



Norwegian University of
Science and Technology

Avkastning på utdanning

Mai 2019

Bacheloroppgave SØK2901

Institutt for Samfunnsøkonomi

Norges teknisk-naturvitenskaplige universitet - Trondheim

INNHOLDSFORTEGNELSE

1.	Innledning	2
1.1	<i>Problemstilling.....</i>	2
1.2	<i>Hypoteser:.....</i>	2
1.3	<i>Begrepsforklaring</i>	2
2	Tidligere studier.....	3
3	Datamateriale:	4
3.1	<i>PIAAC:</i>	4
3.2	<i>Variabler</i>	5
3.3	<i>Deskriptiv statistikk</i>	5
3.3.1	<i>Lønn</i>	6
3.3.2	<i>Menn og kvinner.....</i>	6
3.3.3	<i>offentlig- og privat sektor</i>	7
3.4	<i>Kontrollvariablene</i>	8
3.5	<i>Datasettets styrker og svakheter</i>	9
4	Teori	9
4.1	<i>Regresjonsanalyse</i>	10
4.2	<i>Determinasjonskoeffisienten R^2</i>	11
4.3	<i>Funksjonsformer</i>	11
4.4	<i>Minste kvadratets metode</i>	12
4.5	<i>Mincers regresjonsmodell</i>	14
4.4	<i>Hypotesetest.....</i>	15
5	Regresjonsanalyse	16
5.1	<i>Menn og kvinner.....</i>	17
5.2	<i>Offentlig- og privat sektor</i>	20
6	Konklusjon	22
	litteraturliste.....	23

1. INNLEDNING

I denne empiriske oppgaven skal jeg ta for meg tema lønnsavkastning. Når det er snakk om lønnsavkastning skal det fokuseres på hvordan antall år utdanning påvirker lønn. For å innsnevre temaet skal jeg undersøke om det er en signifikant forskjell i lønnsavkastning mellom kvinner og menn, samt offentlig- og privat sektor. Undersøkelsen gjøres med tall hentet fra PIAAC med fokus på Norge. Datasettet skal analyseres ved bruk av programvaren Stata.

1.1 PROBLEMSTILLING

Er lønnsavkastningen positiv ved ett år økt utdanning?

- Ser vi forskjell på lønnsavkastning mellom menn og kvinner?
- Har offentlig- og privat sektor ulik lønnsavkastning?

1.2 HYPOTESER:

Legger vi problemstilling til grunn vil jeg anta at menn og kvinner har relativt lik lønnsavkastning. Når vi skal ta for oss forskjeller på offentlig- og privat sektor ser man at gjennomsnittslønnen i privat sektor er høyere enn i offentlig sektor, samtidig vil samfunnet nyttes av kunnskapsrike arbeidstakere innenfor offentlig sektoren. Dette kan gi et insentiv for staten å gi høy lønn til høyt utdannede personer. Jeg vil derfor anta at vi ser en høyere lønnsavkastning i offentlig sektor.

1.3 BEGREPSFORKLARING

For å presisere funnene i ulike variabler legger vi ved en begrepsforklaring av relevante begreper som kan være uklare.

Avkastning:

Avkastning er knyttet til vinning av et gjennomført arbeid. I denne oppgaven vil begrepet brukes som nytte av antall år utdanning i form av lønn.

Offentlig- og privat sektor:

«Offentlig sektor defineres som statsforvaltning, kommuneforvaltning og forretningsmessige foretak som eies eller er kontrollert av staten eller kommunene.» «På den andre siden har vi privat sektor som beskrives som alle privateide bedrifter, ideelle organisasjoner, samt alle husholdninger.» (Idsø, 2017)

2 TIDLIGERE STUDIER

Vi ser at stadig flere velger å ta utdanning etter fullført videregående skole, og vi har derfor en økning i antall høyere utdannede. I tråd med dette kan vi anta at arbeidsgivere vil værere interessert i utdannede arbeidstakere. Det er også forståelig at når en tar lang utdanning vil en gjerne vite hvilken lønn man kan forvente. Noe av dette kan være årsaken til at vi finner mye forskning på tema og jeg skal derfor presentere noen tidligere forskningsartikler som tar for seg lønnsavkastning. (SSB, 2018).

Avkastning på lønn er et tema det har blitt forsket mye på. I den publiserte artikkelen «*A review of human capital theory: Microeconomics, 1 utgitt av Kai-Joseph Fleischauer i 2001*», analyseres flere studier som er gjort på tema. Jeg har valgt å presentere noen av disse i tabell 1, vist under. Av de forskjellige forskningene utført ser vi at det variere i prosentvis avkastning. Samtidig må det presesers at disse er utført på ulike vilkår med tanke på land og årstall. Det alle har til felles er at lønnsavkastning viser seg å være positiv, som vil si at økt utdanning gir økt lønn. Det kan derfor forventes at utfallet i denne oppgaven vil se noe lignende ut.

Tabell 1: Tidligere studier

Studie	Estimat
Ashenfelter and Krueger (1994)	12% - 16%
Dearden (1998)	5,5% - 9,3%
Ashenfelter, Harmon and Oosterbeek (1999)	6.6% - 9,3%
Wilson (2001)	5% - 10%
Arias and McMahon (2001)	11,7% - 13,3%

Tallene hentet fra tabell 1 er basert på gjennomsnitt av en populasjon som helhet. Det er derfor ikke tatt i betraktning andre bakenforliggende faktorer som kjønn og sektor.

