



Statistics Sweden

Statistiska centralbyrån

# Bortfallsanalys och kalibreringsrapport

Riksmaten ungdom

SCB, Stockholm  
08-506 940 00

SCB, Örebro  
019-17 60 00

[www.scb.se](http://www.scb.se)



## Inledning

I en urvalsundersökning är alltid skattningarna behäftade med *urvalsfel* beroende på att endast en delmängd (urval) av populationen studeras. Ett annat fel uppkommer om vi inte lyckas få svar från alla personer (bortfall) och om de avviker från de svarande med avseende på undersökningsvariablerna. Detta fel kallas för *bortfallsfel*.

För att underlätta användningen av statistiken är det värdefullt om storleken på felen kan uppskattas. Av nämnda feltyper är det endast storleken på urvalsfelen som kan skattas med hjälp av urvalsinformation. Kunskap om bortfallsfelet kan i regel bara fås på ett indirekt och approximativt sätt genom att utnyttja registervariabler som  $s_k$  hjälpvariabler.

Både urvalsfel och bortfallsfel kan reduceras genom att använda ett effektivt uppräkningsförfarande (vägning som justerar för bortfallet). I undersökningen används tekniken med kalibrering av vikter, vilket är en effektiv och flexibel metod för att reducera dessa fel. I korthet går tekniken ut på att vid uppräkningsutnyttja en hjälpvektor bestående av hjälpvariabler som är kända både för de svarande och bortfallet. Dessa hjälpvariabler ska hellst samvariera med svarsbenägenheten och målvariabler samt avgränsa viktiga redovisningsgrupper.

## Bortfallsanalys

Syftet med bortfallsanalysen är att analysera vilka variabler som bör användas som hjälpinformation vid bortfallssjustering. Detta görs i avsnittet *Analys av bortfallet*.

### Analys av bortfallet

Det centrala arbetet för att få god kvalitet på skattningarna, då kalibreringsestimatorn används, är att använda stark hjälpinformation. Vid val av hjälpvariabler är det tre kriterier som ska beaktas (se Lundström och Särndal 2001):

- Det första kriteriet är att variabeln samvarierar väl med svarsbenägenheten (-sannolikheten). Det är det viktigaste kriteriet eftersom det leder till en minskning av bortfallsskevheten för alla skattningar.
- Det andra kriteriet är att variabeln samvarierar väl med (viktiga) målvariabler. Om så är fallet minskar bortfallsbiasen för de skattningar som byggs upp av dessa målvariabler. Även variansen minskar för dessa skattningar.
- Det tredje kriteriet är att variabeln avgränsar (viktiga) redovisningsgrupper. Det leder framförallt till minskad varians i skattningar för dessa redovisningsgrupper.

### Tänkbara hjälpvariabler

I tabell 1 presenteras de variabler som ingår i analysen och som är kandidater till att användas som hjälpinformation vid bortfallsjusteringen. Variablerna är hämtade från Skolenhetsregistret, Elevregistret avseende oktober 2016, ÅK9 registret 2016, Registret över avgångna från gymnasiet 2016, Registret för totalbefolkningen (RTB), Inkomst och taxeringsregistret (IoT) och Utbildningsregistret.

**Tabell 1. Tänkbara hjälpvariabler**

Variabel (benämning)	Kategorier (koder)	Källa (aktualitet)
ÅRSKURS	5 = Åk 5 8 = Åk 8 2 =Gymn åk 2	Skolenhetsregistret
KOMMUNGRUPP	(grupperad) 1 = Förortskommuner till storstäder 2 = Förortskommuner till större städer/Större städer 3 = Kommuner i tätbefolkad region 4 = Storstäder 5 = Övriga kommuner	Skolenhetsregistret
HUVUDMANSTYP	1 = Kommunal/Landsting 2 = Enskild	Skolenhetsregistret
MERITVÄRDE skolenivå	1 = < 225 2 = 225-244 3 = ≥ 245 4 = Inget värde, okänt	ÅK9 registret 2016
MEDELBETYG skolenivå	1 = < 13,5 2 = 13,5-14,5 3 = > 14,5 4 = Inget värde, okänt	Registret över avgångna från gymnasiet 2016
ANTAL ELEVER PER SKOLENHET OCH ÅRSKURS	1 = < 60 2 = 60-99 3 = ≥ 100	Elevregistret 2016
KÖN	1 = Pojke	RTB 2016-12-31

