

Sofie Bakke og Alexandra Saphira Thoresen

Påvirker velstandsnivået i landet hvordan elever presterer på skolen?

En empirisk analyse av elevprestasjoner og forskjeller mellom land med ulikt velstandsnivå.

Bacheloroppgave i Samfunnsøkonomi

Veileder: Bjane Strøm

Mai 2020

Sofie Bakke og Alexandra Saphira Thoresen

Påvirker velstandsnivået i landet hvordan elever presterer på skolen?

En empirisk analyse av elevprestasjoner og forskjeller mellom land med ulikt velstandsnivå.

Bacheloroppgave i Samfunnsøkonomi
Veileder: Bjane Strøm
Mai 2020

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Innholdsfortegnelse

Kapittel 1 - Innledning	2
1.1 Innledning	2
1.2 Skole i Norge	2
1.3 Skole i Argentina	2
1.4 Skole i Colombia	3
Kapittel 2 - Tidligere studier, teori og økonometrisk modell	3
2.1 Innledning	3
2.2 Tidligere studier	3
2.3 Regresjonsanalyse	4
2.4 Minste kvadraters metode	6
2.5 Korrelasjon	7
2.6 Hypotesetester	8
Kapittel 3 - Presentasjon av data og deskriptiv analyse	9
3.1 Innledning	9
3.2 Om undersøkelsen	10
3.3 Presentasjon av variabler	10
3.5 Deskriptiv statistikk	13
3.6 Fordeler og begrensninger	14
3.6.1 Reliabilitet	15
3.6.2 Validitet	15
Kapittel 4 - Regresjonsanalyse	15
4.1 Innledning	15
4.2 Multippel regresjonsanalyse	16
4.3 Blokkvis regresjonsanalyse	19
4.4 Hypotesetest	21
4.5 Feilkilder	22
Kapittel 5 - Oppsummering og konklusjon	23
5.1 Oppsummering	23
5.2 Diskusjon	24
5.3 Konklusjon	25
Kapittel 6 - Litteraturliste	26

Kapittel 1 - Innledning

1.1 Innledning

I oppgaven vil vi fokusere på sammenhengen mellom velstandsnivå og leseferdighetene til barn. Vi synes det er interessant å se på hvordan barns tilgang på ulike ressurser kan påvirke resultatene deres på skolen. Vil tilgang på pc og bøker påvirke i positiv retning? Har et barn fra et mer økonomisk utsatt hjem større motivasjon til å gjøre det bra på skolen kontra et barn med større sikkerhetsnett? For å se nærmere på dette har vi kommet frem til tre ulike land, Norge, Argentina og Colombia. Landene er plukket ut fra FN sin liste over ulike lands indeks for menneskelig utvikling justert for skjevfordeling (IHDI) fra 2010, da dette er det tidligste målet vi har tilgjengelig. Norge ligger høyest med 0.879, deretter kommer Argentina med 0.656 og nederst Colombia med 0.520. (FN, 2020) Indeksen går fra 0 til 1 og viser den menneskelige utviklingen i form av forventet levealder, utdanning og inntekt for en gjennomsnittsinbygger i et land. Gjennom analyse av datasett fra PIRLS-undersøkelsen har vi funnet flere positive korrelasjoner mellom mål for velstand og prestasjoner, derav blant annet utdanningen til foreldre, tilgang på bøker hjemme og tilgang til PC.

1.2 Skole i Norge

For å få en grunnleggende forståelse for skolesystemene, vil vi gi en kort beskrivelse av hvordan det er i de ulike landene. I Norge har vi gratis, obligatorisk skolegang for barn og ungdom i alderen 6 til 16 år. Ungdom mellom 16 og 19 har også lovfestet rett til videregående opplæring. I 2017 utgjorde andelen barn og unge som utførte opplæring på private skoler 3 % av barn i grunnskole og 8 % av elever på videregående skole. (Thune, Reisegg & Askheim, 2019) I 2001, året for undersøkelsen, brukte Norge 6,73 % av bruttonasjonalproduktet på skole. (UNESCO)

1.3 Skole i Argentina

I Argentina utgjør grunnskolen syv år og videregående skole 5 år, der det er obligatorisk å fullføre 9 år. De fleste fullfører videregående og omtrent halvparten velger å ta høyere utdanning. De private skolene utgjør 20 % av tilbudet, og er av høyest kvalitet. Utdanningens

kvalitet har stor spredning og de offentlige skolene har gjerne lavere utdannede lærere og dårligere utstyr. I år 2000 ble analfabetismen i landet beregnet til å ligge på omtrent tre prosent. (Welle-Strand, 2012) I 2001 utgjorde utgifter knyttet til utdanning 4,83 % av BNP i Argentina. (UNESCO)

1.4 Skole i Colombia

Det er store skiller på skoletilbudet i byene og utenfor i Colombia, noe som samsvarer med den lave IHDI'en i landet. I år 2000 var 18 % av grunnskolene og 32 % av videregående skoler drevet privat. I Colombia har de 10-årig obligatorisk skole, der grunnskolen utgjør fem av disse. På grunnskolenivå starter 90 % av barna på skolen, men bare 70 % av disse fullfører. I år 2000 lå analfabetismen i landet på åtte prosent. (Welle-Strand, 2012) I Colombia ble 3.71 % av BNP brukt på utdanning i 2001. (UNESCO)

Det ser ut til å være tydelig spredning på utdanningens kvalitet i de tre landene vi har valgt ut. For å studere disse forskjellene nærmere, vil vi ta utgangspunkt i den kvantitative undersøkelsen PIRLS fra 2001. Dette er en omfattende undersøkelse besvart av elever, foreldre og lærere i flere land. Denne undersøkelsen vil være grunnlaget for vår analyse.

Kapittel 2 - Tidligere studier, teori og økonometrisk modell

2.1 Innledning

I dette kapittelet vil vi presentere tidligere studier som har gitt oss en pekepinn på valg av interessante variabler til analyse, for å gi oss svar på problemstillingen. Videre vil vi også presentere vårt valg av økonometrisk modell, og valg av teori som vi vil bruke videre i oppgaven.

2.2 Tidligere studier

I undersøkelsen *Indigenous Ecuadorian children: Parental education, parental wealth and children's cognitive ability level* studerte de kognitive evner hos barn i alderen 9 til 14 på tre

skoler i landlige områder i Ecuador. De så hvordan foreldrenes utdanning, velstand, eiendeler, bøker hjemme, oppdragelsesmetode, disiplin i skolen og høyde påvirket barnas resultater på ulike tester. I undersøkelsen har de sett på resultater fra SPM, PIRLS og TIMSS. Individuelle forskjeller i kognitive evner ble best forklart av foreldrenes utdanning, men også høyde, bøker i hjemmet og disiplin hadde positive effekter. Ut fra denne studien velger vi derfor å se på foreldrenes utdanning, og legger dette til grunn da vi tror at elevene vil se økt resultat dersom foreldrene har høyere utdanning.

