

Wist, Christian Jacob Remen Horneman  
Bakken, Sigurd Pedersen  
Reitan, Georg Kristoffer

# En empirisk analyse av prestasjonsforskjeller mellom elever født i Norge og elever ikke født i Norge

Bacheloroppgave i Samfunnsøkonomi

Veileder: Bjarne Strøm

Mai 2020



Wist, Christian Jacob Remen Horneman  
Bakken, Sigurd Pedersen  
Reitan, Georg Kristoffer

# **En empirisk analyse av prestasjonsforskjeller mellom elever født i Norge og elever ikke født i Norge**

Bacheloroppgave i Samfunnsøkonomi  
Veileder: Bjarne Strøm  
Mai 2020

Norges teknisk-naturvitenskapelige universitet  
Fakultet for økonomi  
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden



**SØK2901 – NTNU Trondheim**

**En empirisk analyse av prestasjonsforskjeller mellom  
elever født i Norge og elever ikke født i Norge**

*Bacheloroppgave i samfunnsøkonomi*

Kandidatnummer: 10006, 10016 & 10032

14.05.2020

<b>1. Innledning</b> .....	<b>2</b>
1.1 Oppgavebeskrivelse.....	2
1.2 Våre problemstillinger .....	3
1.3 Sammendrag .....	3
1.4 Begrunnelse av valg.....	3
1.5 Norges innvandringshistorie.....	4
<b>2 Teoretisk rammeverk og tidligere litteratur</b> .....	<b>4</b>
2.1 Innledning .....	4
2.2 Teoretisk rammeverk .....	4
2.3 Tidligere litteratur og resultater .....	5
2.4 Oppsummering .....	6
<b>3 Datamaterialet</b> .....	<b>7</b>
3.1 Innledning .....	7
3.2 Om datamaterialet.....	7
3.3 Deskriptiv statistikk for avhengige variabler.....	9
3.4 Deskriptiv statistikk for kontrollvariabler .....	10
3.5 Oppsummering .....	11
<b>4 Økonometrisk modell</b> .....	<b>11</b>
4.1 Innledning .....	11
4.2 Funksjonsform .....	11
4.3 Teori om hypotesetesting.....	12
4.4 Oppsummering .....	15
<b>5 Empiriske resultater</b> .....	<b>15</b>
5.1 Innledning .....	15
5.2 Empiriske hovedresultater .....	16
5.3 Oppsummering .....	20
<b>6 Diskusjon</b> .....	<b>21</b>
6.1 Resultater.....	21
6.2 Svakheter ved modellen .....	22
<b>7 Oppsummering og konklusjon</b> .....	<b>23</b>
<b>Litteraturliste</b> .....	<b>24</b>
<b>Appendix</b> .....	<b>25</b>

# 1. Innledning

## 1.1 Oppgavebeskrivelse

Utdanning for alle er en viktig del av Norges skolepolitikk, og det er sentralt at alle skal ha samme forutsetninger for å lykkes best mulig, uavhengig av bakgrunn. Retten til utdanning følger direkte av Grunnloven § 109 og Opplæringslova. I 2015 vedtok FNs medlemsland 17 mål for bærekraftig utvikling frem mot 2030. Bærekraftsmål nr. 4 er God utdanning. Målet skal sikre inkluderende, rettferdig og god utdanning og fremme muligheter for livslang læring for alle.

I denne oppgaven skal vi ta for oss landet Norge, og analysere forskjeller i elevprestasjoner mellom:

- Elever som ikke er født i Norge
- Elever som er født i Norge

Dette analyseres ved å studere leseferdigheter til elevene. Norge er et demokrati og regnes som et av de rikeste landene i verden, og i en verden preget av mye migrasjon skal man anta at Norge har forutsetninger for å håndtere innvandring på en god måte. Vi ønsker å se på ulike faktorer som påvirker leseferdighetene til elever, som har betydning for elevprestasjonene generelt. Dette gjøres for at myndighetene kan tilrettelegge et best mulig integreringsprogram for innvandrere som bosetter seg i Norge.

I denne empiriske analysen vil leseferdighetene til to ulike elevgrupper bli analysert: Elever som er ikke født i Norge og elever som er født i Norge. Dette gjøres ved å analysere fire kategorivariabler som påvirker leseferdigheter:

- Kjønn
- Tilgang på bøker i hjemmet
- Foreldrenes inntektsnivå
- Tidlige leseferdigheter

Vi ønsker å finne ulike tiltak som fungerer bedre enn forventet på leseferdigheter for elevgruppen ikke født i Norge. Med “bedre enn forventet” mener vi tiltak med større positiv effekt på elevgruppen ikke født i Norge enn elever født i Norge. Slike tiltak vil bidra til å redusere avstanden i leseferdigheter mellom de to elevgruppene.

## 1.2 Våre problemstillinger

1. Eksisterer det en strukturell forskjell i modellene for leseferdigheter mellom elevgruppene?
2. Har elevgruppen ikke født i Norge gjennomsnittlig dårligere leseferdigheter enn elevgruppen født i Norge?
3. Undersøke om relevante faktorer har ulik effekt på de to elevgruppens leseferdigheter

## 1.3 Sammendrag

Formålet med denne analysen er å undersøke forskjellen i leseferdigheter mellom elever født i Norge og elever ikke født i Norge. Analysen tar for seg hvordan ulike faktorer påvirker leseferdighetene til disse elevene forskjellig, og baseres på et datasett fra 2001.

Problemstillingene vil bli undersøkt ved å gjennomføre en empirisk analyse, som hovedsakelig baserer seg på OLS-regresjon.

Resultatene viser at det eksisterer en signifikant strukturell forskjell i hvordan de forskjellige faktorene påvirker leseferdighetene til de to elevgruppene av interesse.

