

# Avkastning av å ha gått i barnehage på leseprestasjonene hos barneskoleelever

Bacheloroppgave i samfunnsøkonomi

Mai 2020

Av: Ellen Høvstad Vågen, Trine Gillebo og Adrian Lodgaard

## **Sammendrag**

En skoleproduktfunksjon har som mål å vise hvordan testresultat påvirkes av skolefaktorer, familie/elevkarakteristika og medelevkarakteristika. I denne oppgaven er det leseferdigheter blant barn i fjerdeklasse som blir studert med utgangspunkt i disse faktorene. Vi starter med å se på hvordan det å ha gått i barnehage påvirker leseferdighetene til barn, og utvider etterhvert modellen gradvis til å inkludere flere forklaringsvariabler. Tidligere forskning som denne oppgaven baserer seg på er PIRLS-undersøkelsen, og vi bruker dermed et datasett for Norge basert på denne undersøkelsen til å innhente informasjon vedrørende hvilke effekter som påvirker leseferdigheter. Som et redskap i vår analyse bruker vi minste kvadraters metode for å klargjøre den antatte årsakssammenhengen mellom variablene. Analysen avdekker at det er mange faktorer som påvirker leseferdigheter, i tillegg til at vi får bekreftet at barnehage er positivt korrelert med leseprestasjoner.

## **Abstract**

A school product function aims to show how test results are influenced by school characteristics, family/student characteristics and fellow student characteristics. In this paper, reading skills among fourth graders are studied on the basis of these three factors. We start by separately looking at how kindergarden attendance affects childrens reading performance, and further expand the model to include more explanatory variables. This paper is based on the PIRLS study, and we are using a dataset for Norway based on this study to gather information about the effects that is affecting readings skills. As a tool in our analysis, we use the least squares method to clarify the assumed causal relationship between the variables. The analysis reveals that there are many factors that are affecting reading skills, as well as it is confirming that kindergarden attendance is positively correlated with childrens reading performance.

## Innhold

## Innholdsfortegnelse

<b>1. Innledning.....</b>	<b>5</b>
1.1 Motivering .....	5
1.2 Presentasjon av problemstilling .....	5
<b>2. Relevant litteratur og tidligere funn.....</b>	<b>5</b>
2.1 Innledning.....	5
2.2 Litteratur og PIRLS undersøkelsen.....	5
2.3 Oppsummering .....	6
<b>3. Økonometrisk teori.....</b>	<b>7</b>
3.1 Innledning.....	7
3.2 Minste kvadraters metode (OLS).....	7
3.3 Regresjon med flere variabler.....	8
3.4 Hypotesetesting .....	9
3.5 Korrelasjon.....	10
3.6 Valg av funksjonsform.....	10
<b>4. Datamaterialet .....</b>	<b>11</b>
4.1 Innledning.....	11
4.2 Om datamaterialet.....	11
4.3 Oversikt over variabler .....	11
4.4 Deskriptiv statistikk for avhengig variabel.....	13
4.5 Deskriptiv statistikk for interessevariabel.....	13
4.6 Deskriptiv statistikk for kontrollvariabler .....	14
4.7 Korrelasjonsmatrise for datamaterialet .....	15
4.8 Hypotesetesting .....	16
<b>5. Empiriske resultater.....</b>	<b>16</b>
5.1 Innledning.....	16
5.2 Regresjon.....	16
5.3 Bedre leseferdigheter av å gå i barnehage? .....	18
5.4 Kan forskjell i leseferdigheter forklares med familie-/elevkarakteristika?.....	18
5.5 Kan forskjell i leseferdigheter forklares med skolefaktorer?.....	19
5.6 Kan forskjell i leseferdigheter forklares med medelevkarakteristika?.....	19
5.7 Avkastning av barnehage for innvandrerfamilier.....	20

5.8 Begrensninger.....	21
6. Oppsummering.....	22
Referanseliste:.....	23

**Tabelliste**

- Tabell 1: Oversikt over variabler
- Tabell 2: Deskriptiv statistikk for avhengig variabel
- Tabell 3: Deskriptiv statistikk for interessevariabel
- Tabell 4: Deskriptiv statistikk for kontrollvariabler kinderg\_att = 1
- Tabell 5: Deskriptiv statistikk for kontrollvariabler kinderg\_att = 0
- Tabell 6: Korrelasjonsmatrise
- Tabell 7: Regresjon for alle modellene
- Tabell 8: Regresjon med inkludering av samspillvariabler

**Figurliste**

- Figur 1: Fordeling av leseresultat
- Figur 2: Spredning i poengscore
- Figur 3: Differanse i spredning

## 1. Innledning

### 1.1 Motivering

Det er mange ulike syn på hva som er årsakene bak de store forskjellene i prestasjoner blant elever på barneskolen. Et av mange synspunkter på dette er: *“De som har gått i kvalitativt gode barnehager har bedre skolefaglige prestasjoner”*. Denne påstanden synes vi er interessant, så derfor vil vi ta den med oss videre i oppgaven for å studere validiteten i påstanden. Datasettene vi har tilgjengelig begrenser oss til å se på avkastningen barnehage har på leseferdighetene hos barneskoleelever og hvilken forståelse de har for det de leser.

### 1.2 Presentasjon av problemstilling

Barnehagen er en institusjon som har som funksjon å ivareta barnas behov for omsorg og lek, i tillegg til å fremme læring og utdanning som grunnlag for allsidig utvikling. Og nettopp dette med kvalitet i barnehagen er noe regjeringen har stort fokus på. Med tanke på støtteordninger og lovgivning vi har i Norge er det lagt opp slik at de som trenger det skal få ha barn i barnehage. Innenfor hver kommune er det barnehagegaranti i følge barnehageloven §12a, dette innebærer at barn har rett til barnehageplass i den kommunen det er bosatt. Sammenlignet med flere andre europeiske land er det også relativt billig å gå i barnehage i Norge siden det er innført en makspris fra og med 1. januar 2020 på 3 135 kr per måned, og dette er en makspris som gjelder for alle typer barnehager. Dersom familien eller husholdningen har lav inntekt er det mulig å søke om støtte fra kommunen. På grunnlag av disse statlige reguleringene og kommunale tiltakene ser man tydelig at barnehagen blir vurdert som en viktig institusjon. I følge SSB.no har 92,2% av alle barn i alderen 1-5 barnehageplass, men hva med de resterende 7,8% som ikke går i barnehage før skolestart?

Spørsmålet blir dermed om dette vil påvirke det pedagogiske grunnlaget til de barna som inngår i disse 7,8 prosentene, og dermed deres leseforståelse videre på skolen, ettersom barnehagen skal gi et godt grunnlag på disse områdene. På grunnlag av dette spørsmålet er problemstillingen vi har valgt å ta for oss *“Har gode leseferdigheter på skolen en sammenheng med om barna har gått i barnehage?”*. Som en tilleggsproblemstilling har vi valgt å ta for oss hvilke barn som har størst utbytte av å gå i barnehage. Dette skal vi studere ved å bruke data som er samlet inn for Norge i forbindelse med en leserundersøkelse som ble gjort i 2001.

