Bacheloroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

Vegard J. Løwe, Herman M. Blom, Lars M. Johnsen

Impliserer mindre klasser bedre skoleresultater?

En kvantitativ studie av norske elever på 4. trinn

Trondheim 15. mai 2020

NTNU
Kunnskap for en bedre verden

Innhold

1 Innledning

1.1 Motivasjon

Store klasser og få lærere per elev blir ofte brukt for å forklare svake elevprestasjoner. Både elever, foreldre og lærere argumenterer for at mindre klasser gir mer tid til hver elev, og at dette resulterer i bedre elevresultater. “Klassestørrelse har noe å si for læring”, er tittelen på et Si;D-innlegg av Stella Ramborg - 14 år [1]. Klassestørrelse er et av virkemidlene myndighetene har i skolepolitikken og det vakte stor debatt da regjeringen i 2004 fjernet kravet om maks 28 elever i hver klasse [2].

Politikere har følgelig hengt seg på denne debatten. I 2017 skrev Audun Lysbakken et debattinnlegg hvor han beskylder høyresiden for å benekte at klassestrørrelse har noe å si for elevprestasjoner. “Likevel hevder høyresiden at klassestørrelsen ikke har noe si. Alle som har kontakt med barn, vet at det har noe å si om det er 15 eller 30 elever per lærer.” [3]. Utsagnet står ikke uimotsagt, og i et debattinnlegg med tittel “Klassestørrelse har ingenting å si” [4] referer Høyres Mats Kirkebirkeland til en stor undersøkelse utgitt av SSB [5].

Etter mye diskusjon, vedtok regjeringen en lærernorm i grunnskolen som sa at fra og med august 2018 skulle det være maksimalt 16 elever per lærer fra 1. til 4. trinn. I 2019 ble denne grensen senket ytterligere til 15 elever per lærer [6]. Dette er et tiltak som medfører store utgifter. I 2018 ble det bevilget 1,4 milliarder kroner for å øke lærertettheten. Kostnadene ved mindre klasser er også noe Kirkebirkeland er inne på i sitt debattinnlegg. Er det riktig å bruke så mye penger på noe man er usikker på om fungerer?

1.2 Problemstilling

Finnes det en sammenheng mellom klassestørrelse og elevprestasjoner hos norske elever på 4. trinn?

I denne oppgaven skal vi gå nærmere inn på det politikere og andre har diskutert lenge: Er det grunnlag for å si at klassestørrelsen påvirker prestasjonen til en elev? Vil man kunne oppleve økte resultater ved å redusere klassestørrelsen?

Ideelt ville vi gjennomført et eksperiment der elevene hadde blitt tilfeldig plassert i små og store klasser. Et slikt eksperiment er svært kostbart og vanskelig å gjennomføre. Vi må derfor, som mange andre forskere, basere oss på spørreundersøkelser.

Tilleggsproblemstilling: Hvilke andre faktorer påvirker elevresultatene, og bør debatten dreies vekk fra klassestørrelse?

2 Teoretisk rammeverk og tidligere studier

2.1 Innledning

Som vi var inne på under motivasjonen, er det allerede gjennomført flere studier for å kartlegge om det er noen signifikant forskjell mellom skoleprestasjonene til elever i små og store klasser. I denne delen av oppgaven skal vi diskutere og undersøke tidligere studier, samt presentere og forklare skoleproduktfunksjonen.

2.2 Teoretisk rammeverk

Elevprestasjonene kan representeres ved Bonesrønnings skoleproduktfunksjon [7, s. 16]. Skoleproduktfunksjonen er en metode for å presentere skole-data og legger til rette for empirisk forskning og analysering av denne. Funksjonsformen til skoleproduktfunksjonen er "value added", noe som vil si at den ser på hvordan summen av effektene fører til et målbart resultat i form av testscore. Modellen vil bli brukt til å finne spesifikke elementer som påvirker en enkelt elevs prestasjon. Skoleproduktfunksjonen gir oss testscore (T) som en funksjon av elevkarakteristikker (F), medelevkarakteristikker (P) og skoleinnsatsfaktorer (S).

$$T = f(F, P, S) \tag{1}$$

Under skoleinnsatsfaktorer (S) finnes lærerkarakteristikker og skolekarakteristikker. Lærerkarakteristikk avhenger som oftest av utdanningsnivå, erfaring, kjønn og etnisitet, mens skolekarakteristikk trekker inn faktorer som klassestørrelse og administrative utgifter.

Innenfor elevkarakteristikk (F) finner man blant annet familiebakgrunn. Det er fordi elevprestasjonen ofte er avhengig av foreldrenes utdannelse, inntekt og størrelse på familien. Medelevkarakteristikk (P) trekker inn faktorer som skolemiljø og trivsel i læringsmiljøet.

Det mest kontroversielle spørsmålet i skoleproduktfunksjonslitteraturen dreier seg om i hvilken grad klassestørrelsen påvirker elevresultatene [7, s. 16]. Klassestørrelse er en av de største faktorene for driftskostnadene til en skole. Effekten av den er derfor sentral å undersøke for å finne ut hva som er mest nytte-kostnadseffektivt i skolepolitikken.

2.3 Litteratur og tidligere funn

I årene 1985-1989 ble det gjennomført et prosjekt av Department of Education i Tennesse kalt «STAR-project» [8]. Her ble over 6000 elever fra 19 forskjellige skoler tilfeldig plassert i klasser med ulike lærere. De ble plassert i små klasser (13-17 personer), normal klassestørrelse (22-25 personer) og normal klassestørrelse med ekstra hjelp. Elevene ble testet med «the Stanford Achievement Test» eller kort: SAT, for å kartlegge matematikk- og leseferdigheter. Testingen viste at elever i små klasser fikk bedre resultater både i matematikk og lesing. Forskjellen mellom de små og store klassene vedvarte relativt konstant over årene elevene ble testet [8, s. 3]. I et oppfølgingsstudie gjennomført på STAR-elevene flere år senere kunne det vises fordelaktige langtidseffekter av å ha små klasser. Resultatene viste at elever i normale klasser hadde 7% lavere fullføringsgrad enn elevene i små klasser.