## 1.4 Begrunnelse av valg

Tema for oppgaven er å se på forskjeller i leseferdigheter i grunnskolen mellom elevgruppen ikke født i Norge, og elevgruppen født i Norge. Ifølge en artikkel i Aftenposten [\[5\]](#) har andelen elever som ikke får beregnet grunnskolepoeng økt fra 2,6 % til 4,6 % på 10 år. Med dette mener man elever som mangler standpunktkarakterer i mer en 50 % av fagene. Dette er et viktig problem å ta tak i, og vi har valgt å se på leseferdigheter til elever ikke født i Norge ettersom denne gruppen har hatt en negativ utvikling når det kommer til faglige resultater. På de samme ti årene som andelen elever som ikke får beregnet grunnskolepoeng gikk fra 2,6 % til 4,6 % har andelen elever med innvandrerbakgrunn innad i denne gruppen gått fra 17 % til 36 %. Flere av elevene i denne gruppen kommer fra land herjet av krig og med svært lite skolegang fra før av. Mange av de sliter også med norskerferdigheter. På grunn av disse utfordringene, og den negative trenden når det kommer til andelen elever med



innvandringsbakgrunn som får beregnet grunnskolepoeng, ønsker vi å se på hvilke faktorer som påvirker deres leseferdigheter.

## 1.5 Norges innvandringshistorie

Innvandring til Norge har foregått helt siden vi kan snakke om en norsk statsdannelse, fra om lag år 900. Innvandring har hatt stor betydning for Norge. Norge har opp igjennom historien vært under både dansk og svensk styre, før Norge fikk sin egen grunnlov i 1814, og unionen med Sverige ble oppløst i 1905. Under 2. Verdenskrig var Norge okkupert av Tyskland fra 1940 til 1945.

Norge opplevde en kraftig utvandring til Amerika fra 1865 til 1915, da omkring 750 000 nordmenn forlot landet. Fra tidlig på 1960-tallet og frem til innvandringsstoppen i 1975 kom det arbeidsinnvandrere fra Sørøst-Europa, Jugoslavia, Tyrkia og Pakistan. Det kan grovt skilles mellom tre typer innvandrere:

1. Fagfolk som aktivt blir rekruttert til Norge.
2. Arbeidsinnvandrere som kommer på eget initiativ.
3. Flyktninger.

Alle disse tre gruppene har kommet til Norge til alle tider, men hver og en av dem har dominert i bestemte perioder i norsk innvandringshistorie. [1]

## 2 Teoretisk rammeverk og tidligere litteratur

### 2.1 Innledning

I dette avsnittet skal vi presentere ulike årsaker til forskjeller i leseferdigheter til de to elevgruppene av interesse. Vi skal også presentere den generelle skoleproduktfunksjonen som våre modeller tar utgangspunkt i.

### 2.2 Teoretisk rammeverk

Det eksisterer flere årsaker for at vi kan forvente at leseferdighetene vil variere mellom de to elevgruppene. En av årsakene til dette knyttes til ulik beherskelse av det norske språk, der elever som har vokst opp med norsk som morsmål vil ha en naturlig fordel over elever med et annet morsmål. Man kan også forvente at elever som kommer fra et hushold med lite norsk

talespråk i det daglige vil ha et bedre grunnlag for å gjøre det bra på lesetesten ved å ha tilgang til bøker i hjemmet. Dette fører til økt eksponering for språket i tidlig alder gjennom lesing, i tillegg til at flere bøker i hjemmet signaliserer at husstanden har et prioriteringsgrunnlag som knyttes til språk, kunnskap og læring gjennom det å lese.

## 2.3 Tidligere litteratur og resultater

### 2.3.1 Tidligere studier

Det har blitt gjennomført en rekke studier og analyser av skoleprestasjonene til elever med innvandrerbakgrunn i Norge sett opp mot norskfødte elever. En rekke studier har i utgangspunktet fokusert på avgangselever i ungdomsskolen og på videregående, men i vår oppgave vil vi se hvordan de presterer i yngre alder.

Oslo Met publiserte en forskningsartikkel i 2017 der det står: *“innvandrerbarn har større mobilitet, og velger en mer ambisiøs utdanning enn foreldre med lav utdanning. De vil ofte velge lengre profesjonsfag som medisin og odontologi, mens barn av lavt utdannede foreldre ofte velger kortere, mer yrkesrettede fag som sykepleier.”* I studiene artikkelen er basert på, forteller barn av innvandrere at siden foreldrene valgte å flytte, så fikk de helt andre muligheter for utdanning, og at av respekt for foreldrene så vil de gripe den muligheten. [2]

Nasjonale Prøver som gjennomføres på 5. og 8. trinn i lesing er en prøve ikke ulik PIRLS. Vi kommer tilbake til hva PIRLS er senere i oppgaven. I 2012/2013 så fikk hele 19% av innvandrere på 5. trinn fritak fra Nasjonale Prøver i lesing. En så høy andel fritak påvirker tolkningen av resultatene. Det var i tillegg også høy fritaksprosent i prøvene som omhandler engelsk og regning. På 8. trinn derimot var andelen innvandrere som var fritatt halvert, som vil påvirke tolkningen av resultatene i en mer korrekt retning. På 8. trinn var forskjellen i leseferdighetene mellom innvandrere og øvrige større enn på 5. trinn, og noe av forklaringen på dette kan være det høye fritaket 5. trinn elevene opplevde. [3]

Andre faktorer som naturligvis vil påvirke resultatene for elever som ikke er født i Norge er botid i landet. Flere studier viser at unge trenger 5-7 år før de behersker akademiske ferdigheter i språket (Bakken 2007, Cummins 2000). Jo lengre ut i skoleløpet man kommer desto større betydning vil denne type ferdigheter ha for læringsresultatene.

### 2.3.2 Colemans Produktfunksjon For Utdanning

Modellen vi bruker i denne teksten tar utgangspunkt i James Colemans produktfunksjon for utdanning. Med denne funksjonen kan man se hvordan forskjellige faktorer innenfor kategoriene F, P og S påvirker T. [\[12\]](#)

$$T = f(S, F, P) \quad (2.1)$$

Tabell 2: Symbolforklaringer i produktfunksjonen

Symbol	Forklaring
T	Testscore
f	Skoleproduktfunksjon
S	Skolefaktorer (Klassestørrelse, lærerkarateristika)
F	Familie/elevkarakteristika
P	Medelevkarakteristika

Oppgaven kommer til å se på en variant av (2.1) som tilpasses våre kontrollvariabler, slik at vi kan analysere effekten av disse tilpasset de gruppene vi ønsker informasjon om. Dette gjøres ved å estimere verdiene til variablene av interesse ved bruk av minste kvadraters metode. Minste kvadraters metode, også kalt OLS-regresjon, er en statistisk metode for å estimere sammenhengen mellom en avhengig og en eller flere uavhengige variabler. [\[10\]](#)

## 2.4 Oppsummering

I dette avsnittet har vi presentert argumenter til hvorfor vi kan forvente at leseferdigheter vil variere mellom de to elevgruppene vi analyserer. Videre har vi presentert skoleproduktfunksjonen lagt fram av James Coleman.

