

Eira Mørch-Thoresen

Analyse av hvordan kjønn og fødselsmåned påvirker leseferdigheter hos elever i Norge og Sverige

Bacheloroppgave i Samfunnsøkonomi

Veileder: Bjarne Strøm

Mai 2020

Eira Mørch-Thoresen

Analyse av hvordan kjønn og fødselsmåned påvirker leseferdigheter hos elever i Norge og Sverige

Bacheloroppgave i Samfunnsøkonomi
Veileder: Bjarne Strøm
Mai 2020

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Sammendrag

I denne oppgaven utforskes det om det er forskjeller i leseferdighet hos gutter og jenter, og hvorvidt fødselsmåned har noen effekt på leseferdigheter. I tillegg analyseres forskjellen i leseferdighet mellom de to aktuelle landene Norge og Sverige. Dataene som brukes i analysen er hentet fra PIRLS-studien som ble gjennomført på elever i grunnskolen i mange land verden over i 2001. I denne oppgaven benyttes data fra Norge og Sverige. For å undersøke disse problemstillingene har OLS-regresjonsanalyse blitt benyttet. Forskjellige modeller som inkluderer ulike sett av variabler har blitt testet ut. Resultatet av undersøkelsene viser at det er signifikante forskjeller i leseferdighet mellom gutter og jenter, og mellom elever født tidlig eller sent på året. Effekten av fødselsmåned er ikke ulik for gutter og jenter. I tillegg er det en signifikant forskjell i resultatene for Norge og Sverige.

Innhold

Sammendrag	2
1 Introduksjon	5
1.1 PIRLS-studien	5
1.2 Skolesystemet i Norge og Sverige	5
1.2.1 Norge	5
1.2.2 Sverige	6
2 Tidligere studier	6
2.1 Fødselsmåned og prestasjon	6
2.2 Kjønn og prestasjon	7
2.3 Oppsummering	7
3 Data og deskriptiv analyse	7
3.1 Deskriptiv statistikk for avhengig variabel	8
3.2 Deskriptiv statistikk for kontrollvariabler	9
3.3 Korrelasjonskoeffisienter	10
3.4 Oppsummering	10
4 Økonometrisk modell	10
4.1 Teoretisk bakgrunn	11
4.2 Funksjonsform	11
4.3 Hypotesetesting	12
4.4 Oppsummering	12
5 Regresjonsanalyse	13
5.1 Resultater fra regresjonsanalyser	13
5.2 Hypotesetester	13

5.2.1	Test på fødselsmåned	13
5.2.2	Test på kjønn	14
5.2.3	Test på interaksjonsledd	15
5.2.4	Test på forskjeller i leseferdighet mellom Norge og Sverige	15
5.2.5	Test på tilleggsvariabler	16
5.3	Oppsummering	17
6	Diskusjon	17
7	Konklusjon	19
A	Stata outputvindu	i
B	Tabell- og figuroversikt	ii

1 Introduksjon

Når elever skal begynne på skolen er en pågående debatt, og det argumenteres stadig for at elever født sent på året bør få utsatt skolestart, eller at fleksibel skolestart bør innføres i grunnskolen. Disse argumentene begrunnes med at elever som er født sent ikke er modne nok til å motta kunnskap^[1], og at det burde være modenhet og ikke alder som bestemmer tidspunkt for skolestart^[2]. Forskjell mellom kjønnene er også noe som stadig debatteres^[3]. Er det noe i disse påstandene? Denne bacheloroppgaven ønsker å se nærmere på disse sammenhengene mellom leseferdighet, kjønn og fødselsmåned gjennom regresjonsanalyse av dataene fra PIRLS-studien fra Norge og Sverige i 2001.

De tre hovedpunktene denne bacheloroppgaven vil gjøre rede for er

1. Hva fødselsmåned har å si for leseferdighet
2. Om det er noen forskjell mellom gutter og jenters leseferdighet
3. Om det er noen forskjell i leseferdighet mellom Norge og Sverige

1.1 PIRLS-studien

Dataene som blir benyttet i regresjonsanalysen kommer fra PIRLS-undersøkelsen fra 2001. PIRLS (Progress in International Reading Literacy Study) er en internasjonal undersøkelse gjennomført for å teste leseferdighetene til fjerdeklassinger i grunnskolen. Testene ble gjennomført i 35 land og på rundt 150 000 elever i 5777 skoler. I tillegg til at elevene ble testet i leseferdighet, besvarte de et skjema med spørsmål om lesevaner, arbeidsvaner og interesser. Foreldre og lærere fylte også ut skjemaer med relevante spørsmål^[4].

For å kunne analysere leseferdighet (engelsk: literacy) er det nødvendig å kunne definere begrepet. Leseferdighet defineres av Store Norske Leksikon som evnen til “å kunne forstå og tolke skrift”^[5]. Den norske kortversjonen av PIRLS-rapporten skriver at informasjon om leseforståelse kan fås gjennom å teste elevene på å “hente ut informasjon, trekke enkle slutninger, tolke og sammenholde informasjon og vurdere språk, innhold og virkemidler i teksten”^[4]. Prøvene elevene tok inneholdt både litterære tekster og faktatekster, og underveis ble de bedt om å svare på spørsmål om disse. 60% av spørsmålene ble gitt som flervalgsoppgaver, og de resterende 40% hadde åpne svar der elevene selv måtte skrive svaret. De åpne spørsmålene kunne gi mellom 1 og 3 poeng. Resultatet av testene ga et internasjonalt gjennomsnitt på 500 poeng.

1.2 Skolesystemet i Norge og Sverige

1.2.1 Norge

I Norge starter elevene på barneskolen det året de fyller 6, og skolestart skjer etter sommerferien, altså i august. Det vil si at når skolen begynner vil elever som er født i januar allerede ha vært 6

år gamle i et halvt år, mens andre kan ennå ikke ha fylt 6 år. Dermed kan det potensielt oppstå et skille mellom elever hva gjelder utvikling, prestasjon og modenhet. Ellers er utdanning i Norge er gratis og obligatorisk for barn og ungdom mellom 6 og 16 år. Grunnskolen er delt opp i barneskolen (1.-7. trinn) og ungdomsskolen (8.-10. trinn)^[6]. I Norge var grunnskolen 9-årig frem til 1997, da det ble innført 10-årig grunnskole. Det betyr at før 1997 begynte norske elever på barneskolen det året de fylte 7^[7].