3 DATAMATERIALE:

I dette kapitlet skal jeg presentere datamaterialet hentet fra PIAAC. Det vil innledningsvis bli gitt informasjon om datasettet. Videre blir det gitt en forklaring av variablene som blir brukt under analysen senere i oppgaven. Deretter legger vi frem deskriptiv statistikk for å få en oversikt over innholdet i variablene. Avslutningsvis ser vi på styrker og svakheter i datasettet.

3.1 PIAAC:

Datamateriale jeg skal bruke i denne oppgaven er basert på en undersøkelse i regi av OECD, utført i 2011. Undersøkelsen er navngitt «*The Programme for the International Assessment of Adult Competencies*» (PIAAC), der undersøkelsen er gjennomført i over 40 land. Formålet med undersøkelsen var å se på den voksne befolkningens ferdighetsnivå, med fokus på leseferdigheter, tallforståelse og problemløsning (OECD, 2012). Disse variablene ses på som grunnleggende for å ha en videre ferdighetsutvikling. I Norge ble undersøkelsen utført av SSB. Undersøkelsen brukte metodene intervju og selvutfylling. Intervjuene ble utført hjemme hos intervjuobjektet eller på SSB's lokaler. Deltakerne fikk utdelt gavekort på 500kr for å delta på undersøkelsen. (Bjørkeng, 2013).

I denne undersøkelsen skal jeg bruke tall hentet fra Norge og alle variablene som blir brukt stammer fra datasettet PIAAC. For Norge har vi 5 218 observasjoner i undersøkelsen. Dog har ikke alle observasjonen besvart på spørsmål som knytter seg til problemstilling, og vi har derfor filtrert ut disse. Det betyr at vi står igjen med 2 005 observasjoner.

3.2 VARIABLER

Tabell 2. Beskrivelse av variabler

Variabler	Beskrivelse
female	1 hvis kvinne, 0 hvis mann
male	1 hvis mann, 0 hvis kvinne
exper	Antall år arbeidserfaring
exper_sq	Kvadratrotten av arbeidserfaring
wage	Timelønn gitt i NOK
lnwage	logaritmen av timelønn
lit	Leseforståelse
lum	Tallforståelse
edu	Antall år fullført utdanning
pub	1 hvis offentlig sektor
priv_dummy	1 hvis privat sektor
edu_pub	Utdanning * Offentlig sektor
inter	Utdanning * male

Variablene wage og edu vil fungere som hoved variablene i denne undersøkelsen. Disse variablene tar for seg lønnsavkastning. Vi skal også se nærmere på male, female, pub og priv_dummy, da disse vil være relevante for problemstillingen. Utdanningsvariabelen tar for seg 7-21 år med utdanning. Det er dog ikke inkludert i variabelen hvilken type utdanning individene har. Det er også laget en dummy-variabel for privat sektor for bruk i analysedelen. I tillegg har vi to interaksjonsledd edu_pub og inter. Disse er relevante for analysen som vi kommer tilbake til senere i oppgaven.

3.3 DESKRIPTIV STATISTIKK

Under deskriptiv statistikk skal det legges frem data som blir relevant å analysere senere i oppgaven. Det vil være en enkel belysning av tall for lønn og utdanning, samt kjønn og sektor.

3.3.1 LØNN

Tabell 3. Lønn

Lønn Norge	Antall
Gjennomsnittslønn	242,7756
Min	100,1927
Maks	817,1259
Standardavvik	82,51224
Antall observasjoner	2,005

sum wage if edu

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	2,005	242.7756	82.51224	100.1927	874.1259

Tabell 3 viser en oversikt over gjennomsnittlig timelønn. Vi kan se at gjennomsnittslønnen ligger på 242,77 kr i timen. Den laveste timelønnen ligger på 100,19 kr og den høyeste ligger på 817,12 kr timen. Standardavviket er 82,51 som vil si at vi har høy spredning av inntekt blant de 2 005 observerte.

3.3.2 MENN OG KVINNER

Tabell 4. Lønnsforskjell: menn og kvinner.

	Alle	Menn	Kvinner
Gjennomsnitt	242,77	262,52	217,89
Min	100,2	103,44	100,19
Maks	874,12	874,13	598,29
Standardavvik	82,51	93,20	57,75
Antall observasjoner	2,005	1,118	887

Kvinner:

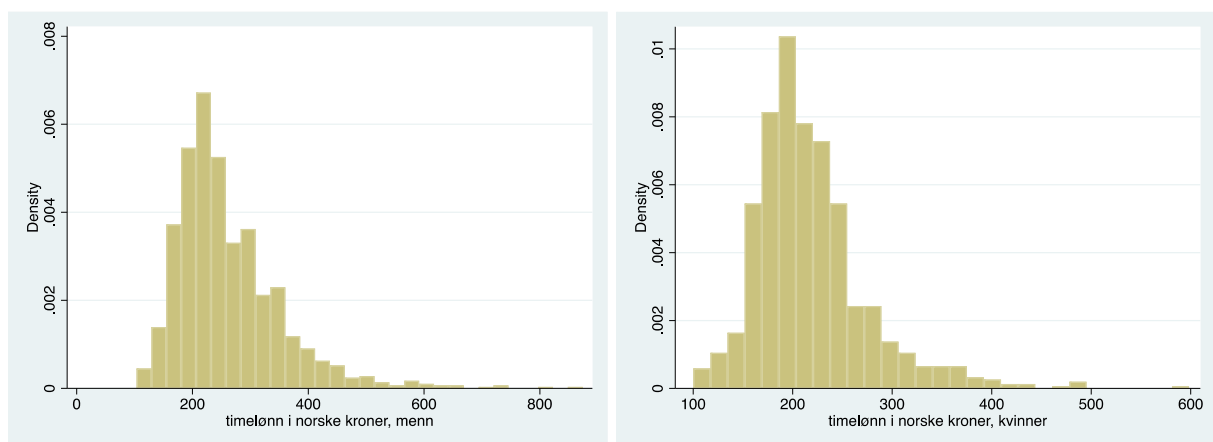
Variable	Obs	Mean	Std. Dev.	Min	Max
wage	887	217.8873	57.7544	100.1927	598.2906

