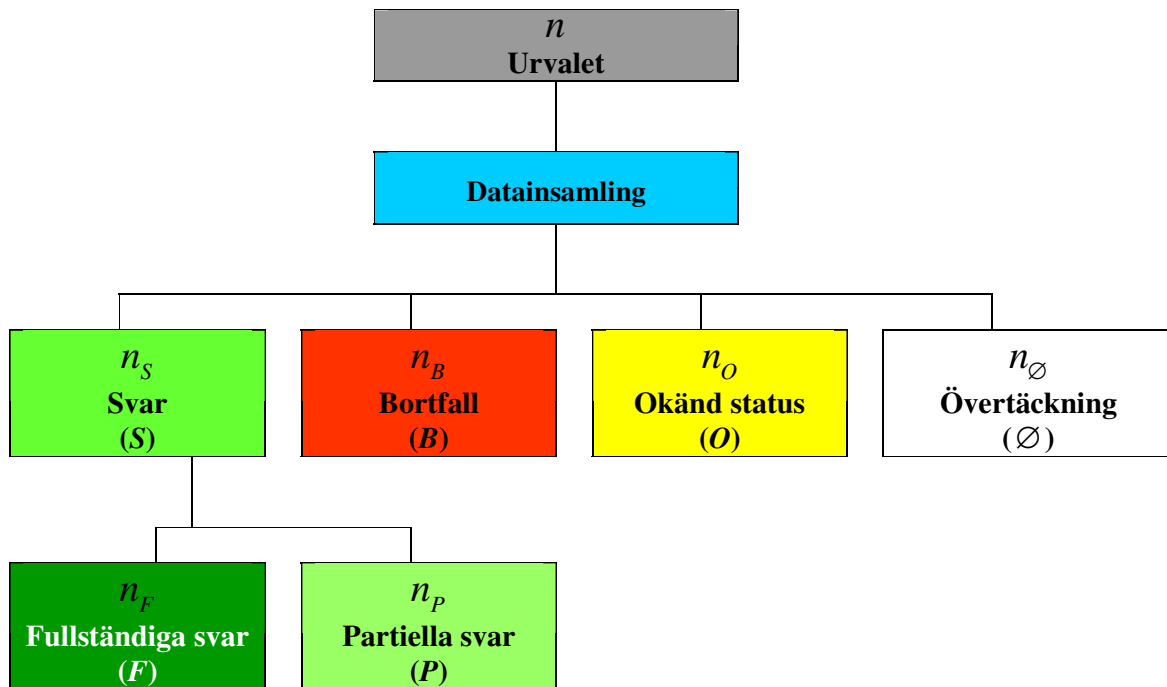


STANDARD FÖR BORTFALLSBERÄKNING



Sektionen för Surveystatistik

<http://www.statistikersamfundet.se/survey/>

Svenska statistikersamfundet

<http://www.statistikersamfundet.se/>

Innehållsförteckning

1. Inledning	1
2. Tidigare och nuvarande standarder	3
3. Arbetets inriktning	5
4. Statistiska undersökningar av surveytyp	6
4.1 Översiktlig beskrivning av en survey och några grundbegrepp	6
4.2 Begrepp kopplade till praktiska problem i en survey	7
5. Grundläggande modell för beräkning av svars- och bortfallsandelar	9
5.1 Inledning	9
5.2 Ovägda svars- och bortfallsmått	10
5.2.1 Grundläggande mått	10
5.2.2 Ytterligare mått	11
5.3 Vägda svars- och bortfallsmått	11
6. Numeriska exempel	14
7. Orsaker till varför man hamnar i olika bortfallskategorier	20
8. Sammanfattning och rekommendationer för rapportering av bortfallet	25
Referenser	26
Bilagor	
Bilaga 1. Ett exempel på en kortfattad teknisk rapport (baserat på exempel 6.2)	
Bilaga 2. Exempel på system för resultatkoder	
Bilaga 3. Bortfallssnurra med resultatkoder	

1. Inledning

Bortfall är en felkälla som förekommer i så gott som alla statistiska undersökningar och som allvarligt kan försämra användbarheten av undersökningarnas resultat. Vare sig man är producent, beställare eller konsument av undersökningar behöver man ett mått på bortfallets storlek för att kunna jämföra olika undersökningar eller för att bedöma kvaliteten i en viss undersökning. Ett problem härvidlag är att det förekommer en rik flora av sinsemellan mycket olika bortfallsmått vilket skapar oklarhet vid kvalitetsbedömningar och nästan kan omöjliggöra jämförelser. Behovet av en standard för hur bortfall skall definieras och mätas i olika typer av undersökningar är därför stort.

Standardiserade mått på bortfallet kräver enhetliga definitioner på ingående begrepp och storheter i alla led i undersökningen. För att en standard skall bli etablerad och i framtiden också kunna utvecklas ytterligare, är det viktigt att den förankras hos såväl de olika sammanslutningarna av kommersiella undersökningsföretag som inom den offentliga sektorn, speciellt inom de informations- och statistikansvariga myndigheterna. Lika angeläget är att de som köper undersökningar förstår och kan använda standarden vid upphandlingar och vid bedömningar av kvaliteten på undersökningarna. Som ett led i detta är det viktigt att den surveystatistiska forskningen och universitetsutbildningen samarbetar med olika statistikproducenter.

Initiativet till detta arbete har tagits av Sektionen för surveystatistik inom Svenska statistikersamfundet. Vid dess styrelsemöte den 26 november 2001 utsågs Gösta Forsman till samordnare av en kommitté med syfte att utveckla en standard för bortfallsberäkningar för opinions- och marknadsundersökningar. Följande personer har därefter utsetts av surveysektionens styrelse att ingå i denna kommitté:

Gösta Forsman, ordförande, adjungerad professor, avdelningen för statistik, Linköpings universitet och chefsstatistiker, Vägverket

Karin Dahmström, universitetslektor, Statistiska institutionen, Stockholms universitet

Jan Hörngren, metodstatistiker, Statistiska centralbyrån, Örebro

Ulf Isander, chefsstatistiker, Research International, representant för SMIF¹

Toivo Sjören, chef, opinionsundersökningar, SIFO Research and Consulting, representant för SMIF

Bengt Swensson, professor, institutionen för ekonomi, statistik och informatik, Örebro universitet

Åke Wissing, seniorkonsult, ISI Wissing AB, representant för FSM², Sverigerepresentant för ESOMAR³

¹ SMIF = Svenska Marknadsinformationsföretag

² FSM= Föreningen Svenska Marknadsundersökningsinstitut

³ ESOMAR=The European Society for Opinion and Marketing Research

2. Tidigare och nuvarande standarder

Ända sedan de första opinionsundersökningarna på 1930-talet har man varit medveten om problemet med bortfall och från och med 1940-talet började det behandlas i den vetenskapliga litteraturen. Som helhet har fokus satts på fyra aspekter:

1. åtgärder för att minska bortfallet,
2. bortfallsredovisning,
3. mätning av bortfallets snedvridande effekt på undersökningsresultaten och
4. metodik för att kompensera för denna effekt, t ex imputation och viktning.

Många organisationer har tagit initiativ till att utveckla standarder för redovisning av bortfallet. Internationellt kan nämnas American Association for Public Opinion Research (AAPOR) och American Statistical Association (ASA), sektionen för Survey Research Methods. Den senare organisationens arbete utvecklades 1977 till den så kallade Panel on Incomplete Data som organiserade en omfattande inventering av metoder för bortfallsbehandling som rapporterades i en artikelsamling i tre volymer. Arbetet utmynnade i 35 rekommendationer, bl a om beräkning av bortfall.

Inom den kommersiella sektorn har bl a Council of American Survey Research Organizations (CASRO) studerat hur företagen rapporterar bortfallet. Resultatet blev att det inte fanns något gemensamt sätt att redovisa detta. Deras förslag blev att man i tabellform skulle redovisa "the final status of all sampled units".

Sedan 1996 har AAPOR utarbetat standarder för "the final disposition of case codes and of various outcome rates (e.g., response rates and cooperation rates) based on these codes". Den senaste versionen utgavs 2004, se AAPOR (2004).

Statistiska centralbyrån har sedan länge verkat för att all publicerad statistik skall förses med en kvalitetsredovisning. Den grundläggande principen för denna har varit som den är framställd i "Policy för användarorienterad kvalitetsredovisning av statistik", se SCB (1983), följande:

Statistikproducenten skall informera användarna om faktorer av betydelse för en korrekt tolkning av statistiken. Informationen skall vara tillgänglig och begriplig för användarna samt i övrigt utformad med tanke på deras behov.

Under de gångna åren har SCB:s riktlinjer för kvalitetsredovisning reviderats ett antal gånger och den senaste är daterad 2001, se SCB (2001). Dessa rekommendationer avser i första hand den officiella statistiken som publiceras i serien Sveriges Officiella Statistik. Vad gäller bortfallet specifikt uppmanas de statistikansvariga myndigheterna "att ange kvantitativa och kvalitativa aspekter på bortfallet; bortfallsandel (vägd och/eller ovägd), uppdelning i "avböjd medverkan" respektive "ej anträffad", kända eller förmodade bortfallsmönster, åtgärder för att minska bortfallet mm. Redovisa bortfallets bedömda konsekvenser för statistikens tillförlitlighet. Beskriv metod(er) för bortfallskompensation i skattningsförfarandet".

Ett stort intresse för bortfallsstandarder har visats av Statistics Canada, som har antagit de standarder som redovisas i Hidioglou m fl (1993).