1.2.2 Sverige

I Sverige praktiserte de 9-årig grunnskole frem til 2018. Før dette var utdanning obligatorisk først fra det året barnet fyller 7 år. Det innebærer at den obligatoriske skolen begynte august året barnet fylte 7 år - ett år senere enn i Norge^[8]. Dette fører til den samme potensielle utfordringen som nevnt i avsnittet over. Likevel har Sverige en litt annen ordning enn vi har i Norge, nemlig at fra 1991 har det vært valgfritt for foreldre om barna skal begynne på skolen når de er 6 år^[9]. Dette kalles fleksibel skolestart. Dermed kan barn som er ekstra utviklet få begynne ett år tidligere på skolen. Dette tar derimot ikke hensyn til de barna som er litt senere utviklet og som kanskje ville hatt bedre av å utsette skolestarten. Fra 2018 ble denne fleksible skolestartordningen som kalles "förskoleklass" (tidligere årskurs 0) gjort obligatorisk^[10], og grunnskolen i Sverige er i dag 10-årig slik som i Norge. Likevel må en huske på at dataene som benyttes i denne oppgaven er fra PIRLS-undersøkelsen fra 2001, og at Sverige på dette tidspunktet praktiserte 9-årig grunnskole. Dermed var fjerdeklassingene som gjennomførte undersøkelsen i snitt ett år eldre enn de norske elevene, noe som også bekreftes av datamaterialet som benyttes i denne oppgaven. I Sverige er grunnskolen gratis på lik linje som i Norge og deles inn i førskoleklassen og ni klassetrinn (1.-9. trinn)^[11].

2 Tidligere studier

I dette avsnittet vil et utvalg tidligere studier som har forsket på problemstillingene nevnt i avsnitt 1 og deres resultater presenteres.

2.1 Fødselsmåned og prestasjon

En forskningsrapport fra Utdanningsdirektoratet kalt *Hva betyr fødselsmåned for skolerestultater og valg av utdanningsløp?*^[12] har kommet frem til at elever født sent på året presterer svakere på nasjonale prøver og at elever født sent på året går ut av grunnskolen med lavere karaktersnitt enn øvrige elever. Her er analysen basert på en sammensetning av resultater fra nasjonale prøver og elevers grunnskolepoeng.

En annen forskningsrapport er *Tjue år med TIMSS og PISA i Norge*^[13] fra UiO publisert i 2018. Der presenteres det de mener er de mest sentrale funnene for norsk skole gjennom 20 år på tvers av ulike undersøkelser som PISA, PIRLS og TIMSS. Resultatene de presenterer omhandler blant annet forskjeller i prestasjon mellom elever født tidlig og sent på året. Her konkluderes det med at aldersforskjell innenfor klassetrinn resulterer i store forskjeller i

skoleresultater, og at selv om disse forskjellene avtar med alderen, var det mulig å se slike tendenser helt frem til fylte 16 år.

2.2 Kjønn og prestasjon

Studien fra UiO^[13] nevnt i avsnitt 2.1 presenterer ett av hovedfunnene sine som: “det er små kjønnsforskjeller i realfagene, men jentene leser mye bedre enn guttene”. Dette resultatet er gjennomgående for de fleste rapporter som undersøker denne problemstillingen. Hvorfor det er slik har blitt forsket mye på, og en hypotese er at måten testene er utformet på er med på å favorisere jentene^[14]. I denne oppgaven forventes det å finne de samme resultatene, nemlig at jentene presterer bedre enn guttene.

2.3 Oppsummering

I dette avsnittet har det kort blitt utredet for et utvalg tidligere studier som har forsket på de samme problemstillingene som denne oppgaven ønsker å gjøre rede for. Disse konkluderer med at fødselsmåned har en innvirkning på leseferdighet og at jenter leser bedre enn gutter. Etttersom disse resultatene er såpass gjennomgående for de fleste forskningsrapporter, forventes de også å bli funnet i analysen som foretas i denne oppgaven.

3 Data og deskriptiv analyse

Datamaterialet som benyttes i denne oppgaven stammer som nevnt fra PIRLS-studien utført på skoleelever verden over i 2001. Studien testet elevers leseferdighet, samt at den samlet inn data om lesevaner, klassestørrelse, kjønn, alder, etc. I denne oppgaven er det benyttet data fra Norge og Sverige for å forsøke å svare på problemstillingen. Variablene som benyttes i regresjonsanalysen er

- *read* - dette er den avhengige variabelen i regresjonsanalysen. Dette er en kontinuerlig variabel som beskriver testscore på lesetesten i PIRLS-undersøkelsen.
- *girl* - dummyvariabel som beskriver om individet er jente eller gutt. Hvis variabelen er lik 1 er individet jente, og hvis den er lik 0 er individet gutt.
- *birthm* - kategorivariabel som beskriver fødselsmåned, der 1 er januar, 2 februar, helt frem til 12 som er desember.
- *notborn* - dummyvariabel som er lik 1 hvis individet ikke er født i landet testen tas i, og lik 0 hvis individet er født i landet.
- *clsiz* - kontinuerlig variabel som beskriver klassestørrelse gjennom antall elever i klassen.

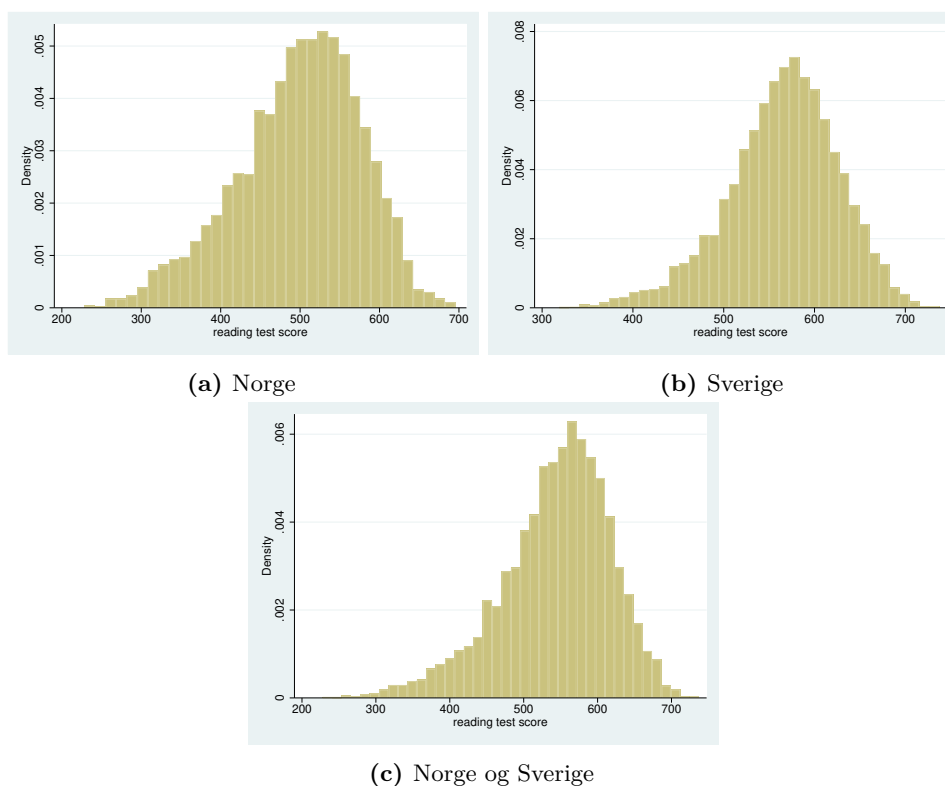
3.1 Deskriptiv statistikk for avhengig variabel

I tabell 3.1 presenteres deskriptiv statistikk for den avhengige variabelen *read*.

