



Metodenotat for værktøjet 'Skolefamilier'

Dette notat beskriver de metodiske overvejelser, der er gjort forud for gruppering af folkeskolerne i klynger med samme rammevilkår, også kaldet skolefamilier. Det er kun folkeskoler, der grupperes i skolefamilier.

Analysens variable

Data er registerbaserede oplysninger på skoleniveau vedrørende elevernes baggrundsforhold i skoleårene 2013/2014-2015/2016 fra Danmarks Statistiks elevregister. Ved at anvende data fra en 3 årig periode sikres det, at klyngerne er mere stabile i forhold til ændret elevsammensætning fra år til år på de enkelte skoler.

Skoler, som ikke har elever i skoleåret 2015/2016, er ikke medtaget i analysen. Skoler, hvor mere end 10 procent af eleverne har manglende oplysninger om forældrenes højeste fuldførte uddannelse eller gennemsnitlige bruttoindkomst, er heller ikke medtaget i analysen.

Inden selve klyngeanalysen er foretaget, bliver skolerne inddelt i undergrupper, de såkaldte strata. Skolerne stratificeres efter:

- Antal klassetrin på skolen
- Andel ikke-vestlige indvandrere på skolen

Stratificeringen sikrer, at skolerne kun grupperes med andre skoler, der næsten har de samme klassetrin på skolen, dvs. en skole, som har eleverne indtil 6. klasse, bliver kun grupperet med skoler, som også har eleverne til og med 6. klassetrin. Endvidere stratificeres der efter, hvor stor en andel ikke-vestlige indvandrere der er på skolen. Her er der valgt en inddeling i forhold til medianen på landsplan: Under 4 procent. ikke-vestlige indvandrere og over 4 procent. ikke-vestlige indvandrere.

Følgende variable er udvalgt til konstruering af klyngerne:

1. Forældrenes højeste fuldførte uddannelse, består af to variable:
 - a. hvor stor en andel af forældrene, der har grundskolen eller en gymnasial uddannelse som højeste fuldførte uddannelse.
 - b. hvor stor en andel af forældrene, der har en mellemlang-, lang videregående uddannelse eller en Ph.d. som højeste fuldførte uddannelse.

Styrelsen for It og Læring
Center for Data og Analyse

Olof Palmes Allé 38
8200 Aarhus N
Tlf. nr.: 89 37 66 66
E-mail: stil@stil.dk
www.stil.dk
CVR nr.: 13223459

6. februar 2017
Sags nr.:17/01371

2. Forældrenes bruttoindkomst, hvor stor en andel af forældrene, der har en gennemsnitslig bruttoindkomst på under 268.865 kr. om året.¹
3. Elevernes herkomst, hvor stor en andel af eleverne, der er af ikke-vestlig herkomst på skolen².
4. Skolestørrelse, målt som det gennemsnitslige antal elever pr. klassetrin på skolen over en 3 årig periode.
5. Specialklasseelever, hvor stor en andel af eleverne på skolen der er specialklasseelever i skoleåret 2015/2016.

De valgte variable har alle en stor sammenhæng med mange af de registreringer, der foreligger om eleverne i grundskolen: karakter fra folkeskolens prøver, elevernes trivsel, elevernes fravær mm. Det vil derfor ofte være interessant for skoler, at kunne sammenligne sig med andre skoler med tilnærmedesvis den samme elevsammensætning. Da skolerne økonomiske randbetingelser påvirkes af de enkelte kommuner, ses disse oplysninger ikke i denne forbindelse som et rammevilkår, og indgår derfor ikke som et parameter i konstruktionen af skolefamilierne.

Den statistiske metode

Den statistiske metode til gruppering af folkeskoler i skolefamilier med samme rammevilkår, der anvendes i analysen, er klyngeanalysen. Klyngeanalyse er en metode til at gruppere skoler på baggrund af mønstre i data i forhold til skolernes særegne karakteristika på de udvalgte variable. En skolefamilie er en gruppe af skoler, der internt ligner hinanden mest muligt og eksternt adskiller sig mest muligt fra skoler i andre skolefamilier. Variablene, der indgår i den statistiske analyse, beskriver skolernes rammevilkår og kan derfor ikke påvirkes direkte af skolerne.

Klyngeanalysen konstruerer klynger på baggrund af distancedata. Inden selve distancemålene beregnes og klyngeanalysen kan gennemføres, skal klyngeanalysens variable standardiseres og transformeres i proceduren Distance i SAS®. Variable med stor varians har en tendens til at have større betydning for skabelsen af klyngerne end variable med lille varians. Variablene, der anvendes til dannelsen af klyngerne, er ratio- og intervalskalerede og derfor bruges metoden STD til at standardisere de intervalskalerede variable og ABSMAX bruges til at standardisere de ratioskalerede variable. Metoden EUCLID anvendes også i proceduren Distance til beregningen af distancemålene, så data kan anvendes i klyngeanalysen.

Til konstruering af klyngerne anvendes proceduren Modeclus i SAS®. Metoden, som er en variant af two-stage clustering, kan bruge forskellige algoritmer baseret på nonparametriske tæthedsestimater, hvor data kan være numeriske koordinater og distancer. Metoden udfører omtrentlige signifikanstest for klyngerne og sammenlægger hierarkisk nonsignifikante klynger i forskellige step: først estimeres tæthedsestimaterne via fixed-

¹ 268.865 kr. om året er 25 percentil-indkomsten for forældrene til de elever, der indgår i analysen.

² Denne variabel anvendes først til stratificering og efterfølgende i klyngeanalyserne for begge grupperinger; Under 4 procent. ikke-vestlige indvandrere og over 4 procent. ikke-vestlige indvandrere.

radius uniform kernels. Indledende klynger skabes via en ”dal-søgende” metode. Så beregnes omtrentlige signifikansværdier for hver klynge ved at sammenligne den estimerede maksimum tæthed i klyngens midte med den estimerede maksimum tæthed på klyngens grænse. Til sidst samles de mindst signifikante klynger med en naboklynge indtil de tilbageværende klynger er statistisk signifikante. Signifikanstestene valideres via simulationer med stikprøver fra 20 til 2000 observationer.

Metoden er egnet til eksplorativ klyngeanalyse, hvor der ikke opstilles noget krav til antallet af klynger, der skal skabes. I stedet defineres et ”smoothing parameter”, som bestemmer det mindste antal observationer, som en skole skal sammenlignes med og danne en klynge med. Antallet af klynger er en funktion af smoothing parameteren. Antallet af klynger har en tendens til at falde, når smoothing parameteren stiger. I valget af smoothing parameter stiles efter, at én skolefamilie indeholder ca. 20-40 skoler.