Smith (2002) ger i sin sammanställning över utvecklingen av standarder för bortfall en bild av ett arbete som under årens lopp kantats med många ansträngningar men inte slagit igenom bland undersökningsföretagen. I verkligheten redovisas ofta inte svarsandelar eller motsvarande bortfallsandelar över huvud taget. Och om det finns några numeriska mått, är dessa inte definierade. Eftersom det inte har funnits några formella krav från några överordnade myndigheter eller organisationer, har inte bortfallet och dess redovisning tagits på ett seriöst sätt. Trots detta är han ändå optimistisk. För att utveckla standarder krävs ett professionellt arbete som går utöver de enskilda företagens intressen. Förhoppningsvis skall ett resultat av gemensamma standarder som accepteras och efterlevs av alla surveyorganisationer kunna uppnås.

Kommittén har som utgångspunkt för sitt arbete framför allt haft koderna och bortfallsmåtten som de är beskrivna i AAPOR (2000). Dessa tycks kunna bli en internationell standard, men behöver ses i ljuset av svenska förhållanden och behov.

3. Arbetets inriktning

En viktig avgränsning i kommitténs arbete är att enbart undersökningar som bygger på sannolikhetsurval behandlas. Med sannolikhetsurval avses här urval där varje enhet i den population som undersöks har en beräkningsbar sannolikhet som är större än noll att komma med i urvalet.

Vi presenterar i det följande

- en översikt över olika fackuttryck som används i surveysammanhang,
- bortfallsmått som är tillämpliga för de flesta typer av surveyer,
- exempel på hur bortfallsmått kan beräknas för fyra vanliga undersökningstyper: postal enkät till individer, företagsundersökning per telefon och två varianter av individundersökning per telefon samt
- förslag på klassificering av bortfall (ett kodsysteem) för dessa fyra olika undersökningstyper.

Rapporten planeras att efter hand utökas med andra undersökningstyper såsom webb-, fax- och mobiltelefonundersökningar.

För att underlätta den praktiska bortfallsberäkningen har en ”beräkningssnurra” utvecklats. Denna bygger på rapportens kodsysteem och beräknar ovägda bortfallsmått. Den återfinns på <http://www.bortfallssnurran.se/>.

4. Statistiska undersökningar av surveytyp

4.1 Översiktlig beskrivning av en survey och några grundbegrepp

När det i dessa rekommendationer talas om en statistisk undersökning av surveytyp, eller kortare en *survey*, menas en statistisk undersökning med följande speciella kännetecken:

- survey**
- a) Man vill studera ett ändligt, ofta stort, antal *element* (alt. *objekt*), som tillsammans kallas för en *population*, eller *målpopulation*. Elementen kan t ex vara individer, hushåll, företag eller arbetsställen, och vi talar då om individ-, hushålls-, företags- respektive arbetsställepopulation. Det finns en klar definition av vilket slags element som ingår i den population som studeras. Undersökningens mål är att ge information om populationen. Oftast vill man också ge sådan information nedbruten på speciella delpopulationer, så kallade *redovisningsgrupper* - t ex "kvinnor" och "män" eller olika åldersgrupper. Dessa grupper definieras med hjälp av olika *bakgrundsvariabler* - t ex kön och ålder.
- element/objekt**
population
målpopulation
- redovisningsgrupp**
- bakgrundsvariabel**
- b) Till varje element i populationen hör ett värde på en eller flera *undersökningsvariabler* - ibland kallade *mätvariabler* eller *observationsvariabler* - t ex inkomst, varumärkeskännedom, köpvanor, partipreferens, omsättning och antal anställda. Populationen kan beskrivas med avseende på dessa undersökningsvariabler med hjälp av olika sammanfattande storheter, kallade *parametrar*, som t ex totaler (total omsättning i en bransch), medelvärden (medelinkomst), proportioner (procentandel med borgerliga partipreferenser), korrelationer (mått på samband mellan två undersökningsvariabler) och index. En survey syftar till att ta fram information om det okända värdet på var och en av ett antal för undersökningen intressanta parametrar.
- undersökningsvariabel/mätvariabel/observationsvariabel**
- parameter**
- c) För att det skall bli möjligt att skapa kontakt med och därmed samla in data om enskilda element i populationen krävs i allmänhet tillgång till en *urvalsram*, eller kortare *ram*, som består av en uppsättning *urvalsenheter*. Idealt har varje element i populationen exakt en länk (koppling) till en urvalsenhet och varje urvalsenhet exakt en länk till ett element. I det enklaste fallet utgörs ramen av poster i en lista (eller datafil), där varje post svarar mot ett visst element i populationen och där posten t ex innehåller namn och adress till detta element (individ vid individpopulation, företag vid företagspopulation); i detta fall svarar urvalsenhet mot element, och det är då möjligt att direkt dra ett urval av element, dvs att göra *direkturval av element*. I mer komplexa fall kan inte urvalet av element dras direkt, utan man måste dra det slutliga urvalet av element via urval i två eller flera urvalssteg, genom att utnyttja en hierarki av urvalsramar och urvalsenheter. Vi talar då om *tvåstegsurval* eller *flerstegsurval*. Man skulle t ex kunna dra ett urval av individer (element) genom att först dra ett urval av kommuner, följt av ett urval av individer i de utvalda kommunerna.
- urvalsram/ram**
urvalsenhet
- direkturval av element**
- tvåstegsurval**
flerstegsurval

- d) I allmänhet genomförs en survey som en stickprovsundersökning, dvs ett stickprov av urvalsenheter dras från ramen, vilket ger ett *stickprov* (alt. *urval*) av element från populationen. Här och i fortsättningen behandlas bara det fall då stickprovet av element är ett *sannolikhetsurval* draget utan återläggning. Vid direkturval av element innebär detta att varje populationselement skall ha en positiv sannolikhet att komma med i urvalet. (Se dock kommentar i avsnitt 4.2 nedan om undertäckning.) Denna sannolikhet kallas *inklusionssannolikhet*. Inverterade värdet av inklusionssannolikheten (1 dividerat med sannolikheten) utgör elementets s k *designvikt*, ibland kallad *urvalsvikt* eller *uppräkningsvikt*. Vid två- eller flerstegsurval av element krävs att urvalet i varje steg är ett sannolikhetsurval.
- e) För elementen i urvalet samlas data in avseende undersökningsvariablerna enligt något väletablerat mätförfarande, t ex via postenkät, telefonintervju, personlig intervju eller webbenkät. Insamlade data används sedan för att beräkna *skattningar* (*estimat*) - ibland kallade *punktskattningar* - av de okända parametrarna. Det är viktigt att punktskattningarna kompletteras med mått på skattningarnas osäkerhet, t ex i form av felmarginaler. Dessa resultat presenteras därefter på lämpligt sätt.

stickprov/urval

sannolikhetsurval

inklusionssannolikhet

designvikt/urvalsvikt/
uppräkningsvikt

skattning/estimat/
punktskattning

4.2 Begrepp kopplade till praktiska problem i en survey

Då en undersökning av surveytyp skall genomföras ställs undersökaren inför ett antal praktiska problem, som kräver att begreppsapparaten i föregående avsnitt utvidgas något. Vi kommer här bara att ta upp sådana begrepp som är av speciell betydelse då bortfall förekommer. Mätfelsproblematik kommer inte att behandlas i dessa rekommendationer.

I huvudsak kommer vi också att begränsa oss till fallet med direkturval av element. I avsnittet om bortfallsmått ges dock anvisningar om hur utvidgningar till två- och flerstegsurval kan göras.

Vid direkturval av element skulle man idealt ha följande ramsituation: Till varje enhet i urvalsramen hör exakt ett element i målpopulationen, och till varje element i målpopulationen hör exakt en urvalsenhet. Detta innebär att ramen inte innehåller någon urvalsenhet som har länk till element som inte tillhör målpopulationen, och inte heller finns det något element i målpopulationen som inte har länk till en urvalsenhet i ramen.

Denna ideala situation föreligger sällan eller aldrig i praktiken. Det vanliga är att den urvalsram man har tillgång till har två brister:

- *Undertäckning*, vilket innebär att det finns element i målpopulationen som inte har någon länk till enheter i urvalsramen. Sådana element har sannolikheten noll att komma med i urvalet. T ex kan nyetablerade företag eller nyinflyttade individer saknas i den ram som urvalet dras ifrån. Man har då inte längre ett sannolikhetsurval från målpopulationen, utan endast från den del som inte utgör undertäckning, och det är

undertäckning

endast till denna del som objektiva statistiska slutsatser kan dras. Ju större undertäckningen är, desto mer begränsade är alltså de slutsatser som kan dras. (Undertäckning skall inte förväxlas med bortfall.)

- *Övertäckning*, vilket innebär att det finns enheter i urvalsramen med länkar till element som inte tillhör målpopulationen. Ofta finns ingen information i urvalsramen som pekar ut dessa enheter, och vi har då *okänd övertäckning*. Sådana enheter kan inte rensas bort från ramen före urvalsdragningen, dvs man riskerar att i urvalet få med element som inte tillhör målpopulationen, vilket också försämrar slutsatsmöjligheterna. Exempel här kan vara då företag som gått i konkurs eller personer som avlidit fortfarande finns med i urvalsramen.

övertäckning

okänd övertäckning

Vid datainsamlingen lyckas man så gott som aldrig samla in data avseende samtliga undersökningsvariabler för samtliga målpopulationselement i urvalet. För vissa element i urvalet lyckas man inte samla in information som gör det möjligt att avgöra om elementet tillhör målpopulationen eller ej – sådana element i urvalet har då *okänd målpopulationsstatus*.

okänd målpopulationsstatus

Bortfall föreligger då man för ett eller flera målpopulationselement i urvalet misslyckats att få användbara observationer på en eller flera undersökningsvariabler. Man brukar tala om två slags bortfall:

bortfall

- *Objektbortfall* (eller *individbortfall*), vilket föreligger om det för ett målpopulationselement i urvalet saknas data för samtliga undersökningsvariabler.
- *Partiellt bortfall* (eller *variabelbortfall*), vilket föreligger om det för ett målpopulationselement i urvalet saknas data för minst en undersökningsvariabel, men inte för alla.