Menn:

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	1,118	262.5215	93.20281	103.4378	874.1259

Tabell 4 viser en oversikt over gjennomsnittslønn fordelt på menn og kvinner. Av tabellen kan vi observere at kvinner har en lavere gjennomsnittslønn. Samtidig kan vi se av standardavviket at menn har større spredning i lønn i forhold til kvinner. Menn har også en høyere maks lønn, mens kvinner har lavest minste lønn.

Histogram 1. Spredning lønn, menn og kvinner



Histogrammene illustrer spredning av inntekt for menn og kvinner. På den vertikale aksene har vi prosentandelen og den horisontale aksene viser timelønn.

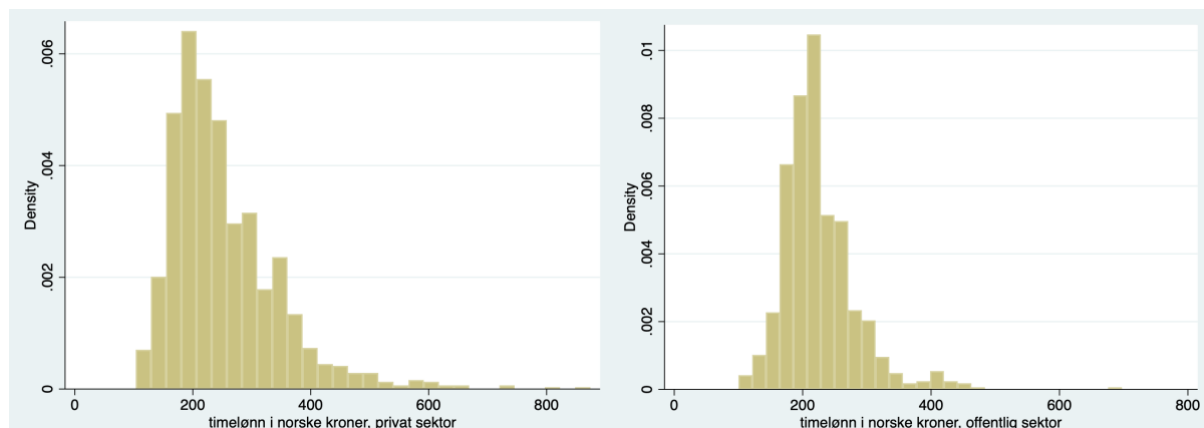
3.3.3 OFFENTLIG- OG PRIVAT SEKTOR

Tabell 5. Lønn offentlig- og privat sektor

	Alle	Off. sektor	Privat sektor
Gjennomsnitt	242,77	225,39	253,94
Min	100,2	100,19	103,44
Maks	874,12	697,12	874,13
Standardavvik	82,51	57,63	93,45
Antall observasjoner	2,005	784	1,221

Tabell 5 viser en oversikt over gjennomsnittslønnen fordelt på offentlig- og privat sektor. Av tabellen kan vi observere at gjennomsnittslønnen for offentlig sektor er lavere enn for privat sektor. Vi kan også se at offentlig sektor har både lavere minstelønn og lavere maks lønn. Ser vi på standardavviket er det større spredning av inntekt i privat sektor.

Histogram 2. Spredning lønn, offentlig- og privat sektor



Histogrammene illustrer spredning av inntekt for offentlig- og privat sektor. På den vertikale aksene har vi prosentandelen og den horisontale aksene viser timelønn.

3.4 KONTROLLVARIABLENE

Legger frem deskriptiv statistikk for kontrollvariablene.

Tabell 6. Deskriptiv statistikk, kontrollvariabler

Variabler	Alle		Kvinner		Menn	
	Gjennomsnitt	Std.	Gjennomsnitt	Std.	Gjennomsnitt	Std.
Utdanning	15,21	2,19	15,43	2,04	15,07	2,28
Arbeidserfaring	21,84	9,17	21,1	8,85	22,42	9,39
Leseforståelse	293,99	38,24	292,5	36,62	295,18	39,45
Tallforståelse	300,97	43,88	293,41	41,56	306,97	44,76
Offentlig sektor	39,1	-	53,54	-	27,63	-
Privat sektor	60,89	-	46,45	-	72,36	-

Av tabellen ser vi at kvinner i gjennomsnitt har lengre utdanning enn menn. Samtidig ser vi at menn skårer bedre på lese -og tallforståelse. Vi kan også observere at 53% av kvinner jobber i offentlig sektor, mens 27% av menn jobber i offentlig sektor. Totalt sett kan vi derfor anta at det er flere kvinner som jobber i offentlig sektor, og flere menn i privat sektor. Observasjonen vi har gjort oss ved fremleggelse av den deskriptive statistikken skal vi analysere senere i oppgaven.