En annen studie, gjennomført av forskningsbiblioteket Campbell Collaboration Library, viser at klassestørrelse har lite å si for elevprestasjonene [9]. I forskningsprosjektet har det blitt gransket 127 ulike studier fra 41 land. Resultatet viser at elever i små klasser presterer noe bedre i lesing, men til gjengjeld dårligere i matematikk. Et problem som fremheves er at flesteparten av de undersøkte studiene har feilkilder, og kun et lite antall kan regnes å ha god nok kvalitet. En feilkilde kan være at ulike faktorer ikke blir tatt høyde for i undersøkelsen, for eksempel trivsel, som spiller inn på personlig utvikling.

Leuven og Løkken gjennomførte i perioden 1978 til 2003 en studie der fokuset lå på langtidseffekter av små klasser [5]. Studien undersøkte om antall elever i klassen på barne- og ungdomsskolen kan påvirke videre karriere og inntekt. I studien ble elevene fulgt over en 25-årsperiode fra de gikk ut av skolen og inn i arbeidslivet. Resultatene viste ingen positive langtidseffekter av små klasser. Disse funnene ser heller ikke ut til å påvirkes av verken foreldrebakgrunn, skole eller kommunekjennetegn.

I likhet med studien utført av Leuven og Løkke viser en studie gjort av Snoen og Grindheim i 2015 [10] at redusert klassestørrelse gir blandede resultater og svært liten effekt. Det konkluderes også med at skoleresultatene ikke kan forsvare kostandene ved å redusere klassestørrelsen.

2.4 Oppsummering

I dette kapittelet har vi presentert og forklart hva som inngår i skoleproduktfunksjonen. Det har blitt vist til flere tidligere studier som har tatt for seg effekten av klassestørrelsen, og det er tydelig uenighet rundt denne.

3 Empirisk strategi

3.1 Innledning

I dette kapittelet presenteres vår empiriske strategi. Her vil funksjonsformen og estimeringsmetoden for parameterne presenteres. Den matematiske fremgangsmåten i minste kvadraters metode og ulike tester vil bli forklart i korte trekk.

3.2 Empirisk strategi

I kapittel 2 ble skoleproduktfunksjonen presentert. Den blir som nevnt brukt til å representere elevprestasjoner ved hjelp av inngående variabler fra S, P og F. For å avdekke den sanne sammenhengen mellom klassestørrelse og testresultater, er det viktig å inkludere alle effektene som kan spille inn og føre til bedre testresultater. Det vil derfor bli valgt ut flere variabler vi mener kan ha en effekt innenfor gruppene S, F og P, i skoleproduktfunksjonen (1). Slike variabler kalles kontrollvariabler og er noe vi kommer tilbake til i kapittel 4. De inkluderes i regresjonsanalysen for å utelukke at en eventuell sammenheng mellom avhengig variabel og interessevariabel skyldes utelatte variabler i analysen. Det ville i så fall vært en spuriøs sammenheng, som er ønskelig å unngå. Avhengig variabel i modellen er elevresultater, mens interessevariabel er klassestørrelse. Med andre ord inkluderer vi kontrollvariabler for å utelukke at en eventuell sammenheng vi finner avhenger av noe annet enn klassestørrelse.

3.2.1 Funksjonsform

Vi velger en lineær approksimasjon av (1) og kommer senere til å diskutere en ikke-lineær utvidelse.

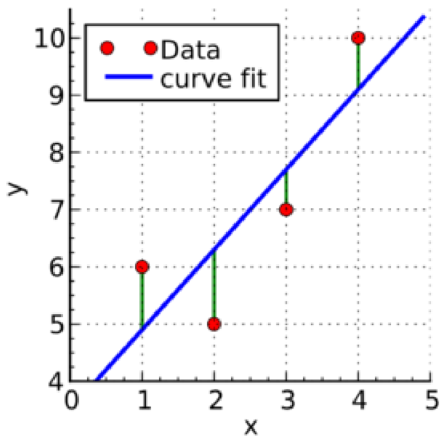
$$\text{testscore} = \beta_1 + \beta_2 \text{klassestørrelse} + \beta_3 X + \varepsilon \tag{2}$$

Her er X et sett av kontrollvariabler med tilhørende koeffisientvektor β. Kontrollvariablene blir presentert senere i oppgaven. Det siste leddet, ε, er et stokastisk restledd som representerer alle de variablene som påvirker elevresultater, men som ikke er spesifisert i ligningen.

3.2.2 Estimeringsmetode

For å finne effekten av klassestørrelse er det interessant å estimere parameteren β_2 i ligning (2). Ulike hypotesetester av denne er også interessant. Vi ønsker å undersøke hvordan estimatet av β_2 endres når vi inkluderer ulike sett av kontrollvariabler. Enkelt forklart vil en β_2 rundt null symbolisere svak eller ingen effekt av klassestørrelse på elevresultater, motsatt vil en høy $|\beta_2|$ symbolisere sterk effekt.

Estimeringsmetoden som brukes er minste kvadraters metode. Mer konkret vil vi bruke multippel lineær regresjon [11, s. 386-396], da modellen vår inkluderer en avhengig variabel samt flere uavhengige variabler. Ideen er å tilpasse rett funksjonsform til datasettet.



Figur 1: Illustrasjon av minste kvadraters metode (OLS)

Fremgangsmåten i modellen er forholdsvis matematisk, men det grunnleggende blir her forklart i korte trekk:
Som det fremgår av navnet på metoden ønsker vi å finne den linjen som minimerer kvadratene skapt av residualer, heretter kalt S.

$$S = \sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i})^2 \tag{3}$$

Vi ønsker å minimere denne likningen med hensyn på β_1, β_2 og β_3. Altså finne de koeffisientvektorene som gir oss minst S. Dette er en matematisk prosess som innebærer å partiellderivere S med hensyn på β_1, β_2 og β_3. Der de deriverte er lik null, vil vi finne minimumspunktet så fremt den dobbeltderiverte er positiv. β_1 estimeres direkte fra $\frac{\partial S}{\partial \beta_1}$.

For å estimere β_2 og β_3 blir det benyttet algebraiske metoder slik at man omsider ender opp med et likningssett som kan løses ved Cramers regel.

Vi antar at feilleddet ε har forventningsverdi lik null, varians lik σ^2 og er normalfordelt. Da vil estimatene av β_1 og β_2 være forventningsrette og variansminimale. Det vil si at forventningsverdien til estimatorene er lik det vi ønsker å estimere, samt at de har lavere varians enn alle andre forventningsrette estimatorer. Siden vi antar at ε er normalfordelt, kan vi teste hypoteser på parameterne i ligning (2).

