

Teknisk beskrivning

Gammal och ny beräkningsmodell för socioekonomiskt index

Teknisk beskrivning av gammal och ny beräkningsmodell för socioekonomiskt index

Sedan 2020 publicerar SKR "Öppna jämförelser: Förskola" för att ge politiker och tjänstepersoner stöd i det systematiska kvalitetsarbetet. En del av detta arbete handlar om att analysera de kommunala förskolornas socioekonomi, där ett socioekonomiskt index tas fram för varje förskola. Indexet används för att förstå och sätta respektive förskolas förutsättningar i ett större sammanhang. Från och med 2024 har SKR tagit fram en ny metod för att beräkna indexet. Denna rapport beskriver förändringarna och hur de påverkar tolkningen av statistiken. För den som vill fördjupa sig ytterligare finns även SCB:s tekniska rapport om statistikarbetet att läsa.

Tidigare beräkningsmodell

Den tidigare modellen för att beräkna indexet byggde på en statistisk metod som kallas standardisering. Detta innebär att varje variabel – som föräldrarnas utbildningsnivå, inkomst, ekonomiskt bistånd och invandringsbakgrund – omvandlades till en skala där värdena mäts i förhållande till medelvärdet och spridningen (standardavvikelsen) för just den variabeln. Med andra ord omvandlades värdena så att de fick ett medelvärde på 0 och en standardavvikelse på 1.

När alla variabler hade standardiserats, vägdes de samman enligt en formel som såg ut så här:

$$(1/3 * \text{Föräldrarnas utbildningsnivå}) - (1/3 * \text{Nyinvandrad eller okänd bakgrund}) + (1/6 * \text{Inkomst}) - (1/6 * \text{Ekonomiskt bistånd})$$

Värdena rangordnades till sist genom en metod som kallas *percentilindelning*. Detta innebär att man delade upp alla värden i hundradelar (percentiler) och placerade varje enhet i en specifik procentnivå, beroende på hur dess indexvärde förhöll sig till övriga enheter.

Denna metod för relativ rangordning gjorde det möjligt att jämföra enheter inom samma år, men den skapade också en utmaning: eftersom indexet byggde på en relativ skala var det svårt att följa trender över tid. Ett värde som såg högt ut ett år kunde vara lågt ett annat år, beroende på hur resten av populationen förändrades, vilket gjorde indexet svårtolkat i ett längre tidsperspektiv.

Ny beräkningsmodell

Den nya modellen för att beräkna indexet har utvecklats för att göra det enklare att förstå och använda, samtidigt som den behåller sin funktion att ge en rättvisande bild av socioekonomiska skillnader. Modellen bygger på samma grund som tidigare, men vissa viktiga förändringar har införts för att förbättra användbarheten och stabiliteten.

Precis som tidigare används fyra variabler för att beskriva den socioekonomiska bakgrunden: föräldrarnas utbildningsnivå, inkomst, ekonomiskt bistånd och nyinvandrad eller okänd bakgrund. En viktig förändring i den nya modellen är att inkomstvariabeln inte längre hanteras som en kontinuerlig variabel. Istället delas den in i kategorier vilket gör modellen mindre känslig för extrema värden och bidrar till en stabilare analys.

För att ytterligare hantera snedfördelningar inom de enskilda variablerna har gränsvärden införts. Dessa gränsvärden används för att justera extrema värden som annars skulle kunna få oproportionerligt stor inverkan på det slutgiltiga indexet.

- Föräldrarnas utbildningsnivå: Låga värden under 1,5 justeras upp till 1,5.
- Andel nyinvandrad eller okänd bakgrund: Höga värden över 0,1 justeras ned till 0,1.
- Andel ekonomiskt bistånd: Höga värden över 0,2 justeras ned till 0,2.

Efter att variablerna har förberetts, omvandlas de till en gemensam skala mellan 0 och 100 genom en metod som kallas omskalning. Detta gör att variablerna blir intuitiva och lätta att tolka, där ett högt värde indikerar en hög nivå av den aktuella variabeln (t.ex. hög utbildningsnivå eller inkomst) och ett lågt värde indikerar en låg nivå. Genom denna omskalning kan alla variabler behandlas på samma enhetliga sätt, vilket underlättar jämförelser och analyser.

