

Elever i årskurs 9 med högutbildade föräldrar presterar bättre

Elever i årskurs 9 med högutbildade föräldrar presterar bättre

Producent SCB, Statistiska centralbyrån
Befolkning och välfärd
701 89 Örebro
010-479 40 00

Förfrågningar Mats Hansson
010-479 65 26
mats.hansson@scb.se

Rasmus Evaldsson
010-479 57 37
rasmus.evaldsson@scb.se

Det är tillåtet att kopiera och på annat sätt mångfaldiga innehållet.
Om du citerar, var god uppge källan på följande sätt:
Källa: SCB, Elever i årskurs 9 med högutbildade föräldrar presterar
bättre

URN:NBN:SE:SCB-2020-UFSAMBR2101_pdf
Denna publikation finns enbart i elektronisk form på www.scb.se
This publication is only available in electronic form on www.scb.se

Förord

För att sprida den officiella statistiken om befolkningens utbildningar tar SCB fram korta analyser där statistik presenteras på ett lättamt sätt för en bred allmänhet. Denna rapport är bakgrundsrapporten till en sådan kort analys.

Rapporten innehåller deskriptiv statistik och en fördjupad analys över hur meritvärdet hos elever i årskurs 9 påverkas av socioekonomiska bakgrundsvariabler. I huvudsak analyseras om föräldrars utbildningsnivå påverkar meritvärdet. För att enbart kunna fånga upp effekten av föräldrars utbildningsnivå har flertalet variabler vägts in i en regressionsanalys.

Rapporten är skriven av Rasmus Evaldsson och Mats Hansson vid enheten för statistik om utbildning och arbete. Johan Löfgren från metodenhet individ Örebro har bidragit med värdefulla synpunkter.

SCB i Örebro 2021-02-08

Magnus Sjöström

Katarina Wizell

Avdelningschef

Enhetschef

Sammanfattning

Rapporten består av två delar, där den första innehåller deskriptiv statistik och den andra en fördjupad analys där meritvärdet utifrån föräldrarnas utbildningsnivå och de övriga socioekonomiska bakgrundsvariabler analyseras.

Den deskriptiva statistiken visar följande mönster:

- Flickor har ett högre genomsnittligt meritvärde än pojkar.
- Elever med eftergymnasialt utbildade föräldrar har ett högre genomsnittligt meritvärde än elever med förgymnasialt utbildade föräldrar.
- Elever med svensk bakgrund har ett högre genomsnittligt meritvärde än de med utländsk bakgrund.
- Elever som lever i ett hushåll med mycket hög disponibel inkomst har högre genomsnittligt meritvärde än elever som lever i hushåll med lägre disponibla inkomstnivåer.
- Det genomsnittliga meritvärdet hos elever som bor med båda sina föräldrar är högre än hos de som endast bor med en förälder.

I den andra delen av rapporten finns den fördjupade analysen där en linjär regressionsmodell används för att förklara meritvärdet. Centrala resultat är att

- flickor presterar generellt bättre än pojkar
- elever med högre utbildade föräldrar presterar i genomsnitt bättre än elever med lågutbildade föräldrar.

Båda delar överensstämmer med vad den deskriptiva statistiken visar.

Modellen visar att utländsk bakgrund inte ser ut att spela någon roll för meritvärdet, medan den deskriptiva statistiken visar på ett högre genomsnittligt meritvärde för eleverna med svensk bakgrund. Anledningen kan vara att elever med utländsk bakgrund har andra egenskaper än elever med svensk bakgrund, med avseende på de variabler som är med i modellen. Elever med utländsk bakgrund har exempelvis till större del lågutbildade föräldrar, föräldrar med låga inkomster och en annan boendesituation, jämfört med elever med svensk bakgrund. När man kontrollerar för detta i modellen, minskar skillnaderna i meritvärde mellan elever med svensk bakgrund och elever med utländsk bakgrund.

Regressionsmodellen används för att separera effekten för respektive variabel, för att på så vis kunna avgöra vilka faktorer som påverkar meritvärdet. Vi har använt en linjär flernivåmodell som tar hänsyn till skillnader i hur det genomsnittliga meritvärdet varierar mellan

skolenheterna. Meritvärdet predikteras sedan utifrån de socioekonomiska bakgrundsvariablerna och föräldrarnas utbildningsnivå.

Inledning

Syftet med analysen är att belysa meritvärdet hos elever i årskurs 9 utifrån föräldrarnas utbildningsnivå tillsammans med andra socioekonomiska variabler. Som en del i analysen har en modell tagits fram som används för att kunna betrakta intressanta variabler och förklara om föräldrarnas utbildningsnivå och andra socioekonomiska variabler spelar roll för det meritvärde eleverna får.

Rapporten ska fungera som ett fördjupande komplement med information som använts vid framställandet av den korta analysen ”Elever i årskurs 9 med högutbildade föräldrar presterar bättre”.

Inledningsvis följer information om den population som använts för att göra analysen, följt av fördelningar av variablerna utifrån kön och genomsnittligt meritvärde. Därefter beskrivs den modell som använts i rapporten och en tabell som inkluderar det slutgiltiga materialet som sedan tolkas under avsnittet resultat.

Lästips

I vilken mån skolan lyckas utjämna skillnader mellan barns förutsättningar är ett ämne som diskuteras återkommande. Flera studier har gjorts de senaste åren som visar hur skillnader i elevers resultat samvarierar med deras bakgrund. Följande aktuella studier analyserar detta ingående:

- Skolverket (Rapport 467 2018) ”Analyser av familjebakgrundens betydelse för skolresultaten och skillnader mellan skolor - En kvantitativ studie av utvecklingen över tid i slutet av grundskolan”¹. Analysen visar bland annat att elevernas socioekonomiska bakgrund har fått en större betydelse för elevernas betygsresultat över tid. Analysen visar även att föräldrarnas utbildningsnivå har den största betydelsen för betygsresultaten.
- SNS rapport ”Studiegapet mellan inrikes och utrikes födda elever”² av Hans Grönqvist och Susan Niknami. De finner att det finns stora skillnader i hur väl elever födda utomlands och elever födda i Sverige klarar sig i skolan och de konstaterar att föräldrarnas socioekonomiska bakgrund är den främsta förklaringen till studiegapet.

