

Bortfallsanalys

Bortfallets effekt

Bortfallsfrekvensen i en undersökning används ofta som en indikator på datakvalitet. En låg bortfallsfrekvens ses som en garanti för att undersökningen är av god kvalitet och på motsvarande sätt ses en hög bortfallsfrekvens som liktydig med dålig kvalitet. Även om detta till viss del stämmer så är det ändå en grov förenkling, eftersom bortfallsfelet också beror på skillnaden i variabelvärden mellan svarande och bortfall.

Bortfall har en snedvridande effekt på skattningarna om de som svarar har andra egenskaper än de som inte svarar. Detta får vi aldrig veta med säkerhet, men däremot kan vi för registervariabler jämföra gruppen svarande med bortfallet. Om registervariablerna är (starkt) korrelerade med målvariablerna ger jämförelsen en indikation på bortfallsfelet.

Nedan följer en redovisning och försök till analys av bortfallets struktur och effekt.

Tabell 1. Analyserade bakgrundsvariabler

Variabel (benämning)	Kategorier
Elevens kön	Pojke; flicka
Regional indelning	Storstäder och förortskommuner; större och medelstora städer; övriga
Elevens födelseland	Födda i Sverige; övriga
Vårdnadshavarens ålder	0-35 år; 36-50 år; 51- år
Vårdnadshavarens födelseland	Födda i Sverige; övriga
Vårdnadshavarens inkomst	0-79 999 kr; 80 000-299 999 kr; 300 000- kr
Vårdnadshavarens civilstånd	Gift och registrerat partnerskap; övriga

Populationen i denna undersökning är sådan att vissa registervariabler inte är användbara. Eftersom alla är (ungefär) lika gamla, alla har samma utbildning och ingen förvärvsarbetar ger inte registervariablerna ålder, utbildning och inkomst någon information. Däremot är det rimligt att titta på en del av föräldrarnas ("vårdnadshavarens") variabler.

Tabell 2. Procentuell andel svarande fördelat på elevens kön.

Kön	Pojke	Flicka
Svarsandel (%)	82,13	82,67

Tabell 3. Procentuell andel svarande fördelat på region.

Region	Storstäder och förortskommuner	Större och medelstora städer	Övriga
Svarsandel (%)	78,87	84,96	90,45

Tabell 4. Procentuell andel svarande fördelat på elevens födelse land.

Födelse land	Sverige	Övriga
Svarsandel (%)	82,72	75,56

Tabell 5. Procentuell andel svarande fördelat på vårdnadshavarens ålder.

Vårdnadshavarens ålder	0-35 år	36-50 år	51- år
Svarsandel (%)	73,09	82,53	83,34

Tabell 6. Procentuell andel svarande fördelat på vårdnadshavarens födelse land.

Vårdnadshavarens födelse land	Sverige	Övriga
Svarsandel (%)	83,29	77,59

Tabell 7. Procentuell andel svarande fördelat på vårdnadshavarens inkomst.

Vårdnadshavarens inkomst	0-79 999 kr	80 000-299 999 kr	300 000- kr
Svarsandel (%)	72,93	82,37	84,49

Tabell 8. Procentuell andel svarande fördelat på vårdnadshavarens civilstånd.

Vårdnadshavarens civilstånd	Gift eller registrerat partnerskap	Övriga
Svarsandel (%)	83,30	81,21

Tabellerna 2-8 visar att (den oviktade) svarsfrekvensen är lägre i storstäder/förortskommuner än övriga landet. Dessutom finns vissa skillnader i svarsfrekvens beroende på vårdnadshavarens ålder, inkomst och födelse land. Bortfallet i enkäten kan därför ha en viss snedvridande effekt på skattningarna. Detta gäller både vid uppskattning av populationens fördelning i form av frekvenstabeller och vid beräkning av medelvärden, totalvärden och andelen element med viss en egenskap.

Storleken på eventuella skevheter orsakade av bortfallet är dock mycket svåra att uppskatta. Den enda rekommendation vi kan göra är att använda ämneskunskap och erfarenheter om materialet i samband med analys av resultaten. Exempelvis bör små skillnader som kan bero på födelse land noggrant analyseras innan slutsatser dras.

Själva viktörfarandet har gått till så att med utgångspunkt från den ordinarie vikten som räknats fram för EP6 har en justering gjorts för det bortfall som skett i samband med enkäten.

Vikten för respektive rektorsområde har justerats med en form av ”rak uppräknings”. Det innebär i princip att man gör ett antagande om att svarssannolikheten är lika inom samma stratum. Ett antagande som givetvis kan ifrågasättas. Vi har dock använt denna ”konventionella” och enkla teknik för justering av bortfall istället för en mer avancerad svarsmodell med motiveringen att vi inte har tillräcklig information för att förbättra denna modell.

Kompensation för känd övertäckning har gjorts. Övertäckningen i undersökningen har antagits vara liten och övertäckningen bland individerna i bortfallet antas vara lika stor som bland de svarande.