3.2.3 Tester

Underveis i oppgaven kommer det til å bli brukt ulike tester for å avdekke sammenhenger i datasettet og regresjonsmodellene. Testene som senere skal brukes og derfor blir forklart her, er: T-test, F-test, Chow-test, Reset-test og VIF-test.

En **T-test** er en statistisk hypotesetest basert på Students t-fordeling [11, s. 158-159]. Den gjennomføres ved at det settes inn estimert verdi for $\bar{X}$ og $\hat{\sigma}$, og tester med verdi for μ. Variabelen $\bar{X}$ er gjennomsnittverdien og $\hat{\sigma}$ er standardavviket (forklart i 5.2.3) til målt parameter. Til slutt sjekkes det om verdien er signifikant eller ikke ved hjelp av student-t fordelingen.

$$t = \frac{Z}{s} = \frac{\bar{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \tag{4}$$

En **F-test** kan brukes for å sjekke signifikans for en enkel variabel eller som en sammensatt hypotesetest med flere variabler [11, s. 415-416]. Et eksempel på en F-test er vist under, hvor nullhypotesen kan være både én eller flere variabler:
H_0: $\beta_1 = 0$,
H_0: $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$.
H_1 vil derimot alltid være å sjekke om minst en variabel har en signifikant effekt på Y (regresjonsvariabelen). Det betyr at dersom det kun testes en variabel blir testen lik som en vanlig t-test. Dersom flere variabler testes samtidig, kan det ikke avdekkes hvilken variabel som gir signifikans, det kan være alle eller kun én. F-test er derfor et nyttig verktøy for å avvise signifikans.

Chow-test kan brukes for å avdekke splittelser i datasettet [12, s. 223]. Et eksempel på dette kan være en regresjonsmodell for boligpriser som inkluderer et signifikant boligprisfall. Her kan det vært lurt å splitte datasettet i to modeller; en modell før og en modell etter boligprisfallet.

Testen gjennomføres ved at man antar to grupper, g=1 og g=2. Det er ønskelig å teste om det er en sammenheng mellom krysningspunkt og stigningstall for de to gruppene. Modellen skrives slik:

$$y = \beta_{g,0} + \beta_{g,1} \cdot x_1 + \beta_{g,2} \cdot x_2 + \cdots + \beta_{g,k} \cdot x_k + u \tag{5}$$

Nullhypotesen er at hver β_i i ligning (5) er identiske mellom de to gruppene. Det blir så brukt en F-test mellom modellene for å måle signifikansnivå.

(Ramsey) Reset-test er en test for feilspesifikasjon i funksjonsform [12, s. 227]. Den tar utgangspunkt i en lineær regresjonsmodell og tester om det finnes en ikke-lineær modell, for eksempel en polynom-form, som kan forklare responsen i dataene bedre.

Testen fungerer ved at det først defineres en enkel lineær regresjonsmodell:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \tag{6}$$

Denne modellen skal testes mot en ikke-lineær regresjonsmodell:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u \tag{7}$$

La $\hat{y}$ være de estimerte variablene fra ligning (6). Variablene: $\hat{y}^2$ og $\hat{y}^3$ kan ses på som funksjoner av x_i-ene. Ligning (7) blir videre testet mot ligning (6) for å avdekke om den mangler viktige ulineariteter.

Det blir så gjennomført en F-test som tester H_0: $\delta_1 = 0$, $\delta_2 = 0$, i ligning (7). Et signifikant resultat betyr at det kan være problemer med funksjonsformen. Testen finner altså en form for manglende perfeksjon, men den forteller imidlertid ikke noe om hvilken modell som kunne passet bedre. Dette er en ulempe ved å bruke Ramsey-testen.

Variance Inflation Factor, eller kort **VIF**, er en test som måler graden av multikollinearitet mellom variabler [12, s. 86]. Dette betyr at det finnes en lineær sammenheng (korrelasjon) mellom kun to forklaringsvariabler, med andre ord: De forklarer hverandre og ikke datasettet. Multikollinearitet kan bli et problem for analyser dersom VIF-testen gir verdier over 5 [13, s. 756]. VIF beregnes fra ligningen:

$$VIF_j = \frac{1}{1-R_j^2} \tag{8}$$

Determinasjonskoeffesienten R^2 er et mål på hvor godt parametrene i modellen sammenfaller i regresjonen [11, s. 273-274] [12, s. 83-86]. Tallet man får fra R^2 kan ligge mellom 0 og 1. $R^2 = 1$ betyr at modellen forklarer all variasjon i Y, altså at alle variablene ligger på regresjonslinja. En $R^2 = 0$ vil si at modellen ikke forklarer noe av variasjonen i Y. R^2 er definert som totalvarians (SST) over forklart varians (SSE):

$$R^2 = \frac{SST}{SSE}$$

Total varians er et mål på den totale variasjonen og viser hvor spredt ut Y_i er i datasettet. Hvis man deler SST på $n-1$ får man varians i Y. Total variasjon er definert som:

$$SST = \textstyle\sum_{i=0}^{n}(Y_i - \bar{Y})^2$$

Forklart varians måler varians i $\hat{Y}_i$. Altså hvordan x forklarer variansen i $\hat{Y}_i$. Forklart varians er definert som:

$$SSE = \textstyle\sum_{i=0}^{n}(\hat{Y}_i - \bar{Y})^2$$

3.3 Oppsummering

I dette kapittelet har vi presentert funksjonsformen og gjort kort rede for hvordan vi estimerer parameterne i analysen. Altså hvordan vi vil estimere effekten av de inkluderte variabelene på elevresultater. Det har også blitt presentert og forklart ulike tester som vil bli brukt for å analysere resultatene i kapittel 5.

4 Datamateriale

4.1 Innledning

I dette kapittelet presenteres datamaterialet som er brukt for å svare på problemstillingen. Dette innebærer kort informasjon om datasettet før vi definerer ulike kontrollvariabler. Det blir presentert deskriptiv statistikk for interesse- og avhengig variabel, før vi gjør rede for hvilke kontrollvariabler vi tar med oss videre til den empiriske analysen. Korrelasjonsmatrisen for datamaterialet blir også kommentert.