**objektbortfall/
individbortfall**

**partiellt bortfall/
variabelbortfall**

I detta sammanhang, slutligen, använder man också följande två begrepp:

- *Fullständiga svar*: Ett målpopulationselement i urvalet har fullständiga svar om användbara observationer har erhållits på samtliga undersökningsvariabler.
- *Partiella svar*: Ett målpopulationselement i urvalet har endast partiella svar om användbara observationer har erhållits på minst en undersökningsvariabel, men inte på alla.

fullständiga svar

partiella svar

5. Grundläggande modell för beräkning av svars- och bortfallsandelar

5.1 Inledning

Vi utgår i detta kapitel från att ett sannolikhetsurval bestående av n element (objekt) dragits via den tillgängliga urvalsramen. När datainsamlingen är avslutad kan dessa n element klassificeras i fyra huvudkategorier:

S : Målpopulationsobjekt för vilka fullständiga eller partiella svar erhållits.

B : Målpopulationsobjekt med inga eller otillräckliga svar, dvs element som utgör *bortfall* (objektbortfall).

O : Element i urvalet med *okänd målpopulationsstatus*, dvs vi vet inte om dessa tillhör målpopulationen eller utgör övertäckning.

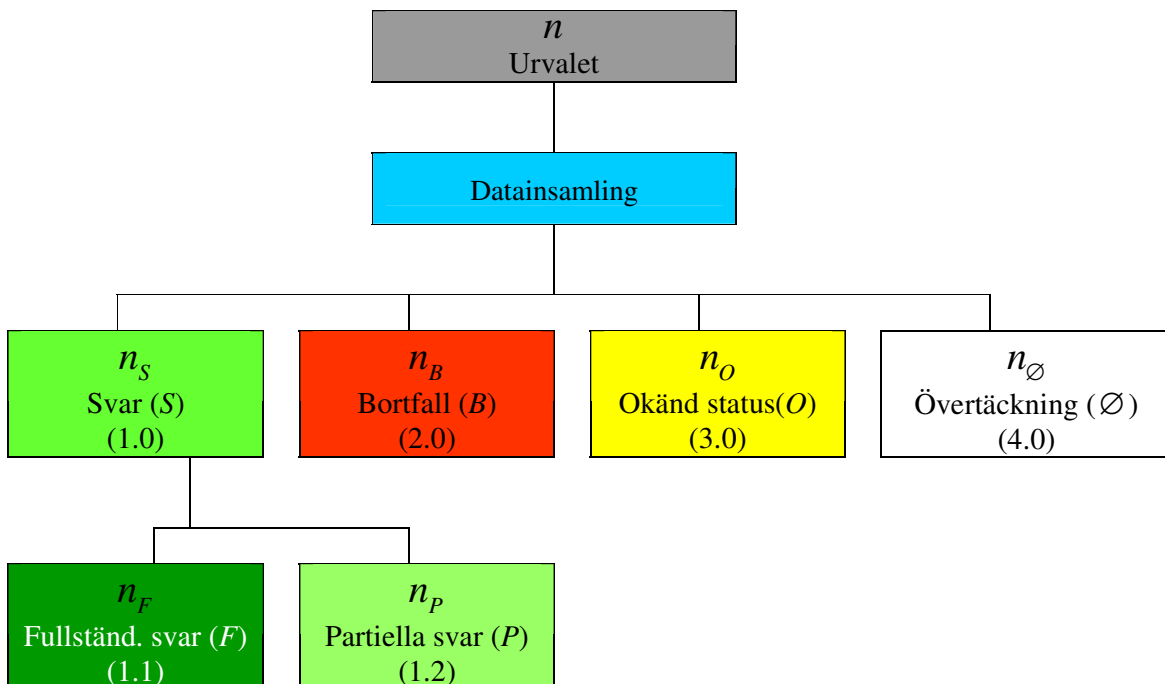
\emptyset : Element i urvalet som utgör *övertäckning*, dvs element som inte tillhör målpopulationen

Elementen i den första huvudkategorin (S) kan i sin tur delas in i två underkategorier:

F : Målpopulationsobjekt för vilka *fullständiga svar* erhållits.

P : Målpopulationsobjekt för vilka endast *partiella svar* erhållits.

Indelningen av ett urval som består av n element och dess uppdelning på de sex kategorierna illustreras i följande figur:



Beräkning av svars- och bortfallsmått kan göras på olika sätt. Vi går här igenom två huvudtyper, som ger olika slags information:

- (i) ovägda svars- och bortfallsmått, och
- (ii) vägda svars- och bortfallsmått.

Ovägda mått kan sägas ge information om hur väl man lyckats/misslyckats med det speciella stickprov för vilket data samlats in, medan vägda mått ger en uppfattning om (uppskattning av) vad man skulle ha uppnått vid datainsamling för en totalundersökning med samma metodik i övrigt som för stickprovsundersökningen.

I avsnitt 5.2 redovisas ett antal ovägda mått, medan vägda mått redovisas i avsnitt 5.3.

5.2 Ovägda svars- och bortfallsmått

Låt antalet element i de sex kategorierna S , B , O , \emptyset , F och P ovan betecknas n_S , n_B , n_O , n_\emptyset , n_F respektive n_P . När man skall beräkna olika bortfallsandelar ger elementen i kategori O problem, eftersom man inte vet om de utgör bortfall (dvs tillhör målpopulationen) eller övertäckning (dvs tillhör inte målpopulationen). Man tvingas då till att göra något antagande om hur stor andel av O -elementen som utgör bortfall. Tre sådana antaganden är:

- (1) Samtliga n_O element i O utgör bortfall, dvs inget av de n_O elementen utgör övertäckning.
- (2) Andelen element i O som utgör bortfall är u , där alltså $0 \leq u \leq 1$.
- (3) Inget av de n_O elementen i O utgör bortfall, dvs samtliga n_O element utgör övertäckning.

5.2.1 Grundläggande mått

Tre grundläggande svars- och bortfallsmått, svarande mot de tre antagandena ovan om hur stor andel av de n_O elementen i O som utgör bortfall, dvs tillhör målpopulationen, är:

- (1) Samtliga n_O element i O utgör bortfall

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_1(\text{ovägt}) = \frac{n_S}{n_S + n_B + n_O}$	$BA_1(\text{ovägt}) = 1 - SA_1(\text{ovägt})$

- (2) Andelen element i O som utgör bortfall är u

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_2(\text{ovägt}) = \frac{n_S}{n_S + n_B + u \cdot n_O}$	$BA_2(\text{ovägt}) = 1 - SA_2(\text{ovägt})$

- (3) Inget av de n_O elementen i O utgör bortfall

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_3(\text{ovägt}) = \frac{n_S}{n_S + n_B}$	$BA_3(\text{ovägt}) = 1 - SA_3(\text{ovägt})$

Ann. 1. Bortfallsmåttet $BA_1(\text{ovägt})$ rekommenderas i första hand såvida inte stark kunskap om fördelningen på målpopulationselement och övertäckningselement bland elementen i O finns. Då $BA_2(\text{ovägt})$ eller $BA_3(\text{ovägt})$ rapporteras bör en motivering till valet också anges.

Ann. 2. $BA_1(\text{ovägt})$ kan ses som ett specialfall av $BA_2(\text{ovägt})$ med $u=1$, medan $BA_3(\text{ovägt})$ kan ses som ett specialfall av $BA_2(\text{ovägt})$ med $u=0$. Motsvarande gäller för svarsmåten $SA_1(\text{ovägt})$, $SA_2(\text{ovägt})$ och $SA_3(\text{ovägt})$.

5.2.2 Ytterligare mått

De tre bortfallsmåtten $BA_1(\text{ovägt})$, $BA_2(\text{ovägt})$ och $BA_3(\text{ovägt})$ tar alla fasta på objekt-bortfall. Om man speciellt vill betona i vilken utsträckning man fått in *fullständiga* svar kan följande svars- och bortfallsandelar vara av intresse (med motsvarande antaganden om övertäckningsgraden).

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_1^F(\text{ovägt}) = \frac{n_F}{n_S + n_B + n_O}$	$BA_1^F(\text{ovägt}) = 1 - SA_1^F(\text{ovägt})$

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_2^F(\text{ovägt}) = \frac{n_F}{n_S + n_B + u \cdot n_O}$	$BA_2^F(\text{ovägt}) = 1 - SA_2^F(\text{ovägt})$

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_3^F(\text{ovägt}) = \frac{n_F}{n_S + n_B}$	$BA_3^F(\text{ovägt}) = 1 - SA_3^F(\text{ovägt})$

5.3 Vägda svars- och bortfallsmått

Vid vägda bortfallsmått utnyttjas de designvikter som hör till de olika element som ingår i stickprovet av storleken n . Antag att stickprovselementen är numrerade $1, 2, \dots, k, \dots, n$ och låt designvikten för element k betecknas d_k .

Vid direkta elementurval är designvikten för element k lika med inverterade värdet av dess inklusionssannolikhet (1 dividerat med dess inklusionssannolikhet). Vid sådana urval beräknas de tre grundläggande vägda svars- och bortfallsmåtten på följande sätt.