¹ <https://www.skolverket.se/download/18.6bfaca41169863e6a65d200/1553967875648/pdf3927.pdf>

² <https://snsse.cdn.triggerfish.cloud/uploads/2020/03/studiegapet-mellan-inrikes-och-utrikes-fodda-elever.pdf>

Beskrivning av populationen & variabler

Population

Populationen som använts för rapporten är hämtad från registret över slutbetygen årskurs 9 för elever som har fått ett betyg enligt det svenska betygsystemet läsåren 2015/2016, 2016/2017 och 2017/2018. Efter bearbetning av materialet med avseende på matchning mot andra register och hantering av elever som är nyinvandrade omfattar populationen 300 200 individer.

Cirka 10 900 elever med tillfälligt personnummer har exkluderats eftersom inga uppgifter går att hämta om deras föräldrars socioekonomiska bakgrund. Dessa elever är dessutom nyinvandrade och har inte hunnit börja skolan i tid för att ha rimliga möjligheter att arbeta upp ett komplett betyg till årskurs 9. Även elever som har hunnit få ett personnummer, men som folkbokförts senast två år innan de gick ut årskurs 9, har tagits bort från materialet (8 900 elever).

Variabler

Valet av variabler i analysen är en subjektiv bedömning av vad som kan påverka elevernas möjligheter att uppnå höga meritvärden och ett resultat av tester på materialet och den effekt olika ingående variabler visar på meritvärdet. Tanken är att de ingående variablerna ska fånga upp så mycket information som möjligt om den socioekonomiska bakgrunden hos eleverna samtidigt som vi har uteslutit variabler som visat på mindre effekt.

I framställning av den slutgiltiga modellen har vi även undersökt andra variabler som inte tas upp här eftersom det visade sig att de inte tillför analysen något. Exempel på dessa variabler är HDI (Human development index), Grund för bosättning, Antal elever per skolenhet och Andel utbildade lärare per skolenhet.

Nedanstående variabler har tagits med i den slutgiltiga modellen eftersom vi anser att de kan förklara meritvärdet och vara till hjälp för analysen. I tabellerna nedan visas deskriptiv statistik i form av fördelningen av antal mellan olika grupper samt det genomsnittliga meritvärdet för dessa grupper.

Meritvärde

Meritvärdet innefattar betyg i moderna språk och som språkval tillsammans med de 16 bästa betygen i övrigt. Betygen översätts till poäng enligt följande sätt:

- F, 3 och 9 = 0
- E=10
- D=12,5
- C=15
- B=17,5
- A=20.

Notera att:

3 = Anges när ett ämne inte lästs på grund av anpassad studiegång eller anpassad timplan enligt 3 kap. 12 och 12 h §§ skollagen (2010:800)

9 = Streck (-) i betygs katalogen, dvs. bedömningsunderlag saknas pga. frånvaro.

Möjliga värden är 0 - 340 poäng. Meritvärde med en decimal är möjligt.

Kön

Uppgift om elevens kön har hämtats från slutbetygen årskurs 9.

Tabell 1. Antal elever och genomsnittligt meritvärde efter kön

| Kön | Antal | Genomsnittligt meritvärde |
|--------|---------|---------------------------|
| Flicka | 146 037 | 244 |
| Pojke | 154 212 | 219 |

Föräldrarnas utbildningsnivå³

Uppgift om föräldrarnas utbildningsnivå (vid tidpunkten då eleven avslutade grundskolan) har hämtats från Utbildningsregistret.

Tabell 2. Antal elever och genomsnittligt meritvärde efter kön och föräldrars utbildningsnivå

| Utbildningsnivå | Totalt antal | Genomsnittligt meritvärde | Flickor | Genomsnittligt meritvärde | Pojkar | Genomsnittligt meritvärde |
|---------------------------------|--------------|---------------------------|---------|---------------------------|--------|---------------------------|
| Uppgift saknas | 1 294 | 158 | 654 | 166 | 640 | 151 |
| Förgymnasial | 18 880 | 170 | 9 195 | 181 | 9 685 | 158 |
| Förgymnasial och gymnasial | 27 120 | 196 | 13 167 | 209 | 13 953 | 184 |
| Gymnasial | 85 812 | 214 | 41 681 | 228 | 44 131 | 201 |
| Förgymnasial och eftergymnasial | 9 992 | 221 | 4 750 | 234 | 5 242 | 209 |

³ I rapporten avses genomgående elevens vårdnadshavare när föräldrarna nämns. Detta för bättre läsbarhet..

| Utbildningsnivå | Totalt antal | Genomsnittligt meritvärde | Flickor | Genomsnittligt meritvärde | Pojkar | Genomsnittligt meritvärde |
|------------------------------|--------------|---------------------------|---------|---------------------------|--------|---------------------------|
| Gymnasial och eftergymnasial | 72 861 | 243 | 35 341 | 257 | 37 520 | 230 |
| Eftergymnasial | 84 290 | 265 | 41 249 | 277 | 43 041 | 253 |

Utländsk bakgrund och nyinvandrad

Nyinvandrad avser senaste invandringsår om eleven invandrat flera gånger. Denna variabel hämtas från Befolkningsregistret. Utländsk bakgrund hämtas också ifrån Befolkningsregistret och anger elevens bakgrund. Dessa två variabler har kombinerats till en variabel Nationell bakgrund.

Tabell 3. Antal elever och genomsnittligt meritvärde efter kön och nationell bakgrund

| Nationell bakgrund | Totalt antal | Genomsnittligt meritvärde | Flickor | Genomsnittligt meritvärde | Pojkar | Genomsnittligt meritvärde |
|--|--------------|---------------------------|---------|---------------------------|---------|---------------------------|
| Född utomlands, invandrat inom 2 - 4 år | 6 973 | 154 | 3 250 | 163 | 3 723 | 146 |
| Född utomlands, invandrat för mer än 4 år sedan | 28 263 | 213 | 14 047 | 226 | 14 216 | 199 |
| Född i Sverige, båda föräldrarna födda utomlands | 29 491 | 226 | 14 540 | 238 | 14 951 | 214 |
| Svensk bakgrund | 235 522 | 236 | 114 200 | 249 | 121 322 | 224 |

Hushållets försörjningsstöd

Denna variabel identifierar om någon av föräldrarna hade försörjningsstöd under året före eleven gick ut årskurs 9. För elever som gick ut årskurs 9 år 2018 är det således försörjningsstöd under 2017 som räknas. Uppgiften har hämtats från Inkomst- och Taxeringsregistret.

