

## Bortfallsanalys UGU 2017 - vårdnadshavare

### Bortfallsanalys på individnivå

I tabellerna nedan analyseras ett antal variabler som bedömts samvariera med svarsandelar eller viktiga undersökningsvariabler. Svarsandelarna har beräknats som antalet svarande elever delat med totalt antal elever i urvalet exklusive känd övertäckning.

Tabell 1. Andel svarande fördelat på vårdnadshavares<sup>1</sup> födelseland

Kön	Svarsandel (%)
Sverige	54,7
Övriga Världen	34,9
Totalt	49,7

Tabell 2. Andel svarande fördelat på vårdnadshavares<sup>2</sup> ålder

Vårdnadshavares ålder	Svarsandel (%)
28-39	33,7
40-45	51,1
46 -	57,1
Totalt	49,7

Tabell 3. Andel svarande fördelat på region<sup>3</sup>

Region	Svarsandel (%)
Storstadskommuner	47,3
Övriga kommuner	50,3
Totalt	49,7

<sup>1</sup> Om det finns flera vårdnadshavare används i första hand kvinnlig vårdnadshavares födelseland. Om vårdnadshavarna har samma kön används vårdnadshavare 1 enligt Befolkningsregistret

<sup>2</sup> Se 1

<sup>3</sup> Kommun där eleven var folkbokförd 2017-03-31, storstadskommuner är Stockholm, Göteborg och Malmö



Tabell 4. Andel svarande fördelat på högsta utbildning  
vårdnadshavare

Utbildningsnivå	Svarsandel (%)
Grundskoleutbildning	18,4
Gymnasial utbildning	37,7
Eftergymnasial utbildning	59,4
Totalt	49,7

Tabell 5. Andel svarande fördelat på elevens kön

Födelseland	Svarsandel (%)
Pojke	49,6
Flicka	49,8
Totalt	49,7

Tabell 6. Andel svarande fördelat på huvudman

Huvudman	Svarsandel (%)
Kommunal	49,5
Fristående	51,5
Totalt	49,7

Tabell 7. Andel svarande fördelat på betyg årskurs sex

Ämnesbetyg	Svarsandel (%)
Godkänd i alla ämnen	54,3
Minst ett F	32,5
Uppgift saknas	29,6
Totalt	49,7

Tabell 8. Andel svarande fördelat på provresultat

Provbetyg	Svarsandel (%)
Godkänd på alla delprov	54,2
Minst ett F	35,6
Uppgift saknas	31,3
Totalt	49,7

I tabellerna ovan ser man tydliga mönster. Exempelvis har vårdnadshavare födda i Sverige har svarat i klart högre grad än vårdnadshavare födda utanför Sverige. Elever som har vårdnadshavare med låg utbildning har deltagit i låg utsträckning. Elever med godkänt betyg i alla ämnen och elever med godkänt på alla prov har deltagit i klart högre grad än elever som har minst ett underkänt ämne eller prov.

Sammantaget tyder bortfallsanalyserna på att bortfallet är skevt och att vårdnadshavare med hög socioekonomi och vårdnadshavare till elever med goda betyg/provresultat är överrepresenterade bland de svarande.

För att komplettera bilden har en regressionsanalys genomförts där sambandet mellan svar/ej svar på individnivå och variablerna ovan modelleras.

Analysen ger resultatet att variablerna vårdnadshavarens födelseland, vårdnadshavarens ålder, vårdnadshavarens utbildningsnivå, betyg och provresultat har ett signifikant samband med svarsbenägenheten hos vårdnadshavarna.

### Slutligt val av variabler för viktning

Utifrån analyserna ovan används vårdnadshavarens födelseland, vårdnadshavarens ålder, vårdnadshavarens utbildningsnivå, betyg och provresultat för viktning.

### Effekter av viktning

Ett sätt att utvärdera effekten av viktningen är att studera skattningar för variabler där vi har ett facit. Den variabel som studeras är provresultat årskurs sex (andel elever som uppnådde godkänt i alla ämnen).

Elever som gick i årskurs tre hösten 2013 hämtas från Skolverkets elevregister. Elever som inte matchar mot Befolkningsregistret 2017-03-31 utesluts. Därefter hämtas provresultat från insamlingen avseende våren 2017.

Tabell 9. Antal elever efter godkänt i alla ämnen

Provresultat	Antal elever
Godkänd i alla ämnen	79 286
Minst ett F	22 638
Uppgift saknas	2 614
<b>Totalt</b>	<b>104 538</b>

Därefter skattas motsvarande uppgifter från UGU-undersökningen. Två olika estimatorer används:

Estimator 1: ingen ytterligare hjälpinformation används vid viktningen utöver stratuminformation (se formel 5.4)

Estimator 2: hjälpvariablerna vårdnadshavarens födelseland, vårdnadshavarens ålder, vårdnadshavarens utbildningsnivå, betyg och provresultat används i en kalibreringsestimator (se formel 5.5).