Fra rapporten *Books are forever: Early life conditions, education and lifetime earnings in europe* har de funnet ut at avkastningen i lønn fra utdanning, varierer mye ut i fra sosioøkonomisk bakgrunn tidlig i livet. De ser at avkastningen er signifikant lavere for de som har få bøker i 10-års alderen. Det er ikke mulig å se en direkte sammenheng mellom antall bøker i husholdningen og foreldrenes utdanning. Men, det å ha tilgang til mange bøker i hjemmet reflekterer til å lettere danne kognitive og sosial-emosjonelle ferdigheter, noe som er en viktig faktor for å oppnå økonomisk suksess i livet. I rapporten blir det også påvist en positiv sammenheng mellom det å ha få bøker og faktorer som tyder på dårlig helse. Blant disse faktorene inngår følgende: vært borte fra skole over en måned på grunn av helseproblemer, alvorlig sykdom, vaksinasjoner og regelmessige tannlegebesøk. (Brunello et al. 2012, s 14) Grunnet denne studien så ønsker vi å se på bøker som et tegn på velstand, og hvordan det påvirker elevene i vår undersøkelse. Felles for begge undersøkelsene er at foreldrenes utdanning påvirker barnas resultater.

2.3 Regresjonsanalyse

Regresjonsanalyse er en analyse av sammenhenger mellom en avhengig variabel og en eller flere uavhengige variabler. I vår oppgave så skal vi sammenligne en avhengig variabel med en interessevariabel og fire kontrollvariabler, som alle vil ha samlebegrepet *uavhengige variabler*. En slik analyse vil vise oss om de uavhengige variablene kan ha en direkte effekt på den avhengige variabelen. Dette er den mest vanlige formen for regresjonsanalyse, nemlig en lineær regresjon. Den avhengig variabelen vi har valgt er en kontinuerlig variabel, og dette er grunnlaget for valget av lineær regresjonsanalyse. Her vises sammenhengen mellom

uavhengig og avhengig variabel som en rett linje. Den lineære regresjonsanalysen kan vises ved ligningen:

$$E(Y) = \alpha + \beta X + \varepsilon$$

Hvor Y er den avhengig variabelen, også kalt eksogen variabel. Alfa representerer konstantleddet, beta representerer stigningstallet, og X er den uavhengige variabelen, også kalt for en endogen variabel. Siste leddet er epsilon som er det såkalte residualen, også kalt for et stokastisk restledd, som fanger opp avvik. Alfa og beta er konstante parametre, mens Y, X og epsilon er spesifikke for hver observasjon. Stigningstallet beta forteller hvor mye Y øker/synker ved en endring i X. Ser at Y er gitt ved E(Y), det vil si forventningsverdien til Y, og dette kan forklares ved at regresjonsanalysen kun gir oss et estimat på den avhengig variabelen Y. Det er her residualen spiller en viktig rolle. Residualen fanger opp avviket mellom forventningsverdien Y og den faktiske Y, og vil derfor gir et mer korrekt estimat. Residualen representerer også faktorer som ikke er inkludert, men som kan ha en effekt på Y.

$$Y = E(Y_i) + \varepsilon_i$$

Utvalgsligningen for hver enkelt observasjon kan derfor skrives som

$$E(Y_i) = \alpha + \beta X_i + \varepsilon_i \quad \text{hvor } i = 1, 2, \dots, n$$

Dette blir kalt for en populasjonsregresjonsligning. Vi har ikke observasjoner for hele populasjonen, men kun et utvalg fra de tre forskjellige landene vi har valgt. Ligningen kan kun anvendes dersom man har data for hele populasjonen. Denne ligningen representerer en ukjent linje vi ønsker å få et best mulig estimat for.

Som nevnt tidligere, så skal vi ha med fem uavhengige variabler, noe som gjør at vi må ta i bruk en *multippel regresjonsanalyse*. Det vil si, det vi vil se på er hvordan flere endogene variabler vil påvirke den eksogene variabelen. En endogen variabel, X, er den som forklarer og den eksogene, Y, er variabelen som blir forklart. En multippel regresjonsanalyse kan vises ved denne ligningen:

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 X_{3i} + \dots + b_{ki} X_{ki} \quad \text{hvor } i = 1, 2, \dots, n$$

(Thomas, 2005, s. 387)

I denne ligningen representerer a og b estimatene på de ukjente parameterne alfa og beta. Videre er X fremdeles en representasjon av de uavhengige variablene. Denne ligningen viser at man kan legge til så mange variabler som er ønskelig uten problemer. Ser også fra ligningen at Y nå er representert av \hat{Y} , det vil si den predikerte verdien av Y . Dette er fordi vi nå bruker et estimat på alfa og beta, og vi kommer frem til en antatt verdi av Y . Avviket mellom Y og \hat{Y} vises gjennom residualen ε . Kan også vises ved ligningen

$$Y_i = \hat{Y}_i + \varepsilon_i \quad \text{hvor } i = 1, 2, \dots, n$$

For å finne dette estimatet på alfa for a og beta for b , så vil vi bruke minste kvadraters metode(OLS).

2.4 Minste kvadraters metode

Minste kvadraters metode (Ordinary least squares/OLS) er en metode for å estimere de ukjente parametre ut fra gitte observasjoner i en regresjonslinje. OLS vil gi oss den beste linjen som gjør at summen av de kvadrerte avstandene mellom datapunktene og linjen er minst mulig. (Ringdal, 2012, s.393) Med andre ord, OLS brukes for å tilpasse en rett linje i et plottdiagram. Summen av de kvadrerte avstandene kan skrives som:

$$= \sum e_i^2 = \sum (-\hat{Y}_i)^2 = \sum (Y_i - b_1X_1 - b_2X_2 - \dots - b_iX_i)^2 \quad \text{hvor } i = 1, 2, \dots, n$$

(Thomas, 2005, s. 266)

OLS viser da skjæringspunktet a mellom regresjonslinjen og y-aksen, og stigningstallet b til linjen, som begge minimerer summen av kvadratene til residualene. Vi kvadrerer fordi residualene kan være både positive og negative, og ved kvadrering blir verdiene utelukkende positive og vi får et resultat som kan brukes i en videre analyse. Vi ønsker å finne de verdiene av a og b som minimerer S . Måten man finner disse verdiene på, er ved å partiellderivere S med hensyn på a og b . Da vil man få disse ligningene for henholdsvis b og a :

$$b = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

$$a = \bar{Y} - b\bar{X}$$

(Ringdal, 2012, s. 395)

Det er kovariansen til X og Y delt på variansen til X som fanger inn den statistiske sammenhengen mellom X og Y. Dersom vi bruker estimatene fra a og b vil vi da ende opp med den predikerte verdien \hat{Y}

Den predikerte verdien vil tilsi den best mulige linjen vi kan oppnå, men vi må også beregne hvor godt regresjonslinja beskriver datasettet. Dette kan vi gjøre ved å benytte

determinasjonskoeffisienten R^2 , som også kalt et føyningsmål. (Thomas, 2005, s. 273)

Føyningsmålet viser hvor stor andel av total variasjon i Y som forklares av variasjonen i X.