	2 = Flicka	
BAKGRUND	1 = Svensk 2 = Utländsk	RTB 2016-12-31
DISPONIBEL INKOMST HUSHÅLL	1 = < 450 000, okänd 2 = 450 000 – 649 999 3 = ≥ 650 000	loT2015
FÖRÄLDRARS UTBILDNINGSNIVÅ	1 = Grundskoleutbildning 2 = Kort eftergymnasial utbildning 3 = Eftergymnasial utbildning	UREG 2016

Kommungrupp, huvudmangrupp och årskurs är på skolenhetsnivå och är stratifieringsvariabler. Meritvärde och medelbetyg avser skolenhetens värde. Antal elever per skolenhet och årskurs avser skolenhetsnivå. Grupperingen utgår från antalet elever per aktuell årskurs på skolenheten. För Årskurs 5 har grupp 2 och 3 slagits ihop till en grupp.

Medelbetyg avser avgångsbetyg från gymnasieskolan, och är aktuellt för elever på gymnasiet årskurs 2. Meritvärde avser slutbetyg i årskurs 9 och är aktuellt för elever i årskurs 5 och 8.

För uppgiften om föräldrarnas utbildningsnivå och inkomst har det största av modern och faderns värde använts.

### Svarsandelar

Sambandet mellan de svarande och bortfallet studeras genom att beräkna den vägda svarsandelen per kategori för de tänkta hjälpvariablerna (se tabell 2-11b). Svarsandelar presenteras både avseende skolenheter i första steget och elever i svarande skolenheter. Vid vägningen används designvikten i första steget respektive designvikten för eleverna hos svarande skolor. Vid stora skillnader mellan svarsandelarna samvarierar svarsbenägenheten för variabeln och den kan antas vara en stark hjälpvariabel enligt kriterium 1.

Den vägda svarsandelen totalt bland skolenheter var 22,2 procent och bland elever 68,7 procent.

Den vägda svarsandelen totalt bland skolenheter (provtagning) var 23,7 procent och bland elever (provtagning) 55,3 procent.

**Tabell 2. Skattad procentuell andel svarande skolenheter och elever fördelat på årskurs**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
Åk 5	19,9	78,3	42	1217
Åk 8	22,6	66,7	43	1198
Gymn åk 2	27,3	62,7	46	1061

**Tabell 2b. Skattad procentuell andel svarande skolenheter och elever fördelat på årskurs (provtagning)**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
Åk 5	18,9	64,7	19	434
Åk 8	27,6	51,5	23	474
Gymn åk 2	29,0	50,9	20	396

**Tabell 3. Skattad procentuell andel svarande skolenheter och elever fördelat på kommungrupp**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
1	19,7	74,0	15	483
2	27,2	70,8	52	1425
3	19,3	62,4	12	301
4	16,7	61,4	17	471
5	20,7	67,9	35	796

**Tabell 3b. Skattad procentuell andel svarande skolenheter och elever fördelat på kommungrupp (provtagning)**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
1	18,7	66,4	5	123
2	29,1	60,0	25	539
3	35,9	56,4	9	184
4	17,8	44,5	9	187
5	18,2	47,3	14	271

**Tabell 4. Skattad procentuell andel svarande skolenheter och elever fördelat på huvudmanstyp**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
Kommunal/Landsting	22,1	72,0	101	2903
Enskild	22,4	56,2	30	573

**Tabell 4b. Skattad procentuell andel svarande skolenheter och elever fördelat på huvudmanstyp (provtagning)**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
Kommunal/Landsting	23,0	58,0	45	1016
Enskild	26,5	43,8	17	288

**Tabell 5. Skattad procentuell andel svarande skolenheter och elever fördelat på meritvärde på skolnivå**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
< 225	21,3	62,2	21	540
225-244	25,3	79,3	19	631
≥ 245	20,0	66,9	18	517
Inget värde, okänt	22,3	67,1	73	1788

**Tabell 5b. Skattad procentuell andel svarande skolenheter och elever fördelat på meritvärde på skolnivå (provtagning)**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
< 225	26,2	40,6	11	177
225-244	19,1	52,7	7	142
≥ 245	18,0	62,4	11	281
Inget värde, okänt	25,2	56,3	33	704