4.2 Om datamaterialet

Datasettet brukt i denne oppgaven er hentet fra PIRLS-undersøkelsen gjennomført i 2001. PIRLS (Progress in International Reading Literacy Study) [14] er en internasjonal leseundersøkelse av leseferdigheter blant elever på 4. trinn. Helt siden 1991 har IEA (The International Association for the Evaluation of Educational Achievement) vært ansvarlig for den internasjonale leseundersøkelsen. PIRLS omfatter 150 000 elever fra 5777 skoler i 35 ulike land. I denne studien vil kun data fra Norge bli benyttet .

For å kunne utføre statistisk inferens for lineær regresjon, er det blitt antatt normalitet i dataene. Ved store data, kan det bevises ved hjelp av sentralgrenseteoremet at distribusjonen til dataen følger en normalfordeling [12, s. 106-107]. Vi antar derfor normalitet for vår data. Dette er også vist i Figur 2, der testresultatene er plottet mot antall elever med samme testresultat.

For å kunne utføre lineær regresjon, antar man at residualene har konstant varians, dvs at man antar homoskedasitet. Etter å ha utført en Breach Pagan-test, oppdaget vi heteroskedasitet. Vi forkastet derfor nullhypotesen for homoskedasitet og brukte robuste standardavvik i våre regresjonsmodeller.

Variabel	Beskrivelse
read:	Testscore i leseundersøkelsen
clsize:	Antall elever i klassen
teacher_exp:	Antall år læreren har jobbet
schoolsize4:	Antall 4.klasselever på skolen
par_edu_low (dummy):	Foresatt med høyest utdanning har kun fullført ungdomsskole = 1
	Høyest utdannet foresatt har høyere utdanning = 0
par_edu_med (dummy):	Foresatt med høyest utdanning har fullført videregående = 1
	Høyest utdannet foresatt har lavere eller høyere utdanning = 0
par_edu_high (dummy):	Foresatt med høyest utdanning har universitetsgrad= 1
	Høyest utdannet foresatt har lavere utdanning = 0
pc_class (dummy):	PC tilgjengelig i klasserommet = 1
	PC ikke tilgjengelig i klasserom = 0
kinderg_att (dummy):	Vært i barnehage = 1
	Ikke vært i barnehage = 0
par_not_born (dummy):	Foresatte ikke født i Norge = 1
	Foresatte født i Norge = 0
sameteacher4_plus (dummy):	Elevene har hatt samme lærer siden de begynte på skolen = 1
	Elevene har ikke hatt samme lærer siden de begynte på skolen = 0
teacher_fem (dummy):	Kvinnelig lærer = 1
	Mannlig lærer = 0
pct_disadv_10 (dummy):	Flere enn 10% av elevene på skolen kommer fra økonomisk svake hjem = 1
	Færre enn 10% av elevene på skolen kommer fra økonomisk vanskeligestilte hjem = 0

Tabell 1: Definisjoner i datasettet

4.2.1 Forklaring av kontrollvariabler

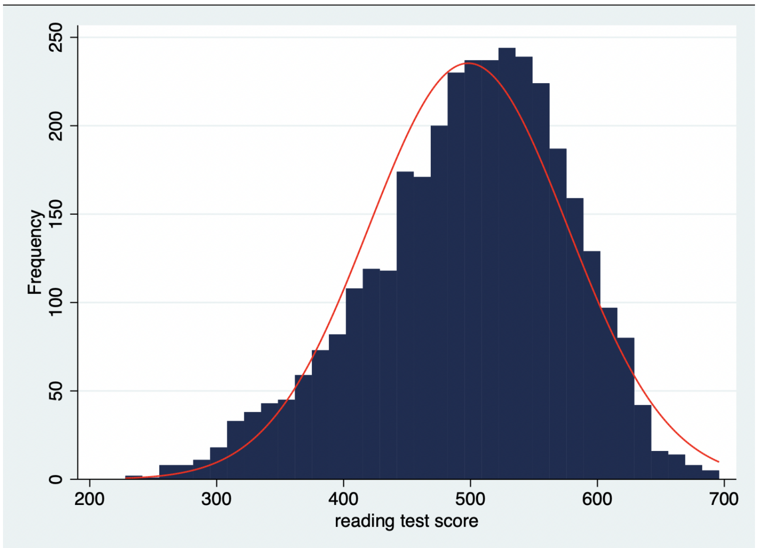
Forklaringsvariabler blir valgt ut fordi de kan ha en direkte påvirkning på leseresultatene samt en indirekte påvirkning via klassestørrelsen. Det vil si at forklaringsvariabelen kan føre til høyere eller lavere leseresultater, samtidig som den kan ha en medvirkende effekt på hvordan en klasse fungerer sammen på bakgrunn av klassestørrelsen. Kontrollvariablene inkluderes for å minimere en uunngåelig utelatt variabelskjevhet fra regresjonen.

I skoleproduktfunksjonen (1) blir det referert til tre ulike variabelkategorier, lærerkarakteristikk (S), elevkarakteristikk (F) og medelevkarakteristikk (P). Variablene teacher_exp, teacher_fem, schoolsize4, sameteacher4_plus og pc_class faller inn under kategorien lærer-karakteristikk. I kategorien elevkarakteristikk kommer variablene: par_edu, par_not_born og kinderg_att. Til slutt finner vi variabelen pct_disadv_10 under medelevkarakteristikk.

4.3 Deskriptiv statistikk for avhengig variabel

Variabel	Observasjoner	Gjennomsnitt	Standardavvik	Minimum	Maksimum
read	3 459	498,2563	78,36616	228,0606	695,8717

Tabell 2: Deskriptiv statistikk, testscore (read)



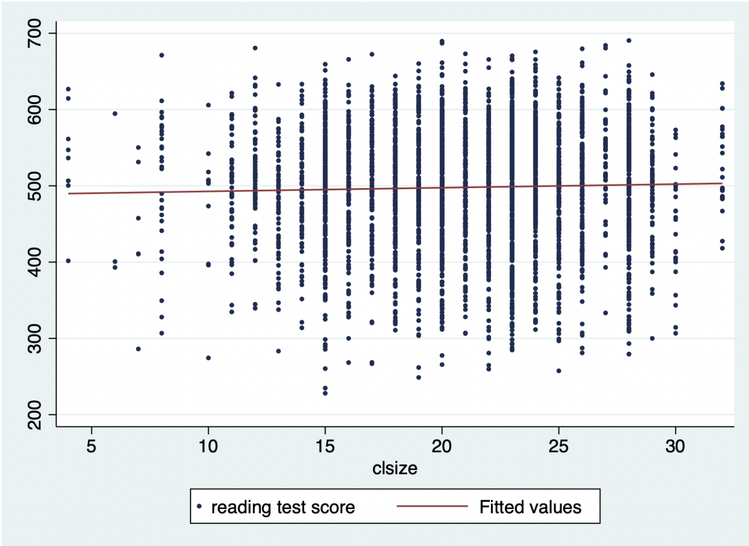
Figur 2: Fordeling av avhengig variabel testscore (read).

