Bacheloroppgåve

NTNU
Noregs teknisk-naturvitskaplege universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

Stine Fjellkårstad
Pernille Brenna Johansen

Ei kvantitativ analyse av elevprestasjonar, fødselsmånad og sosioøkonomiske ressursar

Bacheloroppgåve i Samfunnsøkonomi

Mai 2020

NTNU
Kunnskap for en bedre verden

Stine Fjellkårstad
Pernille Brenna Johansen

Ei kvantitativ analyse av elevprestasjonar, fødselsmånad og sosioøkonomiske ressursar

Bacheloroppgåve i Samfunnsøkonomi
Mai 2020

Noregs teknisk-naturvitskaplege universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

NTNU

Kunnskap for en bedre verden

Innholdsfortegnelse

1.0 Innleiing

1.1 Motivasjon

I Noreg sine offentlige utgreiingar, NOU 2019:3 leia av Camilla Stoltenberg tek dei føre seg elevprestasjonar i skulen i Noreg, og ser på høvesvis forskjellen mellom gutar og jenter. I tillegg til at det viser seg at det er store forskjellar på jenter og gutar i skulen, er det også aningar til at det er forskjell mellom born føydde seint på året og born føydde tidlig på året. Vi ynskjer å undersøkje dette ubunden av kjønn. Vi vil derfor i denne oppgåva legge vekt på fødselsmånadar, og ventar å finne forskjellar. Desse forskjellane ynskjer vi å belyse gjennom denne oppgåva.

1.2 Problemstilling

Vi har på bakgrunn av innleiinga og forventningane vi har til ulikheitene på elevprestasjon i Noreg formulert følgjande problemstilling;

1) *Kva effekt har relativ alder(fødselskvartal) på elevprestasjonar?*

Vi ynskjer også i denne oppgåva å sjå på kva effekt sosioøkonomisk bakgrunn påverkar elevprestasjonar, og har derfor formulert ei andre problemstilling:

2) *Påverkar foreldra sin tilsettingsstatus, utdanningsnivå og inntekt i hushaldet elevprestasjonar?*

Bakgrunnen for val av fleire problemstillingar er at vi ynskjer å sjå på separate effektar av relativ alder og sosioøkonomisk bakgrunn på elevprestasjonar. Vidare ynskjer vi å sjå om sosioøkonomisk bakgrunn har ulik effekt på fødselskvartal.

2.0 Empirisk grunnlag

2.1 Skuleproduktfunksjon

I denne oppgåva vil vi nytte oss av ein skuleproduktfunksjon, som målar elevprestasjonane ved ein testscore. Funksjonen har rot i den meir generelle analysa av humankapital, som har meir fokus på marknadsutfall og lønn, medan skuleproduktfunksjon skil seg ved eit fokus på underliggande fastsetting av ferdigheiter og humankapital.

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

F = Familie/elevkarakteristika

P = Medelevekarakteristika (peer group-effekter),

S = Skolefaktorar (Klassestørrelse, lærerkarakteristika)

Coleman-rapporten frå 1966 var startskotet på ei meir økonomiprega forsking på skule og utdanning. Forventningane til studia var at forskjellar i elevprestasjonar skuldast ressursbruk, men viste seg heller kunne skuldast blant anna sosioøkonomiske forhold (Coleman, 1966). Rapporten finn fleire signifikante koeffisientar ved familie og elevkarakteristika, medelevkarakteristika også kalla *peer group*-effektar. Han fann ingen signifikante funn i skulefaktorar, men det er brei einigheit om at Coleman bomma på blant anna lærarane si betyding (Bonesrønning, 2004).

Modellen er kritisert for å berretelje år på skulen utan å ta føre seg kva som skjer på skulen, og bagatellisera dermed utdanningspolitikk som forsøker å betre kvaliteten på skuletida. Modellen føreset også at eit skuleår produsera den same nytten/ferdigheitene over tid og på tvers av landegrenser. Skuleproduktfunksjon seier også at utdanning er einaste innsatsfaktor i ferdigheitsutvikling (Hanushek, 2020, s. 162).

2.2 Fødselsmånad

Innføringa av reform-97 bestemte at alder ved skulestart skulle vere det året ein fyl seks. Før innføringa av denne reforma var alder ved skulestart det året ein fylde sju. Det at skulestart i Noreg baserar seg på same kalenderår medfører at det kan skilje opptil tolv månadar mellom eldste og yngste eleven i same klasse (NOU 2019:3 s. 126). I NOU 2019:3 ”*Nye sjanser – betre læring: Kjønnsforskjeller i skoleprestasjoner og utdanningsløp*”, leia av Camilla Stoltenberg, viser utvalet til tidlegare studiar. Blant anna eit studie gjennomført av Björnsson og Olsen som tek føre seg fødselsmånader og kva det har og seie for elevprestasjonar. I deira forsking finn dei betydelege forskjellar mellom born føydde i januar og born føydde i desember, men forskjellane utjamnar seg utover i skulelaupet (Björnsson og Olsen, sitert i NOU 2019:3, s. 127).

2.3 Sosioøkonomisk bakgrunn

I denne oppgåva vil vi undersøkje kva effekt sosioøkonomisk bakgrunn har på elevprestasjonar. Og då nyttar vi foreldra sitt utdanningsnivå, tilsettingsstatus og hushaldet si inntekt. Ei statistisk analyse frå statistisk sentralbyrå (SSB), "*Foreldres utdanning avgjørende for barnas skolegang*" skrive av Oddbjørn Raaum og Mona Raabe antyder at born med foreldre med ressursar i form av utdanning eller inntekt gjer det gjennomsnittleg betre enn born med foreldre som ikkje har denne typen ressursar i heimen (Raaum og Raabe, 2004). NOU 2019:3 viser til ein undersøking gjort av statistisk sentralbyrå i 2018 at i den norske grunnskulen var det ein differanse mellom borna som hadde foreldre med høgskule- og universitetsutdanning og foreldre med berre grunnskule på 5,9 grunnskulepoeng. Dersom ein samanliknar ytterpunkta, ser vi at born der begge foreldra har lang høgre utdanning og born der begge foreldra berre har grunnskule, skillast med 12 grunnskulepoeng (Statistisk Sentralbyrå, sitert av NOU 2019:3, s. 13). NOU 2019:3 viser også til ein artikkel av Are Turmo "*Scientific literacy and socio-economic background among 15-year-olds—a Nordic perspective*" der Turmo kan vise til at samanhengen mellom hushaldet si inntekt og elevprestasjon er svak fordi dei økonomiske forskjellane i dei nordiske landa er nokså liten samanlikna med andre land (Turmo 2004, sitert i NOU 2019:3, s. 95). Det kan også vise seg at foreldre med lang høgre utdanning er meir opptatt av at borna deira skal ta høgre utdanning, slik at foreldra legg til rette for dette allereie når borna er små (Raaum, 2003, s. 116).

2.4 Oppsummering

Basert på tidlegare empiri ynskjer vi å sjå på effektar av fødselsmånad, foreldra sitt utdanningsnivå, foreldra sin tilsettingsstatus og inntekta i hushaldet, samt kontrollere for om dei sosioøkonomiske effektane har ulik påverknad på born føydde i de ulike fødselsmånadane. Frå tidlegare empiri kan det vise seg at fødselsmånad har stor effekt på elevprestasjonar og at det utgjer ein betydeleg forskjell med born føydde tidlig og seint på året. I tillegg til dette kan tidlegare empiri vise til at sosioøkonomiske ressursar også har ein betydeleg innverknad.

3.0 Økonometrisk teori

3.1 Innleiing

I dette kapittelet ser vi på økonometrisk teori, og korleis vi skal gjennomføre ei regresjonsanalyse. Grunnleggande økonometrisk teori er naudsynt for at vi skal kunne gjennomføre vår analyse utan unødige feil.

3.2 Regresjon

Regresjon er ein metode for å analysere samanhengen mellom to eller fleire variablar. For å kunne gjennomføre ei regresjonsanalyse går ein ut frå ein lineær årsakssamanheng mellom variablane og forsøker å kvantifisere den lineære samanhengen mellom variablane ein går ut i frå (Thomas, 2005, ss. 259-260). Vi skil mellom enkel lineær regresjon og multippel lineær regresjon. Fyrst skal vi sjå på ein enkel regresjon. I ei enkel regresjonsanalyse har vi ein responsvariabel (avhengig), referert til som Y og en forklaringsvariabel (ubunden), referert til som X (Helbæk, 2011, s. 113).

$$E(Y) = \alpha + \beta X \qquad (3.1)$$

Likninga over blir kalla for populasjonsregresjonslikninga. Denne viser ein lineær samanheng mellom Y og X, der α er konstansleddet og β er stigningstalet og kvantifiserer effekten av X og Y. Stigningstalet fortel oss kor mykje Y endrast dersom det skjer ei endring i X.