Med andre ord, vi ser på andelen forklart variasjon (SSE) i forhold til total variasjon (SST).

Dette kan vises ved ligningen:

$$R^2 = \frac{SSE}{SST} \quad 0 \leq R^2 \leq 1 \quad \text{hvor} \quad \frac{SSE}{SST} = \frac{b^2 \sum (X_i - \bar{X})^2}{\sum (Y_i - \bar{Y})^2}$$

Føyningsmålet ligger altså mellom 0 og 1, og det vil si at forklaringskraften til regresjonslinja vil nesten alltid øke dersom man legger til flere variabler X i modellen. Derfor vil vi i vår

oppgave benytte oss av en justert R^2 i regresjonsanalysen. Denne tar hensyn til at flere

variabler tilføyes modellen og justerer forklaringskraften etter det. (Thomas, 2005, s. 421)

Deretter vil vi benytte oss av R^2 i en blokkvis regresjonsanalyse for å forklaringskraften av hver enkelt variabel når vi legger de til i modellen.

2.5 Korrelasjon

For å se om to variabler har en samvariasjon så ser man gjerne på korrelasjonen mellom de.

Vi vil bruke Pearsons korrelasjonskoeffisient i vår oppgave, også kalt Pearsons r, hvor man

kan få verdi mellom -1 og 1. Dersom man får verdien -1, så er variablene perfekt negativt

korrelert, og 1 er perfekt korrelasjon. Får man 0, så er det ingen korrelasjon. Ved Pearsons r,

så fungerer r som en standardisering av kovariansen mellom to variabler, og er en estimator

for ρ . Grunnen til at vi standardiserer kovariansen er fordi den ikke har noen øvre eller nedre

grense, noe som gjør det vanskelig å skjønne hvor stor samvariasjonen mellom variablene er.

ρ er korrelasjonskoeffisienten til populasjonen. Korrelasjonskoeffisienten måler tendenser til lineære samvariasjoner mellom to variabler. Regner ut med denne ligningen:

$$= r = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2 \Sigma(Y_i - \bar{Y})^2}}$$

(Ringdal, 2012, s. 305)

Ligningen har kovariansen til X og Y i teller, og kvadratroten av variansen i nevner. Dette skal da gi et resultat med korrelasjonsnivå mellom -1 og 1. Vi har nå funnet ut om det er en samvariasjon mellom variablene vi ser på. Selv om det er en korrelasjon, så betyr ikke dette nødvendigvis at det er en samvariasjonen er kausal. Det kan være at det er en spuriøs årsakssammenheng, det vil si at en tredje variabel Z påvirker X og Y, men det ser ut som det er kausalitet mellom X og Y. (Thomas, 2005, s. 258)

2.6 Hypotesetester

Hypotesetest er en metode for å anvende estimatene vi fikk fra OLS stemmer med virkeligheten, hvor vi tar hensyn til at det er en viss usikkerhet. Denne usikkerheten blir målt ved signifikansnivå. Dersom en variabel er signifikant så viser det seg at variablene kan generaliserer og være representative for hele befolkningen. Dette gjør vi fordi vi kun har data på et utvalg, og ikke hele populasjonen. Vi vil starte med en nullhypotese, og teste den opp mot en alternativhypotese. Ut i fra en kritisk verdi, kan vi enten forkaste eller beholde nullhypotesen. Dersom vi må forkaste nullhypotesen, så godtar vi alternativhypotesen. Den hypotesen vi ender opp med, og dermed er signifikant, vil som sagt være et mål som er generaliserbart for populasjonen. Statistisk generalisering forutsetter at utvalget av analyseenheter er trukket tilfeldig. (Ringdal, 2012, s. 367)

Ved en hypotesetest så kan man enten ta en ensidig test eller en tosidig test. Dersom man tar en ensidig test, så vil forkastningsområdet kun være på den ene siden i normalfordelingen. Da ser man kun på en retningseffekt. Ved en tosidig test ser man at forkastningsområdet er på begge sider i normalfordelingen, men her vet man ikke hvilken retning effekten vil ha. (Thomas, 2005, s. 135)

Når vi utfører en hypotesetest vil vi som sagt velge et signifikansnivå, noe som også kan kalles et forkastningsnivå. Signifikansnivået forteller hvor statistisk signifikant et resultat må være for at det skal bli godkjent og dermed kunne bli generaliserbart. Det vanligste innenfor samfunnsstatistikk er å bruke et 5 % signifikansnivå, noe som vil tilsi verdiene 1.64 ved ensidig test, og 1.96 ved tosidig test. Dersom man får en kritisk verdi over disse verdiene så vil man forkaste nullhypotesen og akseptere alternativhypotesen. I vår oppgave har vi benyttet oss av t-test og henholdsvis t-fordeling, og finner den kritiske verdien fra en tabell for t-fordeling. I statistikk sies det at datasett er tilnærmet normalfordelt når antall observasjoner overstiger 30. En t-fordeling vil gå mot normalfordeling ettersom observasjonene går mot uendelig. Datasettene vi har tatt i bruk har vell over 3000 observasjoner hver, og derfor antar vi at t-fordelingen er tilnærmet normalfordelt. Dersom en variabel er normalfordelt, så antas det at det ikke er store avvik i verdiene og dermed kan vi avdekke om effekter er signifikante eller ikke. Deretter tar vi i bruk t-testen for å sammenligne effektene av variablene mot hverandre.

Når man utfører en hypotesetest så kan man utføre type I og type II feil. Type I feil er dersom man forkaster en sann nullhypotese. Type II feil er å beholde en usann nullhypotese. Ofte ser man på P-verdi istedenfor signifikansnivået. P-verdien vil tilsi det laveste signifikansnivået man kan forkaste nullhypotesen på. Er P-verdien mindre eller lik 0.05, så er alternativhypotesen signifikant. Er P-verdien mindre eller lik 0.01 så er alternativhypotesen meget signifikant.

Kapittel 3 - Presentasjon av data og deskriptiv analyse

3.1 Innledning

I dette kapitlet vil vi fortelle litt mer omfattende om undersøkelsen som elevene og foreldrene har svart på. Ut i fra dette vil vi forklare hvilke variabler vi har valgt å bruke og hvorfor. Vi vil så vise den deskriptive statistikken som gir videre utgangspunkt til analysen som kommer i kapittel 5.

3.2 Om undersøkelsen

I denne oppgaven tar vi utgangspunkt i undersøkelsen PIRLS (Progress in International Reading Literacy Study) utført av The International Association for the Evaluation of Educational Achievement (IEA). Meningen med undersøkelsen er å fange opp leseferdighetene til elevene og informasjon om hva leseferdigheten påvirkes av.

Undersøkelsen består av spørreskjema til elevene, foreldrene, rektorene og lærerne. En prøve til elevene som består av flere tekster, flervalgsspørsmål og åpne spørsmål der elevene må svare selv skriftlig. Opplysningene fra undersøkelsen er deretter samlet opp i ulike datasett for hvert land der funnene er formulert som variabler med tallverdi.