Fordelingen viser normalitet i dataplottene slik vi antok utifra sentralgrenseteoremet i avsnitt 4.2

4.4 Deskriptiv statistikk for interessevariabel

Variabel	Observasjoner	Gjennomsnitt	Standardavvik	Minimum	Maksimum
clsize	3 416	20,96165	4,82684	4	32

Tabell 3: Deskriptiv statistikk av klassestørrelse (clsize)



Figur 3: Kryssplot av testscore (read) mot klassestørrelse (clsize) med enkel regresjonslinje

read	Koeffisient	Standardavvik	t	P >| t |	[95% Konf.	Intervall]
clsize	0,4787098	0,2775249	1,72	0,085	-0,0654218	1,022841
_cons	487,9291	5,969574	81,74	0,000	476,2248	499,6334

Tabell 4: Lineær regresjon mellom avhengig variabel (read) og interessevariabel (clsize)

Figur 3 viser at testscore (*read*) avhenger positivt av klassestørrelse (*clsize*). Som det fremgår av Tabell 4 er koeffisienten kun 0.47. Det betyr at en økning på en elev i klassen svarer til 0.47 bedre score på testen. Det er veldig lite tatt tatt i betraktning av at gjennomsnittsscoren er på 498. Sammenhengen kan heller ikke tolkes som en kausal sammenheng, siden det er grunn til å tro at klassestørrelse bare er en av mange faktorer som kan påvirke elevprestasjonere. Dermed er det fornuftig å studere situasjonen med kontrollvariabler.

4.5 Kontrollvariabler

For å gjennomføre en god regresjonsanalyse, er det essensielt å ha forståelse av kontrollvariablene som kan påvirke analysen. Ved å studere sammenhengen mellom testscore og klassestørrelse med utgangspunkt i minste kvadraters metode (OLS), ble det avdekket en svak-positiv trend med økende klassestørrelse. Denne effekten kan potensielt påvirkes av andre variabler som ble beskrevet i seksjon 4.2.

For å sikre at sammenhengen mellom aktuell kontrollvariabel og avhengig variabel er tilstrekkelig troverdig, er det avgjørende å kontrollere signifikansnivået tilknyttet variablene. Det blir derfor gjennomført en ny minste kvadraterts metode, med alle kontrollvariablene inkludert, for å finne hvilke variabler som er signifikante. Øvre signifikansnivå settes her til 0,05, altså 5%. *par_edu_low* og *par_edu_high* måles ut i fra referansevariabelen *par_edu_med*.

I Tabell 5 er variablene merket med én, to eller tre stjerner på hhv. signifikansnivå 5%, 1% og 0,1%. Resultatet viser fem variabler med akseptert signifikans på 5%, disse er: **teacher_exp, par_edu_low, par_edu_high, pc_class, schoolsize4, kinderg_att og par_not_born.**

Det har også blitt gjennomført en VIF-test for å kontrollere for multikollinearitet mellom kontrollvariablene. Det vil si en lineær sammenheng mellom flere forklaringsvariabler i en multippel regresjonsmodell. Resultatet er presentert i tabell 8 i Appendix. For høy verdi av multikollinearitet kan øke regresjonsvariansen og gjøre koeffisientene ustabile. Den høyeste verdien i VIF-testen vår er 1,55, altså godt innenfor grensebetingelsen på 5.

	(1)
	read
clsize	-0.550
	(-1.52)
teacher_exp	-0.468***
	(-3.33)
schoolsize4	0.244*
	(2.51)
par_edu_low	-30.44***
	(-3.41)
par_edu_high	40.70***
	(12.81)
pc_class	-16.38***
	(-3.71)
kinderg_att	11.14*
	(2.46)
par_not_born	-43.48***
	(-6.15)
sameteacher4_plus	3.910
	(1.23)
teacher_fem	-4.605
	(-1.00)
pct_disadv_10	4.090
	(0.92)
_cons	499.3***
	(48.96)
N	2331
R^2	0.122

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tabell 5: Kontrollvariable

4.6 Deskriptiv statistikk for kontrollvariabler

Variabel	Observasjoner	Gjennomsnitt	Standardavvik	Minimum	Maksimum
teacher_exp	3 354	16,51014	11,08948	1	42
par_edu_low	3 098	0,0338928	0,1809825	0	1
par_edu_high	3 098	0,5429309	0,4982339	0	1
pc_class	3 415	0,8497804	0,3573389	0	1
schoolsize4	3 243	41,18409	21,41166	3	192
sameteacher4_plus	3 422	0,3959673	0,4891289	0	1
kinderg_att	3 137	0,8603762	0,3466516	0	1
par_not_born	3 374	0,0583877	0,2345098	0	1

Tabell 6: Deskriptiv statistikk for kontrollvariabler

4.7 Korrelasjonsmatrise for datamaterialet

Korrelasjonsmatrisen, tabell 9 i Appendix, viser korrelasjonen mellom kontrollvariablene, interessevariablene og avhengig variabel. Her har det blitt brukt parvis korrelasjon framfor ordinær korrelasjon. Det er fordi vi ønsker å inkludere all data, ikke kun komplette variabler uten data-mangler. Av korrelasjonsmatrisen ser vi at det ikke finnes noen spesielt korrelerte variabler. De mest korrelerte variablene er *read* og *par_edu_high*, som har en korrelasjon på 0,29.

4.8 Oppsummering

I denne seksjonen har datasettet og deskriptiv statistikk for aktuelle variabler blitt presentert. Behovet for kontrollvariabler ble tydeligjort av kryssplottet mellom avhengig variabel og interessevariabel, før vi valgte kontrollvariabler og begrunnet dette. Korrelasjonsmatrisen for datamaterialet ble også presentert og viste ingen spesielt korrelerte variabler. Gjennomført VIF-test viste sågar liten grad av multikollinearitet.