När variablerna har kategoriserats och omskalats vägs de samman enligt en förbestämd formel, där vissa variabler ges större vikt än andra. Formeln för att beräkna indexet är densamma som i den tidigare modellen:

$$(1/3 * \text{Föräldrarnas utbildningsnivå}) - (1/3 * \text{Nyinvandrad eller okänd bakgrund}) + (1/6 * \text{Inkomst}) - (1/6 * \text{Ekonomiskt bistånd})$$

Den största förändringen jämfört med den tidigare modellen är att indexet nu baseras på faktiska värden och inte längre använder percentilindelning. Detta innebär att indexvärdena är fasta och inte beroende av hur resten av populationen förändras. Därmed blir det möjligt att jämföra enheter både inom samma år och över tid. Jämförelser bör dock göras med försiktighet, då det till stor del rör sig om andra barn. Enligt vår nya statistik om barnomsättning är i genomsnitt 57 % av barnen nya på sin förskolenhet sen den senaste statistiken togs fram för två år sedan.

Skillnader mellan beräkningsmodellerna

Övergången från den tidigare beräkningsmodellen till den nya innebär både metodologiska och praktiska förändringar som påverkar hur indexet fungerar och tolkas. Här lyfter vi fram de viktigaste skillnaderna, styrkorna i den nya modellen samt vad som är viktigt att tänka på vid tolkning av det nya indexet.

Viktiga skillnader

1. Omskalning istället för standardisering

I den tidigare modellen användes standardisering, vilket innebär att variabler omvandlades till en skala där värdena mättes i förhållande till medelvärdet och spridningen. I den nya modellen används istället omskalning, där alla variabler omvandlas till ett intervall mellan 0 och 100.

- *Effekt:* Den nya skalan är mer intuitiv och lättare att tolka. Höga värden (nära 100) representerar en hög nivå av den aktuella variabeln, medan låga värden (nära 0) indikerar en låg nivå. Detta gör att resultatet blir mer direkt användbart för icke-statistiker.

2. Kategorisering av inkomst

Tidigare hanterades inkomst som en kontinuerlig variabel. I den nya modellen delas inkomsten in i kategorier som låg, medel och hög.

- *Effekt:* Kategoriseringen gör modellen mindre känslig för extrema värden, som höga eller låga inkomster, vilket bidrar till en mer stabil analys. Däremot innebär det att vissa små skillnader i inkomst inom en kategori inte längre framgår.

3. Justering av extrema värden

För att hantera snedfördelningar inom variablerna har gränsvärden införts. Dessa gränsvärden klipper av låga eller höga värden som annars skulle kunna få oproportionerligt stor inverkan på indexet:

- Föräldrarnas utbildningsnivå: Låga värden under 1,5 justeras upp till 1,5.
- Andel nyinvandrad eller okänd bakgrund: Höga värden över 0,1 justeras ned till 0,1.

- Andel ekonomiskt bistånd: Höga värden över 0,2 justeras ned till 0,2.
- *Effekt:* Denna justering minskar påverkan från extrema värden och bidrar till att skapa ett mer balanserat och stabilt index. Det gör indexet lättare att tolka, mer robust över tid och mindre känsligt för förändringar i populationens dynamik. Dessutom underlättar detta för användare utan statistisk bakgrund att förstå och använda resultaten.

4. **Slopad percentilindelning**

Den tidigare modellen avslutades med att värdena rangordnades i percentiler, vilket gjorde att varje enhets värde relaterades till resten av populationen. I den nya modellen används inte percentilindelning.

- *Effekt:* Indexvärdena är nu fasta och jämförbara över tid, eftersom de inte längre påverkas av förändringar i resten av populationen. Detta gör det lättare att följa trender och förändringar. Värden som i den tidigare modellen var relativt extrema (dvs, låg väldigt högt eller väldigt lågt) har förändrats mest eftersom de inte längre delas in i jämfördelade percentiler.

5. **Minskad spridning i indexvärden**

Spridningen av indexvärden är mindre i den nya modellen än i den tidigare på grund av ovan nämnda skillnader i beräkning.