**Tabell 6. Skattad procentuell andel svarande skolenheter och elever fördelat på medelbetyg på skolnivå**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
< 13,5	35,5	72,7	18	434
13,5-14,5	19,6	50,0	13	238
> 14,5	29,0	62,9	15	389
Inget värde, okänt	20,6	72,3	85	2415

**Tabell 6b. Skattad procentuell andel svarande skolenheter och elever fördelat på medelbetyg på skolnivå (provtagning)**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
< 13,5	38,7	67,0	6	99
13,5-14,5	27,4	36,3	9	159
> 14,5	19,8	58,8	5	138
Inget värde, okänt	21,9	57,3	42	908

**Tabell 7. Skattad procentuell andel svarande skolenheter och elever fördelat på antal elever per skolenhet och årskurs**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
< 60	23,7	70,9	58	1386
60-99	21,0	70,4	39	1124
≥ 100	17,4	63,6	34	966

**Tabell 7b. Skattad procentuell andel svarande skolenheter och elever fördelat på antal elever per skolenhet och årskurs (provtagning)**

Kategori	Svarsandel		Antal svar	
	Skolenheter	Elever	Skolenheter	Elever
< 60	27,4	56,1	28	550
60-99	19,5	55,6	18	355
≥ 100	16,8	51,8	16	399



**Tabell 8. Skattad procentuell andel svarande elever fördelat på kön**

Kategori	Svarsandel	Antal svar
	Elever	Elever
Pojke	65,8	1635
Flicka	71,4	1841

**Tabell 8b. Skattad procentuell andel svarande elever fördelat på kön (provtagning)**

Kategori	Svarsandel	Antal svar
	Elever	Elever
Pojke	50,1	572
Flicka	59,1	732

**Tabell 9. Skattad procentuell andel svarande elever fördelat på bakgrund**

Kategori	Svarsandel	Antal svar
	Elever	Elever
Svensk	68,9	2739
Utländsk	67,7	737

**Tabell 9b. Skattad procentuell andel svarande elever fördelat på bakgrund (provtagning)**

Kategori	Svarsandel	Antal svar
	Elever	Elever
Svensk	55,4	1020
Utländsk	53,2	284

**Tabell 10. Skattad procentuell andel svarande elever fördelat på hushållets disponibla inkomst**

Kategori	Svarsandel	Antal svar
	Elever	Elever
< 450 000, okänd	67,6	1104
450 000 – 649 999	68,0	1085
≥ 650 000	70,2	1287

**Tabell 10b. Skattad procentuell andel svarande elever fördelat på hushållets disponibla inkomst (provtagning)**

Kategori	Svarsandel	Antal svar
	Elever	Elever
< 450 000, okänd	53,0	424
450 000 – 649 999	53,4	406
≥ 650 000	58,3	474

**Tabell 11. Skattad procentuell andel svarande skolenheter fördelat på föräldrars utbildningsnivå**

Kategori	Svarsandel	Antal svar
	Elever	Elever
Grundskoleutbildning	67,9	348
Kort eftergymnasial utbildning	67,7	1730
Eftergymnasial utbildning	70,2	1398

**Tabell 11b. Skattad procentuell andel svarande skolenheter fördelat på föräldrars utbildningsnivå (provtagning)**

Kategori	Svarsandel	Antal svar
	Elever	Elever
Grundskoleutbildning	53,2	126
Kort eftergymnasial utbildning	53,0	623
Eftergymnasial utbildning	58,0	555

För vissa av variablerna redovisas svarsandelar både på skolenhetsnivå och elevnivå. Resultat uppdelat per årskurs studeras men redovisas inte här. För individvariablerna redovisas endast svarsfrekvenser för eleverna. I b-tabellerna redovisas provtagningsgruppen.

Andelen svarande skolenheter och elever skiljer sig en del mellan de olika kommungrupperna. Storstäder har både lägst andel svarande skolenheter och svarande elever i dessa. Gäller både ordinarie och provtagningsgruppen.

Det är ingen skillnad på svarsandelar mellan huvudmansgrupperna för skolenheterna men högre svarsfrekvens för eleverna i skolenheter med kommunala huvudmän. Gäller både ordinarie och provtagningsgruppen.