3.5 DATASETTETS STYRKER OG SVAKHETER

Variablene som inngår i datasettet er relevant for å svare på oppgaven, noe som er positivt. I analysen av oppgaven ser vi på lønn, utdanning, menn, kvinner, offentlig og privat sektor. Dette gjør at vi kan skille variablene inn i grupper, noe som gjør analysen mulig. Det er derfor essensielt at disse variablene er en del av datasettet.

Selv om datasettet inneholder de riktige variablene finner vi noen svakheter. Ved bruk av de relevante variablene sitter vi igjen med 2 005 observasjoner. Dette er en liten del av den totale befolkningen og det blir derfor vanskelig å si at analysen vil få et representativt resultat. Datasettet gir heller ingen informasjon om hvilken type utdanning individene har, eller hvilken grad som er oppnådd. Vi kan derfor ikke skille mellom en som har tatt høyere utdanning i tre år og en med bachelorgrad. Under variabelen utdanning er det heller ikke gitt informasjon om karaktersnitt. Disse svakhetene i datasettet gjør at det er vanskelig å skille ut ettertraktede arbeidstakere, da grad av utdanning og karaktersnitt ikke er med i analysen.

4 TEORI

I denne delen av oppgaven skal jeg presentere relevant teori som skal brukes i besvarelsen. For å analysere og regne på tallene vi får av datasettet er det en rekke regresjonsmodeller, matematiske metoder og tester en kan gjøre. Teori som benyttes under analysedelen vil derfor presenteres i dette kapitlet og er hentet fra boken Thomas, R. L. (2005). «*Using statistics in economics*».

4.1 REGRESJONSANALYSE

For å analysere empirien er regresjonsanalyse et viktig hjelpemiddel. En regresjonsanalyse viser hvordan de forskjellige variablene avhenger av hverandre både lineært og ikke-lineært. For oppgavens del kan vi se at regresjonsanalyse blir en viktig faktor for å analysere virkningen av utdanning på timelønnen til individer i arbeid. For å se på virkningen av utdanning på lønn bruker vi minste kvadraters metode, forkortet MKM. Minste kvadraters metode viser oss hvordan den endogene variabelen påvirkes gitt en endring i de eksogene variablene. I denne oppgaven vil timelønn representere den endogene variabelen. De eksogene variablene vil variere. En rett funksjon skrives slik:

$$y = \alpha + \beta x$$

Vi forklarer funksjonen. α viser hvor linjen skjærer y-aksen og er konstantleddet. Der vi skal analysere virkningen av utdanning på lønn vil konstantleddet α være gjennomsnittslønnen. β viser den prosentvise endringen i Y. Y vil tilsvare forventet timelønn for den gjennomsnittlige arbeider og X tilsvarer antall år med utdanning. Videre forklart viser linjen en sammenheng mellom forklaringsvariabelen og responsvariabelen. Dette gjør vi for å finne det beste estimatet til den ukjente funksjonen. For å gjøre det må vi estimere α og β ut ifra datasettet. Etter at linjen er estimert får vi regresjonslinjen som skrives følgende:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

Vi kan også skrive regresjonslinjen på en generell måte:

$$Y_i = \alpha + \sum_{i=1}^n \beta_i X_i + \varepsilon$$

Av den generelle regresjonslikningen vil Y tilsvare utfallsvariabelen, α tilsvarer konstantleddet, X_i er forklaringsvariabler og β_i viser effekten av X_i på Y. Vi har også et restledd ε , som fanger opp andre faktorer som påvirker Y. Restleddet er som oftest udefinert.

4.2 DETERMINASJONSKOEFFISIENTEN R^2

Determinasjonskoeffisienten R^2 er et parameter en kan bruke for å se om det er en lineær sammenheng mellom de endogene- og eksogene variablene. R^2 har en verdi mellom 0 og 1. Om parameteren er 1 vil det si at X og Y har en perfekt lineær sammenheng, noe som innebærer at linjen krysser alle observasjonene. Er parameteren på 0 har vi ikke en lineær sammenheng mellom X og Y. Determinasjonskoeffisienten fremstilles slik:

$$R^2 = 1 - \frac{SSR}{SST}$$

4.3 FUNKSJONSFORMER

Ved lineær regresjon har vi flere funksjonsformer og disse vil gi ulike kurver. Viser til tabell 7 illustrert under.

Tabell 7. Funksjoner

Navn	Funksjon
Lineær	$Y = \alpha + \beta X$
Resiprok	$Y = \alpha + \frac{\beta}{X}$
Ekspontiell	$Y = \alpha e^{\beta X}$
Dobbellogaritmisk	$Y = \alpha X^\beta$
Semi-logaritmisk	$Y = \alpha + \beta \ln(\chi)$

Forklarer variablene.

Y-variabelen påvirkes av α og X. α er konstant og sier hvor grafen krysser Y-aksen og X er en eksogen variabel.

Ved Minste Kvadraters Metode tar vi utgangspunkt i en vilkårlig linje og det ligger noen forutsetninger til grunn.