(1) Samtliga n_O element i O utgör bortfall

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_1(\text{vägt}) = \frac{\sum_S d_k}{\sum_S d_k + \sum_B d_k + \sum_O d_k}$	$BA_1(\text{vägt}) = 1 - SA_1(\text{vägt})$

där

$$\begin{aligned}\sum_B d_k &= \text{summan av designvikterna för alla de element i stickprovet} \\ &\text{som tillhör kategori } B, \\ \sum_O d_k &= \text{summan av designvikterna för alla de element i stickprovet} \\ &\text{som tillhör kategori } O, \\ \sum_S d_k &= \text{summan av designvikterna för alla de element i stickprovet} \\ &\text{som tillhör kategori } S\end{aligned}$$

(2) Den andel av $\sum_O d_k$ som kommer från bortfallselementen i O är u

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_2(\text{vägt}) = \frac{\sum_S d_k}{\sum_S d_k + \sum_B d_k + u \sum_O d_k}$	$BA_2(\text{vägt}) = 1 - SA_2(\text{vägt})$

(3) Inget av de n_O elementen i O utgör bortfall

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_3(\text{vägt}) = \frac{\sum_S d_k}{\sum_S d_k + \sum_B d_k}$	$BA_3(\text{vägt}) = 1 - SA_3(\text{vägt})$

Ann. 3. Dessa tre vägda mått blir desamma som motsvarande ovägda mått om man använder en urvalsdesign/urvalsmetod för vilken samtliga element i ramen har samma sannolikhet att komma med i urvalet. Detta är t ex fallet vid obundet slumpmässigt urval och systematiskt urval. Detsamma gäller vid stratifierat urval för den speciella situation då urvalet inom strata fördelas proportionellt mot stratumstorlekarna, dvs då vi har det som kallas ett proportionellt stratifierat urval.

Ann 4. Ovanstående vägda mått kan också användas vid mer komplexa urval än direkta elementurval, t ex flerstegsurval och flerfasurval om bortfallet uppträder först i sista steget/fasen. Designvikten för element k i stickprovet erhålls då på följande sätt. Multiplicera ihop inklusionssannolikheterna associerade med de olika urvalsstegen eller urvalsfaserna. Genom att dividera 1 med resultatet av denna multiplikation får man slutligen elementets designvikt d_k .

Ann 5. På motsvarande sätt som vid de ovägda bortfallsmått rekommenderas i första hand måttet $BA_1(\text{vägt})$ såvida inte stark kunskap motiverar något av de två andra bortfallsmått.

Om man liksom tidigare speciellt vill betona i vilken utsträckning man fått in *fullständiga* svar beräknas de vägda svars- och bortfallsandelarna på följande sätt.

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_1^F(\text{vägt}) = \frac{\sum_F d_k}{\sum_S d_k + \sum_B d_k + \sum_O d_k}$	$BA_1^F(\text{vägt}) = 1 - SA_1^F(\text{vägt})$

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_2^F(\text{vägt}) = \frac{\sum_F d_k}{\sum_S d_k + \sum_B d_k + u \sum_O d_k}$	$BA_2^F(\text{vägt}) = 1 - SA_2^F(\text{vägt})$

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_3^F(\text{vägt}) = \frac{\sum_F d_k}{\sum_S d_k + \sum_B d_k}$	$BA_3^F(\text{vägt}) = 1 - SA_3^F(\text{vägt})$

Här är $\sum_F d_k$ summan av designvikterna för alla element i stickprovet som tillhör kategori F .

Anm 6. I vissa tillämpningar, där elementen har varierande storlek, kan det finnas anledning att förutom vägning med designvikten också bygga in vägning med elementstorlek. Om element k har storleken x_k , så kan man i formlerna ovan ersätta d_k med produkten $d_k x_k$, vilket ger följande mått:

(1) Samtliga n_O element i O utgör bortfall

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_4(\text{vägt}) = \frac{\sum_S d_k x_k}{\sum_S d_k x_k + \sum_B d_k x_k + \sum_O d_k x_k}$	$BA_4(\text{vägt}) = 1 - SA_4(\text{vägt})$

(2) Den andel av $\sum_O d_k x_k$ som kommer från bortfallselementen i O är u

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_5(\text{vägt}) = \frac{\sum_S d_k x_k}{\sum_S d_k x_k + \sum_B d_k x_k + u \sum_O d_k x_k}$	$BA_5(\text{vägt}) = 1 - SA_5(\text{vägt})$

(3) Inget av de n_O elementen i O utgör bortfall

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_6(\text{vägt}) = \frac{\sum_S d_k x_k}{\sum_S d_k x_k + \sum_B d_k x_k}$	$BA_6(\text{vägt}) = 1 - SA_6(\text{vägt})$

Också här rekommenderas i första hand måtten under (1), om inte starkt välmotiverade skäl finns för något av måtten under (2) och (3).

6. Numeriska exempel

Exempel 6.1. För att få en uppfattning om inställningen till trängselskatt hos boende i en av Stockholms innerstadsförsamlingar lät ett politiskt parti genomföra en undersökning på följande sätt.

Med hjälp av SPAR (Statens Person- och Adress-Register) skapades en urvalsram svarande mot de $N = 4000$ församlingsborna i åldern 18-64 år. Därefter drogs ett obundet slumpmässigt urval av storleken $n = 800$. Som datainsamlingsmetod användes postenkät med två påminnelser och telefonkontakt.

Resultat av undersökningen fördelat på de fyra huvudkategorierna i avsnitt 5.1:

Kategori	Antal	
S	550 (n_S)	Svar (fullständiga eller partiella)
B	175 (n_B)	Vägran, tillfälligt bortrest, sjuk, språkhinder
O	50 (n_O)	Ingen kontakt, postretur (adressaten okänd), tom enkät
\emptyset	25 (n_{\emptyset})	Avliden, flyttat från Stockholm
Summa	800 (n)	

I denna undersökning har alla samma sannolikhet att komma med i urvalet (samma inklusionssannolikhet), vilket betyder att de ovägda och vägda svars- och bortfallsmått blir desamma. Svars- och bortfallsandelarna beräknas nu enligt följande (se avsnitt 5.2), där vi alltså måste göra antaganden om de $n_O = 50$ individer som har okänd målpopulationsstatus.

Om *samtliga* 50 antas ingå i målpopulationen och alltså betraktas som bortfall, får vi

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_1 = \frac{550}{550 + 175 + 50} = \frac{550}{775} = 71,0 \%$	$BA_1 = 1 - SA_1 = 1 - 0,710 = 29,0 \%$

Om däremot *ingen* av de 50 tillhör målpopulationen, utan utgör övertäckning, får vi

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_3 = \frac{550}{550 + 175} = \frac{550}{725} = 75,9 \%$	$BA_3 = 1 - SA_3 = 1 - 0,759 = 24,1 \%$

Om man antar att en viss *andel*, u , av de 50 tillhör målpopulationen används SA_2 och BA_2 . Ett tänkbart antagande är att denna andel är densamma som målpopulationsandelen bland dem med känd målpopulationsstatus, dvs här $725/750 = 0,9667$. Vi får då

<i>Svarsandel</i>	<i>Bortfallsandel</i>
$SA_2 = \frac{550}{550 + 175 + 0,9667 \cdot 50} = 71,1 \%$	$BA_2 = 1 - SA_2 = 1 - 0,711 = 28,9 \%$

Exempel 6.2. Ett surveyinstitut har på uppdrag av en tidning undersökt yngre svenskars inställning till att sänka röståldern för allmän rösträtt i Sverige. Uppdragsgivaren var speciellt intresserad av att jämföra åldersklassen 16-17 år med åldersklassen 18-24 år. Det bestämdes därför att urvalsstorleken skulle vara ungefär lika stor i dessa två åldersklasser.

Målpopulation: Sveriges folkbokförda befolkning 16-24 år

Urvalsförfarande: Populationen delades in i två strata efter åldersklass. Inom respektive stratum drogs ett obundet slumpmässigt urval av individer enligt följande:

Stratum (h)	Populationsstorlek (N_h)	urvalsstorlek (n_h)
(I) 16-17 år	202 600	1 000
(II) 18-24 år	802 700	1 000
16-24 år	1 005 300	2 000

Urvalspersonerna blev således dragna med olika sannolikheter beroende av åldersklass och har därmed olika designvikter:

Stratum (h)	Inklusionssannolikhet	Designvikt
(I) 16-17 år	1000/202 600	202,6
(II) 18-24 år	1000/802 700	802,7

En person som är 16 eller 17 år hade alltså ungefär fyra gånger så stor sannolikhet att ingå i urvalet som en person i åldersintervallet 18-24 år. Detta kan också uttryckas som att designvikten är bara ungefär en fjärdedel för 16-17-åringar jämfört med designvikten för 18-24-åringar.

Resultatet av fältarbetet utföll enligt följande tabell:

Kategori	Stratum I	Stratum II	Samtliga (I+II)
	<i>Antal</i>	<i>Antal</i>	<i>Antal</i>
Svar	550	750	1300
Bortfall	450	250	700
därav			
<i>Ej anträffade</i>	390	190	580
<i>Vägrare</i>	50	50	100
<i>Övrigt bortfall</i>	10	10	20
Summa	1000	1000	2000

I denna undersökning förekom ingen känd övertäckning, vilket innebär att de tre ovägda bortfallsmåtten alla är lika, dvs.

$$BA(\text{ovägt}) = 1 - SA(\text{ovägt})$$

där

$$SA(\text{ovägt}) = \frac{n_s}{n_s + n_B}$$

Från tabellen ovan kan vi hämta följande värden:

$$n_s = 1300$$

$$n_B = 700$$

vilket ger

$$\begin{aligned} BA(\text{ovägt}) &= 1 - SA(\text{ovägt}) = 1 - \frac{1300}{1300 + 700} \\ &= 1 - \frac{1300}{2000} = 1 - 0,65 = 0,35 \end{aligned}$$

Det *ovägda* bortfallet blev således *35 procent*. Man beslutade att även redovisa ett *vägt* mått. Eftersom det inte förekommer några individer med okänd populationsstatus, blir också alla de tre vägda måtten lika, dvs.

$$BA(\text{vägt}) = 1 - SA(\text{vägt})$$

där

$$SA(\text{vägt}) = \frac{\sum_S d_k}{\sum_S d_k + \sum_B d_k}$$

Vi får nu

$$\sum_S d_k = 550 \cdot 202,6 + 750 \cdot 802,7 = 713455$$

$$\sum_B d_k = 450 \cdot 202,6 + 250 \cdot 802,7 = 291845$$

vilket ger

$$\begin{aligned} BA(\text{vägt}) &= 1 - SA(\text{vägt}) = 1 - \frac{713455}{713455 + 291845} \\ &= 1 - \frac{713455}{1005300} \approx 1 - 0,71 = 0,29 \end{aligned}$$

Det vägda bortfallet uppgick alltså till *29 procent*, dvs. det vägda bortfallet är 6 procentenheter mindre än det ovägda bortfallet. Det vägda bortfallet kan ses som en skattning av bortfallsandelen i populationen om en totalundersökning hade gjorts.