## 3 Datamaterialet

### 3.1 Innledning

I dette avsnittet foreligger en presentasjon av datamaterialet som oppgaven baserer seg på. Avsnittet skal også presentere deskriptiv statistikk for avhengige variabler og kontrollvariabler.

### 3.2 Om datamaterialet

I 2001 gjennomførte International Association for the Evaluation of Educational Achievement, IEA, en internasjonal leserundersøkelse om leseferdigheter til elever i fjerde trinn. Undersøkelsen ble kalt Progress in International Reading Literacy Study, forkortet PIRLS, og gjennomført i 35 land. Dataene i denne oppgaven er hentet fra Norges resultater i denne undersøkelsen. IEA hentet inn testscore på leseferdigheter, samtidig som elever, foreldre, lærere og skolens ledelse fikk utgitt spørreskjema som kartlegger elevkarakteristika, lærerkarakteristika og skolekarakteristika.

Figur 3: Oversikt over elever som gjennomførte prøven i Norge



#### 3.2.1 Beskrivelser av variablene brukt i analysen

Avhengig variabel: **read**: Vi betegner dette som elevens leseferdigheter. Dette viser hvor høyt hver enkelt elev scoret på testen. I vår analyse representerer nivået på en elevs testscore det generelle nivået på en elevs skoleprestasjoner.

Sentral forklaringsvariabel: **not\_born**:

Not\_born representerer førstegenerasjonsinnvandrere og betegnes som elevgruppen ikke født i Norge.

Dummyvariabel, hvor 1 = ikke født i Norge og 0 = født i Norge.

Kontrollvariabler:

- **income** - Husstanden til elevens samlet inntekt- Kategorivariabel målt 1 - 6 hvor inntekten til foreldrene blir definert:
  1. Mindre enn \$20 000
  2. \$20,000 - \$29,999
  3. \$30,000 - \$39,999
  4. \$40,000 - \$49,999
  5. \$50,000 - \$59,000
  6. \$60,000 eller mer
  
- **books\_home** - Antall bøker hjemme. Kategorivariabel målt 1-5 - hvorav antall bøker blir definert:
  1. 0-10 Bøker i hjemmet.
  2. 11-25 Bøker i hjemmet.
  3. 26-100 Bøker i hjemmet.
  4. 101-200 Bøker i hjemmet.
  5. Over 200 bøker i hjemmet.
  
- **girl** - Er eleven jente - Dummyvariabel, hvor 1=jente og 0=gutt.
  
- **early\_ability** - Hvor tidlig eleven lærte seg å lese. Kategorivariabel målt 1-4 hvorav graden av leseferdigheter blir definert slik:
  1. Ikke i det hele tatt.
  2. Ikke spesielt bra.
  3. Moderat bra.
  4. Veldig bra.

Variablene *income*, *books\_home* og *early\_ability* går alle under kategorien kategorivariabler.

Dette betyr at koeffisienten foran disse variablene representerer endringen i *read* som forekommer av å gå opp eller ned ett nivå i kategorivariabelen.

### 3.3 Deskriptiv statistikk for avhengige variabler

Tabell 3.1: Deskriptiv statistikk for leseferdigheter for gruppene

	Landsgjennomsnitt	Elever født i Norge	Elever ikke født i Norge
Gjennomsnittlig testscore	498.2563	502.7747	463.7903
Standardavvik	78.3661	76.6893	83.4071
Minimum	228.0606	228.0606	234.8922
Maksimum	695.8717	695.8717	673.0267
Antall observasjoner	3459	3050	305

Tabell 3.1 viser deskriptiv statistikk for leseferdigheter for gruppene:

1. Landsgjennomsnittet.
2. Elever født i Norge. Dette er uavhengig av hvor foreldrene kommer fra. Dette er vår referansegruppe.
3. Elever ikke født i Norge. Dette er vår interessegruppe.

**Merk:** Gruppe 1 er gjennomsnittet for hele Norge. Gruppen er i tillegg noe større enn gruppe 2 og 3 til sammen, og det er fordi noen elever ikke har svart på om de er født i Norge eller ikke. Det som vises tydelig er at resultatene til gruppe 1 og 2 er relativt like. Dette gir mening da majoriteten av gruppe 1 består av gruppe 2. Gruppe 3 vil også inneholde elever med norske foreldre, som befant seg utenfor Norge på tidspunktet eleven ble født, samt adoptivbarn.

Datasettet vårt sier ingenting om at gruppen ikke født i Norge (not\_born) også består av adoptivbarn fra utlandet, men det er naturlig å anta. Ved å sjekke SSB sin adopsjonsstatistikk for året 1991, grunnet at de fleste elevene vil fylle 10 år i 2001, det året de gjennomførte undersøkelsen. Det vi ser der er at 833 personer ble adoptert i Norge, hvorav 549 var såkalte utenlandsadopsjoner. Statistikken sier derimot ikke noe om alderen på de som ble adoptert, men grunnet at det ble født om lag 60 000 barn i Norge det året vil antall adoptivbarn trolig

utgjøre under 1% av totalen, og mest sannsynligvis ikke påvirke kategorien *ikke født i Norge* nevneverdig mye. [6] [7] I statistikken vil det likevel være de sosiale foreldrene som er å regne som foreldrene til adoptivbarnet, ikke de biologiske foreldrene. [8]

I tabell 3.1 ser man at gjennomsnittlig leseferdigheter er høyest for elever født i Norge. Deretter følger landsgjennomsnittet med omtrent 4 poeng lavere resultat på testen, og med lavest resultat følger innvandrere med ca 40 poeng lavere resultat enn elever født i Norge. Når det kommer til spredning, ser vi at innvandrere har høyest standardavvik. En mulig forklaring på dette kan være at denne gruppen omfatter individer med forskjellig bakgrunn knyttet til mulighet for skolegang, og kvaliteten på skolegang. Deretter kommer gruppen med landsgjennomsnittet og til slutt elever født i Norge.