Tabell 4. Antal elever och genomsnittligt meritvärde efter kön och hushållets försörjningsstöd

| Försörjningsstöd | Totalt antal | Genomsnittligt meritvärde | Flickor | Genomsnittligt meritvärde | Pojkar | Genomsnittligt meritvärde |
|--------------------------------|--------------|---------------------------|---------|---------------------------|---------|---------------------------|
| Ja, har försörjningsstöd | 19 826 | 169 | 9 487 | 179 | 10 339 | 159 |
| Nej, har inte försörjningsstöd | 280 423 | 235 | 136 550 | 248 | 143 873 | 223 |

Bostadsområdets socioekonomiska status

Det bostadsområde (DeSO-område⁴) som eleven var folkbokförd i, det år eleven gick ut årskurs 9, har klassificerats socioekonomiskt.

Vid klassningen används följande variabler:

- Andelen av befolkningen i bostadsområdet med utländsk bakgrund
- Befolkningens utbildningsnivå (genomsnittlig utbildningsnivå per person, 25-64 år)
- Befolkningens inkomst (median av disponibel inkomst enligt tidigare, 25-64 år).

Vilka variabler som ska användas för att klassificera bostadsområdet socioekonomiskt är delvis subjektivt, men de variabler som valts bör ge en god bild av hur bostadsområdet ser ut socioekonomiskt.

En principalkomponentanalys har sedan tillämpats för att väga samman variablerna till ett socioekonomiskt mått. Principalkomponentanalysen resulterar i en linjärkombination av de ingående variablerna.

Bostadsområdena indelas sedan i fyra lika stora grupper efter vilka egenskaper de har på ovanstående variabler.

Tabell 5. Antal elever och genomsnittligt meritvärde efter kön och bostadsområdets socioekonomiska status

| DeSo-område | Totalt antal | Genomsnittligt meritvärde | Flickor | Genomsnittligt meritvärde | Pojkar | Genomsnittligt meritvärde |
|--|--------------|---------------------------|---------|---------------------------|--------|---------------------------|
| Den svagaste fjärdedelen av områdena | 65 025 | 205 | 31 683 | 218 | 33 342 | 193 |
| Den näst svagaste fjärdedelen av områdena | 70 424 | 221 | 34 279 | 235 | 36 145 | 208 |
| Den näst starkaste fjärdedelen av områdena | 75 623 | 233 | 36 772 | 246 | 38 851 | 220 |
| Den starkaste fjärdedelen av områdena | 89 177 | 256 | 43 303 | 268 | 45 874 | 244 |

Disponibel inkomst per konsumtionsenhet

Uppgifter om föräldrarnas disponibla inkomst hämtas från Inkomst- och Taxeringsregistret. Den disponibla inkomsten består av summan av faktorinkomster (löneinkomst, inkomst av näringsverksamhet och

⁴ <https://www.scb.se/hitta-statistik/regional-statistik-och-kartor/regionala-indelningar/deso---demografiska-statistikomraden/>

kapitalinkomst) samt skattepliktiga och skattefria transfereringar, minus skatt och övriga negativa transfereringar. Uppgiften avser året före eleven gick ut årskurs 9.

Eleven klassificeras utifrån den sammanlagda disponibla inkomsten i det hushåll där eleven och minst en av föräldrarna är folkbokförda. Om ingen av föräldrarna är folkbokförda i det hushåll där eleven är folkbokförd används inkomsten i det hushåll där inkomsten är störst.

Tabell 6. Antal elever och genomsnittligt meritvärde efter kön och disponibel inkomst per konsumtionsenhet

| Disponibel inkomst per konsumtionsenhet | Totalt antal | Genomsnittligt meritvärde | Flickor | Genomsnittligt meritvärde | Pojkar | Genomsnittligt meritvärde |
|---|--------------|---------------------------|---------|---------------------------|--------|---------------------------|
| Mycket låg inkomst (10% lägsta) | 30 016 | 187 | 14 623 | 198 | 15 393 | 176 |
| Låg inkomst (10% näst lägsta) | 30 022 | 203 | 14 482 | 215 | 15 540 | 192 |
| Medelinkomst (60% mellersta) | 180 162 | 232 | 87 565 | 245 | 92 597 | 219 |
| Hög inkomst (10% näst högsta) | 30 038 | 260 | 14 648 | 273 | 15 390 | 247 |
| Mycket hög inkomst (10% högsta) | 30 011 | 269 | 14 719 | 281 | 15 292 | 258 |

Föräldrarnas yrkesnivå

Uppgiften hämtas från RAKS (Registerbaserad aktivitetsstatistik).

Variabeln anger den högsta yrkesnivån hos elevens föräldrar, dvs. om en förälder är chef och den andra föräldern jobbar i ett yrke som kräver elementär utbildning på grundskolenivå så väljs yrkesnivån chef.