## 2. Relevant litteratur og tidligere funn

### 2.1 Innledning

I dette kapittelet presenterer vi relevant litteratur fra blant annet utdanningsdirektoratet, og tidligere funn fra den internasjonale undersøkelsen PIRLS (Progress In International Reading Literacy Study).

### 2.2 Litteratur og PIRLS undersøkelsen

Litteraturen vi tar utgangspunkt i for å utdype grunnlaget til problemstillingen er blant annet §2, §8, §12 og §15 i Barnehageloven som krever henholdsvis at barnehagen skal være av pedagogisk virksomhet. At ethvert barn har rett på barnehageplass i kommunen det er bosatt og at kommunen har plikt til å tilby dette, samt at kongen kan gi forskrifter om søskenmoderasjon, inntektsgradering og maksimalgrense når det kommer til betaling.

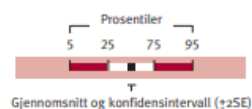
Fra utdanningsdirektoratet kommer det fram at maks prisen er fastsatt til 3135 kroner i måneden uavhengig av om barnehagen er offentlig eller privat, at barn nummer to i familien har krav på 30 prosent avslag i pris og at det tredje barnet og utover har krav på 50 prosent avslag i prisen. Er husholdningens totale inntekt på under 548 500 per år, har husholdningen krav på 20 timer gratis oppholdstid i barnehage i uka.

Når vi skal se på tidligere funn som er gjort i forbindelse med problemstillingen vår, er PIRLS (Progress In International Reading Literacy Study) en relevant undersøkelse å se på da hele datamaterialet vårt er basert på denne undersøkelsen. Dette er en leserundersøkelse som gjennomføres

i regi av IEA (The international Association for the Evaluation of Education Achievement) og dette var den første internasjonale leserundersøkelsen Norge deltok i som på det tidspunktet (1991) omfattet 32 land og både 9-åringer og 14-åringer, men som på slutten av 90-tallet endret målgruppen til bare fjerdeklassinger og 35 deltakerland. Vi ser i denne oppgaven på funnene som ble gjort i 2001 hvor 150 000 elever deltok. Undersøkelsen er en 1,5 timers faglig leseprøve som blir avholdt hvert femte år og foregår på følgende måte: deltakerne av undersøkelsen skal vise leseforståelse av litterære tekster og faktatekster, dette er fordi man leser av to grunner; for fornøynsens skyld og fordi vi er ute etter kunnskap/informasjon. Og gjennom å hente ut informasjon, trekke enkle slutninger, tolke og sammenligne informasjon og vurdere språk, innhold og virkemidler i teksten vektlegges forståelsesaspektet barn har når de leser.

Land	Leseferdighet	Gjennomsnitt	År på skolen	Gjennomsnittsalder
Sverige		561	4	10,8
Nederland		554	4	10,3
England		553	5	10,2
Bulgaria		550	4	10,9
Latvia		545	4	11,0
Canada		544	4	10,0
Litauen		543	4	10,9
Ungarn		543	4	10,7
USA		542	4	10,2
Italia		541	4	9,8
Tyskland		539	4	10,5
Tsjekkia		537	4	10,5
New Zealand		529	5	10,1
Skottland		528	5	9,8
Singapore		528	4	10,1
Russland		528	3/4	10,3
Hong Kong		528	4	10,2
Frankrike		525	4	10,1
Hellas		524	4	9,9
Slovakia		518	4	10,3
Island		513	4	9,7
Romania		512	4	11,1
Israel		509	4	10,0
Slovenia		502	3	9,8
<b>Intern. gj.snitt</b>		<b>500</b>	<b>4</b>	<b>10,3</b>
Norge		499	4	10,0
Kypros		494	4	9,7
Moldova		492	4	10,8
Tyrkia		449	4	10,2
Makedonia		442	4	10,7
Colombia		422	4	10,5
Argentina		420	4	10,2
Iran		414	4	10,4
Kuwait		396	4	9,9
Marokko		350	4	11,2
Belize		327	4/5	9,8

Forklaring fig. 1



Ved å se på figur 1 i den norske kortversjonen av den internasjonale rapporten kan vi studere funn som er gjort i alle deltakerland når det gjelder leseferdigheter og spredning, og dermed se hvordan Norge gjør det i forhold til andre land og det internasjonale gjennomsnittet. Med et internasjonalt gjennomsnitt på 500, ligger Norge rett under gjennomsnittet med 499. Men når det er sagt er også Norske elever yngre enn gjennomsnittet når testen er utført. Av denne figuren kan vi også se at det er en relativt stor spredning mellom de norske elevene som deltok i undersøkelsen, som igjen indikerer store forskjeller mellom elevenes resultater.

### 2.3 Oppsummering

I dette kapitlet presenterte vi relevant litteratur angående barnehageordninger i Norge og vi så på funn som har blitt gjort i forbindelse med PIRLS-undersøkelsen. Sentralt i denne undersøkelsen er at norske barn ligger rett under internasjonalt gjennomsnitt når det gjelder leseforståelse, og at det er store

forskjeller blant elever. Og det er nettopp årsaker bak disse forskjellene som danner grunnlaget for videre undersøkelser i denne oppgaven.

## 3. Økonometrisk teori

### 3.1 Innledning

I denne delen av oppgaven skal vi vise hvordan vi benytter minste kvadraters metode (OLS) som et redskap i vår analyse av datamaterialet. I tillegg presenterer vi skoleproduktfunksjonen som ser på resultater (leseresultater i dette tilfellet) som en funksjon av familiekarakteristika (F), skolefaktorer (S) og medeleverkarakteristika (P). Vi vil også introdusere variablene vi benytter til forklaring, og forventning til hver enkelt variabel.

### 3.2 Minste kvadraters metode (OLS)

I vår empiriske analyse av forskjeller i avkastning på leseferdigheter av å ha gått i barnehage, skal vi benytte oss av minste kvadraters metode, bedre kjent som OLS (ordinary least squares), for å behandle observasjonsmaterialet fra datasettene vi har tatt for oss. OLS er en regresjonsmetode som gjør at vi kan tilpasse en linje til et spredningsplott gjennom å kvantifisere den antatte årsakssammenheng mellom variablene. OLS gir den linja som minimerer summen av kvadrerte avvik mellom faktisk observasjon og predikert verdi. (Thomas, 2005, s.266-267)

Den enkle regresjonsanalysen tar utgangspunkt i to variabler som vi kaller X og Y, der vi antar at Y avhenger av X, at X er en uavhengig variabel og ikke avhenger av Y, og at vi har en lineær funksjonssammenheng. Den endogene variabelen Y og eksogene X, kan forklares som følgende:

$$\hat{Y} = \alpha + \beta X \quad (3.2.1)$$

Ut i fra vår oppgave vil leseferdighetene (*read*),  $\hat{Y}$ , være avhengig av X, om en har gått i barnehage (*kinderg\_att*). I tillegg har vi parameteren  $\beta$  som sier oss hvor mye hvor mye leseferdighetene øker når man har gått i barnehage.

$$Y_i = \alpha + \beta X_i + \varepsilon \quad (3.2.2)$$

$$\hat{Y}_i = a + bX_i \quad (3.2.3)$$

Likning 3.2.3 viser oss viser den predikerte verdien av Y, gitt ved  $\hat{Y}$ . Her har vi brukt a og b, som er de estimerte verdiene på henholdsvis  $\alpha$  og  $\beta$ . Epsilon ( $\varepsilon$ ) er støyledet som tar for seg alt mellom predikert og faktisk verdi. Dette vil dermed være alle de andre tilfeldige faktorene som kan påvirke leseferdighetene. I empirien bruker vi  $e_i$  istedenfor epsilon. Den vil være positiv om  $\hat{Y}$  overgår Y og negativ om den er mindre.