5 Empiriske resultater

5.1 Innledning

I denne delen av oppgaven skal det presenteres analytiske modeller for å trekke sammenhenger mellom interessevariabel, avhengig variabel og kontrollvariabler.

5.2 Resultater

Det er valgt ut seks ulike modeller. Modell (2) - (5) er lineære med ulike modifikasjoner av kontrollvariabler og (6) er kvadratisk for å skille ut eventuelle ikke-lineære tendenser.

5.2.1 Modeller

Lineær grunnmodell:

(1) read = β_1 + $\beta_2$clsize+ε

Lineærgrunnmodell med noen kontrollvariabler:

(2) read = β_1 + $\beta_2$clsize + $\beta_3$teacher_exp + $\delta_1$par_edu_low + $\delta_2$par_edu_high + ε

Lineærgrunnmodell med flere kontrollvariabler:

(3) read = β_1 + $\beta_2$clsize + $\beta_3$teacher_exp + $\delta_1$par_edu_low + $\delta_2$par_edu_high + $\delta_3$kinderg_att + $\delta_4$par_not_born + ε

Lineærgrunnmodell med alle kontrollvariabler:

(4) read = β_1 + $\beta_2$clsize + $\beta_3$teacher_exp + $\delta_1$par_edu_low + $\delta_2$par_edu_high + $\delta_4$kinderg_att + $\delta_4$par_not_born + $\beta_5$schoolsize4 + $\delta_6$pc_class + ε

Samme som (4), men her er *cl_size* delt opp i tre dummyvariabler som hver inneholder $1/3$ av datamaterialet. *Schoolsize*4 hadde tilnærmet null effekt og ble fjernet.

(5) read = β_1 +$\delta_1$clsize_19 + δ_2 clsize_20_23 + δ_3 clsize_24 + $\beta_2$teacher_exp + $\delta_4$par_edu_low + $\delta_5$par_edu_high + $\delta_6$kinderg_att + $\delta_7$par_not_born + $\beta_8$schoolsize4 + $\delta_9$pc_class + ε

Ikke-linearitet ved å ta med kvadratisk ledd i interessevariabelen, *clsize_sq*:

(6) read = β_1 + $\beta_2$clsize + $\beta_3$teacher_exp + $\delta_1$par_edu_low + $\delta_2$par_edu_high + $\delta_4$kinderg_att + $\delta_4$par_not_born + $\beta_5$schoolsize4 + $\delta_6$pc_class + $\gamma_1$clsize_sq + ε

5.2.2 Resultater

	(1)	(2)	(3)	(4)	(5)	(6)
	read	read	read	read	read	read
clsize	0.479	0.296	0.357	0.224		-0.171
	(0.278)	(0.280)	(0.281)	(0.325)		(1.888)
teacher_exp		-0.211	-0.208	-0.353^{**}	-0.262^{*}	-0.263^{**}
		(0.122)	(0.123)	(0.131)	(0.129)	(0.129)
par_edu_low		-35.34^{***}	-27.72^{**}	-27.84^{**}	-28.65^{**}	-28.56^{**}
		(8.252)	(8.521)	(8.886)	(8.637)	(8.620)
par_edu_high		42.07^{***}	40.80^{***}	41.07^{***}	40.22^{***}	40.37^{***}
		(2.792)	(2.822)	(2.939)	(2.944)	(2.845)
kinderg_att			7.567	8.178	7.948	7.955
			(4.153)	(4.244)	(4.199)	(4.206)
par_not_born			-42.28^{***}	-41.65^{***}	-42.43^{***}	-42.25^{***}
			(7.287)	(7.611)	(7.620)	(7.298)
schoolsize4				0.0537		
				(0.0751)		
pc_class				-10.54^{**}	-9.37^{*}	-9.42^{*}
				(4.060)	(3.943)	(4.073)
clsize_20_23					1.559	
					(3.359)	
clsize_24					4.902	
					(3.311)	
clsize_sq						0.0130
						(0.046)
_cons	487.9^{***}	477.7^{***}	473.2^{***}	484.9^{***}	487.5^{***}	487.1^{***}
	(5.954)	(6.499)	(7.636)	(8.954)	(6.962)	(19.2)
N	3416	2983	2912	2723	2892	2892
R^2	0.001	0.092	0.105	0.110	0.110	0.110

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tabell 7: Resultater fra ulike modeller

5.2.3 Kommentar til tabellen

Tabell 7 viser resultatet av en regresjon gjort på modellene (1) - (6). Tabellen viser de estimerte parameterne β, δ og γ for hver av modellene med tilhørende standardavvik i parentes. Som presentert i avsnitt 4.2, er det grunnet heteroskadisitet i datamaterialet valgt å bruke robuste standardavvik.

Standardavviket er definert som kvadratroten av variansen, og viser den gjennomsnittlige avstanden til gjennomsnittet for parameteren. Variansen kan forklares som usikkerheten til de estimerte β_i. Variansen er et mål på hvor mye en observert verdi for parameteren vil variere dersom man gjentar forsøket som den stokastiske variabelen er definert ut fra.

Stjernene i tabellen forteller noe om signifikansnivået til effekten av variablene. Én stjerne svarer til en p-verdi under 0,05, to stjerner til under 0,01 og tre stjerner til en p-verdi under 0,001. P-verdi er definert som sannsynligheten for at det vi observerer er sant, gitt nullhypotesen. I vårt tilfelle er nullhypotesen at koeffisienten er lik null. Det vil si at tre stjerner bak en variabel forteller oss at det er mindre enn 0,1% sjanse for at effekten av variabelen er lik null. Er p-verdien større enn 0,05 konkluderes det gjerne med at nullhypotesen ikke forkastes, og at det dermed ikke finnes en sikker statistisk signifikant sammenheng.

I modell (5) ble klassestørrelsen delt inn i tre nivåer. Liten klasse (0-19 elever), middels klasse (20-23 elever) og stor klasse (24+ elever). Regresjonen av denne modellen ble gjort med liten klasse som referansevariabel. Altså måler vi effekten av større klasser i forhold til en liten klasse.

Alle modellene har en relativ lav determinasjonskoeffesient R^2. Det betyr at de ikke kan brukes til å forklare variansen i Y, *read*. Forklaring av R^2 kan man finne i kapittel 3.2.3.