Tabell 7. Antal elever och genomsnittligt meritvärde efter kön och föräldrarnas yrkesnivå

| Föräldrarnas högsta yrkesnivå | Totalt antal | Genomsnittligt meritvärde | Flickor | Genomsnittligt meritvärde | Pojkar | Genomsnittligt meritvärde |
|--|--------------|---------------------------|---------|---------------------------|--------|---------------------------|
| Uppgift saknas | 39 185 | 193 | 18 917 | 205 | 20 268 | 182 |
| Elementär utbildning på grundskolenivå. Översatt till svenska förhållanden innebär det inga eller låga utbildningskrav | 7 462 | 192 | 3 655 | 205 | 3 807 | 180 |

| | Totalt antal | Genomsnittligt meritvärde | Flickor | Genomsnittligt meritvärde | Pojkar | Genomsnittligt meritvärde |
|--|--------------|---------------------------|---------|---------------------------|--------|---------------------------|
| Föräldrarnas högsta yrkesnivå | | | | | | |
| Utbildningar på gymnasial nivå samt eftergymnasiala utbildningar kortare än 2 år | 89 634 | 210 | 43 704 | 224 | 45 930 | 197 |
| Praktiska eller yrkesspecifika eftergymnasiala utbildningar om 2 till 3 år | 34 289 | 237 | 16 610 | 251 | 17 679 | 225 |
| Teoretiska eller forskarförberedande efter gymnasiala utbildningar samt forskarutbildningar om minst 3 år, normalt 4 år eller längre | 84 617 | 256 | 41 249 | 269 | 43 368 | 244 |
| Chefer | 45 062 | 259 | 21 902 | 272 | 23 160 | 246 |

Föräldrarnas sysselsättningsstatus

Uppgiften hämtas från RAKS (Registerbaserad aktivitetsstatistik).

Vi har valt att gruppera värdemängderna i tre olika kombinationer för att minska antalet möjliga utfall. Ingående värdemängder för förälder 1 och 2 är:

- 1 = Ej förvärvsarbete men har högsta inkomst från arbete
- 2 = Studerande
- 3 = Vård av barn / anhörig
- 4 = Sjuk
- 5 = Arbetslös
- 6 = Förtidspensionär
- 7 = Ekonomiskt bistånd
- 8 = Arbetsmarknadspolitisk åtgärd
- 9 = Ålderspensionär
- 77 = Anställd
- 88 = Företagare
- 99 = Saknar inkomst

Vi valde sedan att kombinera variabeln för föräldrarna 1 och 2 till en variabel. Variabeln som används i regressionerna har följande fördelning:

Tabell 8. Antal elever och genomsnittligt meritvärde efter kön och föräldrarnas sysselsättningsstatus

| Föräldrarnas sysselsättningsstatus | Totalt antal | Genomsnittligt meritvärde | Flickor | Genomsnittligt meritvärde | Pojkar | Genomsnittligt meritvärde |
|---|--------------|---------------------------|---------|---------------------------|---------|---------------------------|
| Uppgift saknas | 1 376 | 162 | 691 | 168 | 685 | 155 |
| De föräldrar elev har, har sysstat (3,4,5,6,7,8,99) | 14 908 | 168 | 7 188 | 179 | 7 720 | 158 |
| En förälder har sysstat (1,2,9) en har (3,4,5,6,7,8,99) | 2 226 | 189 | 1 079 | 200 | 1 147 | 179 |
| De föräldrar elev har, har sysstat (1,2,9) | 2 104 | 193 | 1 019 | 205 | 1 085 | 182 |
| En förälder har sysstat (3,4,5,6,7,8,99) en har (77,88) | 27 115 | 207 | 13 206 | 219 | 13 909 | 196 |
| En förälder har sysstat (1,2,9) en har (77,88) | 9 050 | 227 | 4 473 | 239 | 4 577 | 215 |
| De föräldrar elev har, har sysstat (77,88) | 243 470 | 239 | 118 381 | 252 | 125 089 | 226 |

Bor med föräldrar

Uppgiften hämtas från Befolkningsregistret och variabeln anger om eleven är folkbokförd hos sin/sina föräldrar.

Tabell 9. Antal elever och genomsnittligt meritvärde efter kön och boendesituation

| Boendesituation | Totalt antal | Genomsnittligt meritvärde | Flickor | Genomsnittligt meritvärde | Pojkar | Genomsnittligt meritvärde |
|--|--------------|---------------------------|---------|---------------------------|--------|---------------------------|
| Bor utan föräldrar | 4 332 | 165 | 2 216 | 172 | 2 116 | 158 |
| Bor med en förälder och har endast en förälder | 27 914 | 196 | 13 722 | 209 | 14 192 | 184 |
| Bor med en förälder | 83 497 | 218 | 40 712 | 231 | 42 785 | 205 |
| Bor med två föräldrar | 184 506 | 244 | 89 387 | 257 | 95 119 | 231 |

Den modellbaserade analysen

En modell tillämpas på elevnivå och huvudsyftet med modellen är att skatta vilken effekt de förklarande variablerna har på det meritvärde en elev förväntas uppnå under årskurs 9. Metoden som används är en linjär regressionsanalys. Modellens beroende variabel utgörs av meritvärdet, det vill säga det som modelleras är meritvärdet för respektive elev i modellen.

Regressionsanalys

Regressionsanalys är en statistisk metod för att analysera samband mellan en beroende variabel och olika bakgrundsvariabler.

En sådan analys ger underlag för att bedöma vilka variabler eller faktorer som är viktiga för att förklara den beroende variabeln och hur betydelsefulla faktorerna är i sig och i förhållande till varandra.

En linjär regression kan beskrivas som:

$$Y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad (1)$$

där Y_i är den beroende variabeln och $x_{1i}, x_{2i}, \dots, x_{ki}$ utgör k stycken förklarande variabler för elev i . $\beta_1, \beta_2, \dots, \beta_k$ utgör regressionskoefficienterna, α utgör interceptet och ε_i utgör feltermen.

Koefficienterna anger betydelsen av de olika bakgrunds faktorerna. Dessa bakgrunds faktorer har kodats så att olika grupper lätt kan jämföras med varandra. En grupp är referensalternativet för varje variabel eller faktor. För kön, till exempel, är pojkar referensgruppen med en koefficient som är 0. Koefficienten för flickor anger hur flickor skiljer sig från pojkar i genomsnitt, när vi kontrollerar för de övriga variablerna som ingår i regressionsmodellen. En positiv koefficient för flickor innebär således att flickor förväntas få ett högre meritvärde än pojkar.

En koefficient som är större än 0 anger att eleven beräknas få detta värde som ett ökat värde i sitt meritvärde i genomsnitt, ett värde som är mindre än 0 ger motsats effekt.

Modeller med två nivåer

En regressionsmodell på elevnivå enligt ovan förutsätter oberoende mellan eleverna. Eftersom eleverna som ingår i studien är klustrade i olika grupper (skolor) bör man ta hänsyn till detta vid analysen. Detta kan man göra med hjälp av en flernivåmodell.

