



STYRELSEN FOR
IT OG LÆRING

Den socioøkonomiske reference for grund- skolekarakterer

Den statistiske model

Børne- og Undervisningsministeriet

Den socioøkonomiske reference for grundskolekarakterer
Den statistiske model

2024

Design: Center for Kommunikation og Presse

Denne publikation kan ikke bestilles.

Der henvises til webudgaven.

Publikationen kan hentes på:

www.uddannelsesstatistik.dk

Socioøkonomisk reference for grundskolekarakterer – Om den statistiske model

I følgende kan du læse mere specifikt om den anvendte statistiske model for beregning af socioøkonomiske referencer på Grundskolekarakterer. Ønsker du generel vejledning til læsning og forståelse af resultater henvises til "Den socioøkonomiske reference for Grundskolekarakterer - Vejledning til læsning og tolkning af resultaterne".

Indhold

| | |
|--|----------|
| Den statistiske model | 4 |
| Kodning af socioøkonomiske baggrundsoplysninger | 6 |
| Den statistiske model for socioøkonomisk reference på kommuneniveau | 8 |

Den statistiske model

Den anvendte statistiske model for beregning af socioøkonomisk reference for grundskolekarakterer er en såkaldt multilevel-model:

$$y_{ij} = x_{ij}\beta + u_j + e_{ij} \quad (1)$$

hvor y_{ij} er karakterresultatet for elev i på institution j , x_{ij} er elevens værdier på baggrundsvariable, og β er de tilhørende parameterestimer, u_j er varianskomponenten svarende til variationen mellem skolerne, og e_{ij} er residualen svarende til variationen mellem eleverne på skolen.

Ved anvendelse af modellen fås et estimat for, hvor stor en del af variationen mellem karaktererne der skyldes forskelle mellem eleverne, og hvor stor en del der skyldes forskelle mellem skolerne.

Alle de socioøkonomiske baggrundsvariable beskrevet nedenfor i afsnittet "Kodning af Socioøkonomiske baggrundsvariable" medtages i modellen sammen med interaktionen mellem forældrenes uddannelse og elevens herkomst, interaktionen mellem elevens køn og herkomst samt andelen på skoleniveau af hhv. indvandrere og efterkommere og andelen af elever med forældre med en mellemlang eller lang videregående uddannelse.

Ikke alle elever har fyldestgørende oplysninger på de benyttede baggrundsvariable. Der er i disse tilfælde indsat en værdi, et såkaldt bedste gæt, på alle de steder, hvor oplysningerne er ukendte. For hver elev med ukendt baggrundsoplysning gættes 5 gange, således at modellen bliver estimeret på baggrund af en sandsynlig fordeling. Metoden kaldes multipel imputation. Gættene baseres på fordelingen blandt elever med oplyst baggrundsvariabel, hvor der er taget hensyn til elevens køn, alder og herkomst.

Den beregnede socioøkonomiske reference for hver elev er et gennemsnit over de 5 modelkørsler, der bygger på hvert sit sæt af de 5 sæt af imputationer for forklarende variable.

I den statistiske model for beregning af socioøkonomiske referencer for 3-årsperioden medtages endvidere skoleåret som en forklarende variabel. Herved tages der højde for, at karakterniveauet kan være forskelligt fra år til år¹.

For hver elev beregnes en socioøkonomisk referenceværdi ud fra værdierne af elevens baggrundsvariable og de estimerede parametre for baggrundsvariablene. Herefter beregnes forskellen mellem opnået karakter og den socioøkonomiske reference, det såkaldte residual r_{ij} , for hver elev.

På skoleniveau er den estimerede forskel mellem karaktergennemsnittet og den socioøkonomiske reference i en ordinær regressionsmodel gennemsnittet af disse elevresidualer på den enkelte skole. I de her anvendte to-niveau-modeller, hvor vi både har variation mellem skoler ('Between'), og variation mellem elever inden for skolen ('Within'), bliver residualerne på skoleniveau:

$$\hat{u}_j = c_j \bar{r}_j \quad (2)$$

hvor c_j er den såkaldte 'shrinkage'-faktor, defineret ved

¹ Grundet de varierende regler for prøveafvikling og karaktergivning i skoleårene 2019/2020, 2020/2021 og 2021/2022 er der ikke beregnet 3-årige socioøkonomiske referencer siden skoleåret 2018/2019.

$$c_j = \frac{B}{(B + W/n_j)} \quad (3)$$

hvor B er variationen mellem skoler ('Between'), W er residualvariationen mellem elever inden for skolen ('Within'), og n_j er antallet af elever på den j 'te skole.