- *Effekt:* Den minskade spridningen kan innebära att skillnader mellan enheter inte framträder lika tydligt som tidigare. Däremot blir indexet stabilare och mer robust vid långsiktig analys.

Styrkor i den nya modellen

Den nya beräkningsmodellen har flera styrkor som gör den mer användarvänlig och praktisk:

- **Intuitiv skala:** Genom omskalningen blir indexet lättare att förstå och tolka, även för personer utan statistisk bakgrund.

- **Jämförbarhet över tid:** Eftersom indexvärdena nu är fasta och inte beroende av populationens förändringar, kan man enklare identifiera långsiktiga trender och förändringar i socioekonomisk bakgrund.
- **Mindre känslighet för extrema värden:** Kategoriseringen av inkomst gör att modellen är mer robust och stabil, särskilt i analyser där det finns extrema data.

Viktigt att tänka på vid tolkning

Medan den nya modellen har flera styrkor, finns det också aspekter som läsaren bör ha i åtanke när indexet används:

- **Variation på enhetsnivå:** Även om de övergripande trenderna i sambanden mellan socioekonomi och faktorer som personaltäthet eller andelen förskollärare ser liknande ut i de olika beräkningsmodellerna, kan enheternas individuella socioekonomiska värden skilja sig ganska mycket mellan den gamla och den nya beräkningen. Detta beror i synnerhet på den slopade percentilindelningen där vi inte längre anpassar en enhets värde till resten av populationen.
- **Minskad detaljrikedom i inkomst:** Genom att inkomst delas in i kategorier förloras viss precision jämfört med att använda kontinuerliga värden. Detta innebär att små skillnader inom samma kategori inte framgår i indexet.
- **Mindre spridning i värden:** Den nya modellen gör det lättare att jämföra över tid, men skillnader mellan enheter kan upplevas som mindre framträdande.
- **Tolkning av skalan:** Även om omskalningen gör indexet mer intuitivt, är det fortfarande viktigt att tolka varje variabel i sitt sammanhang. Ett högt indexvärde betyder inte automatiskt att en enhet är "bäst", utan det beskriver de socioekonomiska förutsättningarna i relation till de valda variablerna.

Den nya modellen är utformad för att vara både enklare att använda och att tolka men också att jämföra och vara mer stabil över tid. Genom att förstå de viktigaste skillnaderna och vad man bör tänka på vid tolkning, kan indexet fortsatt användas som ett verktyg för att analysera hur socioekonomiska förutsättningar samvarierar med andra faktorer.

Ett fiktivt exempel: hur beräkningsmodellerna påverkar enheter

För att illustrera skillnaderna mellan den gamla och nya beräkningsmodellen, tänk dig en fiktiv kommun med tre förskoleenheter: Solrosen, Blåklinten och Vallmon. Dessa enheter skiljer sig åt i fråga om socioekonomiska förutsättningar, vilket reflekteras i deras socioekonomiska index.

I den gamla modellen:

- **Standardisering och percentilindelning** gjorde att enheternas indexvärden berodde på hur de förhöll sig till alla andra enheter i kommunen.
- Låt oss säga att Solrosen hade en hög andel föräldrar med hög utbildning och låg andel ekonomiskt bistånd, vilket gav ett relativt högt socioekonomiskt index. Blåklinten, med en stor andel nyinvandrade familjer och låg inkomst, placerades däremot i den lägsta percentilen. Vallmon, som låg i mitten av skalan för de flesta variabler, fick ett indexvärde nära medianen.

I den gamla modellen kunde deras indexvärden se ut så här:

- Solrosen: 85 (percentil 85)
- Blåklinten: 10 (percentil 10)
- Vallmon: 50 (percentil 50)

Eftersom percentilindelning användes, framstår skillnaderna mellan enheterna som väldigt stora, även om de faktiska skillnaderna i variablerna kanske inte var lika dramatiska.

I den nya modellen:

- **Omskalning och slopad percentilindelning** gör att varje enhets värden reflekterar deras faktiska socioekonomiska förutsättningar snarare än deras rangordning i populationen.
- Solrosen, med samma högutbildade föräldrar och låg andel ekonomiskt bistånd, får fortfarande ett högt indexvärde men baserat på sina faktiska variabler snarare än relativ placering. Blåklinten får ett lågt värde, men det är inte lika extremt som i den gamla modellen, eftersom det inte längre relateras till de mest extrema enheterna i populationen. Vallmon,

som ligger nära mitten på skalan, reflekterar också sina faktiska värden utan att "tryckas ihop" till medianen.