I denne oppgaven vil vi ta utgangspunkt i datasettet fra 2001. Denne undersøkelsen ble foretatt av 150 000 elever i 5777 skoler i 35 land. (Senter for leseforskning, 2003, s. 5) I Norge er det statistisk sentralbyrå (SSB) som trekker ut elevene klassevis i samarbeid med IEA. I 2001 deltok 136 av de 160 skolene som var plukket ut. Deltagelsesprosenten av skolene var på 89 og 92 % av elevene utførte undersøkelsen. En prosjektgruppe ved Senter for leseforskning har på oppdrag fra Læringscenteret utført PIRLS 2001 i Norge.

3.3 Presentasjon av variabler

Vi har valgt ut noen variabler vi synes kan representere vår forskning og kan gi oss en god forklaring på om velstand i landet påvirker elevenes prestasjoner. Vi har valgt 6 variabler vi vil ha fokus på, og som vi kan bruke opp mot hverandre og sammenligne for de tre forskjellige landene vi har valgt. De fleste variablene vi har valgt, tar utgangspunkt i økonomien til de forskjellige elevenes familier. Dette fordi vi vil se på ulikheter innad i landene. Vi tar også med variabel som forteller oss om tilgang på ressurser i skolegangen, som vi antar at vil gi oss en pekepinn på hvor mye midler skolene blir bevilget av det offentlige.

Tabell 3.1 - Oversikt over variabler

read	Resultatet fra testen (kontinuerlig)
books_home	Antall bøker hjemme (kategori)
par_emp	Foreldrenes arbeidsstatus (kategori)
pc_class	Pc tilgjengelig = 1 (dummy)
pct_disadv	prosentvis andel fra økonomisk utsatt hjem (kategori)
par_edu	Foreldrenes utdanning (kategori)

For det første så må vi se på den avhengige variabelen *Read*, dette gir oss testresultatet som elevene har fått på lesetesten. Dette er en variabel som er representativ for hvordan eleven presterer på skolen, og vil være utgangspunktet for analysen. Verdien av variabelen er på kontinuerlig form, noe som er en forutsetning for at vi skal kunne utføre en lineær regresjonsanalyse. Resten av våre variable er rangert i kategorier.

Variabelen *pct_disadv* måler hvor stor andel av elevene i en klasse som kommer fra et økonomisk vanskeligstilt hjem. Dette er vår interessevariabel, fordi vi tror denne faktoren påvirker velstanden, og dermed resultatet, til en elev i størst grad. *Pct_disadv* er på ordinalnivå, da prosentandelen har blitt gruppert i 0-10, 11-25, 26-50, og til slutt om mer enn 50 prosent av elevene i klassen kommer fra et økonomisk utsatt hjem. Dette vil si at dersom verdien av *pct_disadv* har verdien 1 så vil 0-10 % av elevene i klassen kommer fra hjem med økonomiske vanskeligheter.

Variabelen *Pc_Class* forteller oss om elevene har tilgang på PC på skolen. Vi antar at dersom det er bevilget penger til at elevene har tilgang på PC på skolen, så tyder dette på at landet har nok velstand i landet til å prioritere mer en kun det nødvendige i opplæringen. *Pc_class* er på nominalnivå, som gir at eleven har tilgang på PC dersom *pc_class*=1 og ikke tilgang dersom *pc_class*=0.

Vi har også valgt variabler som ser på velstanden i hjemmet til eleven. Variabelen *Par_emp* viser først og fremst om foreldrene er i jobb. Variabelen viser også arbeidssituasjonen til

foreldrene, om begge jobber fulltid, deltid eller om kun en av de jobber. Et land med høy arbeidsledighet vil ha lavere velstand totalt sett, og derfor er denne variabelen veldig interessant for vår problemstilling. Dersom foreldrene er arbeidsledige, kan eleven være tvunget til å ha en annen prioritering enn skolegangen sin. Vi tror dette er med på å påvirke prestasjonen til eleven og om eleven fullfører skolegangen. Vi har også valgt å ta med variabelen *Par_edu* som viser foreldrenes gjennomførte skolegang, om de eventuelt har tatt høyere utdanning eller på hvilket nivå de gikk fra skole til arbeid. Vi antar at jo høyere utdanning foreldrene har, jo høyere sannsynlighet er det for at eleven gjør det bra på skolen. Studiet *Indigenous Ecuadorian children* er lagt til grunn for vår antagelse. Vi tror også at velstanden i landet er høyere dersom en stor andel av foreldrene har høyere utdanning.

På grunn av funn i teorien fra studiet *Books are forever*, har vi valgt å inkludere variabelen *Books_Home* som er et mål på hvor mange bøker eleven har tilgang på i hjemmet. Denne variabelen er delt inn i fem kategorier, som henholdsvis viser til antall bøker tilgjengelig for eleven i hjemmet sitt. Jo høyere kategori, dess flere bøker. Vi antar at antall bøker hjemme er en viss indikator på velstand, med tanke på at bøker kan være kostbart.

3.5 Deskriptiv statistikk

Vi vil nå presentere deskriptiv statistikk. Dette gir en oversikt over variablene, og deres egenskaper. Her vektlegges mål på sentraltendenser som gjennomsnitt, standardavvik, minimums- og maksimumsverdier, basert på antall observasjoner vi har tilgjengelig.

Tabell 3.1 - Deskriptiv statistikk for avhengig variabel read.

	Norge	Argentina	Colombia
Gjennomsnitt	498.26	423.86	434.17
Standardavvik	78.37	90.29	75.80
Min	228.06	161.45	200.29
Max	695.87	701.55	654,94
Antall obs	3459	3300	5131

Vi ser fra tabellen at gjennomsnittresultatet på lesetesten er høyest i Norge med 498.26. Resultatet er gjennomsnittlig lavest i Argentina, noe som tilsier at vi ikke har en lineær sammenheng mellom testresultatene og IHDI-scoren i landet per 2010. Vi antok at Norge vil ha høyest resultat, deretter Argentina og Colombia med lavest, ettersom det er slik landene er rangert med tanke på velstand. Likevel ser vi at standardavviket i Argentina er høyere enn hos de andre, som tilsier at de har større spredning i resultatet. Dette ser vi igjen i at de både har resultatet med både høyest og lavest score. Dette kan være en tilfeldighet, men er noe vi ønsker å se nærmere på. Det er ulikt antall observasjoner i de ulike landene, men vi antar at utvalget er plukket ut tilfeldig slik at de kan være representative populasjonen.

Vi har nå dummykodet alle de uavhengige variablene slik at vi kun ser på de med høyest velstand. Det vil si, klassene med færrest elever fra hjem med økonomiske vanskeligheter, de som har tilgang på PC, en eller begge foreldrene i fulltidsjobb, og foreldre med fullført høyere utdanning, og med tilgang på minst 101 bøker hjemme. Vi har dummykodet

pct_disadv slik at vi kun ser på klassene der det er maksimalt 10 % av elevene som kommer fra økonomisk utsatte hjem. På den måten får vi muligheten til å fokusere på de klassene som er best økonomisk stilt i de ulike landene. par_emp er dummykodet slik at vi kun ser på elever som har minimum en forelder i fulltidsjobb. Gjennom å dummykode par_edu ser vi kun på elever som har foreldre med utdanning fra videregående skole eller høyere utdanning. Pc_class er ferdig dummykodet, og vi ser kun på de som har PC tilgjengelig. I books_home ser vi kun på de som har fra 101 eller flere bøker hjemme. Alle variablene ligger dermed på nominalnivå i resten av oppgaven.