Tabell 10. Skattat antal elever efter godkänt i alla ämnen, estimator 1

Provresultat	Antal elever
Godkänd i alla ämnen	86 274
Minst ett F	16 996
Uppgift saknas	1 750
<b>Totalt</b>	<b>105 020</b>

Tabell 11. Skattat antal elever efter godkänt i alla ämnen, estimator 2

Provresultat	Antal elever
Godkänd i alla ämnen	80 766
Minst ett F	22 290
Uppgift saknas	2 472
<b>Totalt</b>	<b>105 528</b>

Om ingen hjälpinformation används överskattas antal elever som uppnått godkänt i alla ämnen. Detta beror på att sådana elever är överrepresenterade i svarsmängden. Vid justering av vikterna för övriga hjälpvariabler så minskar skillnaderna avsevärt. Detta är en indikation på att viktningen fungerar väl.

Tabell 11 innehåller något fler elever än tabell 9. Det beror på att elevpanelerna innehåller elever från skolenheter som inte ingår i individinsamlingen som tabell 11 baseras på. Skattningen av antalet i elever i tabell 11 är också behäftad med en viss osäkerhet.

### Teknisk beskrivning av urval och estimation

Vi har en population  $U$  bestående av  $N$  personer. De parametrar vi är intresserade av är främst 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. Vi kan definiera  $y$  (och även  $z$ ) som en dikotom variabel, d.v.s.

$$y_k = \begin{cases} 1 & \text{om person } k \text{ har studerade egenskap} \\ 0 & \text{för övrigt} \end{cases} \quad (5.1)$$

Det finns givetvis också intresse av parametrar för olika 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

$$Y_d = \sum_U y_{dk} \quad (5.2)$$

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

$Z_d$  bildas på likartat sätt.

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

$$P_d = 100 \frac{\sum_U y_{dk}}{N_d} \quad (5.3)$$

Vi drar ett urval  $s$  av storleken  $n$ , men p.g.a. övertäckning och bortfall har vi endast svarmängden  $r$  av storleken  $m$  att utföra beräkningarna på.

En "konventionell" estimator (för  $Y_d$ ), har följande form:

$$\hat{Y}_d = \sum_r \frac{1}{\lambda_k} w_{1k} w_{2k} y_{dk} \quad (5.4)$$

där

$\lambda_k$  = urvalssannolikhet för den skolenhet som eleven gick på vid urvalsdragningen 2014

$w_{1k}$  = antalet skolenheter i urvalet 2014/ antal skolenheter där det finns minst en svarande elev i vårdnadshavarundersökningen

$w_{2k}$  = antalet elever i svarande skolenheter/ antal elever där vårdnadshavare har svarat på undersökningen

Vikterna  $w_{1k}$  och  $w_{2k}$  justerar för bortfallet av hela skolenheter respektive bortfallet av elever för svarande skolenheter.

I estimator (5.4) används ingen ytterligare hjälpinformation än stratifieringsinformationen. Denna estimationsmetod brukar kallas "rak uppräknning inom strata".

I syfte att erhålla en estimator med mindre urvalsfel och bortfallsskevhet än estimator (5.4) utnyttjar vi hjälpinformation också i estimationen. Vi bildar en hjälpvektor  $\mathbf{x}_k$ , som anger till vilka kategorier av

*Kön + vårdnadshavarens födelseland + huvudman + betyg + provresultat + skolstorlek*

som elev  $k$  tillhör. Från register framställer vi sedan hjälptotalerna  $\sum_s \mathbf{x}_k$ . Vi utnyttjar denna hjälpinformation i en kalibreringsestimator.

Kalibreringsestimatoren för totalen  $Y_d$  har följande utseende:

$$\hat{Y}_{wd} = \sum_r \frac{1}{\lambda_k} w_{1k} w_{2k} v_{2k} y_{dk} = \sum_r w_k y_{dk} \quad (5.5)$$

där

$$v_{2k} = 1 + \left( \sum_s \mathbf{x}_k - \sum_r w_{2k} \mathbf{x}_k \right)' \left( \sum_r w_{2k} \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k \quad (5.6)$$

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

$$w_k = w_{1k} w_{2k} v_{2k}.$$

## Referenser:

Lundström S. och Särndal C.-E. (2001). *Estimation in the Presence of Nonresponse and Frame Imperfection*. Stockholm: Statistics Sweden