Gymnasieskolenheterna har hög svarsfrekvens och grundskolenheter årskurs 5 låg. För eleverna i svarande skolenheter är sambandet det omvända. Liknande mönster i provtagningsgruppen.

Skolenhetens meritvärden eller medelbetyg verkar ha påverkan på svarsfrekvensen både på skolenhetsnivå och elevnivå. Det gäller både huvud- och provtagningsgruppen.

Skolenheter med färre elever per skolenhet (i aktuell årskurs) svarar i större utsträckning än större skolenheter. Eleverna i de större enheterna har något mindre svarsandel. Gäller både huvud och provtagningsgrupp.

Ovan nämnda variabler är därmed intressanta som hjälpvariabler enligt kriterium 1. Skillnader (om än i annan riktning) kvarstår om vi tittar på svarsandelar uppdelat på årskurs.

Svarandelarna för individvariablerna varierar i mindre utsträckning mellan grupperna. Utom möjligen kön i provtagningsgruppen. Det gör dem till svagare kandidater som hjälpvariabler enligt kriterium 1. Dock finns det skillnader om vi tittar på svarsandelar uppdelat på årskurs.

## Fördelningar

För att bedöma representativitet kan fördelningen över de svarande, urval och population studeras<sup>1</sup>. Detta tangerar i viss mån analysen om svarsandelar där hur väl svarsfördelningen stämmer överrens med fördelningen i urvalet. Detta kan också visa på en skev svarsfördelning vilket man bör ta hänsyn till vid beräkningen av kalibrerade vikter. Här studeras även representativiteten mot populationen.

Ur ett designperspektiv är inte urvalet representativt mot populationen av skolenheter, då större skolenheter fick en större inklusionssannolikhet. Urvalet och svarsmängden (speciellt av elever) kan dock fortfarande vara representativ mot populationen med avseende på hjälpvariablerna.

I urvalet översamlades skolenheter från gymn. årskurs 2 och åk 5 undersamlades. Det gör också att skolenheterna med gymnasieklaser är överrepresenterade och skolenheter med femteklassare är underrepresenterade i svarsmängden. I populationen är andelen elever per årskurs jämt fördelade, medan urvalet har en övervikt av gymnasieelever. Svarsmängden är dock mer representativ för populationen.

Andelen skolenheter i urvalet per kommungrupp stämmer väl överrens med populationen, och även relativt väl med fördelningen av skolenheter i svarsmängden. Detta beror delvis beror på att kommungrupp var en stratifieringsvariabel. Fördelningen av elever i urvalet över kommungrupper stämmer också väl överrens med den i populationen. Bland de svarande i provurvalet finns det dock en överrepresntation av skolenheter från kommungrupp 3 och 4 och en underrepresntation av kommungrupp 1. Bland de svarande eleverna är de i provgruppen i kommungrupp 1 och 4 något underrepresenterade och kommungrupp 2 och 3 något överrepresenterade. I ordinarie grupp är kommungrupp 2 något överrepresenterad och kommungrupp 4 något underrepresenterad i förhållande till populationen. De mindre skolenheterna har dock högre svarsandel vilket gör svarsmängden mer representativ på skolenhetsnivå om än fortfarande inte helt. Elever på större skolenheter är överrepresenterade i urvalet, men omvänt är elever i mindre skolenheter bäst representerade bland de svarande. Svarsmängden av elever är då mer reepresentativt för populationen som har liknande förhållande.

Urvalet av skolenheter av stratifierat på huvudmanstyp vilket gör urvalet representativt för populationen av skolenheter. Det gäller också de svarande skolenheterna. Elever på kommunala skolenheter utgör 80 procent av populationen av elever. I urvalet är andelen från enskilda skolenheter något mer representerade. Elever på kommunala skolenheter svarar dock i

---

<sup>1</sup> Inga tabeller redovisas här. Se vidare excelfiler Tabell Fördelningar Skolor, Tabell Fördelningar Elever

högre utsträckning vilket gör svarsmängderna (ordinarie och provtagning) väl representerativa för populationen.

Vad det gäller bakgrund har eleverna samma fördelning såväl i population och urval. Bland de svarande finns något större andel med svensk bakgrund (ordinarie och provtagning). Det finns något fler pojkar än flickor i populationen vilket också är förhållandet i urvalet. Bland de svarande är dock flickorna i majoritet.