Det er ikkje alltid slik at den faktiske verdien på Y er lik forventningsverdien E(Y). For å kunne sjå på differansen mellom disse blir det tillagt eit støyledd i likning, gjeve ved ε (Thomas, 2005, s. 260).

$$Y = E(Y) + \varepsilon \qquad (3.2)$$

Støyleddets si oppgåve er å fange opp kjelda til (i) effekten av alle andre variablar som påverkar Y bortsett frå X og (ii) anna tilfeldig støy (Thomas, 2005, s. 261). Også her må vi gå ut frå at støyleddet er normalfordelt i regresjonsmodellen. Den lineære samanhengen mellom X og Y er gjeve som:

$$Y = \alpha + \beta X + \varepsilon \qquad (3.3)$$

I ei populasjonsanalyse er populasjonslikninga ukjend og den blir ukjend gjennom heile regresjonsanalysa (Thomas, 2005, s. 262). Sidan vi i oppgåva ser på eit utval av ein større populasjon og i tillegg skal sjå på fleire kontrollvariablar og korleis disse påverkar den avhengige variabelen, vil vi bruke multippel regresjonsanalyse. I ein multippel lineær regresjon gjer vi dei same antakingane som i enkel lineær regresjon. For at regresjonsmodellen skal vere gyldig må alle kontrollvariablane (X_i) vere ubundne av kvarandre (Helbæk, 2011, s. 123). Modellen er gjeve som:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \ldots + b_k X_{ki} \qquad i = 1,2,3,..,n \quad (3.4)$$

Her er *a* og *b* respektive variablar av estimatane på α og β. X_i representerer dei ulike forklaringsvariablane som påverkar responsvariablane Y. Ŷ lesast som predikert Y, og predikert Ŷ og faktisk Y er sjeldan samanfallande. Dersom verdien på faktisk Y og predikert Ŷ hadde vore like, ville det bety at alle punkta i spreiingsmålet ville vore på ein rett linje, som så å seie aldri oppstår (Thomas, 2005, s. 263). Verdien på Ŷ går ein ut frå basert på verdiane som ein har gjeve *a* og *b*. Forskjellen mellom predikert Ŷ og faktisk Y er gjeve ved residual e_i (Thomas, 2005, ss. 262-263).

$$Y_i = \hat{Y}_i + e_i \quad (3.5)$$

3.3 Minste kvadratar sin metode (Ordinary least-squares method)

Minste kvadratar sin metode (MKM) er den mest brukte metoden for å tilpasse ei linje i eit spreiingsplott. Vi vil ta for oss minste kvadratar sin metode i ein enkel lineær regresjon. Den best tilpassa linja vil vere der residualen, e_i er lågast, altså der avviket mellom predikert Ŷ og faktisk Y er minst. Vi finn linja ved å minimere summen av den kvadrerte residualen.

$$\min \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \qquad (3.6)$$

Når vi ser på regresjonsanalyse er verdiane av populasjonsparameterane α og β ukjende. Vi brukar derfor MKM til å estimere verdien til α og β ut frå eit datasett. Vi har tidlegare nemnt at estimatane til α og β blir representert ved *a* og *b*. MKM vel skjeringspunktet *a* og stigningstalet *b* i datasettet. Ved å partielle derivere likning (3.6) med omsyn til høvesvis a og

b kvar for seg, og sette disse lik null vil vi finne formelen til skjeringspunktet og stigningstalet (Thomas, 2005, s. 266). Dette utgjer den beste linja i eit spreiingsplott.

$$b = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{Y})^2} \qquad (3.7) \qquad a = \bar{Y} - b\bar{X} \qquad (3.8)$$

Sjølv om MKM gjev oss ei godt tilpassa linje, er vi nøydde til å berekne kor beskrivande regresjonslinja er for datasettet. Dette gjerast ved determinasjonskoeffisienten, også kalla eit føyningsmål. Føyningsmålet er representert ved R^2, altså den kvadrerte av korrelasjonen. Føyningsmålet viser kor mykje av variasjonen i Y som kan forståast av variasjonen i X (Thomas, 2005, s. 276).

$$R^2 = \frac{SSE}{SST} \left(= \frac{b^2 \sum(X_i-\bar{X})^2}{\sum(Y_i-\bar{Y})^2}\right) \quad , \quad 0 < R^2 < 1 \qquad (3.9)$$

Vi ser at vi finn føyningsmålet ved å dividere forklart variasjon (SSE) på total variasjon (SST). Føyningsmålet ligg i eit sett mellom 1 og 0, der 0 tilseier at variasjonen i Y ikkje kan tilskrivast av variasjonen i X, og 1 tilseier at variasjonen i Y kan forklarast av X. Ved å tilleggje modellen fleire forklaringsvariablar (X) vil føyningsmålet stige, dette er fordi variasjonen vil vert fordelt på fleire uavhengige variablar, og vil derfor forklare meir av variasjonen i Y (Thomas, 2005, s. 421).

3.4 Hypotesetesting

Vi anvender hypotesetesting når vi vil sjå om OLS-estimatane stemmer overeins med røynda. For å kunne gjøre dette må vi formulerer to hypotesar; Nullhypotese (H_0) og alternativhypotese (H_A). H_0 er forkastningsgrunnlaget til hypotesen, der H_A er komplementær og den hypotesa ein vil underbygge. Når vi testar med hypotesar kan to ulike feil verte gjort. Disse blir kalla for type I feil og type II feil. Type I feil er når vi forkastar H_0, når det viser seg at H_0 er sann. Type II feil på den andre sida er når vi beheld H_0, når det viser seg at H_0 er usann. Sannsynet for at det blir gjort ein type II feil er ukjend (Thomas, 2005, ss. 137-138).

Signifikansnivået viser sannsynet for å gjere feil når vi avviser H_0, og då type I feil. (Thomas, 2005, s. 129). Normalt sett ved statistikk er signifikansnivået gjeve ved eit 5% signifikansnivå. Det er verdien på testobservatoren (*TS*) som avgjer om ein skal behalde eller

forkaste ei hypotese. Dette gjerast ved at ein samanliknar verdien på *TS* opp mot verdien på kritisk verdi. Kritisk verdi tilsvara verdien på signifikansnivået (Thomas, 2005, ss. 127-129).

Ein kan bruke hypotesetest for å sjekke om det er lineær samanheng i regresjonsmodellen, og då nyttar ein multippel hypotesetest (Helbæk, 2011, s. 127). For å gjennomføre ein slik hypotese må ein bruke F- test. Her seier nullhypotesen at det ikkje er noen lineær samanheng mellom den avhengige variabelen Y og forklaringsvariablane X_i, og at alle β- verdiar lik null (Helbæk, 2011, s. 127). Alternativhypotesen seier at ein- eller fleire av β- verdiane ulik null (Helbæk, 2011, s. 127). Vi avviser nullhypotesa dersom summen av forklart variasjon (SSE) dividert på residual variasjonen (SSR) er større enn kritisk verdi, som vi finn i F-fordelingstabellen gjeve frihetsgrader (Thomas, 2005, s. 416). I tillegg til dette kan vi også sjå på enkeltvariablar og sjekke om disse er signifikante. Dette gjer vi ved ein t-test der vi tek verdien på koeffisienten dividert på standardavviket og måler det opp mot kritisk verdi funne i t-fordelingstabellen. Også her er det slik at vi forkastar nullhypotesen dersom testobservatoren overstig kritisk verdi.

Vi kan også sjå på kor stor betyding ein forklaringsvariabel har i ein regresjonsmodell. Dette gjerast ved at vi dannar ei likning utan restriksjonar, der vi bereknar SSR for heile likninga. Deretter dannar vi ein likning med restriksjonar, der vi utelet den eller dei forklaringsvariablane vi skal undersøkje, og bereknar SSR. Differansen i SSR fortel oss betydinga av forklaringsvariabelen i modellen (Helbæk, 2011, s. 129).

3.5 Korrelasjon

Korrelasjonskoeffisienten ρ er ein måte å normalisere kovariansen slik at vi kan tolke og samanlikne verdiar. Korrelasjonskoeffisienten fortel oss kor sterk samvariasjon det er mellom to eller fleire variablar i eit utval og kor vidt det er ein lineær samanheng. Mål på korrelasjonen for eit utval er definert som:

$$R = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2}\sqrt{\Sigma(Y - \bar{Y})^2}} \quad , \quad -1 \leq R \leq 1$$

Mogeleg verdi for R ligg i mellom -1 og +1 (Thomas, 2005, s. 256). Dersom R=1 tilsei det at det er ein positiv lineær samvariasjon mellom variablane og at spreiingssmåla til utvalet ligg på ein stagande rett linje. Og ved R=-1 fortel det at det er ein negativ lineær samvariasjon

mellom variablane og den rette linja er avtakande. I tilfelle der korrelasjonskoeffisienten R=0, fortel det oss at det er ingen lineær samvariasjon mellom variablane.