I den nya modellen kunde deras indexvärden se ut så här:

- Solrosen: 80
- Blåklinten: 35
- Vallmon: 57

Här ser vi att skillnaderna mellan enheterna är mindre framträdande än i den gamla modellen, vilket speglar en stabilare och mindre extrem fördelning av värden. De relativa positionerna mellan enheterna (Solrosen högst, Blåklinten lägst) är desamma, men de faktiska värdena är nu lättare att tolka och jämföra över tid.

Sammanfattning av skillnaderna i exemplet:

1. **Spridning av värden:** I den gamla modellen var skillnaderna mellan enheterna överdrivna på grund av percentilindelningen. Den nya modellen ger en mer proportionerlig bild av skillnaderna.
2. **Jämförbarhet över tid:** De gamla indexvärdena kunde förändras kraftigt beroende på hur andra enheter i populationen förändrades. I den nya modellen är indexvärdena stabila och kan följas över tid utan att påverkas av populationens dynamik.
3. **Tolkning av värden:** I den nya modellen är skalan (0–100) intuitiv. Ett värde på 80 för Solrosen och 35 för Blåklinten är lättare att förstå och relatera till deras faktiska socioekonomiska förutsättningar.

Ordförklaring

För att underlätta förståelsen av de statistiska termer som används i beskrivningen av indexberäkningen, inkluderas här en kort ordlista:

Standardisering:

En statistisk metod som omvandlar värdena i en variabel till en gemensam skala genom att beräkna z-värden. Detta innebär att varje värde mäts som antalet standardavvikelser från medelvärdet, vilket ger ett medelvärde på 0 och en standardavvikelse på 1. Detta tillvägagångssätt är baserat på statistiska och matematiska principer som kan vara svåra att intuitivt förstå för personer utan bakgrund i statistik.

Omskalning (0-100):

Till skillnad från standardisering, som omvandlar värden till en skala baserad på standardavvikelser, är omskalning en metod som omvandlar variabler till ett enhetligt intervall mellan 0 och 100. Syftet är att skapa en enkel och intuitiv skala där höga värden (nära 100) representerar en hög nivå av den aktuella variabeln (t.ex. hög utbildningsnivå eller inkomst), medan låga värden (nära 0) indikerar en låg nivå. Medan standardisering kan vara svårare att tolka, gör omskalningen det möjligt att jämföra olika värden på ett mer direkt och lättbegripligt sätt, vilket är en central fördel i den nya modellen.

Percentilindelning:

En metod för att dela upp en datamängd i hundradelar, där varje percentil representerar en hundradel av populationen. Till exempel innebär 10:e percentilen att en viss enhet har ett indexvärde som är högre än 10 % av alla andra enheter. Denna metod användes i den tidigare modellen för att relativt rangordna enheter och kan göra det svårt att följa trender över tid, eftersom värdena är beroende av den specifika populationen vid ett visst tillfälle.

Kategorisering:

Processen att omvandla en kontinuerlig variabel, som inkomst, till diskreta grupper eller kategorier (exempelvis låg, medel, hög). Genom att kategorisera variabler minskar känsligheten för extrema värden och gör variabeln lättare att tolka. I den nya beräkningsmodellen har inkomstvariabeln kategoriserats, vilket gör att indexvärdena blir mer stabila och enklare att förstå i praktiska termer.

Teknisk beskrivning

Denna tekniska beskrivning redogör för den tidigare och nya beräkningsmodellen för socioekonomiskt index som används i SKR:s Öppna jämförelser: Förskola. Skriften förklarar de viktigaste förändringarna, varför de har gjorts, samt vad man bör tänka på vid tolkning av indexet. Syftet är att ge en djupare förståelse för metodiken bakom statistiken och hur den kan användas för analys.

Upplysningar om innehållet
David Bross, david.bross@skr.se

© Sveriges Kommuner och Regioner, 2024
Text: David Bross
www.skr.se