	Begge land	Norge	Sverige
Gjennomsnittlig leseferdighet	543.1392	498.2563	564.7048
Standardavvik	74.16086	78.36616	61.31277
Min	228.0606	228.0606	318.6813
Max	737.3258	695.8717	737.3258
Observasjoner	10658	3459	7199

Tabell 3.1: Deskriptiv statistikk for den avhengige variabelen *read*.

Fra dette kan det ses at Sverige har et tydelig høyere gjennomsnittlig leseferdighetsnivå enn det Norge har. Svenske elever scorer i snitt 66.5 poeng bedre enn norske elever. Svenskene ligger også ganske høyt over det internasjonale gjennomsnittet på 500 poeng, mens Norge ligger såvidt under. I tillegg har Sverige et lavere standardavvik enn Norge, noe som kan tyde på at Norge har høyere spredning i leseferdighet, mens svenske elever har en mer konsentrert score rundt gjennomsnittet. Dette underbygges også av at Norge har en større differanse mellom minimum og maksimum verdi enn Sverige (467.8111 mot 418.6445). Dette illustreres også gjennom histogrammene i figur 3.1.



Figur 3.1: Histogram for leseferdighet.

Denne tendensen om lav gjennomsnittscore og høy spredning, og motsatt, nevnes også i den norske kortversjonen av PIRLS-rapporten^[4]. Her vises det til at denne sammenhengen også gjelder mellom klasser internt, og at gode klasser som regel kjennetegnes av mange jevnt gode elever i stedet for noen få som drar opp snittet. Dette kan man også trekke paralleller til på landsbasis, nemlig at svenske elever presterer jevnt bra, mot de norske som har stor spredning i resultat.

3.2 Deskriptiv statistikk for kontrollvariabler

Her presenteres deskriptiv statistikk for kontrollvariablene *girl*, *birthm*, *notborn* og *clsiz*e i tabell 3.2, 3.3 og 3.4 under.

Norge og Sverige					
Variabel	Observasjoner	Gjennomsnitt	Standardavvik	Min	Max
<i>girl</i>	10465	0.4866699	0.4998462	0	1
<i>birthm</i>	10467	6.333333	3.366833	1	12
<i>notborn</i>	10383	0.1113358	0.3145627	0	1
<i>clsiz</i> e	10255	24.71234	8.287167	4	57

Tabell 3.2: Deskriptiv statistikk for kontrollvariabler i Norge og Sverige.

Norge					
Variabel	Observasjoner	Gjennomsnitt	Standardavvik	Min	Max
<i>girl</i>	3401	0.481035	0.4997137	0	1
<i>birthm</i>	3403	6.325889	3.400539	1	12
<i>notborn</i>	3355	0.0909091	0.2875226	0	1
<i>clsiz</i> e	3416	20.96165	4.82684	4	32

Tabell 3.3: Deskriptiv statistikk for kontrollvariabler for det norske datasettet.

Sverige					
Variabel	Observasjoner	Gjennomsnitt	Standardavvik	Min	Max
<i>girl</i>	7064	0.4893828	0.4999226	0	1
<i>birthm</i>	7064	6.33692	3.35071	1	12
<i>notborn</i>	7028	0.1210871	0.3262517	0	1
<i>clsiz</i> e	6839	26.58576	8.989558	8	57

Tabell 3.4: Deskriptiv statistikk for kontrollvariabler for det svenske datasettet.

Tabellene viser at de to landene ligger svært likt når det gjelder andel jenter og gjennomsnittlig fødselsmåned. Begge land har 48% jenter og i snitt er elevene født midt på året. I Sverige er klassene i snitt cirka 6 elever større enn Norge. I tillegg ser vi at Sverige har en høyere andel elever født utenlands enn det Norge har (9% mot 12%).

3.3 Korrelasjonskoeffisienter

Korrelasjonskoeffisienter for de aktuelle variablene presenteres i tabell 3.5. Korrelasjon er et mål på hvor mye to variabler samvarierer. Styrken av korrelasjonen beskrives av korrelasjonskoeffisienten. Denne koeffisienten er et tall mellom -1 og 1 som beskriver graden av samvariasjon^[15].

Norge og Sverige					
	<i>read</i>	<i>girl</i>	<i>birthm</i>	<i>notborn</i>	<i>clsize</i>
<i>read</i>	1.0000				
<i>girl</i>	0.1536	1.0000			
<i>birthm</i>	-0.0968	0.0005	1.0000		
<i>notborn</i>	-0.1677	-0.0308	0.0403	1.0000	
<i>clsize</i>	0.1661	0.0126	0.0060	0.0079	1.0000

Tabell 3.5: Korrelasjonskoeffisienter for datamaterialet.

Fra tabellen presentert over ser vi at kontrollvariablene er korrelert med den avhengige variabelen *read* og interessevariablene *girl* og *birthm*. Dette viser viktigheten av å inkludere andre variabler enn kun de mest interessante, ettersom å utelukke variabler kan føre til et utelatt variabel-problem. Et utelatt variabel-problem oppstår hvis man utelater en variabel fra analysen som er determinant for den avhengige variabelen (koeffisienten er ulik null), og at den utelatte variabelen er korrelert med en av forklaringsvariablene inkludert i modellen. Det vil si at all endring i den avhengige variabelen forklares som en effekt av de inkluderte forklaringsvariablene. Et eksempel på dette kan være en modell som kun bruker én forklaringsvariabel, selv om det er andre variabler som også tydelig påvirker den avhengige variabelen og korrelerer med den inkluderte forklaringsvariabelen. Dette vil føre til en skjevhet i estimatene. Det er derfor viktig å teste for forskjellige kombinasjoner av forklaringsvariabler. Et viktig hjelpemiddel her kan være å se på R^2 som er et mål på hvor mye av variasjonen som forklares av modellen.

3.4 Oppsummering

I dette avsnittet har beskrivelse av de relevante variablene blitt lagt frem. I tillegg har deskriptiv statistikk og korrelasjonskoeffisienter blitt presentert.

4 Økonometrisk modell

I denne seksjonen vil det gås gjennom teoretisk bakgrunn for regresjonsanalyse og hypotesetesting, og det vil i tillegg bli beskrevet hvilke funksjonsformer som vil brukes i regresjonsanalysen.

