



Universitetet  
i Stavanger

HANDELSHØGSKOLEN VED UIS

# BACHELOROPPGAVE

STUDIUM:

Økonomi og administrasjon

OPPGAVEN ER SKREVET INNEN FØLGENDE  
TEMATISKE RETNING:

Samfunnsøkonomi

ER OPPGAVEN KONFIDENSIELL?

(NB! Bruk rødt skjema ved konfidensiell oppgave)

TITTEL:

I hvilken grad kan forskjeller i lønn forklares av variablene kjønn, alder og utdanning?

ENGELSK TITTEL:

To what extent can differences in salary be explained by variables such as gender, age, and education?

FORFATTERE (NB! maks tre studenter pr oppgave):

VEILEDER:

Kandidatnr:

2190

.....

2223

.....

2252

.....

Navn:

Trude Lønseth

.....

Lea Zaviera Cortes Andreassen

.....

Mona Udem

.....

Sara Helene Rønningstad

# Innholdsfortegnelse

<b>Forord</b> .....	<b>4</b>
<b>Sammendrag</b> .....	<b>5</b>
<b>Abstract</b> .....	<b>6</b>
<b>1. Innledning</b> .....	<b>7</b>
<i>Oppgavens oppbygning</i> .....	10
<b>2. Teori</b> .....	<b>10</b>
2.1 Tidligere forskning .....	10
2.2 Begrunnelse for valg av variabler .....	15
2.2.1 Kjønn .....	15
2.2.2 Alder .....	16
2.2.3 Utdanning .....	16
2.2.4 Andre forklaringsfaktorer til lønn .....	17
<b>3. Metode</b> .....	<b>18</b>
3.1 Beskrivende statistikk .....	18
3.2 Kvantitativ metode .....	19
3.3 Økonometri .....	20
3.3.1 Enkel regresjon .....	20
3.3.2 Multippel regresjon .....	21
3.3.3 R <sup>2</sup> – Forklaringskraft .....	22
3.3.4 Hypotesetesting .....	23
<b>4. Datamateriale</b> .....	<b>25</b>
4.1 Definisjon av type variabler .....	26
4.2 Beskrivelse av datasett med avgrensning .....	27
<b>5. Beskrivende statistikk</b> .....	<b>31</b>
5.1 Kjønn .....	31
5.2 Alder .....	37
5.3 Utdanning .....	39

<b>6. Regresjonsanalyse med drøfting .....</b>	<b>42</b>
<i>6.1 Forutsetninger ved Regresjonsanalyse .....</i>	<i>43</i>
6.1.1 Lineær sammenheng .....	43
6.1.2 Uavhengige observasjoner.....	43
6.1.3 Normalfordeling.....	44
6.1.4 Multikollinearitet .....	46
6.2 <i>Hvor stor innvirkning har kjønn på lønn .....</i>	<i>47</i>
6.3 <i>Hvor stor innvirkning har alder på lønn .....</i>	<i>49</i>
6.4 <i>Hvor stor innvirkning utdanning på lønn .....</i>	<i>51</i>
6.5 <i>Multipel regresjon med alle variablene .....</i>	<i>52</i>
<b>7. Avslutning/konklusjon .....</b>	<b>55</b>
<b>Litteraturliste.....</b>	<b>59</b>
<b>Figurliste .....</b>	<b>63</b>
<b>Tabelliste .....</b>	<b>63</b>
<b>Vedlegg.....</b>	<b>64</b>

## Forord

Denne bacheloroppgaven representerer en avslutning på vår utdanning innen Økonomi og Administrasjon på Handelshøyskolen ved Universitetet i Stavanger, og vi håper den vil bidra til å utvide kunnskapen, og nysgjerrigheten, innenfor forhold som påvirker lønn.

Vi vil først og fremst rette en stor takk til vår veileder, Sara Helene Rønningstad for all støtte, veiledning og tålmodighet gjennom prosjektet. Vi ønsker også å rette en takk mot våre forelesere ved Universitetet i Stavanger, som har gitt oss en solid akademisk bakgrunn og inspirert oss til å utforske dette emnet nærmere.