Det er viktig å skilje mellom korrelasjon og kausalitet. Det er viktig å presisere at vi ser på samvariasjonen mellom variablane, og ikkje ein direkte samanheng. Det kan vere ein høg korrelasjon mellom variablane sjølv om det ikkje nødvendigvis er ein kausal samanheng mellom dei. At korrelasjon mellom X og Y er positivt signifikant kan skuldast det vi kallar ein spuriøs årsakssamanheng. Spuriøs samanheng kan også vere skapt av ein tredje variabel, Z, som påverkar både X og Y, og derfor kan det sjå ut som at det er ein årsakssamanheng mellom X og Y (Thomas, 2005, ss. 258-259).

4.0 Presentasjon av data

4.1 Innleiing

For ei betre forståing av regresjonsanalysane i del 5 vil vi no presentere datamaterialet vi skal nytte. Her vil du finne ei utdjuping om datasettet PIRLS 2001 som vi nyttar i analysen. Deretter vil vi forklare korleis vi nyttar variablane i datasettet før vi presentera innhaldet i dei ulike kategorivariablane, deskriptiv statistikk og histogram.

4.2 Presentasjon av datasett

PIRLS er ei forkorting for Progress in International Reading Literacy Study, og er arrangert i regi av The International Association for the Evaluation of Educational Achievement, forkorta IEA. Noreg deltok for fyste gong i 2001 i ei undersøking som omfatta 9- og 14-åringar. Omfanget var på 3459 elevar frå 198 klassar delt på 136 skular, med ein deltakarprosent på 89% og svarprosent hjå elevane på 92%. I tillegg til undersøkinga elevene fekk, svara også foreldre/føresette, lærarar og rektorar på spørjeskjema som karakteriserte rammene til borna.

Vi brukar datasettet PIRLS 2001 til å få ei forståing av elevprestasjonar, og settet tek også føre seg sosioøkonomisk bakgrunn, skulen sine ressursar, elevkarakteristika og karakteristika ved klassen. I denne delen vil vi presentere disse variablane. Den avhengige variabelen vi nytta oss av er *read* som gjev eit mål på elevprestasjonar. Vår sentrale forklaringsvariabel i hovudproblemstilling 1 er fødselsmånad *birthm*, som vi kodar om til fire dummyvariablar der vi slår saman tre og tre månadar for å avgrense og forenkle analysen.

I hovudproblemstilling 2 skal vi ta føre oss effektar eleven sin sosioøkonomiske bakgrunn har på elevprestasjonar. For å måle effekta på dette ser vi på foreldra si utdanning, hushaldet si inntekt og tilsettingsstatus hjå foreldra. Foreldra si utdanning *par_edu* vert målt ved høgaste fullført utdanning hjå eine forelderen. Dette er ein kategorivariabel som vi har koda om til ein dummyvariabel, der vi ser på høgskule-/universitetsgrad mot resten. Å skilje mellom høgre og lågare utdanning er eit naturleg skillje, og kategoriane til variablane hadde også svaralternativ vi ikkje har i Noreg(derav ingen avkryssingar på alternativet *post secondary, not university*).

Hushaldet si inntekt, *income,* er også ein kategorivariabel som målar foreldra si utdanning i per $10 000. Vi har valt å legge skiljet til hushaldet på ei årslønn ved meir enn $40 000 lik 1, under lik 0. Variabelen kallar vi *income_high*, ettersom 0 er lågare del av inntektsskalaen. Hushaldet si tilsettingsstatus, *par_emp,* der vi skil mellom ingen av foreldra jobbar eller

jobbar mindre enn fulltidsstilling er lik 1 og 0 ellers. Denne variabelen kallar vi *par_emp_low,* ettersom 0 er ein eller begge foreldre i fulltidsstilling.

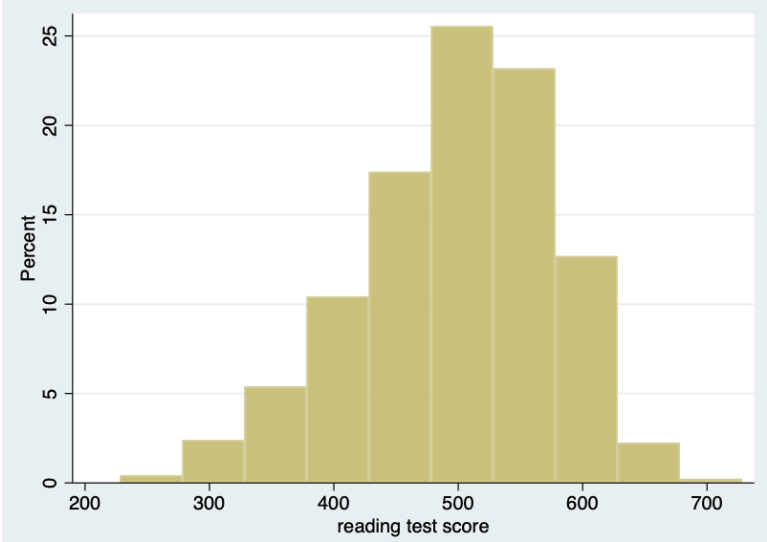
Tabell 4.1: Oversikt og beskrivelse av variablar	
read	Avhengig variabel: mål på elevprestasjonar. Score på lesetest
fødselsmånad	Basert på birthm. Der fødselsmånad januar -mars= 1, april - juni= 2, juli - september = 3, oktober - desember= 4
Fødselmåned1_3	Dummyvariabel, der fødselsmånad januar, februar og mars = 1, resten = 0
Fødselmåned4_6	Dummyvariabel, der fødselsmånad april, mai og juni = 1, resten = 0
Fødselmåned7_9	Dummyvariabel, der fødselsmånad juli, august og september = 1, resten = 0
Fødselmåned10_12	Dummyvariabel, der fødselsmånad oktober, november = 1, resten = 0
par_emp_low	Dummyvariabel for tilsettingssatus, der begge foreldra jobbar ikkje/mindre enn 100% = 1, og 0 ellers
par_edu_high	Dummyvariabel for utdanningsnivå til foreldre der høgre utdanning som universitet og høgskule = 1, medan 0 ellers
income_high	Dummyvariabel for inntektsnivå, der $ 40.000 eller meir = 1, medan null ellers
inchigh_1_3	Interaksjonsvariabel mellom inntektsnivå og fødselskategori 1
inchigh_4_6	Interaksjonsvariabel mellom inntektsnivå og fødselskategori 2
inchigh_7_9	Interaksjonsvariabel mellom inntektsnivå og fødselskategori 3
inchigh_10_12	Interaksjonsvariabel mellom inntektsnivå og fødselskategori 4
paremplow_1_3	Interaksjonsvariabel mellom tilsettingssatus og fødselskategori 1
paremplow_4_6	Interaksjonsvariabel mellom tilsettingssatus og fødselskategori 2
paremplow_7_9	Interaksjonsvariabel mellom tilsettingssatus og fødselskategori 3
paremplow_10_12	Interaksjonsvariabel mellom tilsettingssatus og fødselskategori 4
pareduhigh_1_3	Interaksjonsvariabel mellom utdanningsnivå og fødselskategori 1
pareduhigh_4_6	Interaksjonsvariabel mellom utdanningsnivå og fødselskategori 2
pareduhigh_7_9	Interaksjonsvariabel mellom utdanningsnivå og fødselskategori 3
pareduhigh_10_12	Interaksjonsvariabel mellom utdanningsnivå og fødselskategori 4

4.3 Deskriptiv statistikk

I denne delen av oppgåva vil vi presentere variablane vi skal bruke i regresjonen. Her deskriptiv statistikk.

4.3.1 Lesescore

Vår avhengige variabel, altså variabelen vi ynskjer teste effekta av fødselsmånad og sosioøkonomisk bakgrunn på, er *reading test score.* Vi ser at dei fleste scorar rundt 500 poeng med eit gjennomsnitt på 498.256 poeng.



Figur 4.3.1.1 histogram av lesescore pr. 50 poeng i prosent.