Några viktiga skäl till att använda flernivåmodeller är

- Om antagandet om oberoende mellan elever inte är uppfyllt kan en vanlig regressionsmodell underskatta standardfelen för regressionskoefficienter. Detta kan i sin tur leda till felaktiga slutsatser om vilka variabler som signifikant påverkar den beroende variabeln.
- Det finns risk att effekter på elevnivå (exempelvis effekter av föräldrarnas utbildning och elevens kön) störs av klustereffekterna. Det är svårt att skilja på vad som beror på eleven och vad som beror på elevens omgivning (skolan). I en flernivåmodell kan dessa effekter skattas separat.
- Om det är intressant av att skatta storleken på klustereffekterna, exempelvis skillnader mellan olika skolor, bör man använda sig av en flernivåmodell.

En flernivåmodell används i denna studie för att vi ska kunna uttala oss om vilka variabler som har starkast effekt på meritvärdet. Det är effekterna på elevnivå som är intressanta i denna studie, men vi vill kontrollera för klustereffekterna.

Den variant av flernivåmodell som används i denna tillämpning är en modell där interceptet tillåts variera mellan skolor. På detta sätt tar modellen hänsyn till att eleverna är klustrade i skolor.

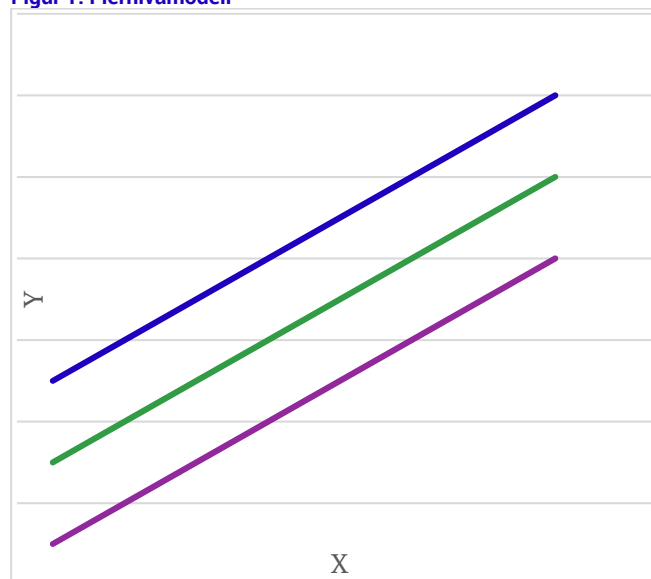
Denna variant av modell brukar kallas för *random intercept model* (Goldstein, 2011). Man kan se det som att en separat regressionsmodell tillämpas på varje skola.

Modellen kan då beskrivas enligt:

$$Y_{ij} = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_j + \varepsilon_i \quad (2)$$

Skillnaden mellan modell (1) och modell (2) är att varje skola j , tillåts ha ett eget intercept ($\alpha + u_j$). Varje skola har därmed en egen regressionslinje. Regressionskoefficienterna $\beta_1, \beta_2, \dots, \beta_k$ antas vara samma i alla skolor även i modell (2). Detta innebär att lutningen på regressionslinjen antas vara samma för samtliga skolor.

Figur 1. Flernivåmodell



Varje linje representerar en skolenhet som presterar på olika nivåer, där Y-axlen är meritvärdet och X-axlen är de förklarande variablerna.

Figuren ovan visar på hur en flernivåmodell fungerar. De tre linjerna ovan representerar regressioner för unika skolenheter j . Modellen tar i beaktning att eleverna är klustrade i skolenheter och att prestationerna skiljer sig åt beroende på skolenhet, vilket medför att intercepten är olika mellan de 3 skolenheterna. Sedan är det regressionskoefficienterna som avgör vilken lutning linjerna ska få, vilka är samma för alla skolenheter och därför följs linjerna åt.

Fixa och slumpmässiga effekter

Modellen har två delar med olika typer av effekter, man brukar benämna dem fixa och slumpmässiga effekter. De fixa effekterna utgörs av $\alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$ och de slumpmässiga av $u_j + \varepsilon_i$. De fixa effekterna antas vara samma för alla elever medan de slumpmässiga effekterna tillåts variera mellan elever och skolor. Därav namnet "Random Intercept Model".

I denna tillämpning är vi intresserade av att skatta de fixa effekterna på elevnivå. De slumpmässiga effekterna är inte intressanta i sig, men vi använder dem för att kontrollera för att eleverna är klustrade i skolor.

De slumpmässiga effekterna fångar upp två olika typer av variation. Feltermen ε_i , fångar upp variationen i meritvärden mellan elever, efter att vi har kontrollerat för de förklarande variablerna. Det skolspecifika interceptet u_j , fångar upp skillnader i meritvärden mellan skolor, efter att vi har kontrollerat för de förklarande variablerna.

I de modeller som vi testat i detta arbete är skillnader i meritvärden mellan skolor relativt stor, efter att man kontrollerat för de förklarande variablerna. De skolspecifika intercepten har därmed ganska stor variation och utgör ungefär 15-20 procent av den totala

residualvariansen (den oförklarade variationen i meritvärden). Detta talar för att det är lämpligt att använda sig av en flernivåmodell.

Modellens förklaringsgrad

För att utvärdera en regressionsmodell är modellens förklaringsgrad (r^2) ofta central. Med förklaringsgrad avses hur stor del av variationen i den beroende variabeln y , som förklaras av modellen (x -variablerna).

När ett intercept är inkluderat i modellen, kan förklaringsgraden för en linjär regression, beräknas genom att kvadrera korrelationskoefficienten r , mellan predikterade och observerade y -värden.

$$r^2 = (r_{y_i, \hat{y}_i})^2$$

Vi använder denna metod för att beräkna förklaringsgraden för de flernivåmodeller som används i denna tillämpning. I de predikterade värdena inkluderas de skolspecifika intercepten, för att fånga upp variationen mellan skolor. Resultatet presenteras i tabell 10.

Fem olika regressionsmodeller

Vi har valt att skatta fem olika modeller, där de förklarande variablerna varierar för varje modell.