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (3.2.4)$$

Restleddet i denne likningen, som er gitt ved epsilon, tar for seg estimeringsfeilen som baserer seg på X, alfa og beta ikke estimerer Y eksakt.

$$\varepsilon_i = Y_i - \hat{Y}_i \quad (3.2.5)$$

Der  $\hat{Y}$  er den estimerte verdien av Y. Vil ligningen ovenfor være differansen mellom estimert verdi og faktisk verdi av  $Y_i$  (Thomas, 2005, s.266-275).

Nøyaktig hvor godt denne regresjonslinjen beskriver datasettet vårt er gitt ved determinasjonskoeffisienten  $R^2$ . Denne variabelen kan tolkes som andel av variasjon i den avhengige

variabelen som kan forklares av variasjonen i de uavhengige variablene.  $R^2 = 1$  indikerer at all variasjon forklares av regresjonen. Flere variabler gir høyere  $R^2$  og den kan dermed bli verre å tolke riktig. En liten  $R^2$  betyr at relevante forklaringsvariabler er utelatt fra modellen. Er  $R^2 = 0.4$  betyr det at 40% av variasjonen i den avhengige variabelen kan forklares av modellen. De resterende 60% forblir uforklarte i restleddet. Siden datasettet vi benytter ikke er fullstendig, kan det forventes en lavere  $R^2$ .

Vi kan finne variansen ved å dele variasjonen i X, men den totale variasjonen i Y. Disse har forkortelsene SSE og SST. Vi har også SSR, som er den uforklarte variasjonen, som vi kan forklare gjennom disse tre ligningene:

$$SST = \sum (Y_i - \bar{Y})^2 \quad (3.2.6)$$

$$SSE = \sum (\hat{Y}_i - \bar{Y})^2 \quad (3.2.7)$$

$$SSR = \sum \varepsilon_i^2 \quad (3.2.8)$$

Den uforklarte variasjonen og forklarte variasjonen blir sammen den totale variasjonen:

$$SST = SSE + SSR \quad (3.2.9)$$

Dermed kan vi definere modellens forklaringskraft ( $R^2$ ):

$$R^2 = \frac{SSE}{SST} = \frac{b^2 \sum X_i^2}{\sum Y_i^2} \quad (3.2.10)$$

(Thomas, 2005, s.273-277)

#### Forutsetninger for å bruke OLS:

- Ikke stokastiske X
- $E(\varepsilon_i) = 0$
- $Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$
- $Cov(\varepsilon_i, \varepsilon_j) = 0$
- $\varepsilon_i \sim N(0, \sigma^2)$

(Thomas, 2005, s.356-359)

### 3.3 Regresjon med flere variabler

Til nå har vi kun sett på regresjon med to variabler, men i vår oppgave kommer vi til regresjon med opptil flere variabler, disse kan vi skrive på følgende form:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_n X_{ni} + \varepsilon \quad (3.3.1)$$

Dermed kan vi ha at hver observasjon av Y,  $Y_i$ , vil vi kunne skrive på følgende form:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_n X_{ni} + \varepsilon \quad (3.3.2)$$

(Thomas, 2005, s.386-387)

Videre må vi ha en antakelse om at ingen av variablene  $X_{ni}$  er en lineær kombinasjon av de andre variablene. Slik som med to variabler, vil OLS finne estimat  $b_n$  for  $\beta_j$ , slik at summen minimeres:

$$S = \sum e_i^2 \quad (3.3.3)$$

(Thomas, 2005, s.389)



Her har vi som tidligere at  $\varepsilon_i = Y_i - \hat{Y}_i$ . Vi hadde tidligere kun to variabler, men i dette tilfellet har vi  $n$  antall variabler. Dermed blir det  $n$  førsteordensbetingelser, der vi deriverer 3.3.2 med hensyn på  $j$ :

$$\frac{\partial S}{\partial b_j} = 2 \sum X_{ji}(Y_i - b_1X_{1i} - b_2X_{2i} - \dots - b_nX_{ni}) \quad (3.3.4)$$

Der  $X_{1i} = 0$ .

Variansen til disse estimatorene er:

$$\sigma_{bj}^2 = \frac{\sigma^2}{\sum X_{ji}^2(1-r^2)} \quad (3.3.5)$$

Der  $r$  fanger opp korrelasjon mellom  $X_{ji}$  og  $\sigma^2$ , variansen til  $\varepsilon_i$ . Men her vet vi sjelden variansen, slik at vi my bruke estimatoren til variansen,  $s^2$ :

$$s^2 = \frac{\sum e_i}{n-k} \quad (3.3.6)$$

Der er  $n$  antall observasjoner og  $k$  antall parametre i bruk (Thomas, 2005, s. 395-396).

### 3.4 Hypotesetesting

Vi vil senere teste ulike hypoteser om estimatorene. Dette gjør vi for å beregne de ulike OLS-estimatorene og se om de samsvarer med de faktiske verdiene, gitt et signifikansnivå. En følgende prosess krever en nullhypotese ( $H_0$ ) og en alternativhypotese ( $H_A$ ), der  $H_0$  er den hypotesen man vil finne et grunnlag for å forkaste, mens man vil prøve å underbygge  $H_A$ . Det som bestemmer om en hypotese skal forkastes eller ei, er testobservatoren (TS) og den kritiske verdien. Det aller vanligste signifikansnivået er 5%, der dette nivået er sannsynligheten for å forkaste en gyldig nullhypotese. En t-test kan for eksempel være om elever har fått bedre leseferdigheter om de har gått i barnehage, enn de som ikke har det.

Fordi vi ikke vet hva variansen er, men samtidig vet at estimatorene er t-fordelt, tester vi kun en og en variabel. Det vil dermed være t-fordelt med  $n - k$  frihetsgrader. Det fører til følgende testobservator:

$$TS = \frac{b}{s_b} \sim n - k \text{ d.f.} \quad (3.4.1)$$

(Thomas, 2005, s.368-369)

Når det kommer til hypoteser med flere variabler som testes samtidig, vil vi bruke en såkalt f-test. Denne kan for eksempel være at flere variabler testes om de har en signifikant effekt på en avhengig variabel. Fra tidligere har vi SSE og SSR, variasjonen i X og den uforklarte variasjonen. Det vil føre til følgende testobservator for en f-test:

$$TS = \frac{\frac{SSE}{(k-1)}}{\frac{SSR}{(n-k)}} \sim f_{(k-1, n-k)} \text{ d.f.} \quad (3.4.2)$$

Der  $h$  er antall variabler vi tester,  $n$  er antall observasjoner og  $k$  er antall variabler totalt (Thomas, 2005, s.416).