Tabell 4.2.1: : Deskriptiv statistikk for avhengig variabel read

	Observasjonar	Gjennomsnitt	Standardavvik	Min	Max
read	3,459	498.256	1.332	228.060	695.872

4.3.2 Fødselsmånad

Vi ser deretter på lesescore ved dei ulike kvartala borna er føydde. Her har vi koda om kategorivariabelen frå tolv månadar til fire kvartal.

Tabell 4.3.2.1: Deskriptiv statistikk for ubunden variablar elevprestasjonar

	Fødselsmånad 1-3	*Fødselsmånad 4-6*	*Fødselsmånad 7-9*	*Fødselsmånad 10-12*
Gjennomsnitt	514.617	500.968	493.089	485.867
Standardavvik	76.053	77.432	77.628	78.265
Minimum	265.669	234.892	257.454	228.060
Maksimum	695.872	689.646	671.260	650.224
Observasjonar	881	887	883	752

Den deskriptive statistikken viser forskjellane i elevprestasjonar blant born føydd i dei fire kategoriane for fødselsmånad vi har koda. Vi les av at born føydde i siste kvartal har ein gjennomsnittleg testscore på 28,75 poeng lågare enn born føydde i fyste kvartal, som dermed vil seie at dei yngste borna presterer gjennomsnittleg 94,413% av det dei eldste borna gjennomsnittleg presterar. Vi les også på standardavvik at resultata varierar meir mellom elevane di seinare på året ein er føydde. Altså ser vi ei jamnare fordeling på score hjå dei eldste borna, medan meir sprik i leseprestasjonar hjå dei yngste.

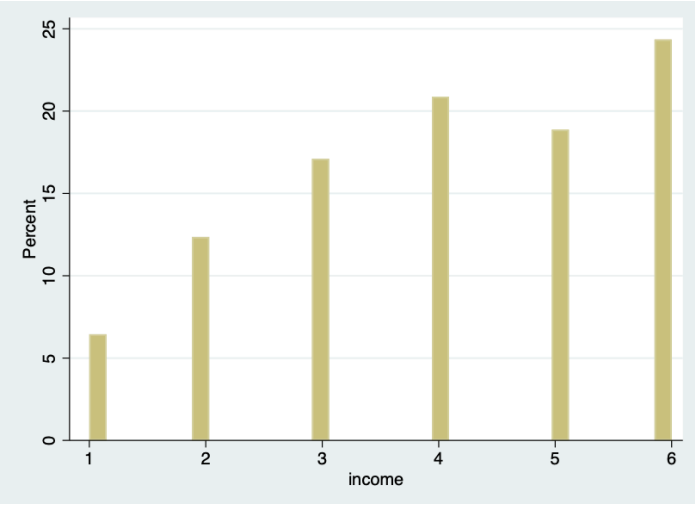
4.3.3 Sosioøkonomisk bakgrunn

Tabell 4.3.3.1: Deskriptiv statistikk for ubunden variablar sosioøkonomisk bakgrunn (før omkoding)

Varaiablar	Observasjonar	Gjennomsnitt	Standardavvik	Min	Max
par_edu	3,098	1.951	1.056	1	5
par_emp	2,975	1.646	0.687	1	4
income	2,994	4.064	1.551	1	6

Når vi skal sjå på elevane sin sosioøkonomiske bakgrunn vel vi sjå på variablane som seier noko om tilsettingsstatus i heimen, foreldra si utdanning og inntekta i hushaldet. Vi les at det er fleire respondentar som manglar informasjon om tilsettingsstatus og inntekt enn utdanning, som kan vere ei svakheit i analysen. Av tabellen kan vi lese at det er større variasjon i lønn enn utdanning, og at det er større variasjon i utdanning enn tilsettingsstatus. Vi ser og at gjennomsnittet av kategorivariablane utdanningsnivå og tilsettingsstatus ligg i kategori 2.

Tabellen viser fordelinga mellom inntektsnivåa, og vi kan sjå at der er ei tydeleg overvekt av respondentar i høginntektsbolken.



Figur 4.3.2 viser kategorivariablen *income* i eit stolpediagram. Her ser vi fordelinga mellom dei seks kategoriane.

Tabell 4.3.3.2: Presentasjon av income, Inntektsfordeling hjå foreldra

Kategori	Inntekt	Frekvens	Prosent
1	Mindre enn $20,00	193	6.45
2	$20,000 - $29,999	370	12.36
3	$30,000 - $39,999	512	17.10
4	$40,000 - $ 49,999	625	20.88
5	$50,000 -$ 59,999	565	18.87
6	$60,000 eller meir	729	24.35

Tabell 4.3.3.5 Presentasjon av par_emp, høgaste utdanning fullført hjå foreldra

kategori	Utdanning	Frekvens	Prosent
1	University	1,682	54.29
2	Post secondary (not university)	0	0
3	Upper secondary	1,311	42.32
4	Lower secondary	96	3.10
5	Not completed lower secondary	9	0.29

Tabell 4.3.3.6: Presentasjon av par_edu, tilsettingssatus hjå foreldra

kategori	Tilsettingssatus	Frekvens	Prosent
1	Begge jobbar fulltid	1,302	43.76
2	Ein jobbar fulltid	1,533	51.53
3	Begge jobbar mindre enn fulltid	30	1.01
4	Jobbar ikkje	110	3.70

Tabell 4.3.3.7: Deskriptiv statistikk for ubunden variablar sosioøkonomisk bakgrunn (etter omkoding)

Variablar	Observasjonar	Gjennomsnitt	Standardavvik
par_emp_low	3,349	0,180	0,385
par_edu_high	3,098	0,543	0,498
income_high	3,459	0,689	0,463

Variablane er opphaveleg kategorivariablar som vi har koda om til dummyvariablar for å enklare kunne nytte dei i interaksjonsledd. Ved å sjå på gjennomsnitta kan vi lese at der er

fleire som er i kategorien med ein eller to foreldre som jobbar fulltid, enn som ikkje gjer det, det er fleire born med ein eller fleire foreldre som har universitetsgrad, enn som ikkje har det og der er fleire born som kjem frå eit hushald med over $40.000 i årsinntekt, enn som ikkje gjer det.

4.4 Kritikk av datasettet

PIRLS 2001 gjev oss mykje informasjon om elevprestasjonar, men å tru datasettet vil gje all relevant informasjon, vil vere urealistisk. Dette er som kjent eit kvantitativ oppgåve, og ein vil gå glipp av nyansar av samanhengar som eit kvalitativt forskingsdesign ville gje. Med dette i bakhovudet medan vi jobbar med analysa, vil biletet analysen dannar vere meir nyansert og realistisk, då ein kjenner aktuelle avgrensingar.

Fordelar:

- Utvalet er nøye valt ut (sampling) som betyr at vi får eit representativt utval, og eit akta datasett styrkar funna sin ytre validitet. Då kan vi i større grad generalisere funna våre.
- Det finnast utdjupande informasjon om korleis undersøkinga er gjennomført på IEA sine sider. Deriblant krava for å verte med i undersøkinga.
- Bruk av datasett gjev også høg kontroll og indre validitet. Datasettet er gjennomsiktig og ryddig, som gjev høg transparens, og minskar forskareffekten, som er bias som påverkar forskinga. Eit transparent datamateriale gjer det også enklare for andre å gjennomgå resultata og etterprøve dei.
- Datasett frå Noreg. PIRLS blir gjennomført i ein rekke land, deriblant i Noreg som aukar moglegheita for generalisering enn om vi skulle brukt tal frå til dømes Sverige.

Ulemper:

- Ved omkoding mistar ein fleire verdiar. Eit døme er som ein kan sjå i tabellane over, før/etter *par_emp, par_edu* og *income*
- Inntektsnivåa er målt i dollar, som når vi avgrensar oppgåva til berre Noreg, er upraktisk.
- Kategorivariabelen *par_edu* gjev eit unyansert bilete av utdanninga hjå foreldra ettersom kategoriane ikkje er tilpassa det norske systemet. Til dømes er det ingen norske som har svara at dei tilhøyrar *post secondary* (2), og det er vanskeleg å skjønne kvar dei som har vidareutdanning eller fagbrev er plassert.
- Datasettet er frå 2001, og dermed kan noko av innhaldet vere utdatert.

- Det er ein svak inntektsforskjell hjå hushaldningane i Noreg som fører med færre respondentar i inntektskategori 1-2 (illustrert ovanfor).

4.5 Oppsummering

Datasettet har eit stort og breitt utval, samstundes som variablane er forenkla slik at det er mogleg for oss å finne fram til ynskja informasjon. Dette gjer det enklare for oss å måle det vi vil måle. Likevel er det viktig å presentere datamaterialet vi skal nytte i analysane for å skape forståing for kva vi faktisk måler, og kva svakheiter til dømes generaliseringa kan ha.

