

Bortfallsanalys UGU 2017 - elever

Bortfall på skolenhetsnivå

I detta avsnitt studeras huruvida skolenhetens egenskaper påverkar benägenheten för en skolenhet att delta i undersökningen. I tabellerna studeras hur stor andel av skolenheterna som skickat in svar för minst hälften av eleverna som ingick i urvalet.

De egenskaper som studeras är skolenheternas storlek, andel elever födda utomlands, andel elever med lågutbildade vårdnadshavare samt hur skolenheterna presterat avseende betyg och ämnesprov i årskurs sex.

Vi börjar med att titta på skolenheterna efter antal elever som ingick i urvalet

Tabell 1. Svarsdata fördelat efter antalet elever i urvalet

Antal elever ¹	Enheter ²	Elever totalt ³	Enheter svar ⁴	Elever svar ⁵	SA skola ⁶	SA elev ⁷
1-2	590	739	220	265	37,3	35,9
3-7	117	492	51	200	43,6	40,7
8-15	78	892	44	483	56,4	54,1
16-25	69	1 406	50	895	72,5	63,7
26-40	65	2 120	42	1223	64,6	57,7
41-55	44	2 065	27	1142	61,4	55,3
56-70	13	851	10	506	71,4	59,5
71-101	13	1 072	7	533	53,8	49,7
Totalt	989	9 572	451	5 247	45,6	54,8

¹ Antal elever som skolenheten har med i urvalet för UGU-undersökningen

² Antal skolenheter

³ Antal elever i UGU-urvalet, exklusive känd övertäckning

⁴ Antal enheter som skickat in svar för minst hälften av sina elever

⁵ Totalt antal svarande elever, inklusive enheter som skickat in svar för mindre än 50 % av sina elever

⁶ Procentuell andel enheter där minst hälften av eleverna svarat

⁷ Procentuell andel elever som svarat, exklusive känd övertäckning



Det är svårt att se några tydliga samband mellan antal elever i urvalet och benägenheten för skolenheten att delta i undersökningen. Bortfallet är ganska jämnt spritt på både små och stora enheter. För skolenheter med väldigt få EP-elever är dock bortfallet av enheter och elever relativt stort.

Tabell 2. Svardsdata fördelat efter andel elever födda utomlands

Andel elever ⁸	Enheter	Elever totalt	Enheter svar	Elever svar	SA skola	SA elev
0-5	244	2 563	108	1 504	44,3	58,7
5-9	240	2 373	115	1 397	47,9	58,9
9-16	236	2 150	92	992	39,0	46,1
16-25	135	1 429	73	760	54,1	53,2
Över 25	116	838	54	418	46,6	49,9
Okänt ⁹	18	219	9	176	50,0	80,4
Totalt	989	9 572	451	5 247	45,6	54,8

Man skulle kunna misstänka att skolenheter med hög andel utrikes födda har svarat i lägre grad, men det finns inget tydligt sådant mönster.

⁸ Andel elever på skolenheten födda utomlands

⁹ Skolenheter där individdata inte finns tillgängligt, exempelvis konfessionella skolenheter

Tabell 3. Svarsdata fördelat efter andel elever med lågutbildade vårdnadshavare

Andel elever ¹⁰	Enheter	Elever totalt	Enheter svar	Elever svar	SA skola	SA elev
0-13	34	520	22	348	64,7	66,9
13-26	169	1 796	68	883	40,2	49,2
26-40	240	2 395	100	1 189	41,7	49,6
40-54	259	2 296	125	1 382	48,3	60,1
Över 54	269	2 346	127	1 269	47,2	54,1
Okänt	18	219	9	176	50,0	80,4
Totalt	989	9 572	451	5 247	45,6	54,8

Här kan man urskilja att skolenheter med den lägsta andelen lågutbildade, har högst svarsandel. Det är dock inte så att svarsandelen minskar med ökad andel lågutbildade. Även skolenheter med hög andel lågutbildade finns representerade ganska väl i svarsmängden.

Tabell 4. Svarsdata fördelat efter totalt antal elever på skolenheten

Antal elever ¹¹	Enheter	Elever totalt	Enheter svar	Elever svar	SA skola	SA elev
6-170	180	784	76	418	42,2	53,3
171-260	211	1 963	107	1226	50,7	62,5
261-350	189	2 051	90	1233	47,6	60,1
351-490	197	1 965	92	1082	46,7	55,1
491 -	194	2 590	77	1112	39,6	42,9
Okänt	18	219	9	176	50,0	80,4
Totalt	989	9 572	451	5 247	45,6	54,8

Det finns en svag tendens att de största skolenheterna har ett större bortfall.