- Modell 1 innehåller endast variabeln *Föräldrarnas utbildningsnivå*, vilken inkluderas i samtliga modeller för att förklara meritvärdet.
- I modell 2 så läggs variablerna *Kön* och *Nationell bakgrund* till.
- För modell 3 läggs det till *Försörjningsstöd* och *Boendesituation*.
- Modell 4 inkluderar också *Disponibel inkomst per konsumtionsenhet*, *Föräldrarnas högsta yrkesnivå* och *Föräldrars sysselsättningsstatus*.
- Modell 5 är sedan den slutgiltiga modellen där vi inkluderat samtliga förklarande variabler, här läggs endast *Boendeområde* till.

Anledningen till att flera regressionsmodeller används är för att kunna se och förklara skillnader som uppstår då fler förklarande variabler läggs till. Vi kan exempelvis se en större effekt av föräldrarnas utbildningsnivå i modell 1 än i modell 5, det beror på att flera andra variabler fångat upp den effekt som tidigare såg ut att tillföras föräldrars utbildningsnivå.

I tabell 11 presenteras regressionskoefficienter för modellerna 1 till 5 där den slutgiltiga modellen inkluderar följande variabler:

- Föräldrars utbildningsnivå
- Kön
- Utländsk bakgrund & nyinvandrad
- Försörjningsstöd
- Boendesituation
- Disponibel inkomst per konsumtionsenhet
- Föräldrarnas högsta yrkesnivå

- Föräldrars sysselsättningsstatus
- Boendeområde.

Resultat

Nedan presenteras först resultatet från Cramer's V som testar för om det förekommer multikollineariteten i modellerna. Därefter presenteras resultatet för respektive modell, där vi redogör för hur föräldrarnas utbildningsnivå påverkar meritvärdet när vi inkluderar fler förklarande variabler i modellen. Dessutom så redovisas skillnader i övriga förklarande variabler för de olika modellerna. I slutet av resultatdelen klargör vi varför modell 5 är den mest lämpliga modellen för att kunna förklara elevernas meritvärde.

Cramer's V

Om de förklarande variablerna i modellen samvarierar starkt, kan det leda till problem med multikollinearitet vilket i sin tur leda till problem med skattningarna av regressionsparametrar. Därför har samvariationen mellan de förklarande variablerna i modellen kontrollerats.

Det mått som använts är Cramer's V, vilket är ett mått på graden av samvariation mellan kategoriska variabler. Cramers' V kan anta värden i intervallet 0 till 1, där 0 indikerar inget samband och 1 innebär ett perfekt samband.

Utifrån kontrollen har det bedömts att de samband som finns, inte bör innebära några stora problem för tolkningen av modellen. Inget samband är högre än 0.5.

Förklaringsgrad modell 1-5

Tabell 10. Förklaringsgrad modell 1-5

| Modell | Förklaringsgrad (%) |
|----------|---------------------|
| Modell 1 | 23,5 |
| Modell 2 | 27,8 |
| Modell 3 | 31,2 |
| Modell 4 | 32,8 |
| Modell 5 | 33,0 |

Modell 1

I modell 1 inkluderas endast variabeln *föräldrars utbildningsnivå*.

Modellen visar att utbildningsnivån hos föräldrarna delvis kan förklara elevens prestation i skolan, där modellen fångar upp 23,5 procent av variationen. Resultatet indikerar även på ett samband där elever med högutbildade föräldrar i genomsnitt får ett högre meritvärde än elever med lågutbildade föräldrar. Modellen visar också att det spelar roll vilken utbildningsnivå båda föräldrarna har. Det visar sig genom att sambandet är högre om elevens båda föräldrar är eftergymnasialt utbildade jämfört med elever med en eftergymnasialt utbildad förälder och en som är lägre utbildad.

Modell 2

I modell 2 läggs även variablerna *Kön* och *Nationell bakgrund* till i regressionen.

Modell 2 fångar upp 27,8 procent av variationen, vilket är högre än modell 1.

När man tar hänsyn till kön och nationell bakgrund så minskar det positiva sambandet mellan föräldrarnas utbildningsnivå och meritvärdet. Det beror bland annat på att nationell bakgrund har ett samband både med den beroende variabeln (meritvärde) och den förklarande variabeln (föräldrars utbildningsnivå). Bland elever med lågutbildade föräldrar så är det en högre andel elever som är födda utomlands och en högre andel elever som är nyinvandrade, jämfört med elever som har högutbildade föräldrar. När hänsyn tas till detta minskar därmed sambandet mellan meritvärde och föräldrars utbildningsnivå.

En vanlig benämning för detta inom statistisk analys är "confounding", man kan säga att effekten av föräldrarnas utbildning är "confounded" av nationell bakgrund. När man tar hänsyn till en "confounder" så förändras effekten av den förklarande variabeln. Det kan finnas fler "confounders" än nationell bakgrund, vilket vi försöker fånga upp i modell 3 till modell 5.

Värden för bland annat eftergymnasialt utbildade föräldrar jämfört med förgymnasialt utbildade föräldrar har sjunkit från 83,0 till 74,9. Med andra ord indikerar resultatet fortfarande att föräldrarnas utbildningsnivå är viktig för meritvärdet men mindre viktigt än i modell 1. Samtidigt visar modell 2 på att ett samband finns mellan kön och meritvärde där flickor i genomsnitt har ett högre meritvärde än pojkar.

Modell 3

I modell 3 läggs variablerna *Försörjningsstöd* och *Boendesituation* till i regressionen.

Förklaringsgraden har nu ökat till 31,2 procent, vilket betyder att de nya variablerna kan förklara en del av variationen som de tidigare variablerna inte fångat upp.

När tar vi tar hänsyn till försörjningsstöd och boendesituation så avtar sambandet mellan föräldrarnas utbildning och meritvärdet ytterligare. Resultatet visar fortfarande att utbildningsnivån hos föräldrarna är viktig för meritvärdet men att koefficienterna nu är något lägre. Koefficienten för eftergymnasialt utbildade föräldrar är fortfarande hög (60,4).