Vi har ikke delt opp elevene som ikke er født i Norge etter landbakgrunn, men likevel er vi inneforstått med at betydningen av landbakgrunn kan være stor. Det er gjennomført flere tidligere studier som har målt prestasjonsforskjellene mellom elever med ulik bakgrunn (F.eks Bakken 2014). Utover det vil også endringer i sammensetningen av innvandrerbefolkningen kunne påvirke resultatene. [9]

### 3.4 Deskriptiv statistikk for kontrollvariabler

Tabell 3.2: Deskriptiv statistikk for våre kontrollvariabler av interesse,

	Landsgjennomsnitt		Elever født i Norge		Elever ikke født i Norge	
	Snitt	St.dev	Snitt	St.dev	Snitt	St.dev
Husstandens samlede inntekt	4.0641	1.5508	4.0932	1.5375	3.7941	1.6900
Antall bøker i hjemmet	4.0329	1.0480	4.0605	1.0273	3.7659	1.2029
Er eleven jente	0.4810	0.5000	0.4825	0.4998	0.4533	0.4986
Nivå på tidlige leseferdigheter	2.6194	0.9467	2.6361	0.9483	2.4590	0.9397

Denne tabellen viser deskriptiv statistikk for våre kontrollvariabler av interesse, avrundet til fire desimaler. Det vi kan tyde her er at samtlige kontrollvariabler vil ha størst betydning for elever født i Norge, sammenliknet med elever ikke født i Norge. Videre tolkning rundt disse resultatene vil vi komme tilbake til videre i oppgaven.

### 3.5 Oppsummering

I dette avsnittet har vi presentert de delene av datasettet PIRLS som er interessant for vår oppgave, samtidig som vi har forklart de ulike variablene. Vi har også presentert deskriptiv statistikk for gjennomsnittlig leseferdigheter til vår referansegruppe og vår interessegruppe, og for kontrollvariabler, som gir informative resultater for videre sammenligning.

## 4 Økonometrisk modell

### 4.1 Innledning

I dette avsnittet skal vi utlede de økonometriske modellene vi vil benytte oss av, vise hvilke faktorer vi vil benytte oss av og tolke de ulike parametrene. Vi skal også utlede den statistiske teorien knyttet til hypotesetesting.

### 4.2 Funksjonsform

Som diskutert i avsnitt 2.3 vil vi benytte oss av ulike varianter av (2.1) i vår analyse, representert ved modellene (A), (B) og (C).

$$read = \beta_0 + \beta_1 girl + \beta_2 books\_home + \beta_3 early\_ability + \beta_4 income + \delta_{not\_born} + \epsilon_i \quad (A)$$

$$read = \beta_0 + \beta_1 girl + \beta_2 books\_home + \beta_3 early\_ability + \beta_4 income + \delta_{not\_born} + \delta_i girl\_not\_born + \epsilon_i \quad (B)$$

$$read = \beta_0 + \beta_1 girl + \beta_2 books\_home + \beta_3 early\_ability + \beta_4 income + \delta_{not\_born} + \delta_i girl\_not\_born + \delta_2 books\_home\_not\_born + \delta_3 early\_ability\_not\_born + \delta_4 income\_not\_born + \epsilon_i \quad (C)$$



(A) er vår basismodell som illustrerer gjennomsnittlig leseferdigheter for elever født i Norge, estimert med våre kategorivariabler. I modell (B) og (C) har vi generert en interaksjonsvariabel som gjør det mulig å skille mellom hvilken effekt variablene i modell (A) har på elever født i Norge og elever ikke født i Norge.

I denne analysen vil vi benytte lineære modeller. Vi kan tenke oss en utvalgslikning

$$Y = \alpha + \beta x + \varepsilon \quad (4.1)$$

Likningen viser en sammenheng som holder for alle observasjoner. I vår analyse estimerer vi alfa og beta ved hjelp av informasjon fra undersøkelsen PIRLS. Dette betyr at vi vil analysere den kvantitative sammenhengen i modellene.

Dersom vi ønsker å estimere Y for en enkelt observasjon vil modellen være uttrykt slik:

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad (4.2)$$

$\varepsilon_i$  representerer et restledd. Leddet representerer altså andre påvirkende faktorer på resultatet.  $\varepsilon_i$  er definert som en stokastisk variabel, representert med følgende egenskaper:

$$(i) E(\varepsilon_i) = 0$$

$$(ii) Var(\varepsilon_i) = \sigma^2$$

$$(iii) Cov(\varepsilon_i, \varepsilon_j) = 0$$

$$(iv) \varepsilon_i = \text{Normalfordelt} \beta$$

## 4.3 Teori om hypotesetesting

### 4.3.1 Teori bak ulike hypotesetester

Hypotesetesting er en statistisk analyse av hypoteser, der man tester hypoteser på bakgrunn av innsamlet datamaterialet for å finne ukjente størrelser. Ved å formulere en nullhypotese,  $H_0$ , og en alternativhypotese,  $H_a$ , kan man på bakgrunn av datamaterialet teste om en nullhypotese stemmer, eller om den må forkastes til fordel for en alternativhypotese. I en slik konklusjonsmetode antar man at hypotesene er komplementære, da man ved forkastning av den ene hypotesen vil føre til akseptering av at den andre hypotesen er sann.

I denne analysen vil vi benytte både *ensidig* hypotesetest og *tosidig* hypotesetest. I tosidig testing formuleres det en nullhypotese om at det ikke er sammenheng mellom variablene i

utvalget, mens den alternative hypotesen sier at det er en sammenheng mellom variablene i utvalget. Ved en ensidig test så undersøker vi ikke bare om det er sammenheng, men også om denne sammenhengen har en bestemt retning, for eksempel at det er en positiv sammenheng mellom to variabler. I det tilfellet blir  $H_0$  at det ikke er en sammenheng i populasjonen eller at den er negativ, mens  $H_a$  er at sammenhengen er positiv.