Exempel 6.3. Vid en marknadsundersökning avseende varumärkeskännetecken i åldersgruppen 65-74 år i Skåne läns kommuner exklusive Malmö, Lund och Helsingborg utnyttjades tvåstegsurval på följande sätt. Antalet kommuner är 30, med en sammanlagd folkmängd i den aktuella åldersgruppen om 60051 individer.

I första steget drogs utan återläggning ett urval om följande fem kommuner med sannolikheter proportionella mot den totala folkmängden (inklusionssannolikheten inom parentes): Lomma (0,14729*), Höganäs (0,19159), Eslöv (0,19850), Kristianstad (0,55478) och Hässleholm (0,37360).

Från var och en av de fem kommunerna drogs sedan ett obundet slumpmässigt urval utan återläggning ur den aktuella åldersgruppen - sammanlagt 1000 individer - med följande resultat (siffrorna inom parentes anger urvalsstorleken och totala antalet individer i åldersgruppen): Lomma (100, 1769), Höganäs (130, 2301), Eslöv (130, 2384), Kristianstad (380, 6663) och Hässleholm (260, 4487). Antal svarande i urvalen blev 95, 117, 113, 267 resp. 208, totalt 800.

Sammanställning i tabellform ger, där sista kolumnen innehåller designvikterna för individerna i de olika kommunerna (för beräkning av designvikterna, se kapitel 5, anmärkning 4):

Kommun	Inklusionssannolikheter		Antal svarande	Designvikt
	1:a steget	2:a steget		
Lomma	0.14729	100/1769 = 0,056529	95	120,102
Höganäs	0.19159	130/2301 = 0,056497	117	92,386
Eslöv	0.19850	130/2384 = 0,054530	113	92,386
Kristianstad	0.55478	380/6663 = 0,057031	267	31,606
Hässleholm	0.37360	260/4487 = 0,057945	208	46,193
Totalt			800	

Det förekommer ingen övertäckning i undersökningen.

De ovägda svars- och bortfallsandelarna blir här

$$SA_1(\text{ovägt}) = \frac{800}{1000} = 0,8 \quad \text{respektive} \quad BA_1(\text{ovägt}) = 1 - 0,8 = 0,2$$

dvs vi har ett ovägt bortfall på 20 %.

De vägda svars- och bortfallsandelarna blir

$$\begin{aligned} SA_1(\text{vägt}) &= \frac{95 \cdot 120,102 + 117 \cdot 92,386 + 113 \cdot 92,386 + 267 \cdot 31,606 + 208 \cdot 46,193}{100 \cdot 120,102 + 130 \cdot 92,386 + 130 \cdot 92,386 + 380 \cdot 31,606 + 260 \cdot 46,193} \\ &= \frac{50705}{60051} = 0,84 \end{aligned}$$

respektive

$$BA_1(\text{vägt}) = 1 - 0,84 = 0,16$$

dvs ett vägt bortfall på 16 %.

* $5 \cdot 1769 / 60051 = 0,14729$, där 1769 är antalet individer 65-74 år i Lomma.

Exempel 6.4. Partisynpatiundersökning av den svenska väljarkåren: Ett tvåstegsurval av röstberättigade genomförs på följande sätt:

Steg 1. Från en förteckning av 4 000 000 telefonnummer (fasta hushållsabonnemang, vi bortser för enkelhets skull från att en del hushåll har flera abonnemang) dras 1000 nummer med obundet slumpmässigt urval utan återläggning. Inklusionssannolikheten för varje hushåll blir $1000/4000000$.

Steg 2. När kontakt etablerats med ett hushåll dras slumpmässigt en röstberättigad person från varje hushåll. För denna dragning måste man ta reda på antal röstberättigade i varje hushåll. Inklusionssannolikheten i steg två för en röstberättigad person i ett hushåll med m röstberättigade blir $1/m$.

Urvalsmetoden ger designvikten $4000m$ för varje röstberättigad som finns i hushåll med m röstberättigade.

Resultat av fältarbetet:

Steg 1.	Kontakt med hushållet	701	
	därav svar	600	(till urvalssteg 2)
	vägran	101	(till röd box: bortfall)
	Ej kontakt, hushållsnummer	200	(till röd box: bortfall)
	Ej kontakt, oklar status	99	(till gul box: okänd status)
	Totalt	<hr/> 1000	
Steg 2.	Urvalsdragning av person genomförd	550	
	därav intervju	500	(till grön box: svar)
	vägran	25	(till röd box: bortfall)
	ej antr	25	(till röd box: bortfall)
	Övertäckning: Ingen röstberättigad i hushållet	50	(till vit box: övertäckn.)

De ovägda svarsandelarna blir här

$$SA_1(\text{ovägt}) = \frac{500}{500 + 351 + 99} \approx 0,53$$

$$SA_2(\text{ovägt}) = \frac{500}{500 + 351 + u \cdot 99}. \text{ Om } u \text{ skattas till exempelvis } 0,8 \text{ blir } SA_2(\text{ovägt}) \approx 0,54$$

$$SA_3(\text{ovägt}) = \frac{500}{500 + 351} \approx 0,59$$

De vägda svars- och bortfallsandelarna kan här inte beräknas eftersom designvikter endast kan beräknas för personer tillhörande hushåll som man etablerat kontakt med i steg 1 och fått uppgift om hushållsstorleken m (det finns alltså endast 600 sådana hushåll i exemplet). Designvikten är därmed okänd för de (icke identifierade) individer i kategori B (bortfall) för vilka bortfallet skett i steg 1 och för alla individer i kategori O (okänd status) som här består av individer för vilka hushållskontakt ej kunnat etableras.

En approximativ beräkning av vägda svars- och bortfallsandelar kan göras med följande resonemang. Hushållsstorlekar varierar måttligt och består till övervägande delen endast av en eller två vuxna (röstberättigade). Om vi ersätter alla hushållsstorlekar m med ett genomsnittsvärde ($m = \bar{m}$) bör vi därför få en hygglig approximation av de vägda måtten. Vi finner genast att om alla $m = \bar{m}$ så blir de vägda andelarna identiska med de ovägda. Under approximationen gäller sålunda att $SA_1(\text{vägt}) = SA_1(\text{ovägt})$, $SA_2(\text{vägt}) = SA_2(\text{ovägt})$ och $SA_3(\text{vägt}) = SA_3(\text{ovägt})$.

En mer förfinad approximation än ovanstående erhålls om man utnyttjar att man har kunskap om m för (i vårt fall) 600 hushåll. Då kan $\sum_S d_k$ och en del av $\sum_B d_k$ beräknas exakt.

7. Orsaker till varför man hamnar i olika bortfallskategorier

Enligt tidigare beskrivna lådmodell kan vi dela in *det kända bortfallet* i tre grupper:

- *Vägran*
- *Ej anträffad*
- *Övrigt*

Dessutom kan elementen kategoriseras som:

- *Okänd status*
- *Övertäckning*

För att konkretisera vanliga bortfallsorsaker och hur de ska definieras ger vi här exempel från fyra olika undersökningar. För var och en av dessa visar vi tänkbara orsaker till bortfall och hur de har grupperats till de fem ovannämnda bortfallskategorierna.

Det kan här vara angeläget att notera att definitionerna inte alltid är självklara och att hanteringen kan skötas mer eller mindre slentrianmässigt eller djupgående.

En distinkt definition av målpopulationen underlättar kategoriseringen av vissa bortfallsorsaker. Ska man göra en läsvärde-mätning av en svensk veckotidning kan det vara motiverat att definiera bort personer som inte är svensktalande. Om vi i stället ska studera hur boende i Sverige värderar svensk socialvård är det angeläget att även ha med exempelvis invandrare med språksvårigheter. I det förstnämnda fallet ska personer som inte talar svenska betraktas som övertäckning och i det andra fallet ska de, om de ej svarar, ingå i övrigt känt bortfall.

På samma sätt kan det för vissa undersökningar finnas anledning att närmare precisera hur man ska definiera personer som är sjuka. En person som är långvarigt/svårt sjuk kan knappast tillhöra målpopulationen i en mätning av köpbenägenhet för en klädesbutikskedjas nuvarande sortiment. Samtidigt är kanske personer som endast är förkylda högst aktuella som målpopulation för en undersökning om halstabletter. Ibland kan det finnas anledning att skilja mellan långvarigt och kortvarigt sjuka. Gränsen ska i så fall bestämmas i samband med att man beskriver syftet med undersökningen och därvid definierar målpopulationen.