Koefficienten "Född utomlands, invandrat för mer än 4 år sedan" (-3,6), närmar sig koefficienten för svensk bakgrund. Det tyder på att det i sig inte är negativt för meritvärdet att vara född utomlands. Att svenskfödda elever får bättre betyg i genomsnitt förklaras till stor del av variablerna föräldrarnas utbildningsnivå, försörjningsstöd och boendesituation.

För de nya variablerna finner vi ett positivt samband för ett högre meritvärde av att bo tillsammans med båda sina föräldrar jämfört med att bo utan föräldrar (50,6), samt ett positivt samband av att inte ha försörjningsstöd (27,6). Boendesituationen ser alltså ut att påverka meritvärdet i ganska stor utsträckning och vi kan se att det är positivt för meritvärdet att bo tillsammans med båda sina föräldrar.

Modell 4

För modell 4 läggs också *Disponibel inkomst per konsumtionsenhet*, *Föräldrarnas högsta yrkesnivå* och *Föräldrars sysselsättningsstatus* till i regressionen.

Modellen förklarar 32,8 procent av variationen, vilket är en mindre ökning än den mellan modell 2 och 3.

De utfall som visar sig ha högst samband med ett högre meritvärde för respektive ny variabel är "*Mycket hög inkomst, (10% högsta)*" (18,2), "*Chefer*" (12,1) och "*Elevens föräldrar har sysselsättningsstatus anställd eller företagare*" (14,3). Det vill säga, modell 4 indikerar att en hög inkomstnivå ger ett högre meritvärde än en låg inkomstnivå, att det finns ett positivt samband för ett högre meritvärde om föräldrar jobbar som chefer jämfört med andra yrkesnivåer samt att om föräldrarna har en sysselsättningsstatus som anställd eller företagare ger det ett högre meritvärde än övriga utfall.

När vi tar hänsyn till föräldrarnas inkomst, yrke och sysselsättningsstatus, så avtar det positiva sambandet för föräldrarnas utbildningsnivå ytterligare. För eftergymnasialt utbildade föräldrar sjunker koefficienten från 60,4 till 47,7. Men det ser fortfarande ut att vara den variabel som påverkar meritvärdet mest. Även boendesituationen har stor påverkan.: att bo tillsammans med båda sina föräldrar har en likvärdig koefficient (44,6).

Det negativa sambandet mellan att vara född utomlands och meritvärdet försvinner helt. Modellen visar en ökning hos samtliga

grupper jämfört med "Svensk bakgrund". Dels går "Född i Sverige, båda föräldrarna födda utomlands" från 0,8 till 7,8 och "Född utomlands, invandrat för mer än 4 år sedan" från -3,6 till 3,3. Det finns dock fortfarande ett negativt samband med meritvärdet av att vara nyligen invandrad, vilket kan antas delvis bero på tiden det tar att komma in i skolsystemet och lära sig det svenska språket.

Anledningen till dessa förändringar är att inkludandet av flera nya variabler fångar upp en del av effekten för de andra variablerna. Exempelvis så är det rimligt att anta att en högre utbildningsnivå medför en högre genomsnittlig inkomstnivå.

Modell 5

I modell 5 läggs *Boendeområde* till i regressionen. Koefficientskattningar för denna variabel indikerar att den har ett samband med elevernas meritvärde. Det är alltså positivt att bo i ett socioekonomiskt starkt område.

I övrigt så är skillnaden mot modell 4 liten. Variabeln boendeområde fångar därför bara upp en liten del av det samband som de övriga variablerna gjort tidigare.

Förklaringsgraden är nu 33,0 procent, en ökning med endast 0,2 procentenheter jämfört med modell 4. Det kan betyda att boendeområdet endast har en liten påverkan på att kunna förklara variationen i modellen.

Vi anser att modell 5 är den lämpligaste modellen att förklara elevernas meritvärde. Modell 5 har den högsta förklaringsgraden samtidigt som koefficienterna endast förändras marginellt i jämförelse med modell 4. Vår sista modell har även flest förklarande variabler som därför kan betraktas i analysen.

För modell 5 kan följande samband utläsas:

- En högre utbildningsnivå hos föräldrarna har ett positivt samband med meritvärdet.
- Ett samband finns mellan kön och meritvärde där flickor i genomsnitt har ett högre meritvärde än pojkar.
- Utländsk bakgrund ser inte ut att vara avgörande för meritvärdet och snarare tendera till att ha ett positivt samband med meritvärdet så länge invandringen gjordes för över 4 år sedan.
- Att ha försörjningsstöd ser ut att ha ett negativt samband med meritvärdet.
- Att bo tillsammans med båda sina föräldrar visar på ett positivt samband med meritvärdet.
- Högre disponibel inkomst har ett positivt samband med meritvärdet.
- Föräldrar som arbetar som chefer eller arbeten som kräver längre utbildningar har ett positivt samband med meritvärdet.

- Att bo i socioekonomiskt starka områden har ett positivt samband med meritvärdet.

Som resultatet tyder på spelar många socioekonomiska faktorer in och påverkar meritvärdet som elever får i årskurs 9. Resultatet visar på ett samband där elever med högre utbildade föräldrar presterar bättre i skolan. Det visar sig även finnas ett samband som visar på att det inte bara är den högst utbildade förälderns som påverkar, utan en kombination av båda föräldrarnas utbildningsnivå. Resultaten pekar också på att nationell bakgrund har ett ganska svagt samband meritvärdet (med undantag för gruppen nyinvandrade). Skillnader i meritvärde mellan olika grupperna av nationell bakgrund, förklaras till stor del av att de socioekonomiska förutsättningarna skiljer sig åt mellan grupperna. Denna bild förstärks sedan av jämförelser från tidigare studier (Rapport, Skolverket 467 2018), som visar på ett liknande resultat, där mer ingående och fördjupade diskussioner förts kring ämnet.