For forskellen mellem opnået karakter og den socioøkonomiske reference for hver skole er der endvidere udregnet et 95%-sikkerhedsinterval. Hvis dette sikkerhedsinterval ligger over 0, opnår skolens elever et statistisk signifikant højere karaktergennemsnit, end elever på landsplan med lignende baggrundsforhold, mens det modsatte er tilfældet, hvis sikkerhedsintervallet ligger under 0. Indeholder 95%-sikkerhedsintervallet værdien 0, da opnår eleverne på den pågældende skole et karaktergennemsnit på niveau med elever på landsplan med lignende baggrundsforhold.

Sikkerhedsintervallet afhænger dels af den estimerede variation og dels af antallet af elever, der indgår i beregningerne for den enkelte skole. Således bliver sikkerhedsintervallet mindre for den enkelte skole, når en 3-årig periode betragtes, idet der her indgår elever for 3 år i beregningerne.

Kodning af socioøkonomiske baggrundsoplysninger

I nedenstående tabel fremgår kodningen af de enkelte socioøkonomiske baggrundsoplysninger:

Tabel 1: Kodning af de enkelte baggrundsoplysninger

| Baggrundsoplysninger | Valgt kodning |
|--|--|
| Køn | Dreng Pige |
| Alder | 15 år eller yngre 15-16 år 16-17 år Ældre end 17 år |
| Herkomst¹ og oprindelsesland² | Dansk herkomst Indvandrere fra Vestlige lande Indvandrere fra Ikke-vestlige lande Efterkommere fra Vestlige lande Efterkommere fra Ikke-vestlige lande |
| Forældrenes højeste fuldførte uddannelse³ | Grundskolen Gymnasial uddannelse Erhvervsfaglig uddannelse Kort videregående uddannelse Mellemlang videregående uddannelse Lang videregående uddannelse (inkl. ph.d. mv.) |
| Faderens arbejdsmarkedsstatus⁴ | Beskæftiget Ledig Øvrige uden for arbejdsmarkedet |
| Moderens arbejdsmarkedsstatus | <i>(som faderens arbejdsmarkedsstatus)</i> |
| Forældrenes gennemsnitlige bruttoindkomst⁵ | 1. kvartil 2. kvartil 3. kvartil 4. kvartil |
| Faderens ledighedsgrad⁶ | 0 % 1 – 50 % 50+ % |
| Moderens ledighedsgrad | <i>(som faderens ledighedsgrad)</i> |
| Familietype⁷ | Par Enlig Ikke hjemmeboende børn |
| Antal børn og placering i børneflokk⁸ | Enebarn Ældste barn i børneflok på højst 4 børn Midterbarn i børneflok på højst 4 børn Yngste barn i børneflok på højst 4 børn Øvrige |
| Forældres kriminalitet⁹ | Mor/far dømt Mor/far <u>ej</u> dømt |
| Sociale foranstaltninger¹⁰ | Social foranstaltning Ej social foranstaltning |

| Baggrundsoplysninger | Valgt kodning |
|---|-------------------------------------|
| Andel indvandrere/efterkommere på skolen | 0-5 % 5-10 % 10-20 % 20+ % |
| Andel forældre med videregående uddannelse på skolen | 0-5 % 5-10 % 10-20 % 20+ % |

Noter:

- 1) Der skelnes mellem indvandrere, efterkommere, og personer med dansk oprindelse. Personer med dansk oprindelse (dansk herkomst) er personer, hvor mindst en af forældrene er dansk statsborger født i Danmark. Efterkommere er født i Danmark, hvor ingen af forældrene er danske statsborgere født i Danmark. Når en eller begge forældre, der er født i Danmark, opnår dansk statsborgerskab, vil deres børn ikke blive klassificeret som efterkommere. Fastholder danskfødte forældre imidlertid begge et udenlandsk statsborgerskab, vil deres børn blive klassificeret som efterkommere. Hvis der ikke findes oplysninger om nogen af forældrene, og personen er udenlandsk statsborger, opfattes personen også som efterkommer. Indvandrere er født i udlandet. Ingen af forældrene er danske statsborgere født i Danmark. Hvis der ikke findes oplysninger om nogen af forældrene, og personen er født i udlandet, opfattes personen som indvandrer.
- 2) Kodningen 'Indvandrere/efterkommere fra Ikke-vestlige lande' indeholder også de meget få, som har uoplyst oprindelsesland, er statsløse m.m.
- 3) Den højest fuldførte uddannelse blandt elevens mor og far målt i det aktuelle år, medmindre eleven er ældre end 13 år. Da indgår oplysningen fra det år, eleven fyldte 13 år.
- 4) Arbejdsmarkedsstatus er et 'øjebliksbillede' målt ultimo november det aktuelle år, med mindre eleven er ældre end 13 år. Da indgår oplysningen fra det år, eleven fyldte 13 år.
- 5) Den valgte kodning er baseret på medianen og nedre og øvre kvartil af bruttoindkomsten blandt forældrene i elevpopulationen, som de socioøkonomiske referencer beregnes på baggrund af. Bruttoindkomsten er fra det aktuelle år, medmindre eleven er ældre end 13 år. Da indgår oplysningen fra det år, eleven fyldte 13 år.
- 6) Grad af ledighed i året beregnes som personens timer i tilstanden (hverdage) i forhold til årets referencetimer, som er antal hverdage i året ganget med 7,4. En ledighedsgrad på 0 % svarer til at personen ikke har været ledig i året. En ledighedsgrad på 50+ % svarer til, at personen har været ledig mellem halvdelen og hele året.
- 7) Familietypen par indeholder ægtepar, registreret partnerskab, samlevende og samboende par.
- 8) Antal børn og placering i børneflokkene er sammensat af to variable: 'Antal børn i familien' og 'Nummer i søskendeflokken'. Antal børn i familien dækker antal hjemmeboende børn under 18 år, som er børn af mindst én af de op til to voksne i familien.
- 9) Variablen angiver hvorvidt mindst en af elevens forældre er dømt for alvorlig kriminalitet i perioden fra elevens fødsel til prøveafleggelse på 9. klassetrin. Alvorlig kriminalitet defineres som overtrædelser af straffeloven, lov om euforiserende stoffer eller våbenloven.
- 10) Variablen angiver om eleven har oplevet en eller flere forebyggende foranstaltninger eller har været anbragt uden for hjemmet i perioden fra barnets fødsel til prøveafleggelse. Forebyggende foranstaltninger gives til børn der har brug for særlig støtte i henhold til udvalgte paragraffer i lov om social service samt lov om bekæmpelse af ungdomskriminalitet.

Den statistiske model for socioøkonomisk reference på kommuneniveau

Opgørelserne på kommuneniveau foretages for gennemsnittet af elevernes karakterer i de bundne prøvefag på baggrund af beregningerne af den socioøkonomiske reference på institutionsniveau. Opgørelsen på kommuneniveau indeholder udelukkende resultaterne fra folkeskolerne.

På kommuneniveau kan forskellen mellem elevernes gennemsnitskarakterresultat og den estimerede socioøkonomiske reference beregnes som det vægtede gennemsnit af de beregnede forskelle på institutionsniveau:

$$(\sum_1^m n_i u_i)/n \quad (4)$$

hvor u_i er den estimerede forskel, n_i er antallet af elever med et karaktergennemsnit af de bundne prøver på i 'te institution, og m er antallet af institutioner i den enkelte kommune. n er antallet af elever i alt med et karaktergennemsnit af de bundne prøver på kommunens folkeskoler.

Tilsvarende beregnes variansen på kommuneniveau²:

$$\text{Var}(\sum_1^m n_i u_i/n) = \sum_1^m n_i^2 \text{Var}(u_i)/n^2 \quad (5)$$

hvor $\text{Var}(u_i)$ er variationen på den estimerede forskel på i 'te institution.

Da residualerne på institutionsniveau, u_j , i model (1) antages at være normalfordelte, kan der efterfølgende beregnes et 95%-sikkerhedsinterval. Hvis dette sikkerhedsinterval ligger over 0, opnår kommunens folkeskoleelever samlet et statistisk signifikant højere karaktergennemsnit end elever på landsplan med lignende baggrundsforhold, mens det modsatte er tilfældet, hvis sikkerhedsintervallet ligger under 0. Indeholder 95%-sikkerhedsintervallet værdien 0, da opnår eleverne i den pågældende kommune samlet et karaktergennemsnit som elever på landsplan med lignende baggrundsforhold.

Sikkerhedsintervallet afhænger dels af den estimerede variation og dels af antallet af elever, der indgår i beregningerne for den enkelte institution og den enkelte kommune. Således bliver sikkerhedsintervallerne mindre for store kommuner, hvorfor relativt små forskelle mellem karaktergennemsnit og den socioøkonomiske reference kan blive statistisk signifikante.

² Under antagelse af uafhængighed mellem institutionerne.



BØRNE- OG
UNDERVISNINGSMINISTERIET
STYRELSEN
FOR IT OG LÆRING