4.1 Teoretisk bakgrunn

For å kunne gjøre analyse på datasettet og de ønskede variablene benyttes lineær regresjon (OLS). Dette krever at man har en økonometrisk modell. Generelt er den økonometriske modellen gitt av

$$Y_i = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \varepsilon_i \quad (4.1)$$

der Y_i er den avhengige variabelen, β_0 konstantledd, X_i forklaringsvariabler og β_i koeffisienter. I denne oppgaven skal det analyseres hvilken effekt fødselsmåned og kjønn har på leseferdigheter. Da er det fornuftig å benytte *read* som avhengig variabel, og *girl* og *birthm* som forklaringsvariabler. Det skal også ses på forskjellen i leseferdighet mellom Norge og Sverige, og det vil derfor bli benyttet en dummyvariabel for Sverige, slik at Norge er referanseland. I tillegg er variabelen ε_i en stokastisk variabel som ofte kalles residualen eller restleddet. ε_i beskriver all variasjon som resten av modellen ikke klarer å fange opp, og har følgende egenskaper

$$E(\varepsilon_i) = 0 \quad (\text{i})$$

$$Var(\varepsilon_i) = \sigma^2 \quad (\text{ii})$$

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \quad (\text{iii})$$

$$\varepsilon_i - \text{Normalfordelt} \quad (\text{iv})$$

For å gjennomføre regresjonsanalysen benyttes programmet Stata. I denne oppgaven vil det kun bli benyttet lineær regresjon, noe som betyr at sammenhengen mellom den avhengige variabelen og kontrollvariablene estimeres som en rett linje. OLS-metoden fungerer slik at den prøver å minimisere differansen mellom observerte data og den estimerte regresjonslinjen. Altså ønsker man å finne de koeffisientene β_i slik at modellen i ligning (4.1) best mulig passer med de observerte dataene fra datasettet.

4.2 Funksjonsform

Generelt skal sammenhengen mellom leseferdighet og kjønn og fødselsmåned analyseres. Dermed kan en generell funksjon utformes som

$$read = f(girl, birthm, X) \quad (4.2)$$

med *girl* og *birthm* som interessevariabler, og der X representerer andre mulige variabler som påvirker *read*. Basert på teorien utledet i avsnitt 4.1 kan ligning (4.2) spesifiseres i følgende utvalg av modeller

$$read_i = \beta_0 + \beta_1 birthm_i + \alpha_1 dSverige_i + \varepsilon_i \quad (\text{A})$$

$$read_i = \beta_0 + \beta_2 girl_i + \alpha_1 dSverige_i + \varepsilon_i \quad (\text{B})$$

$$read_i = \beta_0 + \beta_1 birthm_i + \beta_2 girl_i + \alpha_1 dSverige_i + \varepsilon_i \quad (\text{C})$$

$$read_i = \beta_0 + \beta_1 birthm_i + \beta_2 girl_i + \beta_3 girl_i \cdot birthm_i + \alpha_1 dSverige_i + \varepsilon_i \quad (\text{D})$$

$$read_i = \beta_0 + \beta_1 birthm_i + \beta_2 girl_i + \beta_4 notborn + \beta_5 clsiz + \alpha_1 dSverige_i + \varepsilon_i \quad (\text{E})$$

der variablene er som beskrevet i avsnitt 3. I tillegg har vi variabelen *dSverige* som er en dummyvariabel for Sverige slik at Norge benyttes som referanseland i analysen. De tre første

modellene (A)-(C) inkluderer kun interessevariablene *girl* og *birthm*, og er ganske enkle. I modell (D) inkluderes et interaksjonsledd. Dette er for å teste om effekten av fødselsmåned er forskjellig for gutter og jenter. Til slutt tas det med to andre variabler (*notborn* og *clsiz*) for å se om det er andre ting som påvirker *read*, og om modellen da får større forklaringskraft.

4.3 Hypotesetesting

Videre vil det bli brukt hypotesetester på de estimerte modellene. En hypotesetest er en statistisk testmetode for å undersøke en påstand om egenskaper ved en eller flere populasjoner^[16]. For å kunne utføre en hypotesetest må man utforme en nullhypotese H_0 , og i tillegg en motsigende hypotese H_1 . Når man utfører en hypotesetest undersøkes det om det er grunnlag for å forkaste eller beholde nullhypotesen. Dersom nullhypotesen forkastes, betyr det at den alternative hypotesen aksepteres siden de to hypotesene er komplementære.

De to testene som vil benyttes i denne oppgaven er t-test og F-test. En t-test benyttes når man vil undersøke om en parameter er lik en gitt verdi. For å utføre en slik test brukes det en testobservator (TS) gitt av

$$TS = \left| \frac{b_i - \beta_i}{s_{b_i}} \right| \quad (4.3)$$

der b_i er estimator for β_i , s_{b_i} er estimator for standardavviket og β_i er den verdien vi ønsker å teste for. Dersom testobservatoren er større enn den kritiske verdien $t_{\alpha, n-k}$ (som kan finnes i Tabeller og formler i statistikk^[17]), forkastes nullhypotesen. α er valgt signifikansnivå (i denne oppgaven benyttes $\alpha = 0.05$), n er antall observasjoner og k er antall parametre i modellen.

En F-test benyttes ofte når man vil sammenligne flere økonometriske modeller for samme datasett, for å se hvilken modell som best passer til populasjonen. I en F-test kan flere variabler testes samtidig. Et eksempel på dette kan ses på modell E. Hvis man ønsker å undersøke om de to variablene *notborn* og *clsiz* har noen effekt på *read* kan man formulere følgende hypotesetest:

$$\begin{aligned} H_0 : \beta_4 = \beta_5 = 0 \\ H_1 : \text{Minst én av } \beta_4, \beta_5 \neq 0 \end{aligned}$$

Testobservatoren som benyttes i en F-test er gitt av

$$TS = \frac{(SSR_r - SSR_u)/h}{SSR_u/(n-k)} \quad (4.4)$$

der vi har at SSR_r er sum of squared residuals for modellen med restriksjoner, SSR_u er sum of squared residuals for modellen uten restriksjoner, h er antall restriksjoner, n er antall observasjoner og k er antall parametre. Slik som i en t-test forkastes nullhypotesen hvis testobservatoren er større enn den kritiske verdien. Her finnes den kritiske verdien i Fisherfordelingen i Tabeller og formler i statistikk^[17] og er gitt av $F_\alpha(h, n-k)$.

4.4 Oppsummering

I dette avsnittet har det blitt gjort rede for teorien som brukes i analysen. Den økonometriske modellen og valg av denne har blitt beskrevet, og det har deretter blitt forklart hvilke

hypotesetester som vil bli benyttet og teorien bak disse.