5.3 Kommentar til resultatene

Etter å ha inkludert kontrollvariablene i modellene (2) - (4), får vi en fortsatt positiv effekt av klassestørrelse (*clsize*) på elevresultater (*read*). Når alle kontrollvariabler er med, reduseres dog effekten med over 50% fra et allerede lavt nivå i den enkle modellen (1). Punktestimatet på β_2 i modell (4) er altså så lite som 0,224. Det vil si at en økning i klassestørrelse på fire elever innebærer en økning i testresultat på under ett poeng. Det er veldig lite når vi vet gjennomsnittsresultatet lå på nesten 500 poeng. En økning på fire elever er stor med tanke på at en tredel av alle elevene går i klasser på mellom 20 og 23 elever.

T-test av nullhypotesten $\beta_2 = 0$ i modell (4) konkluderer med at nullhypotesen ikke er i nærheten av å kunne forkastes. Den gir oss en p-verdi på hele 0,49, altså er det nesten 50% sjanse for at $\beta_2 = 0$.

Heller ikke i modell (5), hvor *clsize* er delt inn i tre dummyvariabler, finner vi en signifikant sammenheng. Punktestimatet på δ_3, som representerer effekten på lesetesten ved å gå i stor klasse, er på 3 poeng. En T-test avslører at p-verdien for denne observasjonen er så høy som 0,88. P-verdien for punktestimatet tilhørende δ_2 er også høyt; 0,40. Det vil si at vi heller ikke her er i nærheten av å kunne forkaste nullhypotesene $\delta_2 = 0$ og $\delta_3 = 0$.
Vi forventet at den positive avhengigheten mellom klassestørrelse og testresultater skulle reduseres når vi innførte kontrollvariabler, men de spesifikke punktestimatene på enkelte variabler var overraskende. Vi hadde ikke forventet at *teacher_exp* skulle ha negativ påvirkning på testresultatet. Økt lærererfaring tenkte vi skulle tilsi økte elevresultater, men punktestimatet på β_3 i modell (4) er på -0,352. Det betyr at en lærer med 20 års erfaring skulle tilsi over åtte poeng lavere testresultat sammenlignet med en nyutdannet lærer. En T-test med nullhypotese $\beta_3 = 0$ gir samtidig en p-verdi på minimale 0,007. Det fremgår av tabellen at denne effekten er signifikant ved signifikansnivå 1% (2 stjerner). I ettertid tenker vi dette kan skyldes at nyutdannete lærere har høyere motivasjon og mer effektive læringsmetoder.

Effekten av å ha PC tilgjengelig i klasserommet er interessant, den representeres ved δ_6 i modell (4). Punktestimatet her forteller oss at dersom PCer er tilgjengelig i klasserommet, vil det føre til redusert testscore på hele 10,54 poeng. Her kan også nullhypotesen $\beta_4 = 0$ forkastes ved signifikansnivå 1%. Punktestimatene på *kinderg_att*, δ_4 i modell (4) var som forventet, men her kan ikke en nullhypotese av $\delta_4 = 0$ forkastes på konvensjonelle signifikansnivåer. Likevel viser en T-test at p-verdien ved testen *kinderg_att* = 0 ble 5,4%.

Som forventet er foreldres utdanning og innvandrerbakgrunn det som påvirker leseresultatet i størst grad. Høyt utdannede foreldre betyr gjerne bedre tilgang til leksehjelp, mens utenlandske foreldre har ofte svakere norsk-kunnskaper. Disse faktorene er i modell (4) representert ved henholdsvis δ_1, δ_2 og δ_4. Som det fremgår av stjernene er δ_2 og δ_4 signifikante ved sterkeste konvensjonelle signifikansnivå; 0,1%. En T-test viser for begge at p-verdien faktisk er 0. Det er derfor svært sannsynlig at de påvirker resultatene på lesetesten. Vi merker oss også at forskjellen på lav og høy utdanning hos elevens foresatte er på betydelige 68 poeng.

Den ikke-linære regresjonsmodell (6) er en kvadratisk modell som inkluderer variabelen *clsize_sq*. Dette leddet er inkludert for å fange opp ikke-linære effekter, som for eksempel svingninger. En VIF-test gjort på modell (6), presentert i tabell 10 i Appendix, symboliserer at det kvadratiske leddet blir tilnærmet lineært. Med en VIF-verdi på over 40 for *clsize* og *clsize_sq* fanger testen opp multikollinearitet mellom disse variablene, altså at de forklarer hverandre. Dette kan også sees fra γ-verdien til *clsize_sq*, som er tilnærmet null.

Modell (6) er den eneste modellen som ser en negativ sammenheng mellom *clsize* og *read*, altså at at større klasser gir dårligere testresultater. Men som nevnt i avsnittet over, kan det være preget av den høye multikollineariteteten, og vi kan derfor ikke trekke noen konklusjoner av dette.

Det ble også gjennomført en Ramsey-test av regresjonsmodell (4). Testen ga en p-verdi på 0,43, og vi kan derfor fastslå at det ikke finnes indikasjoner på at en polynomisk funksjon forklarer datasettet bedre enn en lineær modell. Dette resultatet kom også frem i modell (6), som fikk påvist multikollinearitet av gjennomført VIF-test.

5.4 Oppsummering

Vi har her presentert og kommentert resultatene. Tabellen har blitt grundig forklart, og det har blitt vist til T-tester av interessante variabler for å tydeliggjøre signifikansen utover det stjernene i tabellen forteller oss. Effekten av ulike variabler har også blitt tydeligjort og sammenlignet med gjennomsnittscore. Til slutt konkluderte både VIF-testen på modell (6) og Ramsey-testen av modell (4) med at en polynomisk funksjon ikke representerer sammenhengene på en bedre måte enn den lineære funksjonen.

6 Konklusjoner og funn

I denne oppgaven har vi tatt utgangspunkt i en lineær modell mellom avhengig variabel testresultat og interessevariabel klassestørrelse. Ved å legge til ulike variabler som kan antas å påvirke testresultatet, har denne sammenhengen blitt analysert og følgende effekter blitt synliggjort:

- Finner ingen statistisk signifikant effekt av klassestørrelse på elevprestasjoner i 4. klasse. Observert effekt er også minimal. En økning i klassestørrelse på fire elever svarer til en økt testscore på i underkant av ett poeng, altså 0,2% av gjennomsnittsscore. Det kan altså ikke forventes økte elevresultater ved å redusere klassestørrelsen.