Tabell 11. Linjär regressionsanalys med flernivåmodell för skolenheter (regressionskoefficienter)

| | | Modell 1 | Modell 2 | Modell 3 | Modell 4 | Modell 5 |
|---|---|----------|----------|----------|----------|----------|
| Föräldrars utbildningsnivå | Eftergymnasial | 83,0 | 74,9 | 60,4 | 47,7 | 46,9 |
| | Gymnasial och eftergymnasial | 66,3 | 57,0 | 41,0 | 31,8 | 31,3 |
| | Gymnasial | 40,0 | 31,1 | 20,3 | 18,0 | 17,7 |
| | Förgymnasial och eftergymnasial | 45,8 | 38,9 | 24,9 | 19,0 | 18,8 |
| | Förgymnasial och gymnasial | 23,5 | 14,9 | 2,7 | 3,1 | 3,1 |
| | Förgymnasial | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 |
| Kön | Flickor | - | 24,6 | 24,7 | 24,7 | 24,7 |
| | Pojkar | - | 0,0 | 0,0 | 0,0 | 0,0 |
| Nationell bakgrund | Född i Sverige, båda föräldrarna födda utomlands | - | -0,1* | 0,8* | 7,8 | 9,9 |
| | Född utomlands, invandrat för mer än 4 år sedan | - | -10,6 | -3,6 | 3,3 | 4,9 |
| | Född utomlands, invandrat inom 2 - 4 år | - | -53,3 | -43,2 | -32,1 | -30,5 |
| | Svensk bakgrund | - | 0,0 | 0,0 | 0,0 | 0,0 |
| Försörjningsstöd | Nej, har inte försörjningsstöd | - | - | 27,6 | 19,2 | 18,4 |
| | Ja, har försörjningsstöd | - | - | 0,0 | 0,0 | 0,0 |
| Boendesituation | Bor med två föräldrar | - | - | 50,6 | 44,6 | 44,5 |
| | Bor med en förälder | - | - | 31,3 | 28,5 | 28,8 |
| | Bor med en förälder och har endast en förälder | - | - | 20,0 | 18,6 | 19,2 |
| | Bor utan föräldrar | - | - | 0,0 | 0,0 | 0,0 |
| Disponibel inkomst per konsumtionsenhet | Mycket hög inkomst (10% högsta) | - | - | - | 18,2 | 16,3 |
| | Hög inkomst (10% näst högsta) | - | - | - | 15,6 | 14,0 |
| | Medelinkomst (60% mellersta) | - | - | - | 7,8 | 7,0 |
| | Låg inkomst (10% näst lägsta) | - | - | - | 2,5 | 2,4 |
| | Mycket låg inkomst (10% lägsta) | - | - | - | 0,0 | 0,0 |
| Föräldrars högsta yrkesnivå ² | Chefer | - | - | - | 12,1 | 11,8 |
| | Teoretiska eller forskarförberedande eftergymnasiala utbildningar | - | - | - | 11,2 | 11,1 |
| | Praktiska eller yrkesspecifika eftergymnasiala utbildningar | - | - | - | 7,2 | 7,1 |
| | Utbildningar på gymnasial nivå | - | - | - | -4,4 | -4,0 |
| | Uppgift saknas | - | - | - | -9,5 | -8,8 |
| Föräldrars sysselsättningsstatus ³ | Elementär utbildning på grundskolenivå | - | - | - | 0,0 | 0,0 |
| | Elevens båda föräldrar har sysstat (77,88) | - | - | - | 14,3 | 13,9 |
| | Elevens båda föräldrar har sysstat (1,2,9) | - | - | - | 11,7 | 11,5 |
| | En förälder har sysstat (1,2,9) en har (77,88) | - | - | - | 11,7 | 11,4 |
| | En förälder har sysstat (3,4,5,6,7,8,99) en har (77,88) | - | - | - | 4,6 | 4,4 |
| | En förälder har sysstat (1,2,9) en har (3,4,5,6,7,8,99) | - | - | - | 2,3* | 2,5* |
| Boendeområde | Elevens båda föräldrar har sysstat (3,4,5,6,7,8,99) | - | - | - | 0,0 | 0,0 |
| | Den starkaste fjärdedelen av områdena | - | - | - | - | 11,8 |
| | Den näst starkaste fjärdedelen av områdena | - | - | - | - | 7,7 |
| | Den näst svagaste fjärdedelen av områdena | - | - | - | - | 4,1 |
| | Den svagaste fjärdedelen av områdena | - | - | - | - | 0,0 |

1) Där: * = 0,1 > p, de övriga koefficienterna har ett lägre p-värde än 0,0001 > p

2) Fullständig beskrivning av grupperna i föräldrars högsta yrkesnivå:

- Teoretiska eller forskarförberedande eftergymnasiala utbildningar samt forskarutbildningar om minst 3 år, normalt 4 år eller längre
- Praktiska eller yrkesspecifika eftergymnasiala utbildningar om 2-3 år
- Utbildningar på gymnasial nivå samt eftergymnasiala utbildningar kortare än 2 år
- Uppgift saknas
- Elementär utbildning på grundskolenivå. Översatt till svenska förhållanden innebär det inga eller låga utbildningskrav

3) Föräldrars sysselsättningsstatus:

- 1 = Ej förvärvsarbete men har högsta inkomst från arbete
- 2 = Studerande
- 3 = Vård av barn / anhörig
- 4 = Sjuk
- 5 = Arbetslös
- 6 = Förtidspensionär
- 7 = Ekonomiskt bistånd
- 8 = Arbetsmarknadspolitisk åtgärd
- 9 = Ålderspensionär
- 77 = Anställd

88 = Företagare
99 = Saknar inkomst

Referenser

Goldstein, H. (2011). *Multilevel statistical models*. John Wiley & Sons.

Skolverket, (Rapport 467 2018). *Analys av familjebakgrundens betydelse för skolresultaten och skillnader mellan skolor*.

<https://www.skolverket.se/download/18.6bfaca41169863e6a65d200/1553967875648/pdf3927.pdf>

Grönqvist, H & Niknami, S. (2020). *Studiegapet mellan inrikes och utrikes födda elever*. SNS.

<https://snsse.cdn.triggerfish.cloud/uploads/2020/03/studiegapet-mellan-inrikes-och-utrikes-fodda-elever.pdf>

SCB beskriver Sverige

Statistikmyndigheten SCB förser samhället med statistik för beslutsfattande, debatt och forskning. Vi gör det på uppdrag av regeringen, myndigheter, forskare och näringsliv. Vår statistik bidrar till en faktabaserad samhällsdebatt och väl underbyggda beslut.