¹⁰ Andel elever där ingen av vårdnadshavarna har eftergymnasial utbildning

¹¹ Antal elever totalt, årskurs 0-9

Tabell 5. Svardsdata fördelat efter huvudman

Huvudman	Enheter	Elever totalt	Enheter svar	Elever svar	SA skola	SA elev
Kommunal	718	7 268	331	4 147	46,1	57,1
Fristående	271	2 304	120	1 100	44,3	47,7
Tot	989	9 572	451	5 247	45,6	54,8

Fristående skolenheter har svarat i lägre grad än kommunala skolenheter, framför allt saknas en högre andel elever på fristående skolenheter.

Tabell 6. Svardsdata fördelat efter huvudman betyg årskurs sex

Andel elever ¹²	Enheter	Elever totalt	Enheter svar	Elever svar	SA skola	SA elev
0-65	238	1789	110	989	46,2	55,3
65-80	238	2609	115	1370	48,3	52,5
80-89	250	2253	106	1273	42,4	56,5
89-100	228	2689	113	1497	49,6	55,6
Okänt ¹³	35	232	7	118	20,0	50,9
Tot	989	9 572	451	5 247	45,6	54,8

Det finns inget mönster att skolenheter med stor andel "godkända" elever har svarat i högre grad än enheter med lägre andel "godkända" elever. I den starkaste gruppen av skolenheter (där 89-100 % av eleverna har godkänt i alla ämnen) har ca 50 procent av skolenheterna och ca 55 procent av eleverna svarat. I den svagaste gruppen (där 0-65 % av eleverna har godkänt i alla ämnen) har ca 46 procent av skolenheterna och ca 55 procent av eleverna svarat.

¹² Andel elever med godkänt i alla ämnen

¹³ Skolenheter där betygdata saknas

Tabell 7. Svardsdata fördelat efter provresultat årskurs sex

Andel elever ¹⁴	Enheter	Elever totalt	Enheter svar	Elever svar	SA skola	SA elev
0-62	239	2706	118	1558	49,4	57,6
62-76	240	2270	95	1177	39,6	51,9
76-86	237	2 352	113	1292	47,7	54,9
86-100	244	2012	118	1102	48,3	54,8
Okänt ¹⁵	29	232	7	118	24,1	50,9
Tot	989	9 572	451	5 247	45,6	54,8

Det finns inget mönster att skolenheter med stor andel "godkända" elever har svarat i högre grad än enheter med lägre andel "godkända" elever. I den starkaste gruppen av skolenheter (där 86-100 % av eleverna har godkänt på alla prov) har ca 48 procent av enheterna och ca 55 procent av eleverna svarat. I den svagaste gruppen av enheter (där 0-62 % av eleverna har godkänt på alla prov) har ca 49 procent av skolenheterna och ca 58 procent av eleverna svarat.

Det sammanfattande intrycket är att skolenhetens egenskaper inte har någon stor betydelse för om skolenheten har svarat. Bortfallet verkar bero av något annat

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

En sådan analys ger vid handen att variablerna är dåliga på att förklara om en skolenhet har svarat eller inte. Ingen av variablerna har ett signifikant samband med om skolenheten har skickat in minst hälften av sina elever eller inte.

¹⁴ Andel elever med godkänt prov betyg i engelska, matematik och svenska

¹⁵ Skolenheter där provdata saknas

Bortfallsanalys på elevnivå

I det tidigare avsnittet konstaterades att skolenhetens egenskaper inte har någon stor betydelse för om skolenheten har skickat in uppgifter om minst hälften av sina elever inte.

Det kan dock vara så att bortfallet av elever inom skolenheter är skevt fördelat med avseende på elevernas egenskaper. För att undersöka detta analyseras svarsandelar på elevnivå. Svarsandelarna har beräknats som antalet svarande elever delat med totalt antal elever i urvalet exklusive känd övertäckning.

I tabellerna nedan analyseras ett antal variabler som bedömts samvariera med svarsandelar eller viktiga undersökningsvariabler.