Vi vil benytte et 5 % signifikansnivå når vi tester for ukjente størrelser. Ved statistisk signifikans er man sikre på at resultatene i en hypotesetest ikke består av tilfeldige variasjoner. Ved et 5 % signifikansnivå antar man da at en tilfeldig variasjon oppstår 1 av 20 ganger. [\[10\]](#)

En av testene vi kommer til å benytte er T-testen. En T-test benyttes når man ønsker å undersøke om det er tilstrekkelig bevis for om en parameter er lik eller ulik en målverdi. Hvis vi antar at forutsetningene for en stokastisk variabel ((i)-(iv)) holder, vil  $\beta_i$  være tilnærmet normalfordelt uttrykt slik:

$$b_i \sim N(\beta_i, \frac{\sigma^2}{b_i}) \quad (4.3)$$

$b_i$  er den estimerte modellparameteren som man finner ved bruk av OLS. Standardisert vil dette uttrykkes som:

$$\frac{b_i - \beta_i}{\sigma_{b_i}} \sim N(0, 1) \quad (4.4)$$

I flere situasjoner vil populasjonsvariansen være ukjent. I slike tilfeller må en forventningsrett estimator benyttes, og derfor erstattes  $\sigma_{b_i}$  med  $s_{b_i}$ .

$$\frac{b_i - \beta_i}{s_{b_i}} \sim t_{n-k} \quad (4.5)$$

$n$  representerer antall observasjoner, og  $k$  representerer antall parametere. [\[10\]](#)

Vi kan nå uttrykke testobservatoren, som er et uttrykk som lar oss teste hvorvidt nullhypotesen er sann eller ikke. En testobservator når man utfører en t-test er da gitt ved:

$$TS = \frac{b_i - \beta_i}{s_{b_i}} \quad (4.6)$$

Verdien på  $\beta_i$  vil være bestemt av formulert nullhypotese. Nullhypotesen blir forkastet hvis absoluttverdien til TS er større enn absoluttverdien til kritisk verdi, gitt av t-fordelingen.

Dersom vi ønsker å teste om to koeffisienter er ulike på et statistisk signifikant nivå, bruker vi følgende testobservator:

$$TS = \frac{\beta_1 - \beta_2}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} \quad (4.7)$$

$\beta_1$  og  $\beta_2$  representerer henholdsvis de to koeffisientene som blir testet.  $\sigma_1^2$  og  $\sigma_2^2$  representert standardavvikene til de to koeffisientene opphøyd i annen.  $n_1$  og  $n_2$  er størrelsen på populasjonen til de to koeffisientene.

En annen test vi benytter oss av i analysen, er en Chow-test. En Chow-test benyttes ved en undersøkelse om det eksisterer signifikante forskjeller mellom koeffisienten i to lineære regresjoner på forskjellige datasett. Vi kan formulere en generell nullhypotese:

$$H_0: \beta_i = \text{konstant for alle } j$$

$$H_a: \beta_i \neq \text{konstant for alle } j$$

Får følgende TS:

$$TS = \frac{(R_{2u} - R_{2r})/h}{(1 - R_{2u})/(n - k)} \sim f(h, n - k) \quad (4.8)$$

der vi finner  $R_{2u}$  ved å hente  $R_2$ -verdien fra den urestrikterte modellen.

Tilsvarende finner vi  $R_{2r}$  ved å hente  $R_2$ -verdien fra den restrikterte modellen.  $h$  representerer antall variabler fjernet/lagt til,  $n$  representerer antall observasjoner og  $k$  representerer antall parametere i likningen med høyest antall forklaringsvariabler.

Chow-testen følger en f-fordeling med  $(h, n - k)$  frihetsgrader. [\[10\]](#)

#### 4.3.2 Teori knyttet til testing av interaksjonsledd

Ettersom vi ønsker å teste hvorvidt effekten av ulike faktorer som påvirker leseferdigheter til elevgruppen ikke født i Norge er forskjellig fra elevgruppen født i Norge, vil vi i noen tilfeller teste om  $\delta_i$  er større/mindre/ulik 0.

Grunnen til dette er fordi effekten en faktor har på leseferdigheter til elevgruppen ikke født i Norge er:  $\beta_i + \delta_i$ . Effekten en faktor har på leseferdigheter til elevgruppen født i Norge er:  $\beta_i$ .

Som vi ser vil effekten av faktoren være annerledes for elevgruppen født i Norge sammenlignet med elevgruppen ikke født i Norge dersom  $\delta_i$  er større/mindre/ulik 0.

#### 4.4 Oppsummering

I dette avsnittet har vi presentert vår grunnmodell som vi ønsker å bruke som utgangspunkt til å analysere hvilke effekter våre ulike variabler har på leseferdigheter til innvandrerelever.

Videre har vi utledet matematisk teori som gir grunnlag for å tolke parametrene på måten vi gjør. I tillegg har vi illustrert teorien bak T-testen og Chow-testen, som vi senere vil benytte.

### 5 Empiriske resultater

#### 5.1 Innledning

I dette avsnittet foreligger en presentasjon og analyse av regresjonsresultatene som forekommer av OLS-regresjon på datasettet knyttet til innvandrerelever og øvrig norske elever. Videre vil vi gjennomføre hypotesetester med utgangspunkt i teori presentert i del 4.3.

## 5.2 Empiriske hovedresultater

Tabell 5: Regresjonsresultater

VARIABLES	A	B	C
	read	read	read
girl	11.329507 (4.452)	9.778714 (3.692)	9.787163 (3.693)
books_home	16.488796 (12.762)	16.500597 (12.778)	15.661101 (11.529)
early_ability	28.914158 (21.532)	28.947915 (21.569)	28.853382 (20.625)
income	6.691360 (7.736)	6.673722 (7.720)	6.981711 (7.712)
not_born	-19.682754 (-4.203)	-29.146701 (-4.487)	-52.854715 (-2.532)
not_born_girl		19.606651 (2.101)	20.007101 (2.126)
not_born_books_home			9.048459 (2.059)
not_born_early_ability			1.354919 (0.273)
not_born_income			-3.876519 (-1.267)
Constant	329.605887 (52.191)	330.303107 (52.260)	332.701957 (50.328)
R-squared	0.249	0.251	0.252

t-statistics in parentheses

Tabellen viser resultater etter OLS-regresjon på modellene (A), (B) & (C).