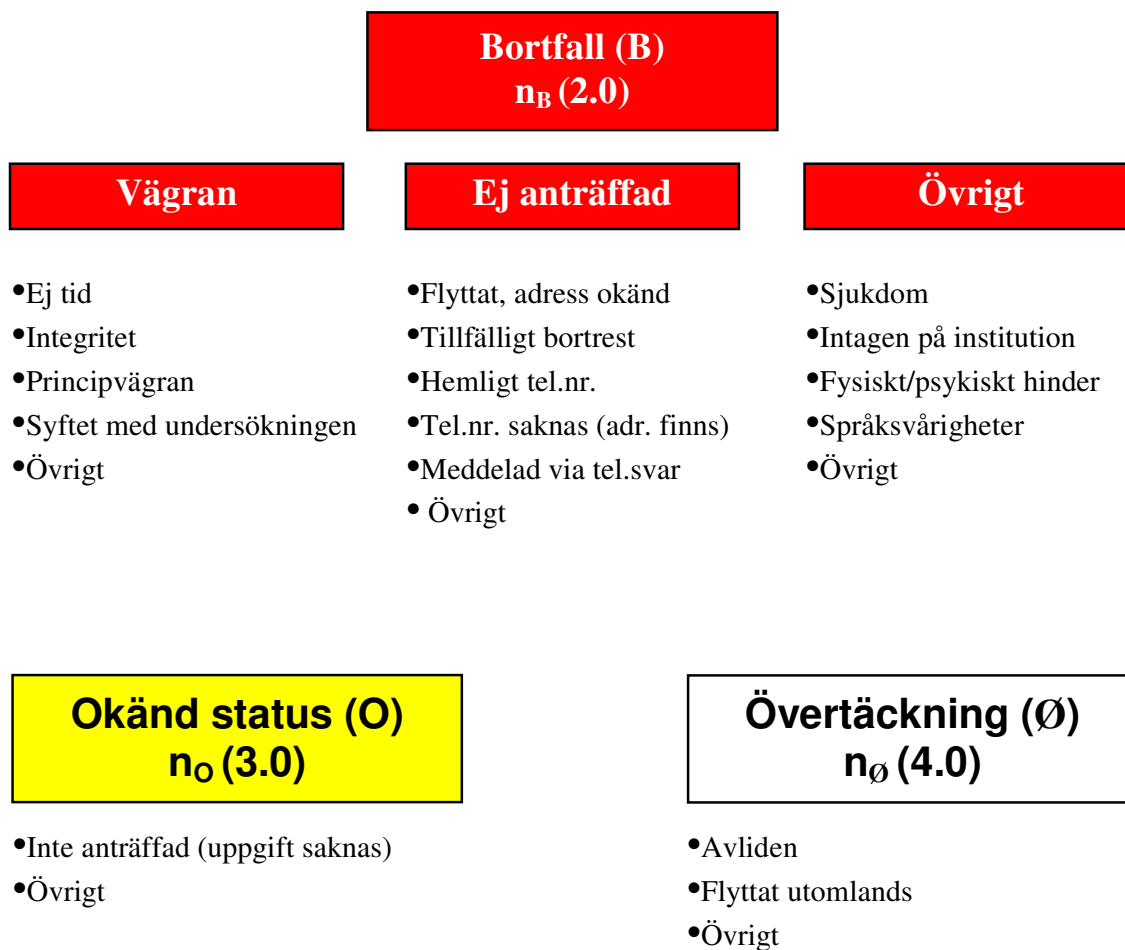
Andra exempel på praktiska problem är hur intensivt man ska arbeta för att härleda bortfall när det gäller t ex postreturer och telefonsökning. Vet man att postreturen beror på att respondenten inte vill svara är det en vägran. Vet man inget annat än att personen är målgruppsaktuell och därför borde ha svarat är det ett övrigt känt bortfall. Annars ingår postreturen under okänd status. När övergår en person som söks per telefon från ej anträffbar till okänd status?

Följande exempel ges för att underlätta förståelsen av hur elementen i en urvalsundersökning kan kategoriseras för en rimlig beräkning av svars- och bortfallsandelar.

Som nämnts i kapitel 3 har en beräkningssnurra för bortfall tagits fram, och denna finns på www.bortfallssnurran.se

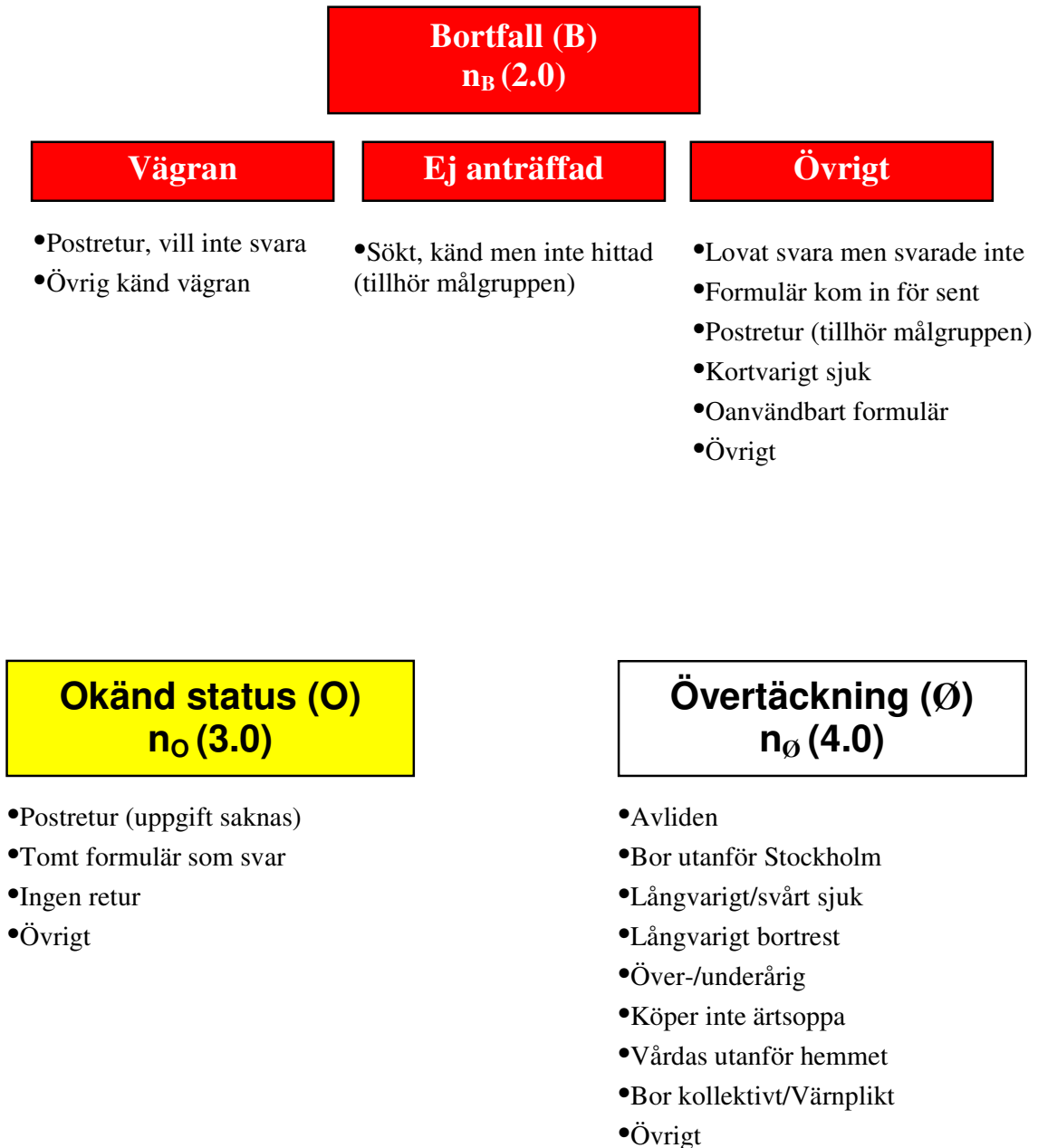
**Exempel 7.1. TELEFONINTERVJUER – sannolikhetsurval av individer.
SCB:s arbetskraftsundersökning.**

Målpopulation: Sveriges folkbokförda befolkning, 15-74 år, (exklusive personer bosatta utomlands > 2 år)



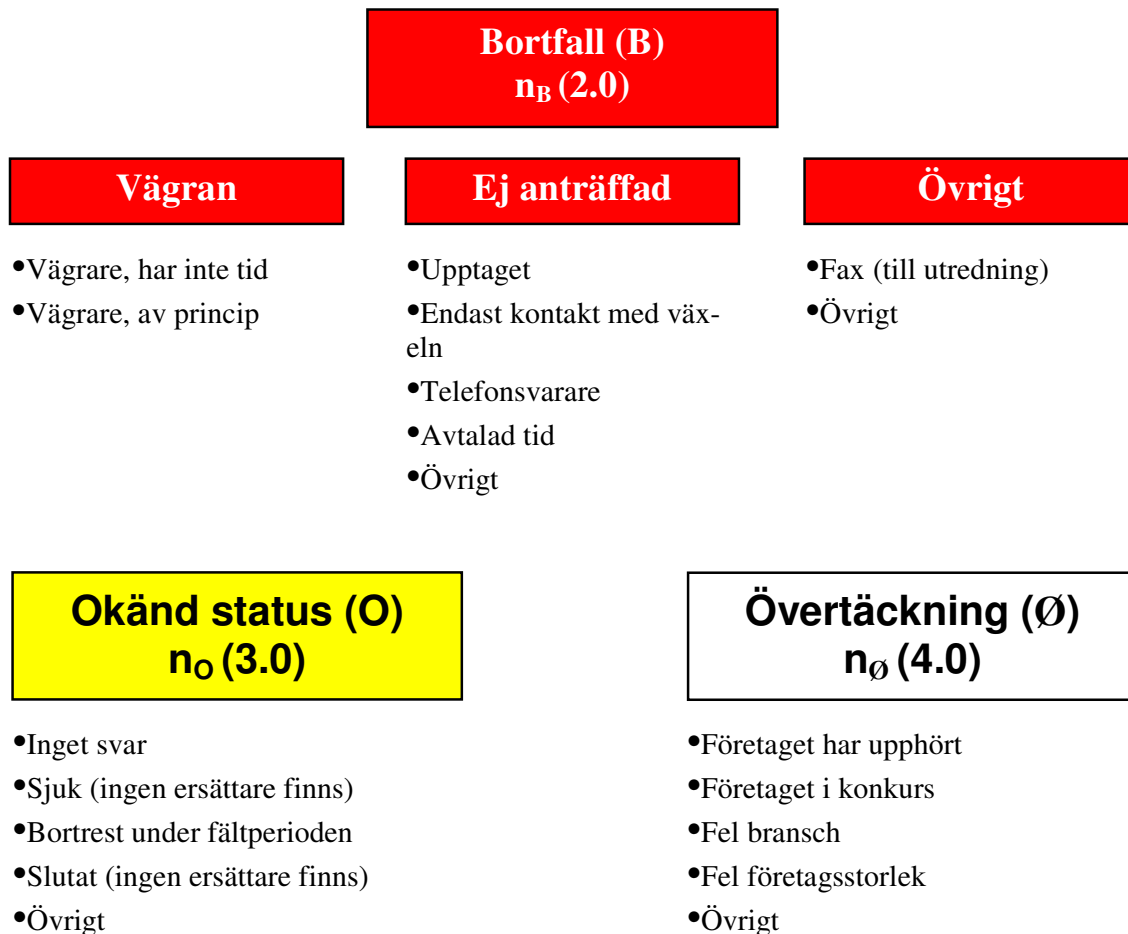
**Exempel 7.2. POSTALA ENKÄTER – sannolikhetsurval av individer.
Findus ärtsoppa – användning och värdering.**

