

# Skolprestationer på kommunnivå med hänsyn tagen till socioekonomi

## 1 Inledning

Sveriges kommuner och landsting (SKL) presenterar varje år statistik över elevprestationer i årskurs 9. Statistiken redovisas på kommunnivå, exempelvis visas genomsnittliga meritvärden på kommunnivå.

För att bättre kunna jämföra betygsvärden på kommunnivå behöver man ta hänsyn till att kommuner är olika avseende elevernas socioekonomiska bakgrund. En kommun med hög andel elever med högutbildade föräldrar kan exempelvis inte utan vidare jämföras med en kommun med hög andel elever med lågutbildade föräldrar. För att möjliggöra jämförelser behöver man därför kontrollera för att kommuner ser olika ut med avseende på elevsammansättning.

Ett sätt att göra detta är att ta fram en modell för skolprestationer. En sådan modell beskriver sambandet i hela riket, mellan skolprestation och socioekonomiska egenskaper. Genom att tillämpa modellen får man fram ett förväntat värde för skolprestationer i kommunen. Detta värde beskriver vad eleverna i genomsnitt borde uppnå, givet den sammansättning av elever som finns i kommunen. Med hjälp av de modellerade värdena får man ett verktyg för att bedöma om eleverna i kommunen har presterat bättre eller sämre än vad som förväntas baserat på ett riksgenomsnitt.

Modellen tas fram på skolenhetsnivå då det i tidigare arbeten visat sig ge den bästa anpassningen till data över skolprestationer.

### 1.1 Dataunderlag

Underlag för beräkningarna utgörs av ett register som innehåller elever som gått ut årskurs nio våren 2016. Registret innehåller uppgifter om skolenhet, kommun, examensår, tre variabler som beskriver elevens resultat och ett antal variabler som beskriver den socioekonomiska bakgrunden.

De variabler som beskriver elevens resultat är

- meritvärde årskurs 9 (17 ämnen)
- om eleven uppnått behörighet till gymnasieskolan eller inte (yrkesprogram)
- om eleven fått godkänt i alla ämnen eller inte

De variabler som beskriver den socioekonomiska bakgrunden är

- Kön
- Föräldrarnas utbildningsnivå
- Nyligen invandrad (enligt Skolverkets definition)
- Ekonomiskt bistånd

Nytt för i år är att det i datamaterialet finns ett stort antal personer med okänd bakgrund. Detta beror på att invandringen varit ovanligt stor samt att skolenheterna i år har ombetts att rapportera in samtliga elever med okänd bakgrund. Elever med okänd bakgrund har även ingått tidigare år, men i mindre omfattning och då inte fått så stor betydelse för skolprestationer som detta år.

## 1.2 Definition av variabler

**Meritvärde** - utgörs av summan av betygsvärdena för de 16 bästa betygen i elevens slutbetyg och för 17 ämnen för elever som läst moderna språk som språkval. Elevens betyg omvandlas till betygsvärdena (E=10, D=12.5, C=15, B=17.5 och A=20) och de bästa 16 ämnena summeras, har eleven godkänt betyg i modernt språk som språkval är det ett 17:e ämne i meritvärdet. Det högsta möjliga meritvärdet för elever som läst 17 ämnen är 340 och för elever som läst 16 ämnen 320. Även om eleven fått betyg i färre ämnen än 16 kan betyget i moderna språk som språkval läggas till meritvärdet, men då kan eleven inte komma upp i det maximala meritvärdet 340 poäng. Lägsta möjliga värde är 0, dvs. ett underkänt betyg i alla ämnen.

Det genomsnittliga meritvärdet beräknas för de elever som fått lägst betyg E i minst ett ämne. Elevernas sammanlagda meritvärde divideras med antal elever som fått lägst betyg E i minst ett ämne. Elever som saknar godkänt betyg i alla ämnen, d.v.s. elever som saknar slutbetyg, ingår ej.

**Behörighet till gymnasieskolan** - godkända betyg i ämnena svenska/svenska som andraspråk, engelska och matematik samt i 5 ytterligare ämnen, totalt 8 ämnen. I beräkningarna ingår elever som fått eller skulle ha fått betyg enligt det mål och kunskapsrelaterade betygssystemet (elever som saknar godkänt betyg i alla ämnen ingår).

**Godkänd i alla ämnen** – godkända betyg i alla ämnen som ingått i elevens utbildning. I beräkningarna ingår elever som fått eller skulle ha fått betyg enligt det mål och kunskapsrelaterade betygssystemet (elever som saknar godkänt betyg i alla ämnen ingår).

**Kön** – pojke eller flicka

**Föräldrarnas utbildningsnivå** – utgörs av medelvärdet av varje elevs biologiska föräldrars utbildningsnivå, faderns respektive moderns högsta utbildningsnivå. Om det finns uppgift om enbart en förälders utbildningsnivå blir det utbildningsnivån. Föräldrarnas utbildning tilldelas 0 poäng om utbildningsnivån är okänd, 1 poäng om förgymnasial utbildning är den högsta utbildningsnivån, 2 poäng om gymnasial utbildning är den högsta nivån och 3 poäng om föräldern har uppnått eftergymnasial utbildning.