Tabell 4.3 Deskriptiv statistikk for uavhengige variabler.

Variabler	Norge		Argentina		Colombia	
	Gj.snitt	Std.avvik	Gj.Snitt	Std.Avvik	Gj.Snitt	Std.Avvik
Dpct_disadv	0.85	0.36	0.06	0.24	0.12	0.33
Dpar_edu	0.97	0.18	0.39	0.49	0.39	0.49
Dpar_emp	0.96	0.19	0.86	0.35	0.87	0.33
Pc_class	0.85	0.36	0.28	0.45	0.17	0.38
Dbooks	0.93	0.25	0.53	0.50	0.39	0.49

Ser at Norge har et høyt gjennomsnitt på alle variablene, med relativt lite spredning. Dette samsvarer med at vi har valgt Norge basert på at det er et land med høy velstand. Argentina har stor variasjon i gjennomsnittene til variablene, og spredningen er høy i alle variablene. Vi ser at det kun er variabelen par_emp, altså om en eller begge foreldrene er i fast jobb, som har et høyt gjennomsnitt. Vi ser for Colombia at de opplever ganske like resultater som Argentina, men med litt lavere gjennomsnittsverdi på pc_class og books_home. Til gjengjeld så har Colombia litt høyere spredning enn Argentina.

3.6 Fordeler og begrensninger

Studiet og undersøkelsen vi bruker for å svare på vår problemstilling er fra 2001, noe som gjør at funnene nødvendigvis ikke er generaliserbar for populasjonen i dag. Valg av land ble

gjort på grunnlag av IHDI, noe som ble introdusert først i 2010. Det er en ulempe for studien vår, da det kan ha vært større endringer på de 9 årene og vi ikke kan anta at forskjellen i indeksen i 2010 samsvarer med 2001.

3.6.1 Reliabilitet

Reliabilitet går på om gjentatte målinger med samme måleinstrument, her utformingen av PIRLS, ville gitt samme resultat. PIRLS undersøkelsen er nøye produsert av erfarne forskere og vi antar at spørsmålene som blir stilt er nøytrale, slik at de ikke påvirker kandidaten. Undersøkelsens form er også tilrettelagt for at den kan utføres anonymt, noe som gir høy grad av personvern. Utvalget er plukket ut for å være representativt for befolkningen i hvert av landene. I Norge er skoler som underviser etter L2L9-former (samisk) og sykehusundervisning blant annet sett bort fra da utvalget er plukket ut.

3.6.2 Validitet

En annen ulempe med datasettet i oppgaven er at det ikke er mange variabler som er tydelige mål for velstand. Det er også manglende verdier på sentrale variabler som inntekt. Denne ulempen påvirker oppgaven vår negativt og svekker muligheten for god validitet. Validitet går på om vi faktisk måler det vi vil måle. Her er vi ute etter å se på hvordan velstand påvirker leseferdighetene til elevene. Selv om datasettene er innholdsrike, er det ikke gitt at variablene vi har plukket ut er representativt for velstand. I velstand legger vi helse, utdanning, rikdom og følelsen av lykke. Vår oppgave tar for seg variabler som i størst grad måler rikdom og utdanning, og inkluderer ikke variabler for helse eller følelsen av lykke.

Kapittel 4 - Regresjonsanalyse

4.1 Innledning

Vi antar i vår oppgave at jo mer velstand et land har, jo bedre vil elevene prestere på skolen. Derfor har vi tatt utgangspunkt i de elevene som har best forutsetninger fra hvert land, for å så kunne sammenligne disse utvalgte gruppene opp mot hverandre. Vi forventer å se at elevene fra Norge, som har best forutsetninger, vil gjøre det bedre på skolen enn elevene fra

Argentina, som igjen vil gjøre det bedre enn elevene fra Colombia. I dette kapittelet vil vi utføre forskjellige analyser for å finne ut om våre antagelser er riktige, og for å trekke en konklusjon til problemstillingen vår.

4.2 Multipel regresjonsanalyse

Vi har nå plukket ut observasjonene vi er interessert i å se på og vil nå utføre en lineær multipel regresjonsanalyse for hvert av landene, ved hjelp av minste kvadraters metode(OLS). Ligningen for regresjonsanalysen blir da :

$$\widehat{Y}(read) = \alpha + {}_1Dpct_disadv + {}_2Dpar_emp + {}_3Dpar_edu + {}_4pc_class + {}_5Dbooks + \varepsilon_i$$

Tabell 4.1 - Multipel regresjonsanalyse

VARIABLER	Norge (Modell 1)		Argentina (Modell 2)		Colombia (Modell 3)	
	Koeffisient	T-verdi	Koeffisient	T-verdi	Koeffisient	T-verdi
Read (endogen variabel)						
Dpct_disadv	0.03	0.01	21.66**	2.65	54.59***	13.64
Dpar_emp	17.7*	2.12	-10.76	-1.67	5.81	1.49
Dpar_edu	48.45***	5.39	43.64***	8.41	30.61***	10.20
Pc_class	-7.39	-1.75	31.31***	6.41	-4.85	-1.32
Dbooks	45.2***	7.55	28.98***	5.49	21.48***	7.06
Observations		2528		1230		2754
Adjusted R-squared		0.04		0.19		0.20

Variablenes signifikansnivå er representert ved * p<0.05, **p<0.01, ***p<0.001

Årsaken til at variabelens signifikansnivå er vist med stjerner ved siden av tallene i tabellen er for å vise om variablene er generaliserbare for populasjonen. Det fant vi ut ifra variabelenes t-verdi under regresjonsanalysen. Jo flere stjerner en variabel har, jo større sikkerhet for at variabelen ikke er et resultat av tilfeldighet. Nederst ser vi modellens adjusted R-squared, hvor mye av variansen i testresultatene som kan forklares av vårt utvalg av eksogene variabler.

Vi starter med å se på modellen for Norge. Vi ser at første og fjerde variablene ikke er en signifikante variabler for resultatene til norske barn og kan dermed ikke generaliseres for populasjonen. Derfor vil vi ikke ta med disse variablene i videre forklaring for testresultatene til norske barn, men fjerner de ikke fra modellen da de er signifikante i modellene for Argentina eller Colombia. Vi vil derfor se på foreldrenes arbeidsstatus, variabelen Dpar_emp. Dersom en eller begge foreldrene har fulltidsjobb, vil eleven få en økning i testresultatet på 17.7 poeng. Dette stemmer med vår antagelse om at elevene kan ha fullt fokus på skole. Variabelen Dpar_edu viser at dersom foreldrene har fullført videregående skole eller høyere utdanning, så vil testresultatet til eleven øke med 48.45. Dette er variabelen med høyest koeffisient i modellen for Norge, noe som stemmer godt med teorien fra studiet *Indigenous Ecuadorian children*. Til slutt, om eleven har tilgang på over 101 bøker hjemme, så vil testresultatet øke med 45.2. Variansen i testresultatet til Norge er svært lavt. Det vil si at vårt utvalg av variabler har svært dårlig forklaringskraft på endringer i resultat hos elevene. Vi antar at det er fordi norske elever allerede har et såpass høyt velstandsnivå, at mer økt velstand vil ikke påvirke noe særlig. Dette kan sees i lys av de små kontrastene i de sosioøkonomiske forholdene i Norge. Det er få som bruker private skoler og de fleste skolene er utstyrt med noenlunde samme midler.

