



STYRELSEN FOR
IT OG LÆRING

Den socioøkonomiske reference for Studen- ternes karakterer

Den statistiske model

Børne- og Undervisningsministeriet



Den socioøkonomiske reference for Studenternes karakterer
Den statistiske model

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 Studenternes karakterer – Om den statistiske model

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

Indhold

Den statistiske model	4
Kodning af socioøkonomiske baggrundsoplysninger	6

Den statistiske model

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

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

hvor y_{ij} er karakteren for student i på afdeling j , x_{ij} er studentens baggrundsvariable og β er de tilhørende parameterestimer, u_j er varianskomponenten svarende til variationen mellem afdelingen, og e_{ij} er residualen svarende til variationen mellem studenterne på afdelingen. For hver kombination af fag, prøveform og niveau foregår beregningerne på individniveau for hver af de gymnasiale uddannelser hver for sig.

Ved anvendelse af modellen fås et estimat over, hvor stor en del af variationen mellem karaktererne, der skyldes forskelle mellem studenter, og hvor stor en del, der skyldes forskelle mellem afdelingerne. Dette kaldes modellens forklaringsgrad.

Tabel 1 i afsnittet "Kodning af socioøkonomiske baggrundsoplysninger" viser kodningen bag de socioøkonomiske baggrundsvariable som er medtaget i den endelige socioøkonomisk reference model. I opbygningen af den endelige model er løbende foretaget en udvælgelse af relevante baggrundsvariable. Baggrundsvariable som konsistent har vist sig ikke at være statistisk signifikante i de multiple modeller er ikke medtaget i den endelige beregning. Fravalgte variable omfatter blandt andet interaktionen mellem forældrenes uddannelse og studentens herkomst samt en række gruppevariable på afdelingsniveau, herunder andelen af elever hvis forældre har lang videregående uddannelse eller andelen af elever hvis forældre har høj/lav indkomst. Disse variable har ikke vist sig statistisk signifikante i de multiple modeller.

Modellerne bag den 1-årige og 3-årige beregning bygger som udgangspunkt på samme baggrundsvariable, med den forskel, at den statistiske model for beregning af socioøkonomiske referencer for 3-års perioden også medtager dimittendåret som en forklarende variabel. Hermed tages der højde for, at karakterniveauet kan være forskelligt fra år til år.

Ikke alle studenter har fyldestgørende oplysninger på de benyttede baggrundsvariable. En ofte anvendt metode til håndtering af dette er 'missing value' indikator metoden, hvor uoplyst medtages som en selvstændig kategori på lige fod med baggrundsvariablenes øvrige kategorier, og der tilføjes en indikatorvariabel, der fortæller hvorvidt, der er tale om en uoplyst værdi. Omfanget af ukendte oplysninger er dog ikke uafhængigt af herkomst. Oplysningsgraden for alle baggrundsvariablene er størst for studenter af dansk herkomst. Der er således ikke tale om tilfældige 'huller' i informationen. Det er f.eks. specielt udtalt blandt indvandrere, at informationen om forældrenes uddannelse mangler. Da uddannelsesniveauet blandt de indvandrere og efterkommere, hvor uddannelsesoplysningerne kendes, er en del anderledes end blandt personer med dansk herkomst, vil der være en stor risiko for, at 'missing value' indikator metoden fejlesterimerer effekten af såvel herkomst som af uddannelse. Tilsvarende problematik eksisterer for de øvrige baggrundsvariable.

Derfor er der i stedet valgt at indsætte en værdi, et såkaldt bedste gæt, på alle de steder, hvor oplysningerne er ukendte. For hver student med ukendt baggrundsoplysning gættes fem gange, således modellen bliver estimeret på baggrund af en sandsynlig fordeling. Metoden kaldes 'multiple imputation'. Gættene baseres på fordelingen blandt studenter med oplyst baggrundsvariabel, hvor der er taget hensyn til studentens køn, alder og herkomst. Den socioøkonomiske reference for hver student er så et gennemsnit af de fem beregnede værdier.

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

På afdelingsniveau er den estimerede forskel mellem karaktergennemsnittet og den socioøkonomiske reference i en ordinær regressionsmodel gennemsnittet af disse studentresidualer på den enkelte afdeling, $\bar{r}_j$. I de her anvendte to-niveaumodeller, hvor vi både har variation mellem afdelinger ('Between'), og variation mellem studenter inden for afdelingen ('Within'), bliver residualerne på hovedinstitutionsniveau:

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

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

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

B er residualvariationen mellem afdelingerne ('Between'), W er residualvariationen mellem studenter inden for afdelingen ('Within') og n_j er antallet af studenter på den j 'te afdelingen.

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

Sikkerhedsintervallet afhænger dels af den estimerede variation og dels af antallet af studenter, der indgår i beregningerne for den enkelte afdeling. Således bliver sikkerhedsintervallerne mindre for den enkelte afdeling, når der betragtes en 3-årig periode, idet der her indgår 3 studenterårgange i beregningerne.