Fördelningen i populationen för hushållets inkomst är jämn över de tre grupperna. Gruppen med hög inkomst är dock överrepresenterade i urvalet. Detta gäller också i viss mån de svarande tillhörande ordinarie grupp medan provtagningsgruppens fördelning mer liknar populationens. Utbildningsgrupp 2 är den största gruppen i populationen och låg utbildningsnivå är den minsta. Denna relation återspeglas i urvalet och även hos de svarande.

Betygspoäng och meritvärde kommenteras inte här då variablerna är mer relevanta för enskilda årskurser (vilket vi inte går in på här).

### Samvariation med målvariabler

Ett antal variabler har valts ut till analysen om samvariation mellan målvariabler och hjälpvariabler. Dessa är förklarade i tabell 12.

**Tabell 12. Målvariabler**

Målvariabel	Förklaring (se även frågeformuläret)
Viktstatus	1: UN/NW (under-/normalvikt)
Lunch_dagligen	1: Inte dagligen
Grönsaker_dagligen	1: Inte dagligen
Fruktbär_dagligen	1: Inte dagligen
Sockersötadedrycker_vecka	1: =<1 g/v
Lightdrycker_vecka	1: =<1 g/v
Fisk_vecka	1: <1 g/v
Sötsaker_vecka	1: =<3 g/v

I tabellerna 13-22 skattas den procentuella andelen elever med egenskapen per målvariabel (enligt tabell 12) inom hjälpvariablernas olika kategorier. Skattningarna görs med hjälp av en bortfallsjusterad, både för skolenheter och klasser, designvikt. Stora skillnader i andelarna mellan kategorierna för de olika målvariablerna tyder på att hjälpvariabeln har ett starkt samband med målvariabeln vilket efterfrågas enligt kriterium 2.

**Tabell 13. Årskurs**

Målvariabel	Åk 5	Åk 8	Gymn åk 2
Lunch_dagligen	12,8	31,5	24,8
Grönsaker_dagligen	60,2	58,9	63,9
Fruktbär_dagligen	70,5	68,9	72,6
Sockersötadedrycker_vecka	76,4	70,6	62,7
Lightdrycker_vecka	86,6	80,4	73,0
Fisk_vecka	37,3	35,9	34,8
Sötsaker_vecka	71,3	62,2	60,6
Viktstatus	77,2	83,4	76,8

**Tabell 14. Kommungrupp**

Målvariabel	1	2	3	4	5
Lunch_dagligen	28,8	19,8	21,2	32,3	20,5
Grönsaker_dagligen	56,2	63,2	64,1	59,5	61,2
Fruktbär_dagligen	71,0	70,7	70,4	74,4	69,1
Sockersötadedrycker_vecka	69,6	67,6	69,6	72,0	70,6
Lightdrycker_vecka	77,7	77,0	83,3	82,4	82,6
Fisk_vecka	34,7	32,4	42,3	41,0	38,9
Sötsaker_vecka	69,6	63,6	62,9	69,4	60,0
Viktstatus	82,4	79,2	76,9	84,5	73,7

**Tabell 15. Huvudmantyp**

Målvariabel	Kommunal/Landsting	Enskild
Lunch_dagligen	22,8	24,5
Grönsaker_dagligen	61,3	61,0
Fruktbär_dagligen	70,7	72,1
Sockersötadedrycker_vecka	69,1	70,7
Lightdrycker_vecka	79,5	78,3
Fisk_vecka	35,1	42,0
Sötsaker_vecka	63,6	69,6

Viktstatus	78,8	79,9
------------	------	------

**Tabell 16. Meritvärde på skolnivå**

Målvariabel	< 225	225-244	≥ 245	Inget värde, okänt
Lunch_dagligen	30,2	25,0	28,7	19,9
Grönsaker_dagligen	68,2	56,1	55,5	63,1
Fruktbär_dagligen	69,5	69,2	69,1	72,0
Sockersötadedrycker_vecka	66,1	73,5	76,2	66,9
Lightdrycker_vecka	78,0	85,3	82,9	76,9
Fisk_vecka	39,3	38,8	31,1	35,5
Sötsaker_vecka	58,2	65,8	71,0	63,4
Viktstatus	76,3	79,9	91,4	76,4