- Det er tydelig at andre faktorer påvirker mer enn klassestørrelse. Resultatene viser at PC tilgjengelig i klasserommet har en signifikant og betydelig negativ effekt på testresultatet. Dette kan være noe som burde bli diskutert og fokusert mer på i skolepolitikken.

- Foreldrenes utdanning og innvandringsbakgrunn er sterkt signifikant og det som påvirker testresultatet i størst grad.

7 Begrensninger og forslag til utvidelser

7.1 Begrensninger

Datasettet er begrenset og det er mulig at relevante elever ikke har vært med i undersøkelsen. Resultatene vil da ikke reflektere den totale situasjonen i Norge. Det er også kun inkludert data fra elever på 4. trinn, muligens er effekten av klassestørrelse sterkere hos yngre eller eldre elever.

Det er mulig å inkludere flere kontrollvariabler med bakgrunn i skoleproduktfunksjonen (1) for å redusere utelatt variabelskjevhet. Det er sannsynlig at flere variabler påvirker testresultatet både direkte og indirekte gjennom klassestørrelse.

Elevene er ikke delt inn i tilfeldige klasser. Det er sannsynlig at variabler som foreldres utdanning og innvandrerbakgrunn spiller inn når klassene og klassestørrelsen blir bestemt på skoler med mer enn én parallell på trinnet. Den svenske forskeren Björn Öckert mener studier som ser på effekten av klassestørrelse må ha en strategi for å kontrollere dette for å være troverdig [15].

7.2 Forslag til utvidelser

- Inkludere kontrollvariabler for distriktet skolen befinner seg.
- Inkludere elevdata fra flere trinn.
- Finne en strategi for å kontrollere for at klassene ikke er tilfeldig inndelt.

Referanser

[1] S. Ramborg, "Lærertetthet har mye å si for elevene." https://www.aftenposten.no/meninger/sid/i/XA2B/laerertetthet-har-mye-aa-si-for-elevene, Mai. 2015.

[2] T. Olsen, "Klassene i barneskolen eser ut." https://www.aftenposten.no/norge/i/kaGBk/klassene-i-barneskolen-eser-ut, Sep. 2014.

[3] A. Lysbakken, "Klassestørrelse har noe å si." https://www.bt.no/btmeninger/debatt/i/PVlnb/klassestoerrelse-har-noe-aa-si, Aug 2017.

[4] M. Kirkebirkeland, "Klassestørrelse har ingen ting å si." https://www.dagbladet.no/kultur/klassestorrelse-har-ingen-ting-a-si/67444441, Mar 2017.

[5] E. Leuven and S. A. Løkken, "Long term impacts of class size in compulsory school," 2017.

[6] A. Bakken, "Innføring av lærernorm i grunnskolen." https://www.regjeringen.no/no/aktuelt/nye-regler-innforing-av-larernorm-i-grunnskolen/id2606134/, Jun 2018.

[7] H. Bonesrønning, "Utforming av utdanningspolitikken," 2004.

[8] J. Boyd-Zaharias, "Project star," p. 6, 1999.

[9] A. Ringgaard, "Klassestørrelse har lite å si for hvordan elevene gjør det på skolen." https://forskning.no/barn-og-ungdom-skole-og-utdanning/stor-studie-klassestorrelse-har-lite-a-si-for-hvordan-elevene-gjor-det-pa-skolen/1294210, Feb. 2019.

[10] J. A. Snoen and J. E. Grindheim, "Lærertetthet og læringsresultater," 2015.

[11] R. L. Thomas, *Using Statistics in Economics*. Mc Graw Hill Education, 2005.

[12] J. M. Wooldridge, *Introductory Econometrics, A modern Approach*. CENGAGE Learning, 2016.

[13] M. Akinwande, H. Dikko, and A. Samson, "Variance inflation factor," 2015.

[14] F. E. T. Ragnar Gees Solheim, “En norsk kortversjon av den internasjonale rapporten om 10-åringers lesekunnskaper,” 2003.

[15] J. Jelstad, “Joda elever i små klasser lærer mer.” https://www.utdanningsnytt.no/joda-elever-i-sma-klasser-laerer-mer/188635, Apr 2015.

8 Appendix

8.1 VIF-test for valg av kontrollvariable

Variable	**VIF**	**1/VIF**
schoolsize4	1,55	0,646277
clsize	1,50	0,666874
teacher_exp	1,14	0,876965
pc_class	1,12	0,896265
par_edu_high	1,10	0,905376
sameteacher4_plus	1,08	0,928927
teacher_fem	1,06	0,943299
pct_disadv_10	1,06	0,945127
par_edu_low	1,06	0,947839
kinderg_att	1,03	0,971245
par_not_born	1,02	0,982005
Mean VIF	1,16	

Tabell 8: VIF-test

8.2 Korrelasjonsmatrise

	read	clsize	teacher_exp	schoolsize4	par_edu_low	par_edu_high	pc_class	kinderg_att	par_not_born
read	1,0000								
clsize	0,0295	1,0000							
teacher_exp	-0,0389	-0,0367	1,0000						
schoolsize4	0,0373	0,4492	0,0198	1,0000					
par_edu_low	-0,1420	0,0108	0,0124	-0,0233	1,0000				
par_edu_high	0,2936	0,1106	-0,0125	0,1244	-0,2041	1,0000			
pc_class	-0,0245	-0,0180	-0,2150	-0,1350	-0,0039	-0,0603	1,0000		
kinderg_att	0,0763	-0,0333	-0,0590	0,0032	-0,0695	0,1349	0,0443	1,0000	
par_not_born	-0,1756	0,0517	-0,0221	0,0465	0,0856	-0,0327	-0,0193	-0,0226	1,0000

Tabell 9: Korrelasjonsmatrise

8.3 VIF-test modell 6

Variable	**VIF**	**1/VIF**
clsize	40.45	0,024723
clsize_sq	40.38	0,024764
par_edu_high	1,08	0,925939
pc_class	1,06	0,941930
teacher_exp	1,06	0,940961
par_edu_low	1,06	0,947478
kinderg_att	1,03	0,969942
par_not_born	1,01	0,987987
Mean VIF	10,89	

Tabell 10: VIF-test modell 6