Målpopulation: Kvinnor 20 – 45 år boende i Stockholm och köper ärtsoppa



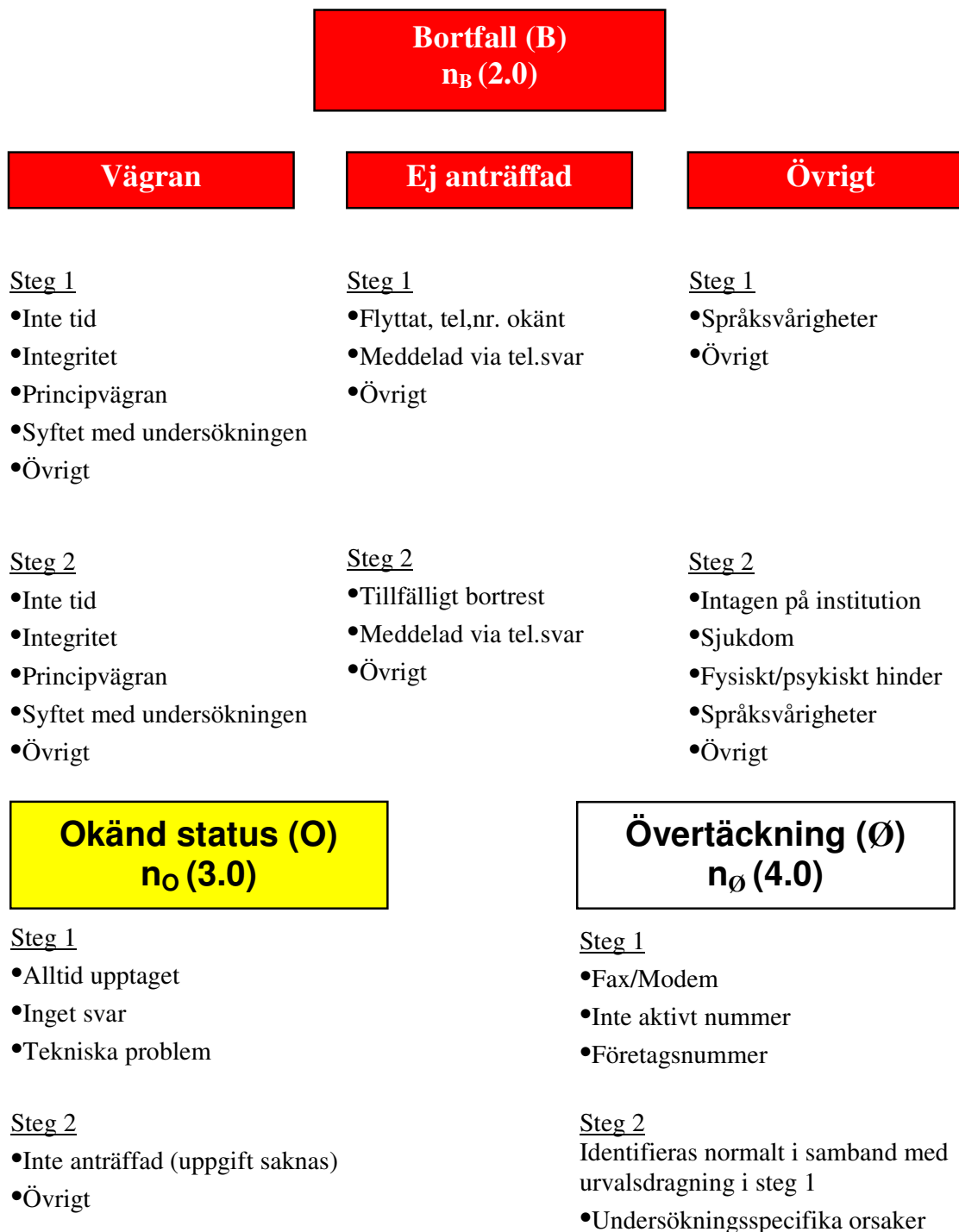
**Exempel 7.3. TELEFONINTERVJUER – sannolikhetsurval av företag.
Konjunkturbarometer för småföretag.**

Målpopulation: Privata småföretag med 1 – 49 anställda



Exempel 7.4. TELEFONINTERVJUER – telefonnummerurval av hushåll och slumpmässigt urval av en person i hushållet.

Steg 1: Kontakt med hushållet. Steg 2: Kontakt med en person i hushållet



8. Sammanfattning och rekommendationer för rapportering av bortfallet

För att producenter och konsumenter av statistiska undersökningar skall kunna bedöma och jämföra bortfallets effekt på undersökningarnas resultat är det viktigt att storleken på och orsaker till bortfallet redovisas på ett standardiserat sätt. Surveysektionens bortfallskommitté har granskat olika förekommande mått och standarder i syfte att arbeta fram och föreslå enhetliga definitioner på viktiga begrepp. Detta arbete beskrivs i rapporten, som innehåller den standard kommittén enats om att föreslå.

Avslutningsvis lämnas här också kommitténs förslag till innehåll i en bortfallsredovisning.

Allmän redovisning:

- Beskrivning av målpopulation i tid och rum inklusive eventuella begränsningar (t ex vid "cut-off" sampling)
- Urvalsram, med redovisning av kända avvikelser från målpopulationen
- Insamlingsmetod (en eller flera), insamlingstidpunkt
- Typ av urvalsförfarande
- Uppgift om totala antalet element i urvalet

Redovisning av svar och bortfall:

- Antal svarande med användbara observationer
- Antal utvalda element med icke användbara observationer
- Val av ett eller flera bortfallsmått och dessas numeriska värden

Om möjligt bör bortfallet redovisas i underkategorier t.ex. "Vägran (Avböjd medverkan)", "Ej anträffad" respektive "Övrigt". På liknande sätt bör element med okänd status och element som utgör övertäckning i möjligaste mån redovisas på underkategorier. Exempel på sådana underkategorier ges dels i exemplen i kapitel 7, dels i resultatkoderna i bilaga 2.

Ovanstående redovisning kan tillämpas såväl på undersökningen i dess helhet som på speciellt intressanta redovisningsgrupper.

Till en kvalitetsredovisning hör också att försöka ange storleken på de systematiska fel som bortfallet kan ha fört med sig samt att beskriva vilka åtgärder som har vidtagits för att minska effekterna av dessa fel.

Referenser

AAPOR (2000). *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys*. 2nd ed. The American Association for Public Opinion Research. Ann Arbor, Michigan.

AAPOR (2004). *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys*. 3rd ed. The American Association for Public Opinion Research. Lenexa, Kansas.

Hidiroglou, M.A., Drew, J.D., and Gray, G.B. (1993). *A Framework for Measuring and Reducing Nonresponse in Surveys*. *Survey Methodology*, 19, pp. 81-94.

SCB (1983). *Policy för användarorienterad kvalitetsredovisning av statistik*. Meddelanden i samordningsfrågor 1983:1. Statistiska centralbyrån, Stockholm.

SCB (2001). *Kvalitetsbegrepp och riktlinjer för kvalitetsdeklaration av officiell statistik*. Meddelanden i samordningsfrågor 2001:1. Statistiska centralbyrån, Stockholm.

Smith, T.W. (2002). *Developing Nonresponse Standards*. Ur *Survey Nonresponse*, eds: Groves, R.M, Dillman, D.A, Eltinge, J.L, and Little, R.J.A. Wiley: New York.

Litteratur som ej nämns i texten:

Eurostat: CODED – The Eurostat Concepts and Definitions Database. Webbadress: <http://forum.europa.eu.int/irc/dsis/coded/info/data/coded/en.htm>

Lynn, P., Beerten, R., Laiho, J., and Martin, J. (2002). Towards standardisation of survey outcome categories and response rate calculations. *Research in Official Statistics*, no. 1/2002.

Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer Verlag, New York

Statistical Policy Office (2001). *Measuring and Reporting Sources of Error in Surveys*. Statistical Policy Working Paper 31, Office of Management and Budget. Washington, D.C.

Vehovar, V., Batagelj, Z., Lozar Manfreda, K., and Zaletel, M. (2002). *Nonresponse in Web Surveys*. Ur *Survey Nonresponse*, eds: Groves, R.M, Dillman, D.A, Eltinge, J.L, and Little, R.J.A. Wiley: New York.