**Tabell 17. Medelbetyg på skolnivå**

Målvariabel	< 13,5	13,5-14,5	> 14,5	Inget värde, okänt
Lunch_dagligen	34,0	12,4	26,9	21,9
Grönsaker_dagligen	71,2	64,1	52,3	59,6
Fruktbär_dagligen	76,1	73,1	66,5	69,7
Sockersötadedrycker_vecka	57,1	66,2	67,2	73,5
Lightdrycker_vecka	66,9	77,2	77,3	83,5
Fisk_vecka	46,9	22,1	31,8	36,6
Sötsaker_vecka	62,6	62,1	55,9	66,8
Viktstatus	72,7	82,3	76,3	80,3

**Tabell 18. Antal elever per skolenhet och årskurs**

Målvariabel	< 60	60-99	≥ 100
Lunch_dagligen	18,7	26,0	25,3
Grönsaker_dagligen	63,0	60,3	60,0
Fruktbär_dagligen	70,6	71,7	70,2
Sockersötadedrycker_vecka	71,1	71,9	64,1

Lightdrycker_vecka	80,8	81,7	75,1
Fisk_vecka	39,0	41,8	25,8
Sötsaker_vecka	67,0	65,8	59,6
Viktstatus	76,3	80,6	80,5

**Tabell 19. Kön**

Målvariabel	Pojke	Flicka
Lunch_dagligen	20,3	25,3
Grönsaker_dagligen	65,9	57,2
Fruktbär_dagligen	77,3	65,1
Sockersötadedrycker_vecka	62,8	74,9
Lightdrycker_vecka	73,8	84,3
Fisk_vecka	32,7	38,8
Sötsaker_vecka	65,3	63,6
Viktstatus	78,6	79,3

**Tabell 20. Bakgrund**

Målvariabel	Svensk	Utländsk
Lunch_dagligen	21,7	28,1
Grönsaker_dagligen	61,1	62,0
Fruktbär_dagligen	72,0	66,4
Sockersötadedrycker_vecka	69,7	67,6
Lightdrycker_vecka	79,3	79,7
Fisk_vecka	34,2	42,8
Sötsaker_vecka	64,6	63,5
Viktstatus	79,2	77,9

**Tabell 21. Hushållets disponibla inkomst**

Målvariabel	1	2	3
Lunch_dagligen	25,0	22,4	21,8
Grönsaker_dagligen	63,0	65,2	56,8
Fruktbär_dagligen	71,8	71,5	69,5



Sockersötadedrycker_vecka	63,4	67,5	75,5
Lightdrycker_vecka	76,3	78,9	82,3
Fisk_vecka	41,1	39,1	29,3
Sötsaker_vecka	65,6	63,9	63,7
Viktstatus	77,2	76,6	82,3

**Tabell 22. Föräldrars utbildningsnivå**

Målvariabel	1	2	3
Lunch_dagligen	29,6	25,3	18,4
Grönsaker_dagligen	61,7	66,6	54,6
Fruktbär_dagligen	62,4	74,3	68,7
Sockersötadedrycker_vecka	68,1	64,1	76,0
Lightdrycker_vecka	78,8	76,1	83,7
Fisk_vecka	41,2	39,4	30,4
Sötsaker_vecka	65,8	63,8	64,6
Viktstatus	80,6	75,9	82,4

Skattningar för eleverna redovisas per hjälpvariabel och målvariabel. Resultat uppdelat per årskurs studeras men redovisas inte här.

Skattningarna för målvariablerna varerar en del mellan årskurserna speciellt *lunch dagligen*. *Fruktbär dagligen* och *fisk vecka* varierar dock mindre.

Hjälpvariabeln Medelbetyg på skolnivå verkar påverka samtliga målvariabler. Variationen för skattningarna är dock inte så stor för hjälpvariabeln Huvudmanstyp.

På liknande sätt varierar skattningarna åtminstone för någon målvariabel för de flesta hjälpvariabler. Det gör dem användbara i en hjälpvektor enligt kriterium 2.

### **Avgränsning av redovisningsgrupper**

Om hjälpvariabeln avgränsar viktiga redovisningsgrupper kan kvaliteten bli bättre i dessa grupper. Framförallt blir skattningarna säkrare om hjälpvariabeln väl avgränsar redovisningsgruppen enligt kriterium 3.