### 3.5 Korrelasjon

Korrelasjonskoeffisienten forteller oss hvor sterk sammenhengen det er mellom to variabler, fra -1 til 1. Der 1 indikerer positiv korrelasjon og -1 negativ korrelasjon, mens 0 sier at to variabler er uavhengige. Korrelasjonen eller korrelasjonskoeffisienten er definert ved:

$$\rho = \frac{\text{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{E[X-E(X)][Y-E(Y)]}{\sqrt{E[X-E(X)]^2}\sqrt{E[Y-E(Y)]^2}} \quad (3.5.1)$$

Der:

Cov = Kovarians

V= Varians

For eksempel, hvis korrelasjonen mellom leseferdigheter og inntekt i husholdning er lik 0,9, kan vi si at det er en sterk positiv sammenheng mellom disse to variablene (Thomas, 2005, s.194-195).

### 3.6 Valg av funksjonsform

I denne oppgaven vil vi benytte flere variabler for å svare på problemstillingen vår, derfor er det hensiktsmessig å bruke multippel lineær regresjon. Vi presenterer skoleproduktfunksjonen, hvor vi lar leseferdighet måles ved T:

$$T = f(S, F, P)$$

Der

S er skolefaktorer

F er familie/elevkarakteristika

P er medelevkarakteristika (peer group effect)

I første omgang vil vi se på hvilken avkastning det å ha gått i barnehage isolert sett har på leseferdighetene hos barneskoleelever:

$$\text{Read4} = f(\text{Kinderg\_att}, X)$$

Ligningen viser leseferdigheter gitt som en funksjon av om barna har gått i barnehage, og et sett kontrollvariabler X (= S, F, P). Kinderg\_att er vår interessevariabel, og X representerer andre relevante forklaringsvariabler vi har valgt å inkludere i modellen. Sammenhengen mellom Read og Kinderg\_att kan leses som forbedret poengsum på lesertesten som årsak av at barnet har gått i barnehage. Multippel regresjon estimerer effekten på leseferdigheter ved å endre en variabel, mens man holder de andre variablene konstant. Denne effekten er uttrykt ved stigningsleddet Beta. Lineær funksjonssammenheng betyr i dette tilfellet at den avhengige variabelen Read er lineær i Beta. Den første likningen vi skal ta for oss er:

$$(1) \text{Read}_i = \beta_1 + \beta_2 \text{kinderg\_att} + \epsilon_i$$

Beta 1 viser gjennomsnittlig score i leseferdigheter når alle variabler er lik null. Beta 2 viser endringen i leseferdigheter når barnet har gått i barnehage. Epsilon er restleddet. Epsilon representerer alle andre faktorer enn barnehage som også påvirker leseferdighetene hos barnet. Etter hvert som vi utvider modellen vil vi se at flere sammenhenger enn å ha gått i barnehage har påvirkning på leseferdighetene og at effekten av å ha gått i barnehage på leseferdigheter endres som resultat av korrelasjon med de andre forklaringsvariablene. At to variabler korrelerer betyr at de samvarierer, altså at det fins en sammenheng mellom variablene innad i modellen.

#### Utvidelse av modellen:

I likning 2 har vi valgt å inkludere familie/ elevkarakteristika. Videre inkluderer vi forklaringsvariabler for skolekarakteristika i likning (3) og inkluderer videre medelevkarakteristika i likning 4:

$$(2) \text{Read}_i = \beta_1 + \beta_2 \text{kinderg\_att} + \beta_3 \text{par\_edulow} + \beta_4 \text{incomehigh} + \beta_5 \text{few\_bookshome} + \epsilon_i$$

$$(3) \text{ Readi} = \beta_1 + \beta_2 \text{kinderg\_att} + \beta_3 \text{par\_edulow} + \beta_4 \text{incomehigh} + \beta_5 \text{few\_bookshome} + \beta_6 \text{teacher\_cert} + \beta_7 \text{clsiz} + \epsilon_i$$

$$(4) \text{ Readi} = \beta_1 + \beta_2 \text{Kinderg\_att} + \beta_3 \text{par\_edulow} + \beta_4 \text{incomehigh} + \beta_5 \text{few\_bookshome} + \beta_6 \text{teacher\_cert} + \beta_7 \text{clsiz} + \beta_8 \text{par\_not\_born} + \beta_9 \text{not\_born} + \beta_{10} \text{girl} + \epsilon_i$$

Modell (1) angir forskjell i leseferdigheter mellom barn som har gått i barnehage og barn som ikke har gått i barnehage når vi ikke kontrollerer for andre faktorer.

Modell (2) kontrollerer i tillegg for familie- og elevkarakteristika. Her har vi inkludert inntekt og utdanning hos foreldrene, samt hvor mange bøker de har hjemme.

Modell (3) utvider ved å kontrollere for skolekarakteristika, som inkluderer klassestørrelse og lærere sertifisert.

Modell (4) tar videre med medelevkarakteristika. Her valgte vi å inkludere om barnet var gutt eller jente, om foreldrene var født i landet eller ikke og om barnet var født i landet.

### Tolkning:

- $\beta_1$  er et konstantledd. Tolkes ofte som den gjennomsnittlige verdien på den avhengige variabelen (leseferdigheter) når alle de andre verdiene er lik 0.
- $\beta$  leses som stigningstall. Disse tolkes som den gjennomsnittlige endringen i leseferdigheter, gitt en enhets økning i X.
- $\beta$  foran en dummyvariabel: Eks  $\beta_4$  viser forventet endring i poengene i lesertesten hvis foreldrene har inntekt kontra lav.
- $\epsilon_i$  vil være restleddet og tolkes ofte som feilen til forklaringen eller prediksjonen.

## 4. Datamaterialet

### 4.1 Innledning

I dette kapittelet som omhandler datamaterialet, vil vi se nærmere hvilket datasett vi har valgt å ta i bruk i vår oppgave. Innholdet i datasettene vi har tilgang til er basert på PIRLS sin leserundersøkelse og omfatter alle de 35 landene som deltok i undersøkelsen i 2001. Den kartleggingen som har blitt gjort i forbindelse med PIRLS, gjør det mulig for oss å enkelt sammenligne elevenes leseforståelse på tvers av land, men vi har begrenset oss til å kun bruke datasettet som omhandler Norge. I dette kapitlet foretar vi også en grundigere presentasjon av de ulike variablene og en forklaring av hvorfor vi inkluderer de variablene vi gjør i den bestemte rekkefølgen.