5 Regresjonsanalyse

I dette avsnittet vil regresjonsanalyse for de forskjellige modellene A-E bli utført. Fullstendig outputvindu fra Stata kan sees i appendiks A. Videre vil det utføres hypotesetester på koeffisientene i modellene.

5.1 Resultater fra regresjonsanalyser

Resultatene fra regresjonsanalysen er oppsummert i tabell 5.1 under.

Variabel	Modell A read	Modell B read	Modell C read	Modell D read	Modell E read
dSverige	65.47372 (1.395991)	65.54619 (1.387161)	65.05996 (1.378493)	65.06656 (1.378507)	64.39823 (1.456546)
girl		22.13863 (1.299872)	22.21397 (1.290546)	24.68145 (2.748855)	21.53935 (1.296179)
birthm	-2.143608 (0.1942299)		-2.141503 (0.1915516)	-1.952617 (0.266854)	-2.007279 (0.1929839)
girlbirthm				-0.3896652 (0.3832845)	
notborn					-41.84632 (2.075153)
clsize					0.3074201 (0.0817788)
Konstantledd	512.6805 (1.680729)	488.1899 (1.299911)	502.2097 (3.053556)	501.0095 (2.128376)	499.4561 (2.461231)
Observasjoner	10467	10465	10447	10447	9942
R squared	0.1815	0.1951	0.2030	0.2031	0.2366

Tabell 5.1: Oppsummering av regresjonsanalyser. Standardavvik i parentes.

5.2 Hypotesetester

I dette avsnittet vil det utføres hypotesetester på de forskjellige modellene for å undersøke problemstillingene nevnt i avsnitt 1.

5.2.1 Test på fødselsmåned

En problemstilling denne oppgaven ønsker å undersøke er om fødselsmåned har noen effekt på leseferdighet. Da er det fornuftig å utføre en hypotesetest på om effekten av *birthm* på *read*

er signifikant. Dette kan formuleres i følgende hypotesetest:

$$\begin{aligned}H_0 : \beta_1 &= 0 \\H_1 : \beta_1 &\neq 0\end{aligned}$$

Begynner med å se på modell A. Testobservatoren blir

$$TS = \left| \frac{b_1 - \beta_1}{s_{b_1}} \right| = \left| \frac{-2.143608 - 0}{0.1942299} \right| = 11.03644701$$

som tydelig er større enn den kritiske verdien $t_{\frac{0.05}{2}, 10467-3} \approx 1.960$ som er funnet i Tabeller og formler i statistikk^[17]. Nullhypotesen kan dermed forkastes, og det kan konkluderes med at effekten av fødselsmåned på leseferdigheter er signifikant.

Med samme fremgangsmåte på modell C har vi

$$TS = \left| \frac{b_1 - \beta_1}{s_{b_1}} \right| = \left| \frac{-2.141503 - 0}{0.1915516} \right| = 11.17977088 \quad \text{og} \quad t_{\frac{0.05}{2}, 10447-3} \approx 1.960$$

slik at nullhypotesen kan forkastes her også.

Modell D gir oss at

$$TS = \left| \frac{-1.952617 - 0}{0.266854} \right| = 7.31717 \quad \text{og} \quad t_{\frac{0.05}{2}, 10447-3} \approx 1.960$$

slik at $TS > t$ og nullhypotesen kan forkastes. Samme resultat kommes frem til i modell E med

$$TS = \left| \frac{-2.007279 - 0}{0.1929839} \right| = 10.401277 \quad \text{og} \quad t_{\frac{0.05}{2}, 9942-3} \approx 1.960$$

5.2.2 Test på kjønn

For å se om det er en signifikant forskjell mellom gutter og jenters leseferdigheter kan det utføres en test på β_2 i modellene. Dette gir følgende hypotesetest:

$$\begin{aligned}H_0 : \beta_2 &= 0 \\H_1 : \beta_2 &\neq 0\end{aligned}$$

Testobservatoren for modell B blir

$$TS = \left| \frac{b_2 - \beta_2}{s_{b_2}} \right| = \left| \frac{22.13863 - 0}{1.299872} \right| = 17.03139232$$

Hvis nullhypotesen skal forkastes må TS være høyere enn kritisk verdi i t-fordelingen. Denne finnes i Tabeller og formler i statistikk^[17] og er $t_{\frac{0.05}{2}, 10465-3} \approx 1.960$. Fra dette kan det ses at $TS > t$ og nullhypotesen forkastes på et 95% signifikansnivå. Dermed kan det konkluderes med at blant norske elever er det en signifikant forskjell i leseferdigheter mellom gutter og jenter.

Den samme testen utført på modell C gir følgende testobservator

$$TS = \left| \frac{22.21397 - 0}{1.290546} \right| = 17.2128463$$

Den kritiske verdien er gitt av $t_{\frac{0.05}{2}, 10465-4} \approx 1.960$, og testobservatoren er tydelig større enn denne. Det betyr at nullhypotesen kan forkastes på et 95% signifikansnivå.

Den samme testen blir utført på modell D og resulterer i

$$TS = \left| \frac{24.68145 - 0}{2.78855} \right| = 8.850998 \quad \text{og} \quad t_{\frac{0.05}{2}, 10447-3} \approx 1.960$$

Den siste modellen gir oss at

$$TS = \left| \frac{21.53935 - 0}{1.296179} \right| = 16.61757 \quad \text{og} \quad t_{\frac{0.05}{2}, 9942-3} \approx 1.960$$

Nullhypotesen om at effekten av kjønn på leseferdighet kan dermed forkastes i alle modellene på 95% signifikansnivå.

5.2.3 Test på interaksjonsledd

Er effekten av fødselsmåned forskjellig for gutter og jenter? Hvis effekten av fødselsmåned er forskjellig for gutter og jenter betyr det at β_3 i modell D er ulik null. Kan da formulere følgende hypotesetest:

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

For å sjekke nullhypotesen benyttes en t-test, og testobservatoren som er gitt av (4.3) blir

$$TS = \left| \frac{b_3 - \beta_3}{s_{b_3}} \right| = \left| \frac{-0.3896652 - 0}{0.3832845} \right| = 1.016647425$$

T-verdien for signifikansnivå $\alpha = 0.05$ er gitt av

$$t_{\frac{0.05}{2}, 10447-4} \approx 1.960$$

og er funnet i Tabeller og formler i statistikk^[17]. Siden vi har at $TS < t$ kan vi ikke forkaste nullhypotesen, og det er dermed rimelig å anta at det ikke er noe forskjell i effekt av fødselsmåned på leseferdigheter hos gutter og jenter. Dette kan konkluderes på 95% signifikansnivå.

