

## Bilaga 1 Bortfallsanalys och viktberäkning

### Bortfallsanalys

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

Följande variabler har analyserats avseende svarsfrekvens:

Tabell 1. Analyserade bakgrundsvariabler

Variabel (benämning)	Kategorier
Elevens kön	Pojke; flicka
Regional indelning	De 54 olika strata som användes i andra i andra steget i urvalsdesignen (de utvalda kommunerna)
Elevens födelse land	Födda i Sverige; övriga
Berättigad till annat modersmål i åk 9	Ja; nej
Svenska som andraspråk i åk 9	Ja; nej
Studievägsgrupp	Ej Gymnasium; Individuellt program; Yrkesförberedande program; Studieförberedande program
Vårdnadshavarens utbildningsnivå	Förgymnasial; gymnasial; eftergymnasial
Vårdnadshavarens ålder	0-35 år; 36-50 år; 51- år
Vårdnadshavarens födelse land	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

De andel svarande som beräknats i tabell 2-10 är beräknade med urvalsvikter (det visar sig dock att svarsandelar beräknade utan vikter inte skiljer sig speciellt mycket åt).

Tabell 2. Procentuell andel svarande fördelat på elevens kön.

Kön	Pojke	Flicka
Svarsandel (%)	48,18	66,05

Tabell 3. Procentuell andel svarande fördelat på om eleven var berättigad till hemspråksundervisning eller ej i åk 9.

Berättigad till hemspråk i åk 9	Ja	Nej
Svarsandel (%)	53,03	57,29

Tabell 4. Procentuell andel svarande fördelat på om eleven hade undervisning i svenska som andraspråk i åk 9

Sv2	Ja	Nej
Svarsandel (%)	45,05	57,40

Tabell 5. Procentuell andel svarande fördelat på studievägsgrupp.

Studievägsgrupp	Ej Gymnasium	Individuellt program	Yrkesförberedande program	Studieförberedande program
Svarsandel (%)	26,57	35,96	56,04	65,51

Tabell 6. Procentuell andel svarande fördelat på vårdnadshavarens utbildningsnivå.

Vårdnadshavarens utbildningsnivå	Förgymnasial	Gymnasial	Eftergymnasial
Svarsandel (%)	43,86	52,45	61,34

Tabell 7. Procentuell andel svarande fördelat på vårdnadshavarens ålder.

Vårdnadshavarens ålder	0-39 år	40-49 år	50- år
Svarsandel (%)	46,77	55,44	62,64

Tabell 8. Procentuell andel svarande fördelat på vårdnadshavarens födelseland.

Vårdnadshavarens födelseland	Sverige	Övriga
Svarsandel (%)	58,62	50,40

Tabell 9. Procentuell andel svarande fördelat på vårdnadshavarens inkomst.

Vårdnadshavarens inkomst	0-79 999 kr	80 000-299 999 kr	300 000- kr
Svarsandel (%)	49,30	57,17	62,25

Tabell 10. Procentuell andel svarande fördelat på vårdnadshavarens civilstånd.

Vårdnadshavarens civilstånd	Gift eller registrerat partnerskap	Övriga
Svarsandel (%)	60,92	51,89

För de 54 olika strata som användes i andra steget i urvalsdesignen (de utvalda kommunerna uppdelat på ”små”/”stora” rektorsområden) varierar den procentuella andelen svar mellan 33 och 85 procent.

#### **Kommentar:**

Det är ganska tydligt av tabellerna ovan att svarsfrekvensen skiljer sig åt mellan olika grupper i urvalet. Det är troligtvis också rimligt att anta att det höga bortfallet snedvrider resultat och analyser vilket kan innebära att felaktiga slutsatser dras. För att korrigera för bortfallets snedvridande effekt kan olika viktjusteringar göras. I korthet kan förfarandet beskrivas som att underrepresenterade grupper kompenseras genom att få en högre vikt.

Av ovan analyserade bakgrundsvariabler är det framförallt elevens kön, studievägsgrupp samt vårdnadshavarens ålder, utbildningsnivå och inkomst som ser ut att kunna samvariera med de okända svarssannolikheterna. Även stratumidentitet i det andra urvalssteget för elevpanelen ser ut att vara ”stark” i avseendet att den förväntas samvariera med okända svarssannolikheterna.

Efter olika vikttester har följande hjälpvariabler används i viktjusteringen:

- Stratumidentitet i andra steget i urvalsdesignen för Elevpanel 5 (dvs. utvalda kommuner uppdelat på storlek på rektorsområden)
- Studievägsgrupp
- Kön

Eftersom bortfallet är relativt stort och vissa delgrupper är små så har bedömningen gjorts att det inte är lämpligt att justera för fler variabler än de som använts. Flera av de analyserade variablerna samvarierar dessutom med varandra. Exempelvis samvarierar vårdnadshavarens ålder, utbildningsnivå och inkomst med elevens studievägsgrupp.

#### **Viktförfarande**

Ansatsen som använts är att svarshomogenitetsgrupper har definierats utifrån de 54 olika strata som användes i andra urvalssteget för Elevpanel 5. Svarshomogenitetsgrupperna består alltså av de utvalda kommunerna uppdelade på två strata efter storlek på rektorsområden inom kommunen. Som nämnts varierade svarsfrekvensen mellan 33 och 85 procent för dessa grupper.

För att försöka justera för de variabler som ansetts starkast i ett bortfallsjusteringsperspektiv har det dessutom kalibrerats mot skattade totaler för studievägsgrupp och kön.

Vikten som använts kan beskrivas enligt följande:

$$\frac{1}{\pi_{i_i}} \frac{1}{\pi_{k|i}} \frac{1}{g_{k|s}} \left[ 1 + \left( \sum_s \frac{x_k}{\pi_{i_i} \pi_{k|i}} - \sum_s \frac{x_k}{\pi_{i_i} \pi_{k|i} g_{k|s}} \right)' \left( \sum_r \frac{x_k x_k'}{\pi_{i_i} \pi_{k|i} g_{k|s}} \right)^{-1} x_k \right] \text{ där}$$

$\pi_{i_i}$  = urvalssannolikheten i första steget (kommunernas urvalssannolikhet)

$\pi_{k|i}$  = urvalssannolikheten i andra steget (rektorsområdets urvalssannolikhet)

$x_k$  = (Studievägsgrupp, Kön) – hjälpvektor i kalibreringsestimatern

$$g_{k|s} = \frac{m_h(k)}{n_h(k)} \text{ där}$$

$m_h(k)$  = Antalet svarande elever i den svarshomogenitetsgrupp som elev k tillhör

$n_h(k)$  = Antalet elever i EP5 i den svarshomogenitetsgrupp som elev k tillhör