### 4.2 Om datamaterialet

I denne internasjonale leserundersøkelsen har det blitt kartlagt leseferdigheter til fjerdeklasser omkring i verden. I vår oppgave har vi valgt å analysere Norge. Datasettet vi bruker heter "data2901-Norway.dta", og dette inneholder den informasjonen vi trenger om landet for å kunne besvare spørsmålene vi har satt for oppgaven.

### 4.3 Oversikt over variabler

Tabell 1

<i>Read</i>	Kontinuerlig variabel. Leseferdigheter hos fjerdeklassinger gitt i poengscore fra 228 til 773
<i>Kinderg_att</i>	Dummy variabel, tar verdien 1 hvis barnet har gått i barnehage og 0 ellers
<i>Incomehigh</i>	Generert dummy, tar verdien 1 hvis inntekten til barnets foreldre er mer enn 30.000.- og 0 hvis lavere
<i>Par_edulow</i>	Generert dummy, tar verdien 1 hvis barnets foreldre har lavere utdanning enn videregående og 0 hvis høyere
<i>Few_bookshome</i>	Generert dummy, tar verdien 1 hvis barnet har færre enn 25 bøker hjemme og 0 hvis høyere
<i>Teacher_cert</i>	Dummy, tar verdien 1 hvis barnets lærer er sertifisert g 0 hvis ikke
<i>Par_not_born</i>	Dummy, tar verdien 1 om barnets foreldre ikke er født i landet og 0 hvis født i landet
<i>Not_born</i>	Dummy, tar verdien 1 om barnet ikke er født i landet og 0 hvis født i landet
<i>Girl</i>	Dummy, tar verdien 1 om barnet er jente og 0 hvis gutt
<i>Clsize</i>	Kontinuerlig variabel, klassestørrelsen varierer fra 4 til 32 elever

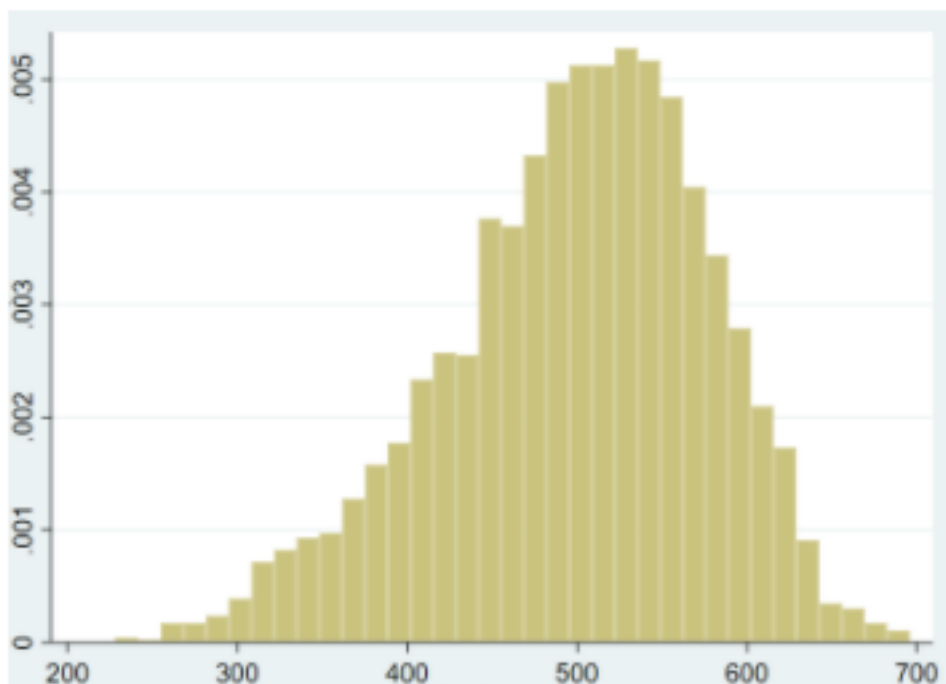
## 4.4 Deskriptiv statistikk for avhengig variabel

Tabell 2

Variable	Obs	Mean	Std. Dev.	Min	Max
read	3,459	498.2563	78.36616	228.0606	695.8717

Tabell 2 viser deskriptiv statistikk for avhengig variabel Read, i Norge. Tabellen gir en oversikt over antall observasjoner, gjennomsnittlig poengscore, standardavvik (altså gjennomsnittlig avvik fra gjennomsnittet), i tillegg til minimums og maksimale verdier (laveste og høyeste poengscore i Norge på lesertesten). Vi ser at det er gjort 3 459 observasjoner, der den laveste poengsummen observert er 228 og den høyeste er 695,9. Standardavviket er på 78,4 som betyr at de aller fleste observasjonene ligger mellom 78,3 poeng over og under den gjennomsnittlige poengscoren på 498,3.

Figur 2: Andre akse viser prosentetsats, mens førsteaksen viser poengscore på lesertesten. Ser tydelig fra figur 2 at de fleste observasjonene ligger mellom 420 og 576,6 8 som er gitt ved standardavvikene til gjennomsnittet.



## 4.5 Deskriptiv statistikk for interessevariabel

Tabell 3

. sum kinderg\_att

Variable	Obs	Mean	Std. Dev.	Min	Max
kinderg_att	3,137	.8603762	.3466516	0	1

Tabell 3 gir oss deskriptiv statistikk for interessevariabelen *kinderg\_att*. Antall observasjoner er her 3 137, der standardavviket på 0,86 forteller oss hvor stor andel av de 3 137 som tar verdien 1, og dermed har gått i barnehage.  $3\ 137 \times 0,86 = 2698$  barn av utvalget på 3 137 går i barnehage (altså 86% av de observerte). Siden dette er en dummyvariabel som kun tar verdiene 0 og 1 som minimum og maksimumsverdi, vil ikke standardavviket på 34,66% gi noen mening for oss.

#### 4.6 Deskriptiv statistikk for kontrollvariabler

Tabell 4 Deskriptiv statistikk for barn som har gått i barnehage, *kinderg\_att = 1*

Variable	Obs	Mean	Std. Dev.	Min	Max
par_edulow	2,699	.4375695	.496179	0	1
incomehigh	2,699	.8251204	.3799345	0	1
few_bookshve	2,688	.0699405	.2550941	0	1
teacher_cert	2,659	.9729222	.1623408	0	1
clsiz	2,659	20.84919	4.826038	4	32
par_not_born	2,646	.0487528	.2153916	0	1
not_born	2,629	.0840624	.2775341	0	1
girl	2,659	.4843926	.4998504	0	1