Kalibreringsestimern ger konsistenta skattningar i den meningen att estimern ger exakta skattningar för utnyttjade registertotaler. Det är därför viktigt att vid val av hjälpvariabler fundera på vilka skattningar man vill få konsistenta.

Årskurs och kön antas utgöra viktiga redovisningsgrupper.

### **Slutligt val av hjälpvektor**

Efter en sammanvägning av analysen kring ovanstående samt efter kontroll av vikternas fördelning används följande hjälpvektor i kalibreringen:

För huvudurvalet:

*Årskurs \* (kommungrupp + huvudmanstyp + antal elever per skolenhet och årskurs+kön)*

För provurvalet:

*Årskurs \* (kommungrupp + antal elever per skolenhet och årskurs+kön)*

I gymnasiet årskurs 2 är kommungrupp 1 och 2 sammanslagna i hjälpvektorn.

Inkludering av ytterligare variabler gav för stora kalibreringsvikter.

## Teknisk beskrivning av urval och estimation

Vi har en population  $U$  bestående av  $N$  individer (elever). 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 elev  $k$  och  $z_k$  värdet på en annan variabel för samma elev. Vi kan definiera  $y$  (och även  $z$ ) som en dikotom variabel, dvs.

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

Det finns givetvis också intresse av parametrar för olika domäner (redovisningsgrupper). Låt oss benämna dessa  $U_1, \dots, U_d, \dots, U_D$ , där  $U = \bigcup_{d=1}^D U_d$ .

Totalen för domän  $d$  kan skrivas

$$Y_d = \sum_U y_{dk} \quad (2)$$

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

$Z_d$  bildas på likartat sätt.

En generell parameter för domän  $d$  ( $d$  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 (1). Om vi låter  $N_d$  vara antalet elever i redovisningsgrupp  $d$ , då kan parametern skrivas

$$P_d = 100 \frac{\sum_U y_{dk}}{N_d} \quad (3)$$

Vi drar ett Pareto  $\pi$ ps urval  $s_{Ih}$  av storleken  $n_{Ih}$  från stratum  $h$  ( $h = 1, \dots, H$ ) av skolenhet ( $I=1, \dots, I$ ) i första steget, men p.g.a. övertäckning och bortfall har vi endast svarsmängden  $r_{Ih}$  av storleken  $m_{Ih}$  att utföra beräkningarna på. Provvurvalet har storleken  $n_{Ih}^p$ . Designvikten är i första steget  $a_1 = \frac{1}{\lambda_{Ih}}$ , där  $\lambda_{Ih}$  är inklusionssannolikheten för varje skolenhet  $I$  inom stratum  $h$ , vilken är proportionell mot antalet elever i respektive årskurs oktober 2015

hos skolenheten (vid tidpunkten för första urvalsdragningen). För provurvalet gäller  $a_1 = \frac{1}{\lambda_{lh}} \cdot \frac{n_{lh}}{n_{lh}^p}$ . De skolenheter vars klasser togs med i urvalet under insamlingen får designvikten 1.

I andra steget dras ett urval om  $n_{l|lh}$  klasser från totala antalet klasser  $N_{l|lh}$   $N_{lh}$  per svarande skolenhet och årskurs, vilket antas vara ett OSU. Designvikten för klass  $l$  i andra steget utgörs av  $a_2 = \frac{N_{l|lh}}{n_{l|lh}}$  (samma vikt för provurvalet). Samtliga elever i utvalda klasser får svara på enkäten. Ett antal klasser blir bortfall och återstår gör  $m_{l|i}$  svarande klasser.

Estimatorn för  $Y_d$  enligt (4) kallas "rak uppräknings" och har följande form:

$$\hat{Y}_d = \sum_r a_{1k} a_{2k} y_k = \sum_r a_k y_k \quad (4)$$

där  $r$  är mängden svarande elever och  $a_{1k} = \frac{1}{\lambda_{lh}} \cdot \frac{n_{lh}}{m_{lh}}$  respektive

$$a_{1k} = \frac{1}{\lambda_{lh}} \cdot \frac{n_{lh}}{n_{lh}^p} \cdot \frac{n_{lh}^p}{m_{lh}^p} \text{ för provurvalet och } a_{2k} = \frac{N_{l|lh}}{m_{l|lh}}.$$