5.0 Regresjonsanalyse

5.1 Innleiing

I denne delen skal vi bruke dataprogrammet STATA for å gjennomføre minste kvadratar sin metode og analysere resultata frå datamaterialet. Vi vil fyrst ta for oss ein enkel modell der vi ser på fødselsmånadar, vidare vil vi leggje til fleire kontrollvariablar i modellen, der vi til slutt vil sjå på ein modell med interaksjonsledd. Ved dette ynskjer vi å analysere korleis testscore påverkast av disse.

5.2 Val av funksjonsform

Vi har valt å bruke ein enkel lineær (lin-lin) funksjonsform for å svare på problemstillinga vår. Dette er fordi det er ein slik funksjonsform som er mest anvendt når ein ser på utdanning. For å kunne gjennomføre dette stillast det strenge føresetnadar til støyleddet. Som nemnd i *3.2 Regresjon*, må støyleddet vere normalfordelt. Støyleddet er gjeve ved:

$$\varepsilon_i \sim N\,(0, \sigma^2) \qquad (5.1)$$

(Thomas, 2005, s. 359)

Funksjonsforma er gjeve ved:

$$Y = \alpha + \beta_i X_i + \varepsilon_i \qquad (5.2)$$

Vidare kjem vi til å legge til den enkle lineære funksjonsforma interaksjonsledd. Dette gjer vi fordi vi vil sjå på korleis effekt av ei endring i ein variabel har på den avhengige variabelen, *read*, gjeve kva fødselskvartal observasjonane tilhøyra. Dette gjev oss moglegheit til å sjå på ei ikkje-lineær samanheng mellom variablane.

5.3 Problemstilling 1

Korleis påverkar alder elevprestasjonar?

For å kunne sjå på dette har vi tatt for oss skuleproduktfunksjon der *read* er vår avhengige variabel Y, og denne representerer poengsummen på lesetesten PIRLS. Vår ubundne variabel er presentert, opphaveleg, ved kategorivariabelen *birthm*, som i skuleproduktfunksjonen representerer F (elevkarakteristika). Sidan problemstillinga vår er korleis alder påverkar elevprestasjonar har vi re-kategorisert variabelen slik at fødselsmånadane er delt inn i fødselskvartal, gjeve ved $fødselsmåned_{1\_3,4\_6,7\_9,10\_12}$. Vi har brukt *fødselsmåned1_3* som referansevariabel. Grunnmodellen vår er gjeve ved:

Modell 1: $read = \beta_0 + \beta_1 fødselsmåned4\_6 + \beta_2 fødselsmåned7\_9 + \beta_3 fødselsmåned10\_12 + \varepsilon_i$

Vidare skal vi bruke STATA til å gjennomføre ei OLS-analyse slik at vi finn estimert parameter β_i og predikert Y. STATA gjev oss følgjande estimater:

$$Read = 514.6 - 13.65 fødselsmåned4\_6 - 21.53 fødselsmåned7\_9 - 28.75 fødselsmåned10\_12 + \varepsilon_i$$

Resultata frå OLS- analysen viser til at born føydde i april, mai eller juni får 13.65 lesepoeng lågare enn born føydde i januar, februar eller mars, når vi held resten av variablane konstant. Vidare kan vi sjå i analysen at born føydde enda seinare på året får enda lågare poengsum. Vi ser at born føydde i juli, august eller september får 21.53 lesepoeng lågare, og at born føydde i oktober, november eller desember får 28.75 lågare lesepoeng enn born føydde i januar, februar eller mars. Vi ser også at desse tendensane har oppstått i tidlegare empiri. Vi vil no gjennomføre ein hypotesetest for å sjå om variablane er signifikante og om fødselsmånad har ein effekt på testscoren. Vi vel å bruke F-test sidan vi skal sjå om det finnast ein lineær samanheng mellom den avhengigvariabelen, *read* og ein eller fleire av de uavhengige variablane (Helbæk, 2011, ss. 127-128). Har følgjande hypotese:

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

H_A: $\beta_1 \neq \beta_2 \neq \beta_3 \neq 0$ (ein eller fleire av variablane er ulik null)

Får følgjande testobservator:

$$TS = \frac{(SSR_R - SSR_{UR})/3}{SSR_{UR}/n-4} \sim F - fordelt\ med\ 3\ frihetgrader\ i\ teljar\ og\ n-4\ frihetsgrader\ i\ nemnar (5.1)$$

SSR_R = Variasjons residual med restriksjonar

SSR_{UR} = Variasjonsresidual utan restriksjonar

n = tal observasjonar,

der 3 og 4 er høvesvis tal restriksjonar gjort i likninga og tal på parameter i regresjonen, inkludert konstanten (*cons_*).

Vel eit 5% signifikansnivå og har kritisk verdi 2.6049 funne i F- fordelingstabellen (Thomas, 2005, s. 589). Dersom verdien overstig kritisk verdi, kan vi forkaste nullhypotesen til eit 5% signifikansnivå. Finn SSR_R og SSR_{UR} i stata, og får følgjande TS:

TS = 51.247562

Vi får at testobservatoren er større enn kritisk verdi, dette vil seie at vi forkastar nullhypotesen ved 5% signifikansnivå og at vi har eit grunnlag for å hevde testscoren påverkast av når på året eleven er føydd. Vi ser også at determinasjonskoeffisienten R^2 er gjeve ved 0.018, dette vil seie at interessevariablane *fødselsmånad* forklarer 1.8% av variasjonen i testscoren.

5.4 Problemstilling 2

Har foreldra sitt tilsettingsstatus, utdanningsnivå og hushaldet inntekt ein påverknad på testscoren?

I denne delen av oppgåva introduserer vi fleire kontrollvariablar. Vi har danna fire nye modellar, der vi tillegg modellane éin og éin ny kontrollvariabel. Disse kontrollvariablane er i utgangspunktet kategorivariablar for høvesvis foreldra sin tilsettingsstatus (*par_emp*), foreldra sitt utdanningsnivå (*par_edu*) og hushaldet si inntekt (*income*). Vi har gjort om disse kategorivariablane til dummyvariablar beskrive tidlegare i oppgåva (seksjon 4.2). Vi har valt å gjere det på denne måten fordi vi skal bruke dummyvariablane til å sjå på interaksjonen med interessevariablane seinare i oppgåva.

5.4.1 Foreldra sitt tilsettingsstatus

Fyst byrjar vi med å sjå på kva effekt foreldra sin tilsettingsstatus har på testscore. Her har vi lagt til dummyvariabel, *par_emp_low*, der begge foreldre er arbeidsledig eller har stilling under 100% tilsvara 1 og foreldre med høgre stillingar enn dette tilsvara 0, i grunnmodellen. Modellen er gjeve ved:

Modell 2: $read = \beta_0 + \beta_1 fødselsmåned4\_6 + \beta_2 fødselsmåned7\_9 + \beta_3 fødselsmåned10\_12 + \delta_1 par\_emp\_low + \varepsilon_i$

Vi gjennomfører regresjonen i STATA og får at born med foreldre med ingen eller stilling under 100% har negativ effekt på testscore *read*. Ein får frå STATA at koeffisienten til *par_emp_low* er gjeve -38.09, når resten av variablane er konstante. Vi ynskjer å gjennomføre ein hypotesetest, i form av ein t-test, på *par_emp_low* for å sjekke om dummyvariabelen er signifikant. Vel også her eit 5% signifikansnivå og kritisk verdi er gjeve ved ±1.96 funnet i t-fordelingstabellen (Thomas, 2005, s. 587). Dersom testobservatoren er lågare enn kritisk verdi beheld vi nullhypotesen og har at foreldra sitt jobbsituasjon ikkje har noko effekt på korleis born presterer i skulen. Vi har følgjande hypotesar:

$H_0: \delta_1 = 0$
$H_A: \delta_1 \neq 0$

Vi får at testobservatoren er ±11.2123 og overstig kritisk verdi. Vi kan dermed forkaste nullhypotesen hjå eit 5% signifikansnivå og kan seie ved 95% sikkerheit at foreldra sin tilsettingssatus har ein effekt på korleis elever presterer på skulen.