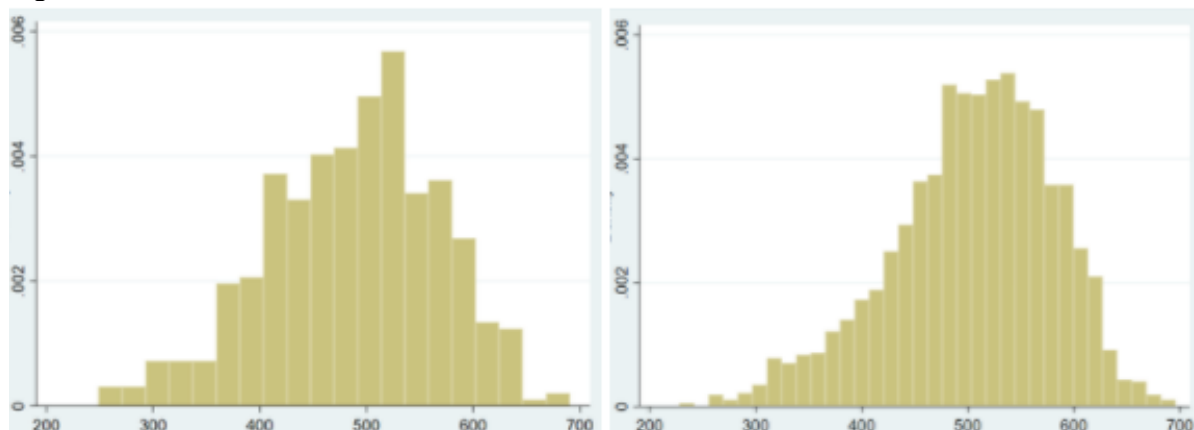
Tabell 5 Deskriptiv statistikk for barn som ikke har gått i barnehage, *kinderg\_att = 0*

Variable	Obs	Mean	Std. Dev.	Min	Max
par_edulow	438	.6324201	.4826975	0	1
incomehigh	438	.7968037	.4028375	0	1
few_bookshve	434	.0852535	.2795808	0	1
teacher_cert	435	.9678161	.1766915	0	1
clsiz	435	21.31494	5.054508	4	32
par_not_born	428	.0630841	.2433987	0	1
not_born	424	.0707547	.2567176	0	1
girl	432	.525463	.4999302	0	1

Sammenligner tabell 4 og 5

- Andelen barn med foreldre med lav utdanning er ganske høy. 44% for barn i barnehage og 63% for barn som ikke går i barnehage. Av dette ser vi at tendensen er at det er flere foreldre med høy enn lav utdanning som sender barna i barnehage.
- Andelen foreldre med høy inntekt er så og si lik mellom barn som går i barnehage og ikke og ligger på ca 80%.
- Andelen med få bøker hjemme er relativt lik og liten uavhengig om man går i barnehage eller ikke.
- Andelen sertifiserte lærere er uavhengig og omtrent lik 1.
- Gjennomsnittlig klassestørrelse er uavhengig og på 21 elever, der de fleste klassene har mellom 16 og 26 elever. Den minste klassen er på 4 elever og den største på 32 elever.
- Andelen foreldre som ikke er født i landet er lav, rundt 5%, men vi ser at den er høyere for barn som ikke har gått i barnehage (altså foreldre som ikke er født i landet sender i mindre grad barnet i barnehage).
- Andelen barn født i utlandet er lav, 6% av barna som ikke har gått i barnehage er født i utlandet og 8,4% av barna som har gått i barnehage.
- Det er like mange gutter som jenter observert i denne testen.
- Forskjellen i leseferdigheter hos barn som har gått i barnehage og barn som ikke har gått i barnehage kan være forklart av utdanningsnivået hos foreldrene.

Figur 3:



Histogrammet til venstre i figur 3 viser spredningen i leseferdigheter for barn som ikke har gått i barnehage, mens histogrammet til høyre viser det samme for barn som har gått i barnehage. Vi ser at for barn som har gått i barnehage er poengscoren mer sentrert rundt 520 poeng, mens for barn som ikke har gått i barnehage er spredningen større.

#### 4.7 Korrelasjonsmatrise for datamaterialet

En korrelasjonsmatrise viser enkle korrelasjonskoeffisienter mellom variablene, altså styrken på den lineære sammenhengen mellom de. Verdien som kommer fram er et standardisert mål ligger mellom -1 og 1. En positiv korrelasjon indikerer en tilnærmet lineær sammenheng, mens en negativ korrelasjon indikerer det motsatte. En korrelasjon nær null vil si at det ikke eksisterer noen sammenheng mellom variablene. Vi ønsker så lav korrelasjon som mulig mellom forklaringsvariablene vi har valgt å ta med når vi skal utføre regresjonsanalyse med OLS. Studerer vi korrelasjonsmatrisen framstilt i tabellen under, ser vi at variablene ikke har høy korrelasjonskoeffisient, og det er dermed enklere for oss å stole på og ikke minst tolke regresjonsanalysen.

Tabell 6

	read	kinder~t	par_ed~w	income~h	few_bo~e	teache~t	clsiz	par_no~n	not_born	girl
read	1.0000									
kinder_g_att	0.0759	1.0000								
par_edulow	-0.2907	-0.1344	1.0000							
incomehigh	0.1192	0.0346	-0.2181	1.0000						
few_booksh~e	-0.1746	-0.0172	0.1938	-0.1731	1.0000					
teacher_cert	-0.0225	0.0066	0.0378	-0.0326	0.0399	1.0000				
clsiz	0.0474	-0.0272	-0.1127	0.0374	-0.0228	-0.0623	1.0000			
par_not_born	-0.1373	-0.0239	0.0423	-0.1485	0.2496	0.0025	0.0429	1.0000		
not_born	-0.1219	0.0202	-0.0011	-0.0615	0.1285	0.0070	0.0254	0.2773	1.0000	
girl	0.1285	-0.0303	-0.0057	0.0058	-0.0032	-0.0139	0.0225	-0.0074	0.0003	1.0000

## 4.8 Hypotesetesting

Videre kommer vi til å teste følgende hypoteser:

- $H_0: \beta_2=0$ ,  $H_A: \beta_2>0$
- $H_0: \beta_3=\beta_4=\beta_5=0$ ,  $H_A: \beta_3\text{ulik}0$  u  $\beta_4\text{ulik}0$  u  $\beta_5\text{ulik}0$
- $H_0: \beta_8=\beta_9=\beta_{10}=0$ ,  $H_A: \beta_8\text{ulik}0$  u  $\beta_9\text{ulik}0$  u  $\beta_{10}\text{ulik}0$
- $H_0: \beta_{11}=\beta_{12}=0$ ,  $H_A: \beta_{11}\text{ulik}0$  u  $\beta_{12}\text{ulik}0$

## 5. Empiriske resultater

### 5.1 Innledning

I dette kapitlet vil regresjon for Norge presenteres i tillegg til hypotesetester. Her vil vi også få de endelige resultatene og svaret på problemstillingen.

### 5.2 Regresjon

Utgangspunktet for regresjonsanalysen er de fire ligningene våre som vi introduserte under kapittel 3.6. Alle modellene blir her presentert samlet i en felles matrise.