- Restleddene forventes lik null.
- Det er en konstant varians til restleddene og den er uendelig for alle verdier.
- Restleddene er uavhengig av hverandre.
- Det er ingen sammenheng mellom restleddet og dens x verdi.
- Alle restleddene er normalfordelt

Er forutsetningene oppfylt vil estimatorene for MKM ha disse egenskapene:

1. De er lineære funksjoner av Y:

$$b = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \sum_{i=1}^n \left(\frac{x_i}{\sum_{i=1}^n x_i^2} \right) Y_i = \sum_{i=1}^n k_i Y_i$$

$$a = \bar{Y} - b\bar{X} = \sum_{i=1}^n \left(\frac{1}{n} - \bar{X} k_i \right) Y_i = \sum_{i=1}^n g_i Y_i$$

2. De er forventningsrette estimatorer for α og β :

$$E(b) = \beta$$

$$E(a) = \alpha$$

3. De har minst varians blant lineære, forventningsrette estimatorer:

$$V(b) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

og

$$V(a) = \frac{\sigma^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n x_i^2}$$

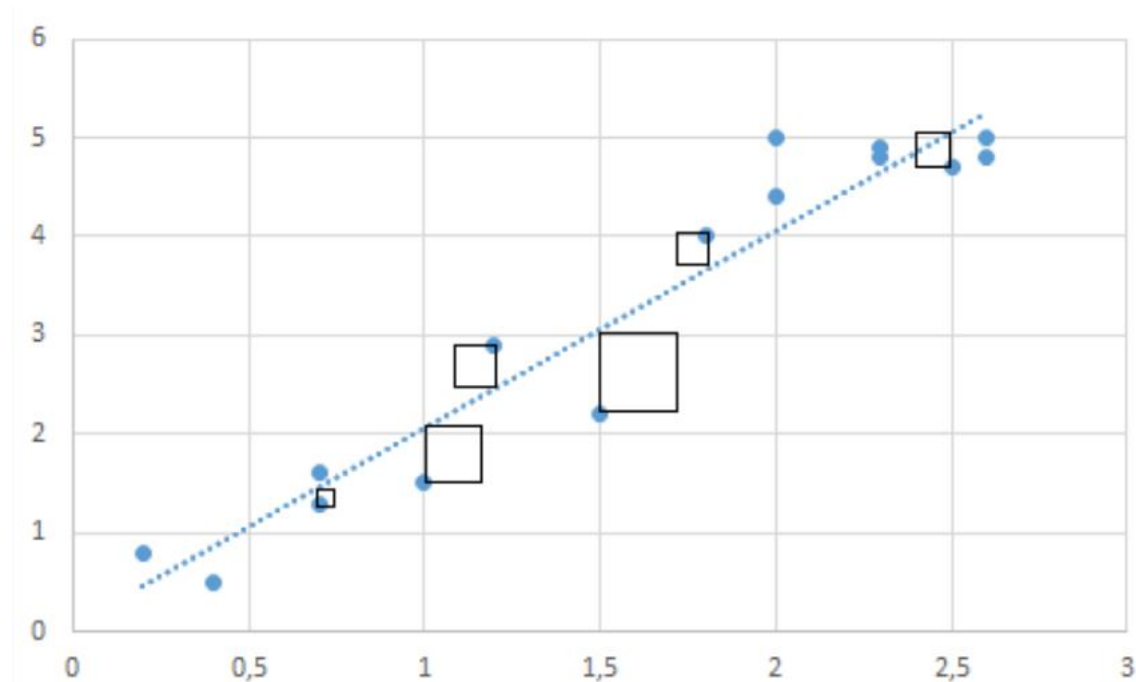
4. De er normalfordelte:

$$b \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right)$$

$$a \sim \left(\alpha \frac{\sigma^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n x_i^2}\right)$$

I funksjonene er x_i og y_i representert som avvik for X og Y. Vi vil finne ut hvilken modell som passer dataen best mulig, dette gjør vi ved å finne avvikskvadratet. Avvikskvadratet finner vi ved å kvadrere avstanden fra hvert punkt til kurven, illustrert i graf 1. Dette gir oss en indikasjon på samlet avstand fra punktene til linjen, gitt residualen e_i . Den laveste kvadratsummen vi kan få er 0,110.

Graf 1. MKM kvadratavvik



Residual:

$$e_i = Y_i - \widehat{Y}_i$$

Setter inn for $\widehat{Y}_i = a + bX_i$:

$$\rightarrow e_i = Y_i - (a + bX_i)$$

Vi ser av graf 1 at avstanden fra linjen og punktet måles vertikalt ettersom vi ønsker å se hvordan de eksogene variablene påvirker den endogene. Det viser da avviket til Y-variabelen. Den modellen som gir lavest residual vil være den mest optimale modellen å benytte. Den totale summen av residualene sitt avvik kan også formuleres som *sum of squared residuals* (SSR). I tillegg har vi residual vi kan forklare *explained sum of squares* (SSE). SSR og SSE summert gir den totale residualen:

$$SSR = \sum_{i=1}^n e_i^2$$

Etter bruk av MKM sammenligner vi SSR og finner den med lavest verdi. Vi antar da vi har funnet den optimale modellen.

4.5 MINCERS REGRESJONSMODELL

Vi har en kjent metode som brukes for å se på avkastning av utdanning, laget av Mincer. I Mincer-modellen ser man på inntekt som en funksjon av erfaring og utdanning. Mincer's metode baseres på at den eneste kostnaden ved å ta ett år ekstra utdanning er inntekten en mister det året man studerer.