5.2.4 Test på forskjeller i leseferdighet mellom Norge og Sverige

Her ønsker vi å undersøke om forskjellen i leseferdighet mellom Norge og Sverige er signifikant. Dette kan gjøre ved å utføre en hypotesetest på koeffisienten α_1 . Dette kan formuleres i følgende hypotesetest:

$$H_0 : \alpha_1 = 0$$

$$H_1 : \alpha_1 \neq 0$$

Dersom nullhypotesen holder, har vi at *read* i modellene er like uavhengig av land. Begynner med å teste på modell A.

$$TS = \left| \frac{a_1 - \alpha_1}{s_{a_1}} \right| = \left| \frac{65.47372 - 0}{1.395991} \right| = 46.90124793$$

Den kritiske verdien i t-fordelingen (funnet i Tabeller og formler i statistikk^[17]) er $t_{\frac{0.05}{2}, 10467-3} \approx 1.960$. Det kan tydelig ses at $TS > t$, og nullhypotesen kan dermed forkastes. Altså er det en signifikant forskjell i leseferdigheter mellom Norge og Sverige, der Sverige har en bedre score enn det Norge har.

For modell B blir testobservatoren

$$TS = \left| \frac{65.54619 - 0}{1.387161} \right| = 47.25204212$$

og den kritiske verdien i t-fordelingen er $t_{\frac{0.05}{2}, 10465-3} \approx 1.960$. Dette leder til samme resultat som for modell A, altså at nullhypotesen forkastes.

Samme prosedyre for modell C leder til

$$TS = \left| \frac{65.05996 - 0}{1.378493} \right| = 47.19643843$$

med kritisk verdi $t_{\frac{0.05}{2}, 10447-4} \approx 1.960$. Her forkastes også nullhypotesen, og vi kommer frem til at forskjellen i leseferdighet mellom de to landene er signifikant.

Modell D gir oss at

$$TS = \left| \frac{65.06656 - 0}{1.378507} \right| = 47.2007469$$

og $t_{\frac{0.05}{2}, 10447-5} \approx 1.960$. Også her forkastes nullhypotesen.

Til slutt har vi at modell E gir

$$TS = \left| \frac{64.39823 - 0}{1.456546} \right| = 44.21297371$$

og $t_{\frac{0.05}{2}, 9942-6} \approx 1.960$. Det kan ses at $TS > t$ og også her kan nullhypotesen forkastes.

I alle modellene kan nullhypotesen forkastes, og det kan dermed konkluderes med 95% sannsynlighet at forskjellen i leseferdighet mellom Norge og Sverige er signifikant.

5.2.5 Test på tilleggsvariabler

I modell E innføres det et knippe nye variabler for å se om det kan føre til et modellen får en større forklaringskraft. Det kan dermed utføres en test på om disse nye innførte forklaringsvariablene er signifikante. Dette gjøres ved hjelp av en F-test som forklart i avsnitt 4.3. De to ekstra variablene i modell E er *notborn* og *clsiz* med tilhørende koeffisienter β_4 og β_5 . Det gir følgende hypotesetest:

$$\begin{aligned} H_0 &: \beta_4 = \beta_5 = 0 \\ H_1 &: \text{Minst én av } \beta_4, \beta_5 \neq 0 \end{aligned}$$

Testobservatoren som benyttes i en F-test er gitt av

$$TS = \frac{(SSR_r - SSR_u)/h}{SSR_u/(n - k)}$$

og inkluderer SSR for modellen med og uten restriksjoner. Modellen med restriksjoner (altså der $\beta_4 = \beta_5 = 0$) er den samme som modell C. SSR_r kan finnes fra Stata-vinduet i figur A.1c i appendiks A, og er gitt av $SSR_r = 45390709.4$. For modellen uten restriksjoner (modell E) er $SSR_u = 41407288.7$ (kan ses i figur A.1e). Testobservatoren kan da beregnes:

$$TS = \frac{(45390709.4 - 41407288.7)/2}{41407288.7/(9942 - 6)} = 477.9263505$$

Kritisk verdi i Fisherfordelingen er gitt av $F_{0.05}(2, 9942 - 6) = F_{0.05}(2, \infty) = 3.00$ ^[17] og vi har dermed at $TS > F$. Altså forkastes nullhypotesen, og det konkluderes med at minst én av β_4, β_5 er ulik null på 95% signifikansnivå. Det betyr at enten klassestørrelse eller om eleven er født i landet (eller begge) har en signifikant innvirkning på leseferdighet.

I tillegg kan man se fra tabell 5.1 at koeffisienten foran $dSverige$ synker fra 65.05669 til 64.39823 når $notborn$ og $clsiz$ inkluderes. Det kan tyde på at leseferdighetene mellom de to landene kan avhenge av systematiske forskjeller i klassestørrelse og om eleven er født i landet eller ikke. Dette underbygges i avsnitt 3.2, der det ses at Sverige både har i snitt større klasser og høyere andel elever født utenlands. Likevel er denne reduksjonen veldig liten, på under ett poeng.

5.3 Oppsummering

I dette avsnittet har resultatene fra regresjonsanalysen på modellene A til E blitt presentert, og det har i tillegg blitt utført hypotesetester for å evaluere problemstillingene nevnt i avsnitt 1.

6 Diskusjon

I avsnitt 5 ble regresjonsanalyser og hypotesetester fremlagt. Når det gjelder effekten av fødselsmåned på leseferdighet har det blitt konkludert med at effekten er signifikant i alle modeller. Fra modell A har vi at marginaleffekten av $birthm$ på $read$ er $\frac{\partial read}{\partial birthm} = -2.144$, og fra modell C er den $\frac{\partial read}{\partial birthm} = -2.142$. Denne negative sammenhengen gjelder også i øvrige modeller der $birthm$ er inkludert. Altså er det en negativ sammenheng mellom fødselsmåned og leseferdighet, og det kan ses at elever som er født sent på året presterer i snitt dårligere på lesetester enn elever født tidligere på året. Dette stemmer også godt med forskningsrapporten fra Utdanningsdirektoratet^[12] sine konklusjoner.

Videre ble effekten av kjønn på leseferdigheter estimert, og det ble funnet ut at forskjellen mellom gutter og jenters leseferdigheter er signifikant. Det er en positiv sammenheng mellom $girl$ og $read$. Dette kan ses ved at $\frac{\partial read}{\partial girl} = 22.14 > 0$ og $\frac{\partial read}{\partial girl} = 22.21 > 0$ i henholdsvis modell B og modell C. De andre modellene der $girl$ er inkludert gir også at $\frac{\partial read}{\partial girl} > 0$. Denne positive

partiellderiverte gir oss informasjon om at jenter har bedre leseferdigheter enn gutter. Også dette stemmer godt overens med tidligere studier nevnt i 2.2.