Tabell 8. Andel svarande fördelat på elevens kön

Kön	Svarsandel (%)
Pojke	53,2
Flicka	55,9
Totalt	54,5

Tabell 9. Andel svarande fördelat på elevens födelseland

Födelseland	Svarsandel (%)
Sverige	55,2
Övriga Världen	46,8
Totalt	54,5

Tabell 10. Andel svarande fördelat på region¹⁶

Region	Svarsandel (%)
Storstadskommuner	48,6
Övriga kommuner	56,1
Totalt	54,5

¹⁶ Kommun där eleven var folkbokförd 2017-03-31, storstadskommuner är Stockholm, Göteborg och Malmö

Tabell 11. Andel svarande fördelat på högsta utbildning
vårdnadshavare

Utbildningsnivå	Svarsandel (%)
Grundskoleutbildning	43,0
Gymnasial utbildning	54,6
Eftergymnasial utbildning	55,6
Totalt	54,5

Tabell 12. Andel svarande fördelat på vårdnadshavares¹⁷ födelseland

Vårdnadshavares födelseland	Svarsandel (%)
Sverige	56,4
Övriga Världen	48,8
Totalt	54,5

Tabell 13. Andel svarande fördelat på vårdnadshavares¹⁸ ålder

Vårdnadshavares ålder	Svarsandel (%)
28-39	53,1
40-45	55,2
46 -	54,5
Totalt	54,5

Tabell 14. Andel svarande fördelat på huvudman

Huvudman	Svarsandel (%)
Kommunal	57,1
Fristående	47,7
Totalt	54,5

¹⁷ Om det finns flera vårdnadshavare används i första hand kvinnlig vårdnadshavares födelseland. Om vårdnadshavarna har samma kön används vårdnadshavare 1 enligt Befolkningsregistret.

¹⁸ Se 17

Tabell 15. Andel svarande fördelat på betyg årskurs sex

Ämnesbetyg	Svarsandel (%)
Godkänd i alla ämnen	56,8
Minst ett F	47,9
Uppgift saknas	21,8
Totalt	54,5

Tabell 16. Andel svarande fördelat på provresultat

Provbetyg	Svarsandel (%)
Godkänd på alla delprov	56,8
Minst ett F	48,6
Uppgift saknas	31,5
Totalt	54,5

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

Sammantaget tyder bortfallsanalyserna på att skolenhetens egenskaper inte påverkar om enheten deltagit eller inte. Däremot så verkar de elever som deltagit på skolenheterna avvika från de elever som inte deltagit.

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 kön, vårdnadshavarens födelseland, huvudman, betyg och provresultat har ett signifikant samband med svarsbenägenheten hos individerna.

Slutligt val av variabler för viktning

Utifrån analyserna ovan används variablerna kön, vårdnadshavarens födelseland, huvudman, betyg, provresultat och skolstorlek 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). Denna variabel bör samvariera starkt med provresultatet i UGU-undersökningen. Om skattningar för denna variabel blir bra, bör skattningar av UGU-resultaten också bli bra.

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 17. 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
Totalt	104 538

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 kön, vårdnadshavarens födelseland, huvudman, betyg och skolstorlek används i en kalibreringsestimator (se formel 5.5).

Tabell 18. Skattat antal elever efter godkänt i alla ämnen, estimator 1

Provresultat	Antal elever
Godkänd i alla ämnen	82 843
Minst ett F	20 797
Uppgift saknas	1 729
Totalt	105 369

Tabell 19. Skattat antal elever efter godkänt i alla ämnen, estimator 2

Provresultat	Antal elever
Godkänd i alla ämnen	79 552
Minst ett F	22 904
Uppgift saknas	2 989
Totalt	105 446

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 försvinner skillnaderna nästan helt. Detta är en indikation på att viktningen fungerar väl.

Tabell 19 innehåller något fler elever än tabell 17. Det beror på att elevpanelerna innehåller elever från skolenheter som inte ingår i individinsamlingen som tabell 17 baseras på. Skattningen av antalet i elever i tabell 19 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 svarsmängden 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

λ_k = urvalssannolikhet för den skolenhet som eleven gick på vid urvalsdragningen 2014

w_{1k} = antalet skolenheter i urvalet 2014/ antal skolenheter där minst en elev svarat på UGU-undersökningen

w_{2k} = antalet elever i svarande skolenheter/ antal svarande elever

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 uppräknings 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 + proovesultat + skolstorlek

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

Kalibreringsestimatorn 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