BILAGOR

Bilaga 1. Ett exempel på en kortfattad teknisk rapport (baserat på exempel 6.2)

Bilaga 2. Exempel på system för resultatkoder

Bilaga 3. Bortfallssnurra med resultatkoder

Ett exempel på en kortfattad teknisk rapport (baserat på exempel 6.2).

1 Syfte

Huvudsyftet med undersökningen var att jämföra attityder till sänkt ålder för allmän rösträtt i Sverige, i första hand eventuella skillnader i attityder mellan åldersklassen 16-17 och åldersklassen 18-24 år. Det bestämdes därför att urvalsstorleken skulle vara ungefär lika stor i dessa två åldersklasser.

2 Population och urval

Målpopulationen utgjordes av Sveriges folkbokförda befolkning 16-24 år i september 2003. Som urvalsram användes SPAR (Statens Person- och Adress-Register). Rampopulationen delades in i två strata efter åldersklass. Inom respektive stratum drogs ett obundet slumpmässigt urval av personer enligt följande:

Stratum	Antal i populationen	Antal i urvalet	Antal svar
I) 16-17 år	202 600	1 000	550
II) 18-24 år	802 700	1 000	750
16-24 år	1 005 300	2 000	1 300

Urvalspersonerna blev således dragna med olika sannolikheter beroende av vilken åldersklass de tillhörde. En person som är 16 eller 17 år hade ungefär fyra gånger så stor sannolikhet att ingå i urvalet som en person i åldersintervallet 18-24 år.

3 Undersökningens genomförande

Undersökningen utfördes under perioden 14 september – 30 september 2003. Datainsamlingen genomfördes med datorstödda telefonintervjuer av institutets intervjuare. Intervjuerna gjordes huvudsakligen under kvällstid (efter 18.00). Antalet kontaktförsök för respektive urvalsperson uppgick som mest till sex.

4. Resultat av fältarbetet

Resultatet av fältarbetet utföll enligt följande tabell:

Kategori	Stratum I		Stratum II		Samtliga (I+II)	
	<i>Antal</i>	<i>Procent</i>	<i>Antal</i>	<i>Procent</i>	<i>Antal</i>	<i>Procent</i>
Svar	550	55 %	750	75 %	1300	65 %
Bortfall	450	45 %	250	25 %	700	35 %
därav						
<i>Ej anträffade</i>	390	39%	190	19 %	580	29 %
<i>Vägrare</i>	50	5%	50	5 %	100	5 %
<i>Övrigt bortfall</i>	10	1%	10	1 %	20	1 %
Summa	1000		1000		2000	

Totalt svarade 1 300 av 2 000 personer i urvalet, medan 650 personer i urvalet resulterade i bortfall, vilket motsvarar en bortfallsandel på 35 procent av samtliga urvalspersoner. Bortfallsandelen var betydligt större för åldersklassen 16-17 år (stratum I) än för åldersklassen 18-24 år (stratum II). Den *vägda* bortfallsandelen, som tar hänsyn till att personerna har olika sannolikheter att komma med i urvalet, var betydligt lägre och uppgick till 29 procent. Den vägda bortfallsandelen kan ses som ett mått på hur stor bortfallsandelen skulle ha blivit i populationen om en totalundersökning hade gjorts. De ovägda och vägda bortfallsmåten beräknades enligt rekommendationerna i *En standard för beräkning av bortfall* (Sektionen för surveystatistik, Svenska statistikersamfundet, 2004).

De skevheter i undersökningsresultaten som kan ha uppkommit till följd av bortfallet är okända.

Exempel på system med resultatkoder							
1. Svar (S)							1.0
	<i>Fullständiga svar (F)</i>						1.1
		Enkät inkommen					1.10
		Telefonintervju					1.11
			svenska				1.111
			annat språk				1.112
		Besöksintervju					1.12
			svenska				1.121
			annat språk				1.122
		Observation i fält					1.13
	<i>Partiella svar (P)</i>						1.2
		Enkät inkommen					1.20
		Telefonintervju					1.21
			svenska				1.211
			annat språk				1.212
		Besöksintervju					1.22
			svenska				1.221
			annat språk				1.222
		Observation i fält					1.23
2. Bortfall (B)							2.0
	<i>Vägrare</i>						2.1
		Ej tid					2.10
		Integritet					2.11
		Principvägran					2.12
		Frivilligheten/vill ej svara					2.13
		Ogillar undersökningens syfte					2.14
		Tycker att man inte passar in i målgruppen					2.15
		Övrigt					2.16
	<i>Ej anträffade (känt målpopulationsobjekt)</i>						2.2
		Tillfälligt bortrest					2.20
		Hemligt telefonnummer					2.21
		Telefonnummer saknas					2.22
		Telefonsvarare					2.23
		Flyttat - adress okänd					2.24
		Känd men inte hittad					2.25
		Endast kontakt med växel					2.26
		Upptaget					2.27
		Uteblivit från avtalad tid					2.28
		Annat					2.29
	<i>Övrigt bortfall</i>						2.3
		Postretur (tillhör målgruppen)					2.30
		Lovat svara					2.31
		Sjukdom					2.32
		Intagen på institution					2.33
		Fysiskt/Psykiskt hinder					2.34
		Språksvårigheter					2.35
		Oanvändbara svar					2.36
		För sent inkomna svar					2.37
		Fax					2.38
		Övrigt					2.39
3. Okänd status (O)							3.0
		Ej anträffad (information saknas)					3.1
		Postretur (information saknas)					3.2
		Uppgiftslämnare slutat (ingen ersättare)					3.3
		Övrigt					3.4
4. Övertäckning (Ø)							4.0
		Avliden					4.10
		Sjukdom (som gör att personen inte tillhör målpopulationen)					4.11
		Inte aktivt nummer (vid telefonnummerurval)					4.12
		Företaget har upphört/konkurs					4.13
		Företaget tillhör ej målpopulationen (bransch, storlek, etc)					4.14
		Tillhör inte målpopulationen (kön, ålder, geografi, innehav, köp, etc)					4.15
		Felaktigt nummer (fax, modem, företagsnr, etc - vid tel.nr.urval)					4.16
		Dubblett - samma urvalsenhet finns tidigare i urvalet					4.17
		Övrigt					4.18

Bortfallssnurren hittar Du på: www.bortfallssnurren.se

Bortfallssnurra med resultatkoder					Resultat
1. Svar (S)				1.0	n_s = 0
	Fullständiga svar (F)			1.1	
			Enkät inkommen	1.10	
			Telefonintervju	1.11	
			svenska	1.111	
			annat språk	1.112	
			Besöksintervju	1.12	
			svenska	1.121	
			annat språk	1.122	
			Observation i fält	1.13	
	Partiella svar (P)			1.2	
			Enkät inkommen	1.20	
			Telefonintervju	1.21	
			svenska	1.211	
			annat språk	1.212	
			Besöksintervju	1.22	
			svenska	1.221	
			annat språk	1.222	
			Observation i fält	1.23	
2. Bortfall (B)				2.0	n_B = 0
	Vägrare			2.1	
			Ej tid	2.10	
			Integritet	2.11	
			Principvägran	2.12	
			Frivilligheten/vill ej svara	2.13	
			Ogillar undersökningens syfte	2.14	
			Tycker att man inte passar in i målgruppen	2.15	
			Övrigt	2.16	
	Ej anträffade (känt målpopulationsobjekt)			2.2	
			Tillfälligt bortrest	2.20	
			Hemligt telefonnummer	2.21	
			Telefonnummer saknas	2.22	
			Telefon svarare	2.23	
			Flyttat - adress okänd	2.24	
			Känd men inte hittad	2.25	
			Endast kontakt med växel	2.26	FÖ
			Upptaget	2.27	
			Uteblivit från avtalad tid	2.28	
			Annat	2.29	
	Övrigt bortfall			2.3	
			Postretur (tillhör målgruppen)	2.30	
			Lovat svara	2.31	
			Sjukdom	2.32	
			Intagen på institution	2.33	
			Fysiskt/Psykiskt hinder	2.34	
			Språksvårigheter	2.35	
			Oanvändbara svar	2.36	
			För sent inkomna svar	2.37	
			Fax	2.38	
			Övrigt	2.39	
3. Okänd status (O)				3.0	n_O = 0
	Ej anträffad (information saknas)			3.1	
	Postretur (information saknas)			3.2	
	Uppgiftslämnare slutat (ingen ersättare)			3.3	FÖ
	Övrigt			3.4	
4. Övertäckning (Ø)				4.0	n_Ø = 0
	Aviden			4.10	
	Sjukdom (som gör att personen inte tillhör målpopulationen)			4.11	
	Inte aktivt nummer (vid telefonnummerurval)			4.12	
	Företaget har upphört/konkurs			4.13	FÖ
	Företaget tillhör ej målpopulationen (bransch, storlek, etc)			4.14	FÖ
	Tillhör inte målpopulationen (kön, ålder, geografi, innehav, köp, etc)			4.15	
	Felaktigt nummer (fax, modem, företagsnr, etc - vid tel.nr.urval)			4.16	
	Dubblett - samma urvalsenhet finns tidigare i urvalet			4.17	
	Övrigt			4.18	
FÖ=specifikt för företagsundersökningar					
Skattning av u = Hur stor del av de med Okänd status (O) tillhör målgruppen?					
u kan skattas genom antaganden, metodstudier eller liknande					
	u = 0	Ingen av de med okänd status antas tillhöra målgruppen			
	u = 1	Alla av de med okänd status antas tillhöra målgruppen			
Skattning:	u =	1,00			
SA = Svandsandel:		n _s			
		n _s +	n _B +	u·n _O	
SA = Svandsandel:	SA				
BA = Bortfallsandel:	BA				