**Nyligen invandrad** – elever som kommit till Sverige för första gången under de senaste fyra åren. De har inte bott i Sverige eller gått i svensk skola tidigare.

**Ekonomiskt bistånd** – om någon av föräldrarna hade ekonomiskt bistånd (minst 100 kr) under året före personerna gick ut årskurs 9. För elever som gick ut årskurs nio 2016 är det således bistånd under 2015 som räknas.

**Okänd bakgrund** – elev med okänd bakgrund. Det finns väldigt lite bakgrundsinformation om dessa elever. Det som är känt är huruvida eleven är en pojke eller flicka.

## 2 Deskriptiv analys

I detta avsnitt beskrivs de variabler som ingår i analysen med hjälp av tabeller. I tabell 2 och 3 ingår elever som fått eller skulle ha fått betyg enligt det mål och kunskapsrelaterade betygssystemet (elever som saknar godkänt betyg i alla ämnen ingår). I övriga tabeller ingår endast elever som fått lägst betyg E i minst ett ämne.

Tabellerna visar på att variablerna kön, föräldrarnas utbildningsnivå, nyligen invandrad, ekonomiskt bistånd och okänd bakgrund samvarierar med meritvärdet.

**Tabell 1 Meritvärde**

År	Bakgrund	Antal elever	Medelvärde
2016	Okänd bakgrund	2 839	45,5
2016	Känd bakgrund	98 990	229,2
2016	Totalt	101 829	224,1

**Tabell 2 Behörighet till gymnasieskolan (yrkesprogram)**

År	Bakgrund	Antal elever	Andel behörig	Andel ej behörig
2016	Okänd bakgrund	5 397	3,6	96,4
2016	Känd bakgrund	100 376	87,3	22,7
2016	Totalt	105 733	83,1	16,9

**Tabell 3 Godkänd i alla ämnen**

År	Bakgrund	Antal elever	Andel godkänd i alla ämnen	Andel som inte är godkänd i alla ämnen
2016	Okänd bakgrund	5 397	2,5	97,5
2016	Känd bakgrund	100 376	78,1	21,9
2016	Totalt	105 733	74,2	25,8

**Tabell 4 Meritvärde efter kön**

År	Kön	Antal elever	Genomsnittligt meritvärde
2016	Okänd bakgrund	2 839	45,5
2016	Flicka	47 833	242,6
2016	Pojke	51 157	216,6
2016	Totalt	101 829	224,1

**Tabell 5 Meritvärde efter föräldrarnas utbildningsnivå**

År	Poäng	Antal elever	Genomsnittligt meritvärde
2016	Okänd bakgrund	2 839	45,5
2016	0	2 158	166,0
2016	1	5 885	162,1
2016	1,5	10 210	197,0
2016	2	31 486	217,1
2016	2,5	23 868	243,2
2016	3	25 383	265,0
2016	Totalt	101 829	224,1

**Tabell 6 Meritvärde efter nyligen invandrad**

År	Kategori	Antal elever	Genomsnittligt meritvärde
2016	Okänd bakgrund	2 839	45,5
2016	Nyligen invandrad	4 626	130,3
2016	Ej nyligen invandrad	94 364	234,0
2016	Totalt	101 829	224,1

**Tabell 7 Meritvärde efter ekonomiskt bistånd**

År	Bistånd	Antal elever	Genomsnittligt meritvärde
2016	Okänd bakgrund	2 839	45,5
2016	Har haft bistånd	7 703	167,1
2016	Har ej haft bistånd	91 287	234,4
2016	Totalt	101 829	224,1

## 2.1 Elever med okänd bakgrund

Elever med okänd bakgrund påverkar resultaten i skolan ganska kraftigt. Exempelvis är andelen behöriga till gymnasieskolan bland elever med känd bakgrund 87,3 procent. Om elever med okänd bakgrund inkluderas är andelen behöriga 83,1 procent. En förklaring till de stora skillnaderna är att elever med okänd bakgrund oftast är elever på flykt från krigsdrabbade länder och de har ofta befunnit sig väldigt kort tid i svensk skola och har därför haft små möjligheter att hinna få undervisning alla ämnen.

Eftersom det är stora skillnader i resultat mellan elever med okänd bakgrund och elever med känd bakgrund finns det anledning att fundera över hur de ska hanteras i den modell som ska tas fram. Elever med okänd bakgrund påverkar onekligen resultaten och man kan därför argumentera för att de ska vara med i modellen. Det finns dock ett flertal motiv för att exkludera dem.