Testen på interaksjonsleddet konkluderte med at det ikke er grunnlag for å si at effekten av fødselsmåned er forskjellig for gutter og jenter. Altså vil gjennomsnittlig lesescore synke like mye for gutter og jenter for hver måned senere de er født. Siden det ble konkludert med at interaksjonsleddet ikke er nødvendig, kan sammenhengen mellom fødselsmåned og kjønn oppsummeres i tabell 6.1 og 6.2 under for modell C.

	Jan	Feb	Mar	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Des
Gutter	500.1	497.9	495.8	493.6	491.5	489.4	487.2	485.1	482.9	480.8	478.7	476.5
Jenter	522.3	520.1	518.0	515.9	513.7	511.6	509.4	507.3	505.2	503.0	500.9	498.7

Tabell 6.1: Oppsummering av modell C avrundet til 4 gjeldende siffer for Norge.

	Jan	Feb	Mar	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Des
Gutter	565.1	563.0	560.8	558.7	556.6	554.4	552.3	550.1	548.0	545.9	543.7	541.6
Jenter	587.3	585.2	583.1	580.9	578.8	576.6	574.5	572.4	570.2	568.1	565.9	563.8

Tabell 6.2: Oppsummering av modell C avrundet til 4 gjeldende siffer for Sverige.

Her kan det ses at for begge land scorer jentene cirka 22 poeng høyere enn guttene, og at for hver fødselsmåned synker lesescore med omtrent 2 poeng. Det resulterer i en forskjell mellom elever født i januar og desember på hele 23 poeng. Ser man på forskjellen mellom en jente født i januar og en gutt født i desember blir den gjennomsnittlige differansen enda større, faktisk på rundt 45 poeng.

Det neste som ble sett på var forskjellen i leseferdighet mellom Norge og Sverige. Det ble gjort gjennom å utføre en t-test på α_1 . Her kunne nullhypotesen forkastes for samtlige modeller (A-E), og det kunne dermed fastslås med 95% sikkerhet at leseferdighetene i de to landene er forskjellige. Her er det viktig å få med seg at elevene i Sverige begynte ett år senere på skolen enn de norske elevene (7 år kontra 6). Det underbygges også i datamaterialet der gjennomsnittlig alder på de svenske elevene er 10.79 år mot den norske gjennomsnittsalderen på 9.97 år. Siden de svenske elevene er ett år mer utviklet, er det også å forvente at de presterer bedre.

Til slutt ble det utført en test på tilleggsvariablene *notborn* og *clsiz* for å se om modellen fikk større forklaringskraft. Etttersom resultatet av F-testen konkluderte med at effektene av nevnte variabler var signifikant, kan det fastslås at minst én av disse også påvirker leseferdighet. Effekten av *notborn* på *read* er gitt av $\frac{\partial read}{\partial notborn} = -41.8 < 0$. Det betyr at elever som er født utenlands vil i snitt prestere ganske mye dårligere enn elever født i landet testen tas i. Effekten av klassestørrelse derimot er mye lavere ($\frac{\partial read}{\partial clsiz} = 0.31$).

Selv når kontrollvariablene innføres i modell E, endres ikke koeffisientene foran interessevariablene *girl* og *birthm* noe særlig. I tillegg viser t-testene utført på nevnte interessevariabler at effekten fortsatt er signifikant. Dette tyder på at det ikke har vært et utelatt variabel-problem, noe som også underbygges av de lave korrelasjonskoeffisientene mellom kontrollvariablene og interessevariablene i tabell 3.5. Det kan likevel være andre variabler i datasettet som gir et

utelatt variabel-problem. Dette kan være verdt å utforske i fremtidig arbeid.

7 Konklusjon

I denne oppgaven har det blitt kommet frem til følgende resultater

1. Effekten av fødselsmåned på leseferdigheter er signifikant. Jo senere eleven er født på året, desto dårligere presterer han/hun.
2. Jenter har bedre leseferdigheter enn gutter.
3. Effekten av fødselsmåned er ikke forskjellig for jenter og gutter.
4. Svenske elever har bedre leseferdigheter enn norske.
5. De ekstra variablene for klassestørrelse og om eleven er født i landet har en signifikant innvirkning på leseferdighet.

Disse konklusjonene har blitt kommet frem til gjennom regresjonsanalyse og hypotesetester på 95% signifikansnivå. I tillegg er deskriptiv statistikk av de relevante variablene lagt frem. For fremtidig forskning kunne en mulig problemstilling være å utforske hvorvidt effekten av de ulike variablene er ikke-lineær eller om det er andre økonometriske modeller som passer bedre enn de foreslått i denne oppgaven. En annen forbedring kan være å inkludere flere land i analysen for å komme frem til et mer generelt resultat, eller å teste for flere ulike kombinasjoner av kontrollvariabler. Disse ulike settene av kontrollvariabler kunne igjen vært gruppert med variabler som hører sammen, slik som for eksempel etnisitet eller skoleforhold.

A Stata outputvindu

```
. reg read birthm d_Sverige
```

Source	SS	df	MS			
Model	10383236.6	2	5191618.28	Number of obs =	10,467	
Residual	46833063.4	10,464	4475.6368	F(2, 10464) =	1159.97	
Total	57216300	10,466	5466.87369	Prob > F =	0.0000	
				R-squared =	0.1815	
				Adj R-squared =	0.1813	
				Root MSE =	66.9	

	read	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
birthm		-2.143608	.1942299	-11.04	0.000	-2.524336 -1.762881
d_Sverige		65.47372	1.395991	46.90	0.000	62.73732 68.21013
_cons		512.6805	1.680729	305.03	0.000	509.3859 515.975

(a) Outputvindu regresjon modell A

```
. reg read girl d_Sverige
```

Source	SS	df	MS			
Model	11200069.1	2	5600034.56	Number of obs =	10,465	
Residual	46212549.9	10,462	4417.18122	F(2, 10462) =	1267.78	
Total	57412619.1	10,464	5486.67996	Prob > F =	0.0000	
				R-squared =	0.1951	
				Adj R-squared =	0.1949	
				Root MSE =	66.462	

	read	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
girl		22.13863	1.299872	17.03	0.000	19.59063 24.68662
d_Sverige		65.54619	1.387161	47.25	0.000	62.82709 68.26529
_cons		488.1899	1.299911	375.56	0.000	485.6418 490.7379

(b) Outputvindu regresjon modell B

```
. reg read birthm girl d_Sverige
```

Source	SS	df	MS			
Model	11539725	3	3853241.67	Number of obs =	10,447	
Residual	45390709.4	10,443	4346.5201	F(3, 10443) =	886.51	
Total	56950434.4	10,446	5451.88919	Prob > F =	0.0000	
				R-squared =	0.2030	
				Adj R-squared =	0.2027	
				Root MSE =	65.928	

	read	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
birthm		-2.141503	.1915516	-11.18	0.000	-2.516981 -1.766026
girl		22.21397	1.299546	17.21	0.000	19.68425 24.74368
d_Sverige		65.05996	1.378493	47.20	0.000	62.35795 67.76207
_cons		502.2097	1.770935	283.58	0.000	498.7384 505.6811