Mincers regresjonsmodell:

$$\ln(Y_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2^2 + \varepsilon_i$$

Forklarer variablene:

Y tilsvarer individets forventede inntekt. X_1 tilsvarer antall år med utdanning, og X_2 viser et mål for arbeidserfaring. Vi har også et støyledd ε_i , som representerer andre faktorer timelønnen påvirkes av. Kvadrert X_2 viser at avkastningen på arbeidserfaring i form av økt timelønn vil øke, men samtidig avtakende. β_1 beskriver avkastningen på lønn per ekstra år utdanning.

Viser fremgangsmåten:

$$\text{Eksponentiell funksjon } Y = \alpha e^{\beta X}$$

$$\ln(Y) = \ln(\alpha) + \ln(e^{\beta X})$$

$$\ln(Y) = \ln(\alpha) + \beta X * \ln(e)$$

$$\ln(Y) = \ln(\alpha) + \beta X$$

Vi forklarer uttrykkene. $\ln(\alpha)$ viser der funksjonen krysser Y-aksen, som betyr at det er konstantleddet. I Mincer-modellen vil dette være lik β_0 . Siden det er en lineær funksjon kan vi legge til flere eksogene variabler og derav se virkningen av arbeidserfaring og utdanning. Vi setter da svaret som eksponenten til e for å finne timelønnen.

4.6 HYPOTESETEST

Vi bruker hovedsakelig hypotesetest for å teste om resultatene fra regresjonsanalyser er statistisk signifikante. Når vi utfører en hypotesetest setter vi opp en nullhypotese H_0 , og en alternativ hypotese H_A . Når man utfører hypotesetesten ønsker vi å finne ut om datamaterialet gir grunnlag for å beholde eller forkaste nullhypotesen. Ved bruk av hypotesetest må vi se på signifikansnivået. Signifikansnivået sier hva sannsynligheten er for å forkaste H_0 feilaktig. Med et signifikansnivå på 5% vil det være 5% sannsynlighet for at vi forkaster nullhypotesen feilaktig. Man kan også teste nullhypotesen med 1% signifikansnivå, som vil si at det er mindre sannsynlig at vi feilaktig forkaster nullhypotesen.

I denne oppgaven vil en hypotese eksempelvis ta for seg hvordan ett år ekstra utdanning påvirker lønnen. Her kan vi teste om lønnen i privat sektor endrer seg mer enn lønnen i offentlig sektor, derav om det er mer lønnsomt for de i privat sektor å ta ett år ekstra utdanning enn de i offentlig sektor.

En slik hypotesetest ser følgende ut:

$$H_0: \beta_{\text{offentlig sektor}} = \beta_{\text{privat sektor}}$$

$$H_A: \beta_{\text{offentlig sektor}} < \beta_{\text{privat sektor}}$$

I oppgaven operer vi med et stort utvalg av individer og derfor tar vi i bruk kritiske verdier fra t-tester når vi utfører hypotesetestene. De kritiske verdiene bestemmes av størrelsen på utvalget, signifikansnivå og om vi har en ensidig eller tosidig test. En tosidig test ser man om

verdiene er signifikant forskjellig, mens i en ensidig test ser man om verdiene er signifikant større eller mindre.

Testene i denne oppgaven vil være med signifikansnivå 1% og 5%, da får vi følgende verdier:

$t_{0,01}(> 120) = 2,326$ en ensidig med 1% signifikansnivå
 $t_{0,05}(> 120) = 1,645$ en ensidig med 5% signifikansnivå
 $t_{0,005}(> 120) = 2,576$ en tosidig med 1% signifikansnivå
 $t_{0,025}(> 120) = 1,96$ en tosidig med 5% signifikansnivå

Når vi utfører testene velger vi å se på verdien av β , ut ifra regresjonsanalysen. β -verdien forteller oss hvor mye timelønnen øker om vi øker utdanningen med gitt antall år. Vi vil derfor se hvordan β -verdien virker inn på to ulike utvalgsgrupper når vi endrer samme variabel. I denne oppgaven vil det være menn og kvinner, samt offentlig- og privat sektor. Ved regresjon av variablene mann og kvinne kan vi se hvordan deres forventede timelønn avhenger av utdanning. Vi må da utføre to regresjonsanalyser, en for kvinner og en for menn. Vi kan da regne ut testobservatoren ved hjelp av β -verdiene.

$$TS = \frac{b - \beta}{Sb}$$

Der vi har følgende verdier:

b = betaverdien for kvinner

β = betaverdien for menn

S_b = Standardavviket til betakoeffisienten for kvinner

Nullhypotesen forkastes dersom testobservatoren er større en de kritiske verdiene.

5 REGRESJONSANALYSE

I denne delen av oppgaven skal vi utføre regresjonsanalyse. Vi skal ta for oss avkastning av utdanning mellom menn og kvinner, samt se om det er forskjeller mellom offentlig og privat sektor. For å undersøke dette skal jeg bruke Mincers-metode som ser på effekten av utdanning og arbeidserfaring på lønn. I fremleggelsen av deskriptiv statistikk ser vi fra tabell 4 og 5 at

utdanning gir en positiv lønnsavkastning. Dette gjelder om man er mann og kvinne, samt om man jobber i offentlig og privat sektor. Videre i kapittelet skal vi analysere om det er forskjeller mellom disse gruppene.