Tabell 7

	(1) read	(2) read	(3) read	(4) read
kinderg_att	17.14 (4.002)	7.619 (3.847)	7.117 (3.870)	9.686 (3.858)
par_edulow		-40.36 (2.776)	-39.55 (2.812)	-39.51 (2.798)
incomehigh		9.673 (3.562)	9.680 (3.585)	5.744 (3.600)
few_bookshome		-35.74 (5.245)	-34.80 (5.277)	-26.43 (5.443)
teacher_cert			-3.503 (8.094)	-2.280 (7.933)
clsize			0.262 (0.276)	0.300 (0.273)
par_not_born				-26.20 (6.517)
not_born				-25.29 (5.041)
girl				19.58 (2.648)
_cons	486.9 (3.712)	508.5 (5.079)	506.2 (11.38)	499.8 (11.29)
N	3137	3122	3079	2950
R <sup>2</sup>	0.006	0.106	0.104	0.135

Standard errors in parentheses

I **modell (1)** har vi den generelle avkastningen av barnehage på leseferdigheter, gitt ved `kinderg_att`, som har en betakoeffisient lik 17,14. Dette betyr at hvis barnet går i barnehage kan man forvente at de scorer 17,4 poeng bedre i lesertesten enn om det ikke hadde gått i barnehage, med et standardavvik på 4 poeng.

I **modell (2)** inkluderer vi familie/elevkarakteristika. Vi ser at hvis barnets foreldre har lav utdanning påvirker dette leseferdighetene i veldig stor grad i negativ retning (`par_edulow` = -40,36). Vi ser også at en høy inntekt hos foreldrene heller resulterer i bedre leseforståelse for barnet (`incomehigh` = 9,673). Vi vet også at å ha få bøker hjemme gjerne assosieres med å ikke lese mye, og denne testen viser også at leseferdighetene hos barnet kan forventes å reduseres kraftig hvis de ha få bøker i huset (`few_bookshome` = -35,74).

I **modell (3)** har vi kontrollert for skolefaktorer, som ikke ser ut til å ha noen innvirkning av betydning når det kommer til barnets leseferdigheter, men vi stiller oss kritisk til resultatet av regresjonen da vi tenkte at sertifiserte lærere skulle ha en positiv påvirkning på barnas leseferdigheter.

I **modell (4)** inkluderer vi til slutt medelevekarakteristika, her ser vi at barn som kommer fra innvandrersfamilier har betydelig dårligere leseferdigheter enn barn som er født i landet og har norske foreldre. (par\_not\_born = -26,20 og not\_born = - 25,29). Vi ser også at jenter presterer vesentlig bedre enn gutter i denne testen, jenter scorer i snitt 19,58 poeng bedre på lesertesten enn hva guttene gjør.

Ser at avkastningen av å ha gått i barnehage avtar betraktelig når vi utvider modellen til å inkludere flere relevante forklaringsvariabler. Modellene viser at barn som har gått i barnehage har bedre leseferdigheter, i snitt 17,14 poeng bedre i modell 1, mens når vi inkluderer flere variabler varierer snittet i poengscore mellom 7,117 og 9,686 poeng bedre enn de som ikke har gått i barnehage, men altså holder seg ganske stabil gjennom alle regresjonsmodellene. Spørsmålet videre blir nå: Kan vi si for sikkert at å gå i barnehage gir en vesentlig høyere avkastning på leseferdigheter enn å ikke gå i barnehage?

#### **Kommentar R<sup>2</sup>:**

Modellens forklaringskraft kan vi skrive som R<sup>2</sup>. Som i kapittel 3.2 antok vi at denne kan vil å bli lav i vår modell, dette fordi det er mange variabler som inkluderes. I tabell 7 kan vi finne R<sup>2</sup> på henholdsvis 0.006, 0.106, 0.104 og 0.135 i de ulike modellene. Dette betyr at hvert punkt som er estimert i OLS, er langt fra regresjonslinjen. Dermed kan vi si at denne modellen ikke passer godt til datasettet vårt.

### 5.3 Bedre leseferdigheter av å gå i barnehage?

Vi formulerer en nullhypotese slik at det ikke eksisterer noen forskjell i leseferdighetene hos barneskoleelever mellom barn som ikke har gått i barnehage og barn som har gått i barnehage. Vi formulerer nullhypotesen:

$$H_0: \beta_2=0, H_A: \beta_2>0,$$

og velger et signifikansnivå  $\alpha=0,05$ , som vi også benytter til hypotesetester senere i oppgaven. En ensidig test vil da ha kritisk verdi for  $t=1,645$  (TS vil være lik siden antall frihetsgrader er så stor i alle modeller).

1.  $TS = (17,14 - 0) / 4,002 = \underline{4,2828}$
2.  $TS = (7,619 - 0) / 3,847 = \underline{1,9805}$
3.  $TS = (7,117 - 0) / 3,870 = \underline{1,8390}$
4.  $TS = (9,686 - 0) / 3,858 = \underline{2,5106}$

Vi ser at vi kan forkaste nullhypotesen i alle modellene, og implisitt konkludere med at barnehage har påvirkning på leseferdighetene hos barneskoleelever. Videre vil vi nå se om foreldrenes inntekt og utdanning har noen innvirkning på om de velger å sende barna i barnehage eller ikke, og senere se på andre faktorer som kan ha større betydning på leseferdigheter.

### 5.4 Kan forskjell i leseferdigheter forklares med familie-/elevkarakteristika?

Formulerer nullhypotese, der vi antar at familie- og elevkarakteristika ikke har noen påvirkning på barnas leseferdigheter:

$$H_0: \beta_3=\beta_4=\beta_5=0, H_A: \beta_3\text{ulik}0 \text{ u } \beta_4\text{ulik}0 \text{ u } \beta_5\text{ulik}0,$$

braker F-test med hhv 3 og 2940 frihetsgrader, og vi har da en kritisk verdi  $F(3, 2940) = 2.6049$ , gitt et signifikansnivå på 5%. Fra STATA får vi TS:

```
( 1)  par_edulow = 0
( 2)  incomehigh = 0
( 3)  few_bookshome = 0
```

```
F( 3, 2940) = 93.62
Prob > F = 0.0000
```

TS er større enn kritisk verdi, og vi forkaster nullhypotesen. Altså kan vi si at familie- og elevkarakteristika har påvirkning på leseferdigheter. Vi ser også av modell (2) tabell 7 at avkastningen av å ha gått i barnehage på leseferdigheter har blitt redusert med 10 poeng. Vi ser også at koeffisienten foran utdanningsvariabelen og bøker hjemme er betraktelig høyere enn koeffisienten foran både inntekts- og barnehagevariabelen.

### 5.5 Kan forskjell i leseferdigheter forklares med skolefaktorer?

Ønsker å teste om leseferdighetene ikke påvirkes av skolefaktorer, og formulerer nullhypotesen:

$H_0: \beta_6 = \beta_7 = 0$ ,  $H_A: \beta_6 \text{ ulik } 0 \text{ u } \beta_7 \text{ ulik } 0$ ,

```
( 1)  teacher_cert = 0
( 2)  clsiz = 0
```

```
F( 2, 2940) = 0.66
Prob > F = 0.5158
```

Ser her at  $TS = 0.66$  er mindre enn den kritiske F-verdien på  $F(2, 2940) = 2.9957$ . Vi kan ikke forkaste nullhypotesen, og dermed kan vi si at skolefaktorer ikke påvirker leseferdigheter i vesentlig grad. Ser av ligning (3) tabell 7 at avkastningen av å ha gått i barnehage på leseferdigheter heller ikke har endret seg fra modell (2).