5.4.2 Utdanningsnivå hjå foreldra

Vidare skal vi sjå på om utdanningsnivå hjå foreldra har noko effekt på testscore. I denne modellen har vi lagt ved dummyvariabelen *par_edu_high*, der foreldre med høgre utdanning er lik 1 og foreldre utan høgre utdanning er lik 0. Modellen er gjeve ved:

Modell 3: $read = \beta_0 + \beta_1 fødselsmåned4\_6 + \beta_2 fødselsmåned7\_9 + \beta_3 fødselsmåned10\_12 + \delta_1 par\_emp\_low + \delta_2 par\_edu\_high + \varepsilon_i$

Vi gjennomfører ein regresjon i STATA. Får frå STATA at koeffisienten til par_edu_high er 42.63. Dette vil seie at born med foreldre som har høgre utdanning har ein positiv effekt på testscoren, gjeve at resten av variablane er konstante. Vi vil også her gjennomføre ein t-test for å sjå om resultatet er signifikant. Vel 5% signifikansnivå og kritisk verdi er gjeve ved 1.96. Har følgjande hypotesar:

$H_0: \delta_2 = 0$
$H_A: \delta_2 \neq 0$

Får her at testobservatoren er lik 15.9067 og overstig kritisk verdi. Dette vil igjen seie at vi kan forkaste nullhypotesen hjå eit 5% signifikansnivå og kan med 95% sikkerheit seie at foreldre med høgre utdanning har ein positiv effekt på testscoren.

5.4.3 Hushaldets inntekt

I denne modellen har vi tillagt dummyvariabelen *income_high*. Denne viser til hushaldets inntekt der *income_high* = 1 tilseier at hushaldet har ei inntekt på $40.000 eller meir i året, og lik 0 dersom inntekta er mindre $40.000.

Modell 4: $read = \beta_0 + \beta_1 fødselsmåned4\_6 + \beta_2 fødselsmåned7\_9 + \beta_3 fødselsmåned10\_12 + \delta_1 par\_emp\_low + \delta_2 par\_edu\_high + \delta_3 income\_high + \varepsilon i$

Resultat frå STATA viser at koeffisienten er 11.33. Dette vil seie at elevar der hushaldet har ein inntekt på $40.000 eller meir scorar 11.33 testscore poeng meir, gjeve at resterende variablar haldast konstant. Gjennomfører ein t-test med 5% signifikansnivå og kritisk verdi 1.96.

H_0: $\delta_3 = 0$

H_A: $\delta_3 \neq 0$

T-testen gjev testobservator lik 3.8296 og vi forkastar nullhypotesen ved eit 5% signifikansnivå og kan med 95% sikkerheit seie at inntekt har ein positiv effekt på testscore.

Tabell 5.4.1 : Resultat frå regresjonen for modell (1)-(4)

	Modell 1	Modell 2	Modell 3	Modell 4
Tilsettingsstatus		x	x	x
Utdanningsnivå			x	x
Hushaldets inntekt				x
Differansen til referansevariabelen *fødselsmåned1_3*				
Fødselsmåned4_6	- 13.65	- 13.61	- 13.56	- 13.66
Fødselsmåned7_9	- 21.53	- 22.23	- 22.30	- 22.23
Fødselsmåned10_12	- 28.75	- 29.60	- 28.54	- 28.47
R^2	0.018	0.053	0.108	0.113

Ser frå tabellen at når vi legg til fleire faktorar som må takast omsyn til, at det ikkje gjev tilsynelatande store endringar på testscore- forskjellen mellom fødselskvartalen. Dette tilseier at ved å legge til fleire faktorar vil ikkje det påverke differansen på testscore mellom fødselskvartalen i noko særleg stor grad. Ser derimot at føyningsmålet R^2 har endrar seg stort etter kvart som det blir tillagt fleire faktorar. Dette er fordi at variasjonen i den avhengige variabelen, *read*, er fordelt på fleire variablar enn ved tidlegare modell.

5.5 Tilleggsproblemstillingar

I denne delen av oppgåva ynskjer vi å sjå på interaksjonen mellom fødselskvartala og hushaldets inntekt, foreldra sin tilsettingsstatus og foreldra sitt utdanningsnivå. Vi vel å køyre interaksjonen mellom disse variablane kvar for seg slik at vi har danna tre ulike likningar, der kvar likning kunn tek føre seg kvar enkelt interaksjon mellom fødselskvartal og kvar enkelt familiekarakteristika.

Det fyste vi ynskjer å sjå på er korleis effekt av ein endring i inntekt har på testscoren, gjeve kva fødselskvartal eleven er føydd. Deretter skal vi sjå på foreldra sin tilsettingsstatus og til slutt foreldra sitt utdanningsnivå.

Har hushaldets inntekt ulik effekt på testscore gjeve når på året eleven er føydde?

Her har vi tillagt modellen interaksjonsledd gjeve ved alle fødselskvartala og dummyvariabelen *income_high*. Som sagt ynskjer vi å sjå på kva effekt ein endring i inntekt har på testscore, når observasjonane er føydde i januar-mars, april-juni, juli-september eller oktober-desember. Vi har ein referansevariabel som vi har valt ved å utelate variabelen frå likning, referansevariabelen er (*fødselsmåned1_3*income_high*). Vi har modellen er gjeve ved:

$$\textbf{Modell 5}: read = \beta_0 + \beta_1 fødselsmåned4\_6 + \beta_2 fødselsmåned7\_9 + \beta_3 fødselsmåned10\_12 + \delta_1 par\_emp\_low + \delta_2 par\_edu\_high + \delta_3 income\_high + \gamma_1 (fødselsmåned4\_6 * income\_high) + \gamma_2 (fødselsmåned7\_9 * income\_high) + \gamma_3 (fødselsmåned10\_12 * income\_high) + \varepsilon i$$

Fødselkvartalet si endring i testscore ved ein eining endring i hushaldets inntekt er høvesvis $(\delta_3 + \gamma_1)$, $(\delta_3 + \gamma_2)$, $(\delta_3 + \gamma_3)$. γ_i representera den ytterlegare effekt ein endring i inntekt har på testscore, gjeve fødselskvartal. Vi gjennomfører ein regresjon i STATA, og får følgjande

estimatar for $(\delta_3 - \gamma_1) = 13.30 - 0.824 = 12.48$,$(\delta_3 - \gamma_1) = 13.30 - 9.371 = 3.93$, $(\delta_3 + \gamma_3) = 13.30+3.083 = 16.38$.

Vi gjennomfører ein hypotesetest for disse estimatane for å sjekke om dei er signifikante. Her blir H_0: $\gamma_i = 0$, der (i=1,2,3) og H_A: $\gamma_i \neq 0$. Intuisjonen bak denne hypotesetesetesten er at dersom nokre av koeffisientane er ulik null betyr det at born føydde i ulike fødselskvartal får ulik endring i testscore gjeve ein endring i hushaldets inntekt. Dersom det viser seg at koeffisienten er null betyr det at ein endring i hushaldets inntekt har same effekt på testscoren uansett fødselskvartal. Vi brukar ein t-test for å sjekke kvar enkelt parameter. Ser frå hypotesetesten at ingen av koeffisientane overstig kritisk verdi og vi beheld nullhypotesen. Dette betyr at vi ikkje har noko bakgrunn for å kunne seie at hushaldets inntekt påverkar testscore-resultata ulikt for dei forskjellige fødselskvartala. Sjekkar også for F-test og vi beheld også nullhypotesen her.

Har foreldra sitt tilsettingsstatus ulik effekt på testscore gjeve når på året eleven er føydde?
Vidare ynskjer vi å sjå på korleis ein endring i foreldra sin tilsettingsstatus påverkar testscore, *read*, gjeve dei ulike fødselskvartala. Også her har vi valt ein referansekategori ved (*par_emp_low*fødselsmåned1_3*). Vi har følgjande modell:

Modell 6: $read = \beta_0 + \beta_1 fødselsmåned4\_6 + \beta_2 fødselsmåned7\_9 + \beta_3 fødselsmåned10\_12 + \delta_1 par\_emp\_low + \delta_2 par\_edu\_high + \delta_3 income\_high + \gamma_1 (fødselsmåned4\_6 * par\_emp\_low) + \gamma_2 (fødselsmåned7\_9 * par\_emp\_low) + \gamma_3 (fødselsmåned10\_12 * par\_emp\_low) + \varepsilon i$

Fødselskvartalet si endring i testscore ved ein eining endring i foreldra sin tilsettingsstatus er også her høvesvis $(\delta_1 + \gamma_1)$, $(\delta_1 + \gamma_2)$, $(\delta_1 + \gamma_3)$. Der γ_i representerer ein ytterlegare effekt. Vi har same intuisjon som likninga ovanfor, og ynskjer også her å sjå på ein t- og F-test for å sjå om koeffisientane er signifikante. Vi ser frå t-testen at ingen av koeffisientane overstig kritisk verdi og vi kan derfor ikkje forkaste nullhypotesane. Også frå F-testen viser det seg at koeffisienten ikkje er signifikant. Vi beheld derfor nullhypotesane ved eit 5% signifikansnivå.