5.1 MENN OG KVINNER

Hypotese: Avkastning på utdanning er lik for menn og kvinner:

Når jeg skal teste om avkastning på utdanning er lik for menn og kvinner velger jeg å bruke en Mincer-modell. Modellen er forklart i teorikapittelet 4.5.

$$\ln W = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2$$

$\ln W$ står for logaritmen til lønn og er i tabellen skrevet som \ln lønn. S tilsvarer antall år fullført utdanning, og X er antall år arbeidserfaring. Estimerer modellen i Stata og får følgende tall:

Tabell 8. Regresjonsresultat 1 modell

Variabler	ln lønn
edu	0,0521*** (-0,00292)
exper	0,0177*** -0,00305
exper_sq	-0,00027 (-6,56E-05)
Konstant	4,418*** (0,0585)
Observasjoner	2,005
R-squared	0,147

T verdier i parentes

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Den estimerte modellen viser en signifikant avkastning på utdanning. Ved økt utdanning med ett år, øker lønnen med 5,2%. Arbeidserfaring har en noe lavere avkastning på lønn, der lønnen øker med 1,7% ved økt arbeidserfaring med ett år. Den kvadrerte arbeidserfaringsvariabelen er negativ og viser den avtakende avkastningen av arbeidserfaring. Ved økt arbeidserfaring med ett år avtar derfor effekten på lønnsavkastning med 0,027%.

Modellen utvides med variabelen male, slik at estimatet inkluderer menn og kvinner.

Modellen har samme notasjon men vi legger til leddet D_m , som er male.

$$\ln W = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2 + \delta_0 D_m$$

Tabell 9. Regresjonsresultat 2 modeller

Variabler	(1) ln lønn	(2) ln lønn
edu	0,0521*** (-0,00292)	0,0554*** (0,00277)
exper	0,0177*** (0,00305)	0,0190*** (0,00288)
exper_sq	-0,00027 (-6,56E-05)	-0,000312*** -6,21E-05
male		0,182*** (0,0118)
Konstant	4,418*** (0,0585)	4,260*** (0,0563)
Observasjoner	2,005	2,005
R-squared	0,147	0,237

T verdi i parentes:

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Vi sammenligner de to Mincer-modellene. Avkastningen på utdanning og arbeidserfaring er forholdsvis ganske uendret, men vi ser en liten økning i begge variablene. Avkastning på utdanning er 5,5% og avkastning på arbeidserfaring er 1,9%. Den kvadrerte arbeidserfaringen viser 0,031%, som vil si at arbeidserfaringens påvirkning vil avta mer enn i den forrige modellen. Variabelen male er positiv som vil tilsi at menn har en høyere startlønn enn kvinner.

Utvider modellen videre og legger til variabelen SD_m , som er en interaksjonsvariabel for å kunne sammenligne avkastning av utdanning mellom kvinner og menn.

$$\ln W = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2 + \delta_0 D_m + \delta_1 SD_m$$

Tabell 10. Regresjonsresultat 3 modeller

Variabler	(1) ln lønn	(2) ln lønn	(3) ln lønn
edu	0,0521*** (-0,00292)	0,0554*** (0,00277)	0,0552*** (0,00438)
exper	0,0177*** (0,00305)	0,0190*** (0,00288)	0,0190*** (0,00289)
exper_sq	-0,00027 (-6,56E-05)	-0,000312*** (6,21E-05)	-0,000311 (6,22e-05)
male		0,182*** (0,0118)	0,179** (0,0852)
Interaksjon			0,000239 (0,00552)
Konstant	4,418*** (0,0585)	4,260*** (0,0563)	4,262*** (0,078)
Observasjoner	2,005	2,005	2,005
R-squared	0,147	0,237	0,237

t verdier i parentes:

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Modellen er ferdig estimert og det er lagt til et interaksjonsledd for å skille avkastning på utdanning mellom kvinner og menn. Vi kan se av tabellen at kvinner har en avkastning på 5,52% og menn har en avkastning på 5,54%. Vi bemerker oss derfor at det ikke er en signifikant forskjell på avkastning på utdanning mellom menn og kvinner. Vi kan også se at den endelige estimerte modellen viser en litt høyere avkastning på arbeidserfaring, men nokså lik. Av R^2 ser vi en liten økning, som vil tilsi at vi har en bedre lineær sammenheng mellom de eksogene og endogene variablene.

5.2 OFFENTLIG- OG PRIVAT SEKTOR

Hypotese: Avkastning på utdanning er større om du jobber i privat sektor.

Tabell 11. Regresjonsresultat offentlig- og privat sektor

reg lnwage edu pub edu_pub

Source	SS	df	MS	Number of obs	=	2,005
Model	31.1179739	3	10.372658	F(3, 2001)	=	139.93
Residual	148.324969	2,001	.074125422	Prob > F	=	0.0000
Total	179.442942	2,004	.089542386	R-squared	=	0.1734
				Adj R-squared	=	0.1722
				Root MSE	=	.27226

lnwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
edu	.0587628	.0034963	16.81	0.000	.0519061	.0656196
pub	.0398027	.0962512	0.41	0.679	-.1489604	.2285658
edu_pub	-.0124153	.0061388	-2.02	0.043	-.0244543	-.0003762
_cons	4.612165	.0522156	88.33	0.000	4.509763	4.714568

Vi ser på betakoeffisienten for privat sektor som er 0,0587628. Dette forteller oss at lønnen øker med 5,9% pr år utdanning man tar om man jobber i privat sektor. For å sjekke om betakoeffisienten er større enn null setter vi opp en nullhypotese, om den er lik null. Og en alternativ hypotese om at beta er større enn null.