Kodning af socioøkonomiske baggrundsoplysninger

I nedenstående tabel fremgår kodningen af de enkelte socioøkonomiske baggrundsoplysninger, som er medtaget i beregningen af de socioøkonomiske referencer.

Tabel 1: Kodning af de enkelte baggrundsoplysninger

Baggrundsoplysninger	Valgt kodning
Karaktergennemsnit fra 9. klasse	Karaktergennemsnit i bundne prøver ¹ .
Særlige prøveregler i grundskolen grundet Covid-19 ²	Særlige prøveregler (2019/2020) Særlige prøveregler (2020/2021 og 2021/2022) Normal prøveafvikling
Køn	Dreng Pige
Alder på tidspunkt for afslutning af uddannelsen	18 år eller yngre 19 år 20 år 21 år eller ældre
Herkomst ³	Dansk Indvandrer Efterkommer
Forældrenes højeste fuldførte uddannelse ⁴	Grundskolen Gymnasial uddannelse Erhvervsfaglig uddannelse Kort videregående uddannelse Mellemlang videregående uddannelse Lang videregående uddannelse (inkl. ph.d. mv.)
Forældrenes gennemsnitlige bruttoindkomst ⁵	1. kvartil 2. kvartil 3. kvartil 4. kvartil
Familiestatus ⁶	Hjemmeboende hos par Hjemmeboende hos enlig Udeboende
Adgangsvej ⁷	Folkeskole Friskole og privat grundskole Efterskole Anden
10. klasse ⁸	Nej Ja

Noter:

- 1) Eleven skal have karakter i mindst 3 ud af 5 bundne prøver for at få beregnet et gennemsnit. De bundne prøver i 9. klasse består af hhv. skriftlig dansk (3 delprøver), mundtlig dansk, mundtlig engelsk, skriftlig matematik (2 delprøver) samt fællesprøven i fysik/kemi, biologi og geografi. I beregningen af det samlede resultat anvendes den officielle vægtning af de skriftlige delprøver i dansk og matematik.
- 2) Variablen tager højde for de særlige prøveregler i grundskolen som følge af Covid-19 situationen i perioden 2019/2020 til 2021/2022. I skoleåret 2019/2020 blev alle prøver aflyst, og eleverne fik i stedet ophøjet deres afsluttende standpunktskarakterer til prøvekarakterer. I skoleåret 2020/2021 og 2021/2022 blev prøverne delvist aflyst. Der blev således kun afholdt prøver i de bundne prøver i dansk, matematik og engelsk i 9. klasse. For de aflyste prøver fik eleverne, ligesom i skoleåret 2019/20, ophøjet deres afsluttende standpunktskarakterer til prøvekarakte-

rer. For de afholdte prøver blev det endvidere besluttet af regeringen og et bredt flertal i Folketinget, at eleverne fik ophøjet deres afsluttende standpunktskarakter til prøvekarakter, hvis standpunktskarakteren var højere end den opnåede prøvekarakter. Derudover var der ved de ikke-aflyste prøver mulighed for at undlade at formulere spørgsmål til dele af pensum, som alene var gennemgået under nødundervisning uden fysisk fremmøde.

- 3) Der skelnes mellem indvandrere, efterkommere og personer med dansk oprindelse. Personer med dansk oprindelse (dansk) er personer, hvor mindst en af forældrene er dansk statsborger og født i Danmark. Efterkommere er født i Danmark. Ingen af forældrene er danske statsborgere og født i Danmark. Når en eller begge forældre, der er født i Danmark, opnår dansk statsborgerskab, vil deres børn ikke længere blive klassificeret som efterkommere. Fastholder dansk-fødte 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 og født i Danmark, opfattes personen også som efterkommer. Indvandrere er født i udlandet. Ingen af forældrene er danske statsborgere og født i Danmark. Hvis der ikke findes oplysninger om nogen af forældrene, og personen er født i udlandet, opfattes personen som indvandrer.
- 4) Den højest fuldførte uddannelse blandt elevens mor og far oplyst ved start på uddannelsen. Hvis uoplyst, da indgår oplysningen fra det år, eleven fyldte 13 år.
- 5) Den gennemsnitlige bruttoindkomst for elevens forældre ved start på uddannelsen. Bruttoindkomst er her beregnet som personindkomst i alt ekskl. beregnet lejeværdi af egen bolig. Den valgte kodning er baseret på medianen og nedre og øvre kvartil af bruttoindkomsten blandt forældrene på årgangen i den elevpopulation, som de socioøkonomiske referencer for gymnasiekarakterer beregnes på baggrund af.
- 6) Familiestatus i året for start på uddannelsen er en kombination af familietype og plads. Familietypen par indeholder ægtepar, registreret partnerskab, samlevende og samboende par. Plads er enten hjemmeboende eller udeboende.
- 7) Seneste grundskoletype (9. klasse eller afsluttet 10. klasse) inden start på uddannelsen.
- 8) Gennemført 10. klasse eller ej.



BØRNE- OG
UNDERVISNINGSMINISTERIET
STYRELSEN
FOR IT OG LÆRING