### 5.2.1 Test på effekt av innvandrerstatus og kjønn på leseferdigheter

Vi ønsker å undersøke om elevgruppen ikke født i Norge har gjennomsnittlig lavere leseferdigheter enn elevgruppen født i Norge.

Fra Tabell 5 modell A, ser vi at elever som ikke er født i Norge har gjennomsnittlig lavere lesescore enn elever født i Norge. Dette ser vi på koeffisienten til “not\_born”, not\_born = -19,68275. Vi ønsker å teste om det er statistisk signifikant om elever ikke født i Norge er statistisk signifikant.

Basert på utledningen gjort i del (4.3), vil vi her gjennomføre en T-test. Vi setter opp følgende nullhypotese og alternativhypotese:

$$H_0 : \beta_5 = 0$$

$$H_A : \beta_5 < 0$$

Bruker her TS oppgitt i modell (4.6):

$$TS = \frac{-19.68275 - 0}{4.683018} = -4.2030054123$$

$$t_{0,05} = 1.645$$

$$TS < -t_{0,05}$$

Vi ser at vi forkaster  $H_0$  på et 5 % signifikansnivå. Dette betyr at vi med 95 % sikkerhet kan si at elever som ikke er født i Norge har gjennomsnittlig lavere leseferdigheter enn elever som er født i Norge.

Videre ønsker vi å se på om det er gjennomsnittlig ulike resultater for jenter og gutter i gruppen elever ikke født i Norge sammenlignet med gjennomsnittlig ulike resultater for elevgruppen født i Norge.

Vi benytter oss da av modell (B) og basert på teorien i avsnitt 4.3 benytter vi her en T-test. Vi får følgende nullhypotese og alternativhypotese:

$$H_0 : \delta_i = \beta_i$$

$$H_A : \delta_i > \beta_i$$

Bruker her TS oppgitt i modell (4.7):

$$TS = \frac{19.60665 - 9.778714}{\sqrt{\left(\frac{(9,330744)^2}{3326} + \frac{(2,648286)^2}{3401}\right)}} = 58.484850783$$

$$t_{0,05} = 1.645$$

$$TS > t_{0,05}$$

Vi ser at vi forkaster  $H_0$  på et 5 % signifikansnivå. Vi kan konkludere med at koeffisienten til gruppen jenter ikke født i Norge er større enn koeffisienten til gruppen jenter født i Norge.

### 5.2.2 Test av interaksjonsledd

Tallene i dette avsnittet er hentet fra modell C.

Som vi ser i avsnitt 5.2.1 har vi bekreftet at elever ikke født i Norge har gjennomsnittlig lavere leseferdigheter enn elever født i Norge. På grunn av dette ønsker vi å finne ulike tiltak som har god positiv effekt for leseferdighetene til denne gruppen.

Ifølge tabell 5 modell C, er koeffisienten til inntektsnivået til husstanden for elevgruppen ikke født i Norge lik -3.876519. Ettersom denne koeffisienten er ulik 0, er det rimelig å anta at effekten av husstandens inntektsnivå på elevens leseferdigheter er ulik for elevgruppene. Vi ønsker å teste om dette er statistisk signifikant.

Setter opp en T-test med følgende null og alternativ hypotese:

$$H_0 : \delta_4 = 0$$

$$H_A : \delta_4 \neq 0$$

Fra (4.4) får vi følgende testobservator:

$$TS = \frac{-3.876519 - 0}{3.0585159} = -1.267449704$$

$$t_{0,05/2} = 1.960$$

$$TS > - t_{0,05/2}$$

Vi kan ikke forkaste  $H_0$  på et 5 % signifikansnivå. Dette betyr at vi ikke kan konkludere med at effekten av husstandens inntektsnivå på elevens leseferdigheter er ulik for elevgruppene.

Videre kan vi se fra Tabell 5 modell C at koeffisienten til antall bøker i hjemmet for elevgruppen ikke født i Norge er lik 9.048459. Vi ønsker her å teste en alternativhypotese om at effekten av antall bøker i hjemmet er større for elevgruppen ikke født i Norge sammenlignet med elevgruppen født i Norge.

Setter opp en T-test med følgende null- og alternativhypotese:

$$H_0 : \delta_2 = 0$$

$$H_A : \delta_2 > 0$$

Fra (4.6) får vi følgende testobservator:

$$TS = \frac{9.048459-0}{4.394316} = 2.059127974$$

$$t_{0,05} = 1.645$$

$$TS > t_{0,05}$$

Vi forkaster  $H_0$  på et 5 % signifikansnivå. Vi kan derfor konkludere med at effekten av flere bøker i hjemmet er større for elevgruppen ikke født i Norge enn elevgruppen født i Norge.

Til slutt i dette avsnittet ønsker vi å teste hvorvidt nivået på leseforståelse i tidlig alder har ulik effekt på leseferdigheter for elevgruppen ikke født i Norge sammenlignet med elevgruppen født i Norge. Vi ser av Tabell 5 at koeffisienten som representerer tidlig leseforståelse for elevgruppen ikke født i Norge er lik 1.354919. Ettersom koeffisienten er ulik 0, tyder dette på at effekten er ulik for de to elevgruppene. Setter opp en tosidig T-test med følgende null-og alternativhypotese:

$$H_0 : \delta_3 = 0$$

$$H_A : \delta_3 \neq 0$$

Fra (4.9) får vi følgende testobservator:

$$TS = \frac{1.354919-0}{4.963363} = 0.2729840634$$

$$t_{0,05/2} = 1.960$$

$$TS < t_{0,05/2}$$

Vi kan ikke forkaste  $H_0$  på et 5 % signifikansnivå. Dermed kan vi ikke konkludere med at tidlig leseforståelse har ulik effekt på leseferdigheter for de to elevgruppene.



### 5.2.3 Chow test

Som vi har bekreftet gjennom denne empiriske analysen er noen av koeffisientene til kontrollvariablene forskjellig for de to gruppene. Dette betyr at likningen for elevgruppen født i Norge ( $read_H$ ) er ulik for likningen til elevgruppen ikke født i Norge ( $read_U$ ). Vi ønsker derfor å teste om det er en strukturell forskjell i modellen for leseferdigheter mellom elevgruppen født i Norge og elevgruppen ikke født i Norge. Her benyttes Chow-testen for strukturell forskjell. Vi refererer til (4.6).