Har foreldra sitt utdanningsnivå ulik effekt på testscore gjeve når på året eleven er føydde?
I den siste modellen vi har konstruert, ynskjer vi å sjå på korleis ein endring i foreldra sitt utdanningsnivå påverkar testscore, *read*, gjeve dei ulike fødselskvartala. Her er referansevariabelen (*par_edu_high*fødselsmåned1_3*). Modellen er gjeve ved:

Modell 7: $read = \beta_0 + \beta_1 fødselsmåned4\_6 + \beta_2 fødselsmåned7\_9 + \beta_3 fødselsmåned10\_12 + \delta_1 par\_emp\_low + \delta_2 par\_edu\_high + \delta_3 income\_high + \gamma_1 (fødselsmåned4\_6 * par\_edu\_high) + \gamma_2 (fødselsmåned7\_9 * par\_edu\_high) + \gamma_3 (fødselsmåned10\_12 * par\_edu\_high) + \varepsilon_i$

Frå *modell 7* har vi at fødselskvartalet si endring i testscore ved ei eining endring i foreldra sitt utdanningsnivå er representert ved $(\delta_2 + \gamma_1)$, $(\delta_2 + \gamma_2)$, $(\delta_2 + \gamma_3)$. Der γ_i også her representere ein ytterlegare effekt. Også i denne modellen har vi same intuisjon for ein hypotesetest som i de to føregåande modellane. Vi gjennomfører også her fyrst ein t-test gjeve 5% signifikansnivå, før vi ser på ein F-test. Resultata frå t-testen viser at ingen av koeffisientane er høgre enn kritisk verdi og vi kan derfor ikkje forkaste nullhypotesen. Det same gjeld for F-testen, og vi beheld nullhypotesen ved eit 5% signifikansnivå.

5.6 Tolking av resultat

For å kunne svare på første problemstilling konstruerte vi *modell 1*. Frå *modell 1* ser vi at born føydde seint på året har tendensar til å gjere det dårlegare på PIRLS- testen, og har ein betraktelig lågare testscore enn born føydde tidlig på året. Vi har også sjekka for om estimatane er signifikante, og kan derfor seie at born føydde seinare på året gjer det generelt dårlegare enn born føydde tidlig på året og testscore reduserast etter kronologisk rekkefølge av fødselskvartelene. Dette er tilfellet når vi ikkje kontrollerer for andre potensielle forklaringsfaktorar.

I den andre problemstillinga konstruerte vi tre nye modellar. I disse modellane la vi til ein ekstra forklaringsvariabel som representerte ulik familiekarakterstikk for kvar ny modell. Grunnen til at vi gjorde dette er fordi vi ønska å sjå på kva effekt familiekarakterstikk har som påverknad på elevprestasjon, og om det hadde ulik effekt på fødselskvartala.

Frå *modell 2* ser vi tendensar til at born med foreldre som korkje har arbeid, eller jobbstillingar under 100%, har ein lågare testscore enn born med foreldre som har ein eller fleire foreldre med 100% stillingar. Vi ser også at når vi legg til i modellen ein ny forklaringsvariabel ikkje gjev store utslag på koeffisientane til fødselsmånad, og disse held seg relativt stabile.

Under *modell 3* har vi lagt til variabelen *par_edu_high*. Desse modellane tydar på at foreldra si utdanning har ein effekt på testscore. Ser at respondentane som har foreldre med høgre utdanning har testscore på 42.63 poeng høgre enn respondentar med foreldre utan høgre utdanning. Ved tillegg av enda ein forklaringsvariabel ser vi på *modell 3* at det ikkje gjev noko større utslag på koeffisienten til fødselskvartala, og at dei også her held seg ganske stabile samanlikna med *modell 1*.

I den *fjerde modellen* har vi lagt til enda ein ny variabel. Denne gangen ser vi på inntekt, og kan også her sjå at respondentar som kjem frå eit hushald med nokså høg inntekt gjer det betre enn respondentar frå hushald med lågare inntekt. Vi kan sjå frå dei andre modellane at inntekt ikkje har like stor effekt på testscore som foreldra sitt utdanningsnivå og tilsettingsstatus. Det er også viktig å presisere her at utvalet viser til at det er veldig få observasjonar frå datamaterialet som kjem frå eit hushald der inntekta er lågare enn $39.999. Også her er det heller ingen store utslag på koeffisientane til fødselskvartala.

Når det gjeld modellane vi har konstruert for tilleggsspørsmåla, altså *modell 5, 6* og *7*, finn vi ikkje noko bevis for at respondentane i dei forskjellige fødselskvartala får ulikt utbytte av dei forskjellige familiekarakterstikkene – tilsettingsstatus, utdanningsnivå og inntekt. Også når vi køyrte regresjonsanalysen i STATA såg vi at interaksjonsledda hadde høg p- verdi. Dette kan tyde på at estimatane ikkje er signifikante, noko som også stemte når vi gjennomførte både t-test og F-test for interaksjonsvariablane.

6.0 Oppsummering og konklusjon

6.1 Oppsummering

Vi har i oppgåva forsøkt å finne samanheng mellom lesepresentasjonar og innverknad frå relativ alder og sosioøkonomisk bakgrunn. Fyst forsøkte vi kartlegge tidlegare forsking, og såg at det var betydeleg meir forsking og empiri på sosioøkonomisk bakgrunn enn fødselskvartal. Vi tydde til Coleman sin skuleproduktfunksjon, og der valte variablar som kom innanfor elev-/og familiekarakteristika. Datasettet vi tok i bruk var PIRLS-undersøkinga frå 2001 som tok føre seg elevar på 4. og 5. årstrinn. For å finne effekt av relativ alder koda vi om fødselsmånad (*birthm*) til fødselskvartal og nytta utdanning hjå foreldra (*par_edu)*,

tilsettingsstatus foreldra (*par_emp*) og inntekta til familien (*income*) som mål på sosioøkonomisk bakgrunn.

Vi nytta minste kvadratars metode (OLS) og enkel lineær-regresjon. Fyst tok vi føre oss berre fødselskvartal, og såg etter ein lineær samanheng mellom fødselskvartala (*fødselsmåned1_3, fødselsmåned4_6, fødselsmåned7_9* og *fødselsmåned10_12*) og lesescore (*read*). Deretter la vi til i analysane fleire dummyvariablar kvar gong i følgande rekkefølge: *par_emp_low, par_edu_high* og *income_high* som gjev måla på tilsettingsstatusen, utdanninga og inntekta. Til slutt nytta vi interaksjonsledd med dei fødselskvartala og dei ulike familiekarakteristika kvar for seg i modell 5, 6 og 7.

6.2 Konklusjon

I vår analyse ser vi at fødselskvartal har betydeleg påverknad på elevane sine prestasjonar på lesetesten. Også foreldrekarakteristikaen vi testa for i problemstilling 2 ga signifikante funn, der borna sin sosioøkonomiske bakgrunn har påverknad for leseprestasjonane. Når vi såg på samspelet mellom relativ alderseffekt og sosioøkonomisk bakgrunn, fann vi ingen signifikante verdiar. Så vi kan derfor ikkje konkludere med at born føydde seint på året blir meir eller mindre påverka av sosioøkonomisk bakgrunn enn born føydde tidlegare på året. Det einaste vi med sikkerheit kan lese av analysane vi har gjort, og med analysa sine føresetnader, er at relativ alder og sosioøkonomisk bakgrunn påverkar elevresultat kvar for seg.

7.0 Referanser

Bonesrønning, H. (2004). Utforming av utdanningspolitikken – Hva kan økonomene bidra med? *samfunnsøkonomene, 58*(3), s. 17.

Coleman, J. S. (1966). *Equality of educational opportunity.* Washington: Departement of Health Education and Welfare.

Hanushek, E. A. (2020). *Education production functions.* London: Academic Press: Economics of Education.

Helbæk, M. (2011). *Statistikk: Kort og godt* (3.utg.). Oslo: Universitetsforlaget.