I estimator (4) används ingen ytterligare hjälpinformation än stratifieringsinformationen. I syfte att erhålla en estimator med mindre urvalsfel och bortfallsskevheter utnyttjar vi hjälpinformation inte bara i urvalsdesignen utan också i estimationen. Vi bildar en hjälpvektor  $\mathbf{x}_k$ , som anger till vilka kategorier av *Årskurs \* (kommungrupp + huvudmanstyp + antal elever per skolenhet och årskurs+kön)* som elev  $k$  tillhör. Från urvalsramen av skolenheter tillsammans med elevuppgifter för 2016 framställer vi sedan hjälptotalerna  $\sum_{U_d} \mathbf{x}_k$ , där  $\mathbf{x}_k$  är en hjälpvektor med värden avseende antalet elever. Vi utnyttjar denna hjälpinformation i en kalibreringsestimator. För provurvalet används hjälpvektorn *Årskurs \* (kommungrupp + antal elever per skolenhet och årskurs+kön)*.

Kalibreringsestimatorn för totalen  $Y_d$  har följande utseende:

$$\hat{Y}_{wd} = \sum_r g_k a_k y_{dk} \quad (5)$$

$$\text{där } g_k = 1 + \left( \sum_U \mathbf{x}_k - \sum_r a_k \mathbf{x}_k \right)' \left( \sum_r a_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k$$

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

Denna kalibreringsvikt uppfyller kalibreringsvillkoret:  $\sum_r w_k \mathbf{x}_k = \sum_U \mathbf{x}_k$ , vilket innebär att om vikterna läggs på variabler som ingår i hjälpvektorn summeras dessa upp till de hjälptotaler som använts.

## Användning av vikter vid estimation

För varje svarande elev har en vikt beräknats. Syftet med detta är att kunna redovisa resultat för hela populationen och inte bara för de svarande. Vikten kallas därför även för uppräkningsstal.

Vikterna kan beskrivas med formeln enligt ovan:

$$w_k = g_k a_k$$

där  $w_k$  = vikt/uppräkningsstal för elev  $k$

$g_k$  = kalibreringsvikt för elev  $k$  baserad på hjälpinformation

$a_k$  = designvikt för elev  $k$  justerad för bortfall

Designvikten är den del av vikten som beror på urvalsdesignen. Vid bortfall kan det vara så att vissa grupper av urvalet svarar i större utsträckning än övriga, t.ex. kan kvinnor svara i högre grad än män. Om de grupper som svarat i högre grad har en annan fördelning med avseende på undersökningsvariablerna än övriga kan detta ha en snedvridande effekt på resultatet. För att kompensera för detta har kalibreringsvikter beräknats. Vikterna bygger på antagandet att urvalsramen återspeglar populationen väl och därmed att över- och undertäckningen är liten.

Vikterna multipliceras med objektens variabelvärden för att skapa statistikvärden gällande för populationen. Om vikterna inte används så kan resultaten bli missvisande. Vikterna kompenserar för objektsbortfallet men inte för det partiella bortfallet.

För beräkning av skattningar av totaler används följande formel:

$$\hat{Y} = \sum_r w_k y_k$$

där  $w_k$  = vikt/uppräkningsstal för elev  $k$

$y_k$  = variabelvärde för elev  $k$

och summering sker över de svarande elever  $r$ .

För beräkning av skattningar av medelvärden används följande formel:

$$\hat{Y} = \frac{\sum_r w_k y_k}{\sum_r w_k}$$

där  $w_k$  = vikt/uppräkningsstal för elev  $k$

$y_k$  = variabelvärde för elev  $k$

och summering sker av de svarande elever  $r$ .

Observera att vikterna är avsedda att användas för statistisk inferens, det vill säga vid skattning av målstorheter i populationen, exempelvis totaler och medelvärden. Vid sambandsanalyser (exempelvis regressionsanalys) bör det uppmärksammas att det inte är givet hur vikterna ska hanteras. Uvalets design samt bortfall är komplikationer som kan störa analyser av samband. För råd vid användning av vikter med statistiska analyser se dokumentet *Tips för analyser med regressionsmodeller på statistikdata*.

## Referenser

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

Martin Ribe, SCB. *Tips för analyser med regressionsmodeller på statistikdata*