Setter opp en tosidig test med følgende nullhypotese og alternativhypotese:

$$H_0 : read_H = read_U$$

$$H_A : read_H \neq read_U$$

Fra (4.8) får vi følgende testobservator:

$$TS = \frac{(0.2517 - 0.2449)/5}{(1 - 0.2517)/2836} = 5.1542973006$$

$$F_{0,05/2} = 2.5665$$

$$TS > F_{0,05/2}$$

Vi forkaster  $H_0$  på et 5% signifikansnivå. Vi kan derfor konkludere med at det er en strukturell forskjell i modellene for leseferdigheter mellom elevgruppene.

## 5.3 Oppsummering

I dette avsnittet har vi presentert resultatene fra OLS-regresjon av våre modeller, samt gjennomført ulike hypotesetester som illustrerer hvilke effekter ulike variabler har på leseferdigheter.

## 6 Diskusjon

### 6.1 Resultater

Vi startet analysen med å undersøke om gjennomsnittlig leseferdigheter for elevgruppen ikke født i Norge var lavere enn for elevgruppen født i Norge. Utgangspunktet for denne testen var resultater lagt frem fra deskriptiv statistikk i (Tabell 3.1) og (Tabell 3.2.) som fortalte at elevgruppen ikke født i Norge har gjennomsnittlig lavere leseferdigheter enn elevgruppen født i Norge. Dette fikk vi så bekreftet med 95 % sikkerhet ved gjennomført hypotesetest i (avsnitt 5.2.1). Dette ga grunnlaget for videre undersøkelser om ulike variablers effekt på leseferdigheter for de to elevgruppene.

Videre undersøkte vi om avstanden for leseferdigheter mellom jenter og gutter for elevgruppen ikke født i Norge avvek fra elevgruppen født i Norge. Tabell 3.2 sier at jenter har gjennomsnittlig høyere leseferdigheter. Hensikten med å teste om avstanden mellom jenter og gutters leseferdigheter er høyere i elevgruppen ikke født i Norge sammenlignet med jenter og gutter i elevgruppen født i Norge, er å undersøke om gutter ikke født i Norge ligger mer bak jenter ikke født i Norge enn det man forventer at gutter naturlig gjør. Hypotesetesten gjennomført i avsnitt (5.2.1) bekrefter med 95 % sikkerhet at avstanden i gjennomsnittlig leseferdigheter mellom jenter og gutter er større i elevgruppen ikke født i Norge enn i elevgruppen født i Norge. På bakgrunn av dette resultatet, kan man argumentere for at det er hensiktsmessig å bruke mer ressurser på elevgruppen gutter ikke født i Norge, slik at avstanden i leseferdigheter mellom gutter og jenter blir avtagende.

Tabell 3.2 viser at foreldrenes inntektsnivå påvirker elevenes leseferdigheter, der høyere inntektsnivå tilsvarer bedre leseferdigheter. Vi testet om effekten av en inntektsendring gir ulikt utslag på leseferdigheter i de to elevgruppene av interesse. Hypotesetesten viser at man ikke kan bekrefte at effekten av en inntektsendring vil gi ulikt utslag på leseferdigheter mellom de to elevgruppene ved et 5 % signifikansnivå. Dersom testen hadde latt oss forkaste nullhypotesen til fordel for alternativhypotesen, hadde en analyse om hvordan effekten av inntektsnivå påvirker de to gruppene ulikt vært relevant. Grunnen til at vi er interessert i effekten av inntektsnivå på elevgruppen ikke født i Norge sine leseferdigheter, er undersøkelser fra SSB som viser at inntektsnivået blant innvandrere er gjennomsnittlig lavere enn inntektsnivå blant den øvrige befolkningen.[\[11\]](#) Ettersom disse tallene tilsvarer at

elevgruppen ikke født i Norge har et dårligere utgangspunkt for å utvikle gode leseferdigheter vil det være relevant å se på hvor mye inntektsnivå påvirker denne elevgruppen direkte.

Videre viser Tabell 3.2 at det å ha flere bøker i hjemmet gir et positivt utslag på leseferdigheter. Et argument for dette positive utslaget kan være at mange bøker i hjemmet tyder på en god lesekultur som igjen vil påvirke barnas syn på lesing. Ved å teste om flere bøker i hjemmet gir større positiv effekt for elevgruppen ikke født i Norge sammenlignet med elevgruppen født i Norge, vil vi få svar på om dette er et tiltak som fungerer bedre enn forventet for å øke elevgruppens leseferdigheter. Etter gjennomført hypotesetest forkaster vi nullhypotesen ved et 5 % signifikansnivå. Dette konkluderer at effekten av bøker i hjemmet er større for elevgruppen ikke født i Norge enn elevgruppen født i Norge.

Den siste variabelen vi ønsket å undersøke, var om nivå på tidlig leseforståelse påvirker elevgruppene av interesse ulikt når det kommer til leseforståelse i fjerde trinn. Vi ønsket å undersøke dette for å se om utvikling av god tidlig leseforståelse hos elevgruppen ikke født i Norge vil være et godt tiltak for å redusere forskjellen i leseferdigheter mellom de to elevgruppene av interesse. Hypotesetesten ga oss ikke grunnlag til å forkaste nullhypotesen på et 5 % signifikansnivå. Dette betyr at vi ikke kan bekrefte at nivå på tidlig leseferdigheter påvirker de to elevgruppens leseferdigheter i fjerdeklasse forskjellig.

På bakgrunn av hypotesetestene gjennomført i analysen, opparbeidet vi grunnlag for å anta at de to modellene for elevgruppens leseferdigheter er ulike. I avsnitt 5.2.3 gjennomførte vi en Chow test for å om det er en strukturell forskjell i modellen for leseferdigheter mellom de to elevgruppene av interesse. Resultatet av testen bekreftet med 95 % sikkerhet at det er en strukturell ulikhet i de to modellene.