- Elever med okänd bakgrund inte haft någon realistisk chans att uppnå ett resultat i skolan. Att de har svaga resultat beror inte på socioekonomi utan orsakerna är troligen andra (kort tid i svensk skola).
- Detta år och möjligen några kommande år är undantag från det normala. Elever med okänd bakgrund har inte varit lika många tidigare och kommer antagligen inte att vara lika många i framtiden, beroende av situationen i världen och hur flyktingmottagningen ser ut framöver. Modellen som tas fram bör vara stabil tid och inte påverkas alltför kraftigt av tillfälliga företeelser.
- Det är osäkert hur komplett rapporteringen av elever med okänd bakgrund är. Skolorna har ombetts rapportera in alla elever, men det finns ingen möjlighet att kontrollera om de verkligen gjort det.

Dessa skäl sammantaget motiverar att utesluta elever med okänd bakgrund ur analyserna.

## 3 Modeller

Analysen görs med hjälp av linjär regressionsanalys. Vid linjär regression kan regressionsmodellen beskrivas som

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

där  $x_{1i}, x_{2i}, \dots, x_{ki}$  utgör  $k$  stycken förklarande variabler för skolenhet  $i$ .  $\beta_1, \beta_2, \dots, \beta_k$  utgör regressionskoefficienterna och  $\alpha$  utgör interceptet. Termen  $\varepsilon_i$  kallas residual och den fångar upp de variationer i  $y$  som inte kan förklaras av  $x$ -variablerna.

Variablerna aggregeras till skolenhetsnivå. För exempelvis behörighet till gymnasieskolan så används andel behöriga per skolenhet som beroende variabel. För meritvärdet används genomsnittligt meritvärde per skolenhet. Skolenheter med färre än 15 elever i årskurs 9 har uteslutits ur analyserna. Detta för att erhålla en stabilare modell. Vid aggregeringen har elever med okänd bakgrund uteslutits.

### 3.1 Beräkning av förväntade värden

Förväntade värden för respektive skolenhet beräknas enligt följande

$$\hat{y}_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (2)$$

Förväntade värden på kommunnivå beräknas därefter genom att aggregera upp förväntade värden på skolenhetsnivå till kommunnivå.

Nedan visas regressionskoefficienter, t-värden<sup>1</sup> och p-värden<sup>2</sup> för de modeller som används för att beräkna förväntade värden på skolenhetsnivå.

### Tabell 8 Skolmodell 1

**Beroende variabel: Meritvärde (17)**

Förklarande variabel	Koeff.	t-värde	p-värde
Andel pojkar	-37,2	-6,9	<0,0001
Föräldrarnas utbildningsnivå	61,0	24,1	<0,0001
Andel nyinvandrad	-90,3	-12,2	<0,0001
Andel haft bistånd	-25,7	-3,6	0,0004

Modellens förklaringsgrad<sup>3</sup> (R<sup>2</sup>-värde): 0,629

### Tabell 9 Skolmodell 2

**Beroende variabel: Andel behöriga till gymnasieskolan**

Förklarande variabel	Koeff.	t-värde	p-värde
Andel pojkar	0,01	0,4	0,7265
Föräldrarnas utbildningsnivå	0,22	20,9	<0,0001
Andel nyinvandrad	-0,61	-19,3	<0,0001
Andel haft bistånd	-0,10	-3,3	0,0011

Modellens förklaringsgrad (R<sup>2</sup>-värde): 0,646

### Tabell 10 Skolmodell 3

**Beroende variabel: Andel godkända i alla ämnen**

Förklarande variabel	Koeff.	t-värde	p-värde
Andel pojkar	-0,06	-2,1	0,0383
Föräldrarnas utbildningsnivå	0,31	24,2	<0,0001
Andel nyinvandrad	-0,45	-11,6	<0,0001
Andel haft bistånd	-0,28	-7,4	<0,0001

Modellens förklaringsgrad (R<sup>2</sup>-värde): 0,643

<sup>1</sup> Avstånd i standardavvikelse, mellan de observerade värdet i datamaterialet och hur värdet förväntas se ut om nollhypotesen vore sann. Ju större värde på teststatistikan, absolut sett, desto mindre sannolikhet att nollhypotesen är sann. Nollhypotesen i detta fall är att den förklarande variabeln inte har någon effekt på den beroende variabeln.

<sup>2</sup> Sannolikheten att observera ett minst lika extremt värde som det värde vi observerat, givet att nollhypotesen är sann. Ett lågt p-värde indikerar att det är väldigt låg sannolikhet att den förklarande variabeln inte påverkar den beroende variabeln.

<sup>3</sup> Förklaringsgraden är ett mått på hur nära modellen följer de observerade utfallen, i modellen (1). Detta mått är ett tal mellan 0 och 1, och det anger kvoten mellan de predikterade och de observerade y-värdenas varianser. Ett värde nära 1 betyder att modellen nära fullständigt förklarar variationerna i y, medan ett värde nära 0 betyder att modellen nästan inte alls förklarar något av variationerna i y.