I Argentina er alle korrelasjonene signifikante på 0.01 eller 0.001-nivå utenom variabelen par_emp, noe som gjør at vi ser bort fra effekten av variabelen i denne modellen. Gjennom resultatene kan vi se at variablene Dpct_disadv, altså det å være i en klasse med få elever fra økonomisk utfordrende familier, har en positiv effekt på 21.66. I motsetning til i Norge, så er denne variabelen signifikant, og noe som stemmer med våre antagelser. Dess mindre fattigdom, jo bedre testresultater. Det å ha foreldre med høyere utdanning eller at de har gjennomført videregående skole påvirker testresultatet positivt med 43.64. Det er denne

variabelen som gir størst utslag i analysen for Argentina, slik som i Norge. Det å ha tilgang på pc i klasserommet øker testresultatet med 31.31, noe som er særegent for Argentina. Vi antar at dette har en sammenheng med at landet har en lavere velstand enn Norge, og tilgang på PC ikke er like vanlig i Argentina ettersom offentlige skoler ofte er dårlig utstyrt. Derfor antar vi at dersom eleven har tilgang på PC så går eleven på en god skole, som viser til velstand, og derfor bidrar til bedre testresultater. Tilgang til over 101 bøker i hjemme øker resultatet med 28.98. Vi ser alt i alt at det å ha tilgang på midler og det å i snitt være fra velstående familier gir positive utslag for testresultatet. Variansen i testresultatet er på 0.19, noe som forteller oss at variablene våre bare forklarer 19 % av variansen i testresultatet. Dette vil si at det kan være andre variabler utenfor vår analyse som påvirker resultatene i større grad enn vårt utvalg av variabler.

I modellen for Colombia så er ikke regresjonen mellom read, Dpar_emp og pc_class signifikant. Resten av variablene er signifikante på 0.01-nivå. I motsetning til de andre landene, så er det Dpct_disadv som gir størst økning i testresultatet for elevene i Colombia. Dette er med en økning på 54.59. Vi antar at dette er fordi Colombia er landet med lavest velstand, og blir i større grad påvirket av økonomiske utfordringer i motsetning til de andre landene. Dette stemmer da med vår antagelse at dersom færre i klassen kommer fra økonomisk vanskeligstilte hjem, så ser vi en økning i testresultatene. Vi ser at også i Colombia så påvirker foreldrenes utdanning i positiv retning, men her med 30.61. Dette er mindre enn i Norge og Colombia, noe som kan stemme med antagelsen om at i et land med mindre velstand, så har færre av foreldrene høyere utdanning. Vår siste signifikante variabel, Dbooks, påvirker resultatet med 21.48. Modellen til Colombia har en forklaringskraft på 20 %, som da er den beste modellen av våre analyser. Vi antar det er fordi dette landet har størst fattigdom, og en økt velstand påvirker elevprestasjonen mer i dette landet i forhold til de to andre.

Vi kunne se fra studiet *Books are forever* at et høyt antall bøker i hjemmet er faktor som spiller inn på utdanningen til elevene. Vi har nå sett på analysen i alle tre landene, hvor både foreldrenes utdanning og tilgang på bøker hjemme er signifikante og øker testresultatene til

elevene en del. Vår analyse stemmer dermed overens med de tidligere studiene vi har tatt utgangspunkt i.

4.3 Blokkvis regresjonsanalyse

Vi vil nå se på hvor stor forklaringskraft de ulike variablene har, og dette gjør vi ved å kjøre en blokkvis regresjonsanalyse. Ved å ta en blokkvis regresjonsanalyse får vi muligheten til å se hvor mye R^2 forandrer seg ved å legge til en ekstra variabel. Dette gjør vi for å sjekke om sammenhengen mellom variabler øker forklaringskraften til modellen, og dermed om valget av variabler er hensiktsmessige for vår problemstilling. Vi får også muligheten til å sammenligne hvor mye forklaringskraft variablene har i de ulike landene. Grunnlaget for valg av blokkvis regresjon istedenfor flere regresjonsmodeller er for å begrense oppgaven vår slik at den ikke blir for omfattende.

Tabell 4.2 - Blokkvis regresjonsanalyse

Modell	Norge		Argentina		Colombia	
	R2	$\Delta R2$	R2	$\Delta R2$	R2	$\Delta R2$
1	0.00		0.03***		0.13***	
2	0.00*	0.0024	0.03	0.0023	0.13***	0.0034
3	0.02***	0.0164*	0.14***	0.1150	0.18***	0.0531
4	0.02	0.0017	0.18***	0.0337	0.18	0.0006
5	0.04***	0.0216	0.20***	0.0198	0.20***	0.0146

Variablenes signifikansnivå er representert ved * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

$$\widehat{Y}(\text{read}) = \alpha + {}_1Dpct_disadv + \varepsilon_i$$

Modell 1 er en regresjonsanalyse mellom variablene read og Dpct_disadv. Vi ser at forklaringskraften til modellen i Norge ikke er signifikant, noe som forteller at vi ikke kan anta at det å være fra et økonomisk vanskeligstilt hjem i Norge gir forklaring på endring i leseferdigheter. Både i Argentina og Colombia er endringen signifikant på 0,001-nivå, og viser til en positiv økning i variabelen read. Det vil si at dersom få av elevene kommer fra

økonomisk vanskeligstilte hjem, så har det en positiv effekt på resultatene. Forklaringskraften er størst i Colombia med en økning på 13 %.

$$\widehat{Y}(\text{read}) = \alpha + {}_1Dpct_disadv + {}_2Dpar_emp + \varepsilon_i$$

I modell 2 har vi nå lagt til variabelen *Dpar_emp*. Denne modellen er signifikant på 0.05-nivå både i Norge og Colombia. Variabelen er ikke signifikant i Argentina og derfor ikke hensiktsmessig å ta med som forklaring på endring deres testresultater. Variabelen forklarer bare 2,3-3,4 % av endringen i testresultatene i Norge og Colombia. Vi kan av det anta at arbeidsstatusen til foreldrene har en positiv, men veldig liten effekt på hvor bra elevene presterer innenfor lesing.

$$\widehat{Y}(\text{read}) = \alpha + {}_1Dpct_disadv + {}_2Dpar_emp + {}_3Dpar_edu + \varepsilon_i$$