## 6.2 Svakheter ved modellen

En mulig svakhet i vår analyse er at vi har tolket at elevgruppen ikke født i Norge kommer fra hjem hvor det snakkes relativt lite/ingenting norsk. Grunnen til at dette er en mulig svakhet, er at denne gruppen vil inneholde elever med norskfødte foreldre til tross for at de selv er født i utlandet. Disse elevene antas å ha like forutsetninger som elevgruppen født i Norge når det kommer til utvikling av leseferdigheter.

En annen svakhet ved modellen er at den kun tar opp en interessegruppe. I utgangspunktet ønsket vi å utlede modeller for to interessegrupper og en referansegruppe. Her var referansegruppen elever født i Norge med norskfødte foreldre, interessegruppe 1 var elever ikke født i Norge med ikke norskfødte foreldre og interessegruppe 2 var elever født i Norge med foreldre som ikke var født i Norge. En slik utledning gav riktignok svært få signifikante resultater. En mulig årsak til dette er det at populasjonsandelen til begge våre interessegrupper var svært liten. En fordel som ville eksistert ved en slik utledning er at kartleggingen av interessegruppene hadde vært mer konkret, da interessegruppene hadde vært mer definert. For eksempel ville problemet med at elevgruppen ikke født i Norge kan inneholde elever med norskfødte foreldre vært løst ved å heller definert gruppen som “elever ikke født i Norge med ikke norskfødte foreldre”. Vi ville også fått informasjon om to interessegrupper framfor kun en.

En annen ting som har betydning for hvor varsomt vi kan trekke sterke konklusjoner, er at det er tiltakende segregering i for eksempel Oslo-skolene, altså raske endringer som vi enda ikke kan se konsekvensene av. Datasettet vi har operert med er basert på resultater fra 2001, som betyr at det har skjedd mye i skolesystemet siden.

## 7 Oppsummering og konklusjon

Etter å ha presentert deskriptiv statistikk og utført relevante hypotesetester kan vi med 95 % sikkerhet konkludere med at det eksisterer en strukturell forskjell i modellene for interessegruppens og referansegruppens leseferdigheter. Deskriptiv statistikk viser at elevgruppen ikke født i Norge har gjennomsnittlig dårligere leseferdigheter enn elevgruppen ikke født i Norge, som også kan bekreftes med 95 % sikkerhet etter gjennomført hypotesetest. Videre kan vi konkludere med at faktorene kjønn og antall bøker i hjemmet har ulik effekt på leseferdigheter til interessegruppen og referansegruppen.

Så lenge det foreligger strukturelle ulikheter for skoleprestasjoner mellom ulike elevgrupper, mener vi at man kan si at alle elever i norsk skole ikke har et likt grunnlag for å opparbeide seg nødvendig kunnskap som behøves for videre utdanning. For å sikre alle elever i norsk skole like muligheter, mener vi det er viktig å arbeide for å redusere disse avstandene så mye som mulig.

## Litteraturliste

### Internettkilder:

Gursli-Berg, Gunhild; Myhre, Jan Eivind: *Norsk innvandringshistorie i Store norske leksikon* på snl.no. Hentet 01.05.2020 fra: [www.snl.no/Norsk\\_innvandringshistorie](http://www.snl.no/Norsk_innvandringshistorie)<sup>[1]</sup>

Forskningsartikkel skrevet av Elise Koppang Frøjd. Publisert: 05.07.19<sup>[2]</sup>:  
[www.oslomet.no/forskning/forskningsnyheter/innvandrere-hoyere-utdanning-enn-foreldre](http://www.oslomet.no/forskning/forskningsnyheter/innvandrere-hoyere-utdanning-enn-foreldre)

Barn, unge og voksne med innvandrerbakgrunn i grunnopplæringen. Publisert av UDIR  
01.08.17: <sup>[3]</sup>: [www.udir.no/globalassets/upload/statistikk/innvandrere-i-grunnoppleringen-2014.pdf](http://www.udir.no/globalassets/upload/statistikk/innvandrere-i-grunnoppleringen-2014.pdf)

NTB 05.09.2018. Stadig flere innvandrere faller utenfor på ungdomsskolen. <sup>[5]</sup>*Aftenposten*.  
Hentet fra:  
[www.aftenposten.no/norge/i/6nRk8O/stadig-flere-innvandrerelever-faller-utenfor-paa-ungdomsskolen?fbclid=IwAR05NI-SRSZBdExp623SQYjbfelMCQDIKI1PoHzvIf3D-BqGRPUNqjP7puI](http://www.aftenposten.no/norge/i/6nRk8O/stadig-flere-innvandrerelever-faller-utenfor-paa-ungdomsskolen?fbclid=IwAR05NI-SRSZBdExp623SQYjbfelMCQDIKI1PoHzvIf3D-BqGRPUNqjP7puI)

Data generert hos Statistikkbanken til SSB 02.05.2020:  
06683: Adopsjoner, etter adopsjonstype 1986 - 2018. Statistikkbanken <sup>[6]</sup>

Fødte, år 2000. Statistikk hentet hos SSB 02.05.2020  
<https://www.ssb.no/befolkning/statistikker/fodte/arkiv/2001-05-29> <sup>[7]</sup>

Lødrup, Peter: *adopsjon i Store norske leksikon* på snl.no. Hentet 02.05.2020 fra  
<https://snl.no/adopsjon> <sup>[8]</sup>

Bakken, A. (2014). Ulike perspektiver på skoleresultatene til barn og unge med innvandringsbakgrunn. Oslo, NOVA. <sup>[9]</sup>

SSB 16.11.2017. Lavere lønn blant innvandrere.  
<https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/lavere-lonn-blant-innvandrere><sup>[11]</sup>

Samfunnsøkonomene, Økonomisk Forum Nr. 3. 2004:

<https://samfunnsokonomene.no/wp-content/uploads/2019/05/Trykkutgave-4-2012.pdf> [12]

### **Artikler:**

Bakken, Anders (2007): Virkninger av tilpasset språkopplæring for minoritetsspråklige elever. En kunnskapsoversikt. NOVA-rapport 10/07.

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual Children at the Crossfire*. Clevedon: Multilingual Matters Ltd.

### **Bøker:**

Thomas, R L. Using statistics in economics. McGraw-Hill Education. 2005 [10]

## Appendix

Do-file: Do-file Bacheloroppgave

Datsett: data2901-Norway.dta