### 5.6 Kan forskjell i leseferdigheter forklares med medelevkarakteristika?

```
( 1)  par_not_born = 0
( 2)  not_born = 0
( 3)  girl = 0
```

```
F( 3, 2940) = 36.68
Prob > F = 0.0000
```

Her har vi  $TS = 36.68 > F(3, 2940) = 2.6049$ , og vi forkaster nullhypotesen. Medelevkarakteristika har påvirkning på leseferdigheter. Fra tabell 7 ser vi at avkastningen av barnehage på leseferdigheter har gått opp med 2,5 poeng, og med dette resonnerer vi oss frem til at barn fra innvandrerfamilier har større utbytte av å gå i barnehage enn majoritetsbefolkningen. Til slutt ønsker vi dermed å teste denne påstanden.

### 5.7 Avkastning av barnehage for innvandrerfamilier

Utvider nå modell 4 til å inkludere interaksjonsvariabler mellom `kinderg_att` og henholdsvis `par_not_born` og `not_born` for å se om det er forskjell i avkastningen av barnehage på leseferdigheter mellom barn som ikke har innvandrerbakgrunn og barn med innvandrerbakgrunn.

$$(5) \text{Readi} = \beta_1 + \beta_2 \text{Kinderg\_att} + \beta_3 \text{par\_edulow} + \beta_4 \text{incomehigh} + \beta_5 \text{few\_bookshome} + \beta_6 \text{teacher\_cert} + \beta_7 \text{clsize} + \beta_8 \text{par\_not\_born} + \beta_9 \text{not\_born} + \beta_{10} \text{girl} + \beta_{11} \text{par\_not\_born\_kinderg\_att} + \beta_{12} \text{not\_born\_kinderg\_att} + \varepsilon_i$$

Da vil  $\beta_2$  angi forskjell i leseprestasjoner for barn som har gått i barnehage og barn som ikke har gått i barnehage,  $\beta_8$  og  $\beta_9$  vil vise forskjell i leseprestasjoner for innvandrere og majoritetsbefolkningen og  $\beta_{11}$  og  $\beta_{12}$  vil vise avkastningen av barnehage på leseprestasjoner for barn med innvandrerbakgrunn. Ved hjelp av F-test kan vi nå teste en hypotese om at avkastningen av barnehage på leseferdigheter for innvandrere er ulik for de som ikke har innvandrer bakgrunn.

$$\begin{aligned} (1) \quad & \text{par\_not\_born\_kinderg\_att} = 0 \\ (2) \quad & \text{not\_born\_kinderg\_att} = 0 \end{aligned}$$

$$\begin{aligned} F(2, 2938) &= 2.23 \\ \text{Prob} > F &= 0.1081 \end{aligned}$$

$TS=2.23 < F(2, 2938) = 2.9957$  og vi kan ikke forkaste hypotesen om at avkastningen av barnehage er ulik for majoriteten av barna kontra barn med innvandrerbakgrunn.

Tabell 8

	(1) read	(2) read	(3) read	(4) read	(5) read
kinderg_att	17.14 (4.002)	7.619 (3.847)	7.117 (3.870)	9.686 (3.858)	7.980 (4.047)
par_edulow		-40.36 (2.776)	-39.55 (2.812)	-39.51 (2.798)	-39.65 (2.798)
incomehigh		9.673 (3.562)	9.680 (3.585)	5.744 (3.600)	5.978 (3.601)
few_bookshome		-35.74 (5.245)	-34.80 (5.277)	-26.43 (5.443)	-25.66 (5.453)
teacher_cert			-3.503 (8.094)	-2.280 (7.933)	-2.826 (7.934)
clsiz			0.262 (0.276)	0.300 (0.273)	0.292 (0.273)
par_not_born				-26.20 (6.517)	-56.19 (15.82)
not_born				-25.29 (5.041)	-18.79 (15.13)
girl				19.58 (2.648)	19.41 (2.648)
par_not_born_kinderg_att					35.83 (17.16)
not_born_kinderg_att					-7.219 (16.04)
cons	486.9 (3.712)	508.5 (5.079)	506.2 (11.38)	499.8 (11.29)	501.8 (11.35)
<i>N</i>	3137	3122	3079	2950	2950
<i>R</i> <sup>2</sup>	0.006	0.106	0.104	0.135	0.136

Standard errors in parentheses

### 5.8 Begrensninger

Analysen tar ikke hensyn til at ulike barnehager kan ha ulikt pedagogisk opplegg, og det er ikke umulig å tenke seg at nemlig dette kan stå for forskjellen i barnas leseferdigheter. Det er heller ikke gjort noen analyse rundt tidlige litterære evner eller om barnet snakker språket (norsk).

På grunn av at punktene fra minste kvadraters metode er langt fra regresjonslinjen, kan vi vise til at den ikke passer godt til vår modell. Det betyr at modellens forklaringskraft er relativt lav, som begrenser våre slutninger ut i fra modellen vår, slikt sett kan det altså være en annen modell som passer bedre til datasettet.

Observasjonene fra deskriptiv statistikk for interessevariabelen og avhengig variabel viser at antall observasjoner er i overkant av 3 tusen. Vi må dermed understreke at dette er et veldig lite utvalg. Dette betyr at utvalget som er brukt i datasettet ikke nødvendigvis blir representativt.

## 6. Oppsummering

Det er mange faktorer som har innvirkning på leseferdighetene hos barneskoleelever. Ut i fra histogrammet for deskriptiv statistikk ser vi at elever som har gått i barnehage har bedre leseferdigheter. Regresjonsanalyser og hypotesetester av datamaterialet gir samme resultat. I tillegg er det tydelig at barn med innvandrerbakgrunn har større utbytte av å gå i barnehage enn barn uten innvandrerbakgrunn. Samtidig er det viktig å betrakte modellens lave forklaringskraft, som gjør det vanskelig å trekke konklusjoner fra resultatene. Det gjør at vi ikke kan si det med sikkerhet, men vår analyse kan tyde på en positiv sammenheng mellom å ha gått i barnehage og barnas leseferdigheter.

## Referanseliste:

- Barnehageloven <https://lovdata.no/dokument/NL/lov/2005-06-17-64>
- Foreldrebetaling <https://www.udir.no/regelverk-og-tilsyn/barnehage/foreldrebetaling/>
- <https://lovdata.no/dokument/NL/lov/2005-06-17-64>
- <https://www.utdanningsforbundet.no/medlemsgrupper/barnehage/barnehageoffensiven2/dette-viser-forskningen/>
- <https://www.ssb.no/utdanning/statistikker/barnehager>
- Using Statistics in Economics: R.L. Thomas, 2005