$$H_0 : \beta_P = 0$$

$$H_A : \beta_P > 0$$

Resultatet viser om avkastningen er positiv i privat sektor. Vi ser av regresjonsanalysen at vi har t-verdien 16,81. Siden tallet er så stort velger vi å forkaste nullhypotesen som vil tilsi at vi har en signifikant endring i timelønn dersom utdanning øker med ett år i privat sektor. I tillegg ser vi at $P > |t| = 0,000$ som forsterker beslutningen om å forkaste nullhypotesen. Vi bemerker oss at hypotesen kan forkastes med 5% signifikansnivå og 1% signifikansnivå. Derfor kan vi si med sikkerhet at det er en positiv økning i lønnsavkastning ved ett år ekstra utdanning i privat sektor.

For å finne betakoeffisienten til et individ i offentlig sektor ser vi på variabelen `edu_pub` og legger til betaen for privat sektor. Får da $0,0587628 + (-0,0124153) = 0,0463475$. Svaret viser oss at en person i offentlig sektor får 4,64% høyere lønn som følge av ett år ekstra utdanning.

Videre skal vi se om det er en forskjell i avkastning på utdanning mellom offentlig- og privat sektor. Nullhypotesen forteller oss at de to betakoeffisienten er like, mens alternativhypotesen forteller oss at betakoeffisienten er høyere for de som jobber i privat sektor.

$$H_0 : \beta_P = \beta_O$$

$$H_A : \beta_P < \beta_O$$

T-verdien i regresjonsanalysen for `edu_pub` er -2,02. Den representerer forskjellen på offentlig og privat sektor. Ser vi i tabellen $p > |t|$ viser den 4,3% sannsynlig at vi beholder nullhypotesen. Med 5% signifikansnivå så kan vi forkaste nullhypotesen. Derav kan vi si at betakoeffisienten til en person som jobber i privat sektor er større enn for en person som jobber i offentlig sektor. R^2 viser en varians på 17,2%. Det vil si at 17,2% av variansen til \ln wage forklares av de eksogene variablene.

6 KONKLUSJON

I denne oppgaven har jeg sett på forskjeller ved lønn og avkastning på utdanning mellom kvinner og menn, samt offentlig- og privat sektor i Norge. Analysen er gjort ved bruk av datasettet PIAAC, som er en internasjonal undersøkelse. I oppgaven kommer det frem at lønnsavkastningen er positiv med antall år utdanning. Videre analyserte vi forskjeller mellom menn og kvinner, samt offentlig- og privat sektor. Vi skal videre i kapittelet oppsummere resultatene som er funnet.

Av analysen mellom menn og kvinner ser vi at gjennomsnittlig har menn en høyere startlønn, og høyere timelønn. Samtidig kan vi se at menn har større ulikhet i inntekten. Menn og kvinner har i gjennomsnitt nokså likt antall år utdanning, og den gjennomsnittlige avkastningen på utdanning ligger på 5,2%. Analysen viser at menn og kvinner har relativt lik avkastning på utdanning, med forholdsvis 5,52% og 5,54%. Vi kan derfor i tråd med analysen konkludere med at hypotesen stemmer.

Av analysen mellom offentlig- og privat sektor finner vi en høyere gjennomsnittslønn i privat sektor, men også større spredning i inntekten. Analysen viser også at privat sektor har en lønnsavkastning på 5,9% per år utdanning, mens offentlig sektor har en lønnsavkastning på 4,63% per år utdanning. Ved utførelse av hypotesetest bekrefter vi at privat sektor har en høyere lønnsavkastning. Vi kan derfor anta at et år ekstra utdanning gir en høyere lønnsavkastning i privat sektor. Resultatet er noe overraskende da offentlig sektor består av mange forskjellige yrker som krever høy utdanning. Vi har blant annet leger, og andre yrker som gir høy lønn, men som samtidig krever lang utdanning. Siden samfunnet vil nytte av at disse har høy kompetanse vil det være positivt å gi innbyggerne et insentiv til å ta lang utdanning. I lys av oppsummeringen kan vi konkludere med at hypotesen ikke stemmer.

LITTERATURLISTE

Bjørkeng, B (2013). «*Ferdigheter i voksenbefolkningen*». Hentet fra:
<https://www.ssb.no/utdanning/artikler-og-publikasjoner/attachment/141211?ts=1416e80e8e0>

Fleischauer, Kai-Joseph (2007). «*A review of human capital theory: Microeconomics*». Hentet fra: <https://core.ac.uk/download/pdf/6710654.pdf>

Hægland, T (2003). «*Økonomisk avkastning av utdanning*». Hentet fra:
<https://www.ssb.no/a/publikasjoner/pdf/sa60/kap-10.pdf>

Idsø, J. (2017). *SNL*. Hentet fra:
https://snl.no/privat_sektor

Organisation for Economic Cooperation and Development: Survey of Adult Skills (PIAAC)
<http://www.oecd.org/skills/piaac/>

Løvås. (2013). «*Statistikk – for universiteter og høyskoler*», 3. utgave. Oslo: universitetsforlaget.

SSB (2018). Befolkningens utdanningsnivå. Hentet fra:
<https://www.ssb.no/utniv/>

Thomas, R. L. (2005). «*Using statistics in economics*». UK: McGraw-Hill Education.