I modell 3 har vi lagt til variabelen *Dpar_edu*. Her er alle modellene signifikant på 0.01-nivå. Endringen i Norge er på 1.6 %, 11.5 % i Argentina og 5.3 % i Colombia. Totalt sett er det denne modellen som viser til størst forklaringskraft på testresultatene totalt sett i vår analyse. Dette betyr at det å ha foreldre med universitetsgrad eller fullført videregående har den mest omfattende betydningen for økt resultat på lesetesten i vår analyse. Det vil si, dette er den variabelen som har størst innvirkning på testresultatet til elevene og kan konkludere med at dette er den viktigste av variablene som vi har inkludert i vår analyse.

$$\widehat{Y}(\text{read}) = \alpha + {}_1Dpct_disadv + {}_2Dpar_emp + {}_3Dpar_edu + {}_4pc_class + \varepsilon_i$$

I modell 4 er variabelen *pc_class* lagt ved. Modellen er fremdeles ikke signifikant for Norge eller Colombia, til tross for at den gir en liten økning i forklaringskraften. Vi antar dette enten kommer av ren tilfeldighet, eller at vi har en manglende forklaringsvariabel. I Argentina er modellen signifikant på 0.01-nivået og viser 3.4 % av endringen i testresultatet, og dette stemmer overens med det vi fant ut i den multiple regresjonsmodellen.

$$\widehat{Y}(\text{read}) = \alpha + {}_1Dpct_disadv + {}_2Dpar_emp + {}_3Dpar_edu + {}_4pc_class + {}_5Dbooks + \varepsilon_i$$

I modell 5 har vi tatt med variabelen *Dbooks*, og har dermed inkludert alle variablene som vi har brukt i tidligere analyser. Modellene er signifikante for alle de tre landene. Vi ser at

modellene fra Argentina og Colombia totalt sett har en forklaringskraft på 20 %, mens hos Norge ligger den bare på 4 %. Dette vil si at variablene våre kun kan forklare 4 % av variasjonen i testresultatet i Norge. Dette ser vi i sammenheng med at velstanden i Norge er god, og at det er små forskjeller blant skoler. Vi ser at variablene vi har brukt for å representere velstand har større forklaringskraft i land med mindre utvikling og større ulikhet, og vi ser dermed at økt velstand faktisk vil påvirke elevprestasjoner i disse situasjonene.

4.4 Hypotesetest

Etter å ha tatt en estimert regresjonsanalyse vil vi nå teste kvaliteten ved modellene våre. Det er mulighet for at vi kan ha feilestimert OLS, og derfor vil vi nå ta en hypotesetest for å sjekke om resultatene vi har kommet frem til kan anvendes for populasjonen. Det er en viss usikkerhet rundt det at vi kan ha tatt med variabler som egentlig ikke er relevante, eller at vi har utelatt viktige variabler. Hypotesetesten vi velger å gjennomføre kalles for en Breusch-Pagan test, og er en av de mest brukte for å teste for heteroskedasitet.

Heteroskedasitet vil si at man ikke har konstant variasjon i restleddet, noe som er en forutsetning for lineær regresjon.

Tabell 4.3 - Hypotesetest

Breusch-Pagan test	Norge	Argentina	Colombia
Chi2 (kritisk verdi)	1.37	1.60	3.08
Prob>chi 1	0.2416	0.2064	0.0793

$H_0 : \sigma_i^2 = 0$ - Restleddet er homoskedastisk

$H_A : \sigma_i^2 \neq 0$ - Restleddet er heteroskedastisk

Vi ser at for Norge og Argentina så er den kritiske verdien utenfor forkastningsnivået på et 5 % signifikansnivå, og vi beholder nullhypotesen. Dette betyr at restleddet i regresjonen er homoskedastisk, altså har restleddet en konstant varians. Vi kan derfor konkludere med at vi har fått riktige verdier på våre estimat, og valget av variabler er forsvarlig for analysen.

Colombia derimot, er innenfor forkastningsnivået, og vi må forkaste H_0 . Det betyr at vi har

fått feil estimater under OLS i variablene i modellen til Colombia. I utgangspunktet så vil dette ugyldiggjøre resultatene våre, og selve oppgaven vil være mindre troverdig. Derfor ser vi på robuste standardfeil, og ser at justeringen på standardavvikene er svært små. Heteroskedasitet må være veldig stor for at det skal gjøre alvorlige utslag på standardavvikene. Velger derfor å beholde modellen for Colombia, og konkluderer med at modellen er gyldig.

Tabell 4.4 - Robust regresjon for Colombia

Variabler	T-verdi	T-verdi r.	Endring	Koeffisient	Koeffisient r.	Endring
Dpct_disadv	13.64	13.18	0.46	54.58821	54.71879	0.13053
Dpar_emp	1.49	1.55	0.06	5.808644	6.283466	0.474822
Dpar_edu	10.23	10.42	0.19	30.61351	32.35161	1.7381
pc_class	-1.32	-1.33	0.01	-4.851447	-5.093534	0.242087
Dbooks	7.06	6.86	0.2	21.47587	21.61489	0.13902

4.5 Feilkilder

I forskningen vår har i vi flere mulige feilkilder. Først og fremst er det mulige datafeil i datasettet som kan komme fra trykkfeil eller de kan ha fått oppgitt opplysninger som er feilaktige. En annen større kilde for feil i oppgaven vår er begrepsvaliditeten, altså om vi faktisk får målt velstanden gjennom variablene vi har valgt. I oppgaven har vi valgt å se på de som scorer høyest på våre variabler for velstand i tre ulike land med ulik utvikling. Velstand forklares gjerne som en kombinasjon av rikdom, helse og lykke. Vi har ingen konkrete variabler for lykke og helse, noe som har ført til at vi har fokusert mer på økonomi, utdanning og skoletilbud. Dette kan være årsaken til at modellene vi har valgt har så lav forklaringskraft.

I oppgaven vår så valgte vi å kjøre en blokkvis regresjonsanalyse istedenfor å kjøre flere multiple regresjonsanalyser. Det hadde vært mer fordelaktig å kjøre regresjonsanalyser og se hvordan variablene påvirkes av hverandre ved å inkludere disse gradvis. Derfor får vår

oppgave en mindre helhetlig forklaring på variablenes oppførsel, men dette valget ble tatt bevisst da oppgaven ville blitt altfor omfattende ettersom vi sammenligner tre land og seks variabler. Grunnet en omfattende oppgave har vi også dummykodet alle variablene våre og ser på de med de beste forutsetningene. Dette kan føre til at vi får mindre spredning i analysen vår, noe som kunne vært relevant for å svare på problemstillingen vår.