(c) Outputvindu regresjon modell C

```
. reg read birthm girl girlbirthm d_Sverige
```

Source	SS	df	MS			
Model	11564217.5	4	2891054.36	Number of obs =	10,447	
Residual	45386217	10,442	4346.50613	F(4, 10442) =	665.14	
Total	56950434.4	10,446	5451.88919	Prob > F =	0.0000	
				R-squared =	0.2031	
				Adj R-squared =	0.2028	
				Root MSE =	65.928	

	read	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
birthm		-1.952617	.266854	-7.32	0.000	-2.475702 -1.429532
girl		24.68145	2.748855	8.98	0.000	19.29317 30.06973
girlbirthm		-.389652	.3832845	-1.02	0.309	-1.140976 .3616456
d_Sverige		65.06656	1.378507	47.20	0.000	62.36443 67.7687
_cons		501.0095	2.128376	235.40	0.000	496.8375 505.1815

(d) Outputvindu regresjon modell D

```
. reg read girl birthm not_born clsiz d_Sverige
```

Source	SS	df	MS			
Model	12831937.8	5	2566387.56	Number of obs =	9,942	
Residual	41407288.7	9,936	4167.40023	F(5, 9936) =	615.82	
Total	54239226.5	9,941	5456.11372	Prob > F =	0.0000	
				R-squared =	0.2366	
				Adj R-squared =	0.2362	
				Root MSE =	64.555	

	read	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
girl		21.53935	1.296179	16.62	0.000	18.99058 24.08013
birthm		-2.007279	.1929839	-10.40	0.000	-2.385567 -1.628991
not_born		-41.84632	2.075153	-20.17	0.000	-45.91404 -37.77859
clsiz		.3074201	.0817788	3.76	0.000	.1471172 .4677231
d_Sverige		64.39823	1.456546	44.21	0.000	61.5431 67.25335
_cons		499.4561	2.461231	202.93	0.000	494.6315 504.2806

(e) Outputvindu regresjon modell E

Figur A.1: Stata-vindu fra regresjonsanalyse.

B Tabell- og figuroversikt

Figurer

3.1	Histogram for leseferdighet.	8
A.1	Stata-vindu fra regresjonsanalyse.	i

Tabeller

3.1	Deskriptiv statistikk for den avhengige variabelen <i>read</i>	8
3.2	Deskriptiv statistikk for kontrollvariabler i Norge og Sverige.	9
3.3	Deskriptiv statistikk for kontrollvariabler for det norske datasettet.	9
3.4	Deskriptiv statistikk for kontrollvariabler for det svenske datasettet.	9
3.5	Korrelasjonskoeffisienter for datamaterialet.	10
5.1	Oppsummering av regresjonsanalyser. Standardavvik i parentes.	13
6.1	Oppsummering av modell C avrundet til 4 gjeldende siffer for Norge.	18
6.2	Oppsummering av modell C avrundet til 4 gjeldende siffer for Sverige.	18

Referanser

- [1] Kjell G. Salvanes. Når bør barna begynne på skolen? <https://www.aftenposten.no/meninger/kronikk/i/A2AQx/naar-boer-barna-begynne-paa-skolen?>, 2014. [Aksessert 7.5.20].
- [2] Åshild Bruun-Gundersen. Når alder blir barnas største fiende. <https://www.dagbladet.no/kultur/nar-alder-blir-barnas-storste-fiende/69893541>, 2018. [Aksessert 7.5.20].
- [3] Astrid Meland. Ja, det er forskjell på kjønnene. <https://www.vg.no/nyheter/meninger/i/G1We56/ja-det-er-forskjell-paa-kjoennene>, 2019. [Aksessert 7.5.20].
- [4] Ragnar Gees Solheim og Finn Egil Tønnessen. *PIRLS. En norsk kortversjon av den internasjonale rapporten om 10-åringers lesekunnskaper*. Senter for leseforskning, 2003.
- [5] Store Norske Leksikon. Leseferdighet. <https://snl.no/leseferdighet>, 2016. [Aksessert 7.5.20].
- [6] Store Norske Leksikon. Skole og utdanning i norge. https://snl.no/Skole_og_utdanning_i_Norge, 2019. [Aksessert 17.3.20].
- [7] Wikipedia. Grunnskole i norge. https://no.wikipedia.org/wiki/Grunnskole_i_Norge, 2015. [Aksessert 7.5.20].
- [8] Wikipedia. Grundskolan i sverige. https://sv.wikipedia.org/wiki/Grundskolan_i_Sverige, 2019. [Aksessert 19.3.20].
- [9] Store Norske Leksikon. Skole og utdanning i sverige. https://snl.no/Skole_og_utdanning_i_Sverige, 2013. [Aksessert 17.3.20].
- [10] Wikipedia. Førskoleklass. <https://sv.wikipedia.org/wiki/Förskoleklass>, 2020. [Aksessert 7.5.20].
- [11] Nordisk samarbeid. Grunnskole i sverige. <https://www.norden.org/no/info-norden/grunnskole-i-sverige>, 2019. [Aksessert 7.5.20].
- [12] Utdanningsdirektoratet. Hva betyr fødselsmåned for skolerresultater og valg av utdanningsløp? <https://www.udir.no/tall-og-forskning/finn-forskning/tema/hva-betyr-fodselsmaned-for-skolerresultater-og-valg-av-utdanningslop>, 2019. [Aksessert 17.3.20].
- [13] UiO Julius Kristjan Björnsson og Rolf Vegar Olsen. Tjue år med timss og pisa i norge. https://www.duo.uio.no/bitstream/handle/10852/71257/tjue_aar_med_timss_og_pisa_i_norge.pdf?sequence=2&isAllowed=y, 2018. [Aksessert 24.3.20].
- [14] Hildegunn Støle Katrin Schulz-Heidorf. Gender differences in norwegian pirls 2016 and epirls 2016 results at test mode, text and item format level. <https://nordicliteracy.net/index.php/njlr/article/view/1270/3007>, 2018. [Aksessert 7.5.20].

- [15] Store Norske Leksikon. Korrelasjon. <https://snl.no/korrelasjon>, 2019. [Aksessert 6.5.20].
- [16] Wikipedia. Hypotesetest. <https://no.wikipedia.org/wiki/Hypotesetest>, 2017. [Aksessert 20.4.20].
- [17] Jan Terje Kvaløy og Håkon Tjelmeland. *Tabeller og formler i statistikk. Institutt for matematiske fag, NTNU*. Fagbokforlaget, 2001.