NOU 2019:3. (2019) *Nye sjansen – bedre læring: Kjønnsforskjeller i skoleprestasjon og utdanningsløp*. Henta frå: https://nettsteder.regjeringen.no/stoltenbergutvalget/files/2019/02/nou201920190003000dddpdfs.pdf

Raaum. O. & Raabe M. (2004). *Foreldres utdanning avgjørende for barnas skolegang*. 2019) *Nye sjansen – bedre læring: Kjønnsforskjeller i skoleprestasjon og utdanningsløp.* Henta frå: https://www.ssb.no/utdanning/artikler-og-publikasjoner/foreldres-utdanning-avgjorende-for-barnas-skolegang

Raaum. O. (2003). *Familiebakgrunn, oppvekstmiljø og utdanningskarrierer*. 2019) *Nye sjansen – bedre læring: Kjønnsforskjeller i skoleprestasjon og utdanningsløp*. Henta frå: https://www.ssb.no/a/publikasjoner/pdf/sa60/kap-6.pdf

Ringdal, K. (2018) *Enhet og mangfold*. Bergen: fagbokforlaget

Thomas, R. L. (2005). *Using Statistics in Economics*. Berkshire: McGraw- Hill Education

Vedlegg:

Tabell V1: Fullstendig oversikt over estimerte resultater i modell (1)-(7)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	read	read	read	read	read	read	read
f√∏dselsm√•ned4_6	-13.65***	-13.61***	-13.56***	-13.66***	-13.12*	-13.92***	-16.92**
	(3.677)	(3.612)	(3.680)	(3.672)	(6.321)	(3.862)	(5.418)
f√∏dselsm√•ned7_9	-21.53***	-22.23***	-22.30***	-22.23***	-16.02*	-21.79***	-15.83**
	(3.682)	(3.616)	(3.676)	(3.667)	(6.309)	(3.844)	(5.519)
f√∏dselsm√•ned10_12	-28.75***	-29.60***	-28.54***	-28.47***	-30.46***	-28.66***	-28.51***
	(3.838)	(3.771)	(3.822)	(3.813)	(6.506)	(4.003)	(5.645)
f√∏dselsm√•ned1_3		0					
		(.)					
par_emp_low		-38.09***	-23.12***	-19.51***	-19.22***	-19.41*	-19.11***
		(3.397)	(4.646)	(4.730)	(4.734)	(8.981)	(4.739)
par_edu_high			42.63***	39.89***	39.95***	39.96***	41.19***
			(2.680)	(2.768)	(2.769)	(2.775)	(5.308)
income_high				11.33***	13.30*	11.36***	11.40***
				(2.958)	(5.620)	(2.961)	(2.959)
inchigh_4_6					-0.824		
					(7.766)		
inchigh_7_9					-9.371		
					(7.756)		
inchigh_10_12					3.083		
					(8.032)		
paremplow_4_6						2.699	

						(12.52)	
paremplow_7_9						-5.344	
						(12.93)	
paremplow_10_12						2.145	
						(13.24)	
pareduhigh_4_6							6.223
							(7.374)
pareduhigh_7_9							-11.33
							(7.383)
pareduhigh_10_12							0.124
							(7.668)
_cons	514.6***	521.8***	497.6***	491.2***	489.9***	491.2***	490.5***
	(2.605)	(2.637)	(3.063)	(3.473)	(4.680)	(3.545)	(4.233)
N	3403	3403	3054	3054	3054	3054	3054
R^2	0.018	0.053	0.108	0.113	0.113	0.113	0.114

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Vedlegg V2: Hushaldet si inntektsfordeling, income- variabelen

Verdi	Inntekt
1	Mindre enn $20.000
2	$20.000 – $29.999
3	$30.000-$39.999
4	$40.000-$49.999
5	$50.000-$59.999
6	$60.000- eller mer

Vedlegg V3: Foreldrenes utdanningsnivåfordeling, par_edu- variabelen

Verdi	Utdanningsnivå
1	Fullført universitet/ høgskole
2	Post secondary (denne er ikke i noway.dta)
3	Videregående skole
4	Ungdomsskole
5	Ikke fullført ungdomsskolen

Vedlegg V4: Foreldrenes jobbstatusfordeling, par_emp- variabelen

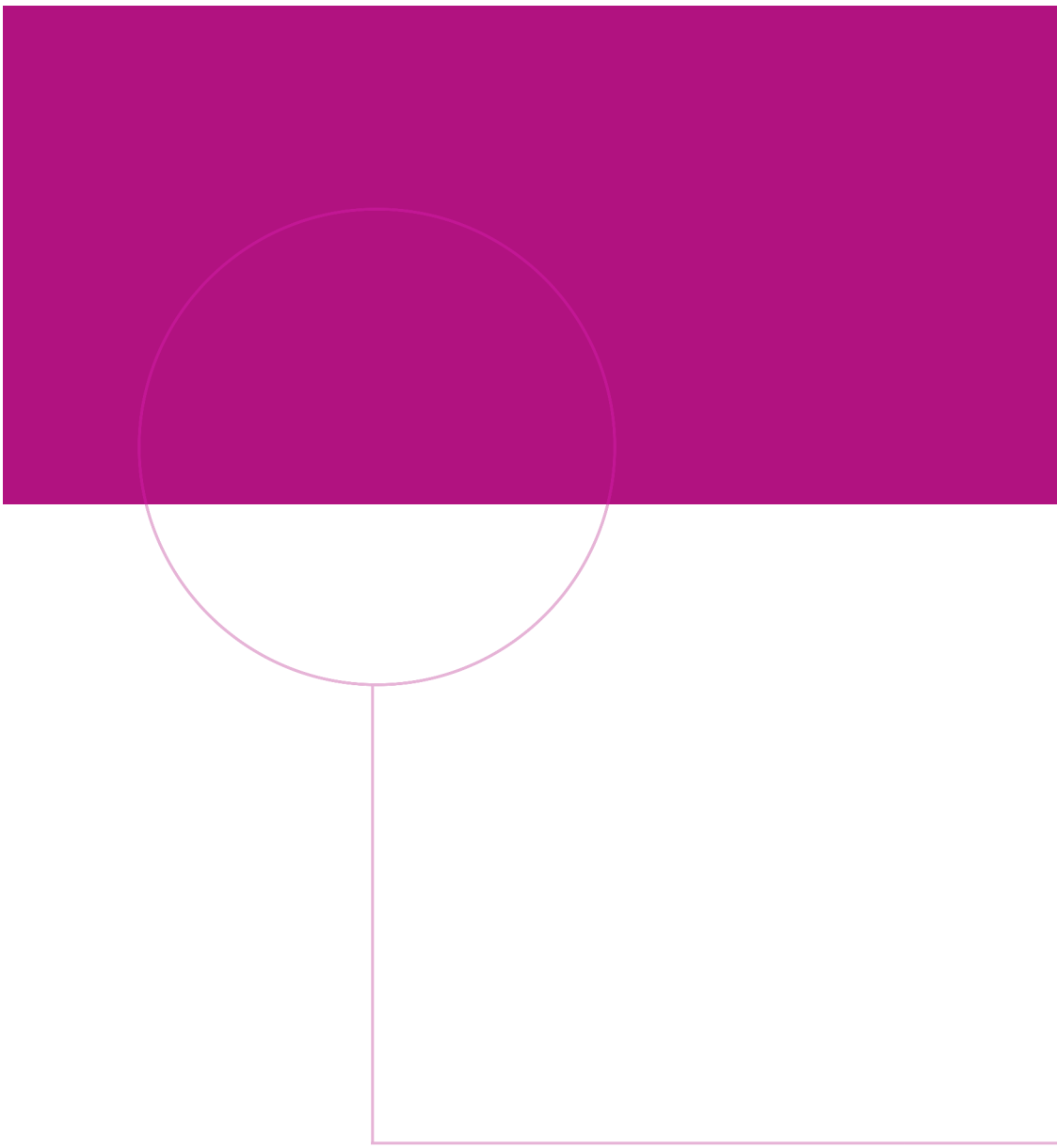
Verdi	Jobbstatus
1	Begge foreldre jobber fulltid
2	Én av foreldrene jobber fulltid
3	Begge foreldre jobber mindre enn fulltid
4	Begge foreldre er arbeidsledig

Vedlegg V5: Fødselsmåneder fordeling, birthm- variabelen

Verdi	Fødselsmåned
1	Januar
2	Februar
3	Mars
4	April
5	Mai
6	Juni
7	Juli
8	August
9	September
10	Oktober
11	November
12	Desember

Vedlegg V6: Korrelasjonsmatrise modell (1)-(4)

	read	fødselsmåned1_3	fødselsmåned4_6	fødselsmåned7_9	fødselsmåned10_12	par_emp_low	par_edu_high	income_high
read	1.0000							
fødselsmåned1_3	0.1210	1.0000						
fødselsmåned4_6	0.0098	-0.3471	1.0000					
fødselsmåned7_9	-0.0427	-0.3483	-0.3513	1.0000				
fødselsmåned10_12	-0.0921	-0.3152	-0.3179	-0.3190	1.0000			
par_emp_low	-0.1228	0.0080	0.0108	-0.0161	-0.0028	1.0000		
par_edu_high	0.2873	0.0081	-0.0211	0.0236	-0.0112	-0.1399	1.0000	
income_high	0.1590	0.0012	0.0050	0.0024	-0.0092	-0.2283	0.2808	1.0000



NTNU

Kunnskap for en bedre verden