Datasettene vi bruker for analysen har fått deler av infoen ut fra at foreldrene har svart på en spørreundersøkelse. Det kan være en del problemer med en slik datainnsamling, da det finnes mange insentiver til å svare uærlig. Enkelte individer vil bare svare fortest mulig, og ender opp med å gi feilaktige svar. Andre vil ikke ønske å svare ærlig, da de vil fremstille seg selv litt bedre enn hva de faktisk er, altså ut ifra hva som er sosialt ønskelig

Kapittel 5 - Oppsummering og konklusjon

5.1 Oppsummering

Denne oppgaven har forsøkt å svare på om et lands velstand- og utviklingsnivå har en effekt på hvordan elever presterer på skolen. Vi valgte tre land som ligger spredt innenfor indeksen for menneskelig utvikling (IHDI). Norge har en høy indeks som viser til at det gjennomsnittlige mennesket er høyt utdannet, BNP per capita er høy og har høy forventet levealder. Argentina scorer gjennomsnittlig og Colombia scorer lavt på denne indeksen. I oppgaven analyserer vi data hentet fra PIRLS-undersøkelsen fra 2001, der vi ser etter variansen i testresultatene til elevene. Deretter tok vi en blokkvis regresjon for å se forklaringskraften til variablene hver for seg, for å sjekke hvilke variabler som påvirker elevene mest.

Til slutt tok vi en hypotesetest for å sjekke om estimeringen av modellen er av god kvalitet. Dette ble gjort ved en Breusch-Pagan test, noe som tester modellen for heteroskedasitet. Modellene for Norge og Argentina er homoskedastiske, altså så har støyleddet i regresjonslikningene deres en konstant varians. Colombia derimot, havnet innen for forkastningsnivået og denne modellen er derfor heteroskedastisk. Vi tok derfor en robust

regresjon, og konkluderte med at vi likevel kan beholde modellen fordi det er såpass liten endring i standardavviket.

5.2 Diskusjon

Et interessant funn er at vi trodde at andelen i klassen som kom fra økonomisk vanskeligstilte hjem ville påvirke resultatet mest, men det viste seg at foreldrenes utdanningsnivå har størst forklaringskraft. Vi ser på høy utdanning som et mål på velstand, ettersom man må investere i seg selv når man tar høyere utdanning. Ved et lavt velstandsnivå, så har man ikke like god mulighet til en slik investering. Derfor er det en god indikasjon på at velstand påvirker elevprestasjon ettersom Norge ser størst forbedring i testresultatene.

Vi antok også at tilgang på bøker hjemme ville være betydelig for resultatet til elevene, etter å ha tatt utgangspunkt i studiet *Books are for life*. Vi kunne se fra analysen at alle landene vi har sett på, har fått en positiv og signifikant økning i resultatene til elevene dersom de har tilgang til et visst antall bøker. Vårt funn stemmer overens med hva de har funnet i tidligere studier. I studiet *Indigenous Ecuadorian children* kunne vi se at både bøker og foreldrenes utdanning har en positiv effekt på resultatene til elevene. Dette er noe vi også kom frem til i vår analyse.

De to gjenværende kontrollvariablene, foreldrenes jobbstatus og tilgang på PC, er variabler som har varierende betydning for testresultatene. Foreldrenes jobbstatus er kun signifikant for Norge, og tilgang på PC er kun signifikant for Argentina. Derfor blir ikke disse variablene brukt for å få en helhetlig forståelse for om velstand påvirker elevprestasjon, ettersom det ikke er gjennomgående signifikant for alle de tre landene.

I den blokkvise regresjonsanalysen kunne vi se at det er variasjon på hvilken variabel som har størst forklaringskraft på modellene totalt sett. Hvilken variabel som har størst betydning for økning i resultat varierer fra land til land. Tilgangen på bøker forklarer mest i Norge, foreldrenes utdanning har høyest forklaringskraft i Argentina og det at få elever i klassen kommer fra økonomisk vanskeligstilte hjem forklarer mest i Colombia. Vi kan forstå det som at alle disse variablene er viktige for å forstå økningen i resultatene til elevene. Vi antar at

grunnen til at det er forskjell på hvilken variabel som er mest forklarende er forutsetningene til elevene, med tanke på at vi ser på tre land med stor variasjon i form av forventet levealder, utdanning og inntekt for en gjennomsnittsinbygger.

5.3 Konklusjon

Resultatet og konklusjonen vi har kommet frem til er at variabler som representerer økt velstand har en lav forklaringskraft på resultatene til elevene i Norge. Derimot ser vi at i Argentina og Colombia så ble det en betydelig økning i testresultatene ved økt velstand. Det vil si, Norge påvirkes ikke noe særlig ved høyere velstand, noe vi antar er fordi norske elever allerede har en såpass høy velstand fra før. Dette stemmer også overens med det vi så i den deskriptive statistikken, ettersom norske elever har høyest gjennomsnittlig resultat. I Argentina og Colombia så kunne vi se en betydelig økning i testresultatene da vi la inn variabler som viste til økt velstand, og konkluderer med at velstand er en signifikant faktor som spiller inn når man ser på elevprestasjoner dersom man ikke opplever høy velstand fra før. Vi vil derfor konkludere med at økt velstand fører til økt elevprestasjon, der det er foreldrenes utdanning som har størst effekt.

Kapittel 6 - Litteraturliste

Brunello, G., Weber, G., Weiss, C. (2012) *Books are forever: Early life conditions, education and lifetime earnings in europe*, The Institute of Social and Economic Research Osaka University, Japan. Hentet fra:

<http://web.b.ebscohost.com/ehost/detail/detail?vid=0&sid=7b894b37-f55b-47de-9434-589b97bd9310%40sessionmgr103&bdata=JnNpdGU9ZWwhvc3QtbGl2ZQ%3d%3d#AN=1640454&db=eoh>

Forente nasjoner. (2020). *IHDI (2010)*. Hentet fra:

<https://www.fn.no/Statistikk/IHDI-forskjeller-i-et-land>

Rindermann, H., Carl, Noah. (2017). Indigenous Ecuadorian children: Parental education, parental wealth and children's cognitive ability level. *Learning and individual differences*, 2017(54.), 202-209. Hentet fra:

<https://www.sciencedirect.com/science/article/pii/S1041608017300328>

Ringdal, K. (2012). *Enhet og mangfold* (3. utg.). Bergen: Fagbokforlaget

Solheim, R. G., Tønnessen, F. E./Senter for leseforskning. (2003) *En norsk kortversjon av den internasjonale rapporten om 10-åringers lesekunnskaper*. Hentet fra:

https://lesesenteret.uis.no/getfile.php/13108402/Lesesenteret/PIRLS_Norsk_kortversjonpdf.pdf

Thomas, R. L. (2005). *Using statistics in economics*. London: Mcgraw Hill Higher Education

Thune, Taran; Reisegg, Øyvind; Askheim, Svein. (1.oktober 2019) *Skole og utdanning i Norge*. Hentet fra: https://snl.no/Skole_og_utdanning_i_Norge

UNESCO. *Education : Expenditure on education as % of GDP (from government sources)* Hentet fra: <http://data.uis.unesco.org/Index.aspx#>

Welle-Strand, Anne. (18.januar 2012). *Skole og utdanning i Argentina*. Hentet fra:

https://snl.no/Skole_og_utdanning_i_Argentina

Welle-Strand, Anne. (31.januar 2012) *Skole og utdanning i Colombia*. Hentet fra:

https://snl.no/Skole_og_utdanning_i_Colombia