Vi ønsker også å takke Jonathan Muringani for bidrag, støtte og gode diskusjoner rundt temaet, noe som har bidratt til inspirasjon og skriveglede.

Vi ønsker også å takke familie og partnere som har støttet oss gjennom hele prosjektet. Deres oppmuntring, tålmodighet og støtte har vært uvurderlig og uten dette hadde ikke oppgaven vært mulig å gjennomføre.

Til slutt vil vi takke alle som har bidratt til denne oppgaven, både direkte og indirekte. Takk til medstudenter som har bidratt til en hyggelig studenttilværelse.

Vi håper at denne oppgaven vil bidra til videre forskning innen forhold som påvirker lønn, og at den kan inspirere andre til å utforske feltet videre.

## Sammendrag

Formålet med denne bacheloroppgaven er å undersøke lønn og dens mulige påvirkningsfaktorer på individnivå, som er relevant da lønn spiller en viktig rolle i arbeids- og karrierevalg. Relevant litteratur og forskning på området har blitt gjennomgått og data er innhentet fra Statistisk Sentralbyrå.

En multippel regresjonsanalyse ble utført på tre utvalgte faktorer- kjønn, alder og utdanning- og resultatene viser at kjønn har en signifikant positiv effekt på lønn, men at kjønn består av flere underliggende forhold. Når det gjelder alder så viser regresjonen at det er en økning i lønn i takt med alder, dog viser datamateriale at det er en økning fram til omtrent 50 års alderen, men etter dette skjer det en reduksjon fram til 67 år. Videre viser regresjonsanalysen at utdanning også har en signifikant positiv effekt på lønn. Resultatene viser også at effekten av variablene øker når variablene blir satt sammen i en multippel regresjonsanalyse, noe som tyder på at det er samspill mellom variablene. Analysen gir en dypere forståelse av hvordan ulike faktorer kan påvirke lønnsnivået til en person, og hvilke faktorer som er mest relevante i ulike kontekster.

## Abstract

The purpose of this bachelor's thesis is to examine salary and its potential influencing factors at an individual level, which is relevant as salary plays an important role in career and job choices. Relevant literature and research in the field have been reviewed, and data has been collected from Statistisk Sentralbyrå.

A multiple regression analysis was performed on three selected factors - gender, age, and education - and the results show that gender has a significant positive effect on salary, but that gender is composed of several underlying factors. Regarding age, the regression shows that there is an increase in salary as age increases, although the data shows an increase up to around the age of 50, but after that, there is a decrease until the age of 67. Furthermore, the regression analysis shows that education also has a significant positive effect on salary. The results also show that the effect of the variables increases when the variables are combined in a multiple regression analysis, indicating that there is interaction between the variables. The analysis provides a deeper understanding of how various factors can affect an individual's salary level, and which factors are most relevant in different contexts.

## 1. Innledning

Formålet med denne oppgaven har vært å gjennomføre en grundig analyse og kritisk belyse faktorene som påvirker lønnsstrukturen. Ambisjonen er å bidra til en dypere forståelse av de underliggende mekanismene i arbeidsmarkedet. Disse faktorene kan operere på ulike nivåer, enten det er på samfunnsnivå eller individnivå.

På samfunnsnivå kan lønnsnivået påvirkes av økonomiske forhold som inflasjon, tilbud og etterspørsel etter arbeidskraft, samt politiske beslutninger knyttet til regulering av lønnsnivåer og skattesystemet. Endringer i lønnsnivået kan ha konsekvenser for økonomisk vekst og forbruk, og kan også påvirke arbeidsledighetsraten og økonomisk stabilitet.

På individnivå påvirkes lønnen av faktorer som utdanningsnivå, erfaring, kompetanse, prestasjoner og forhandlingsstyrke. Disse faktorene kan ha innvirkning på en arbeidstakers evne til å forhandle om høyere lønn, få forfremmelser og øke inntekten over tid. I denne oppgaven er hovedfokuset rettet mot de individuelle faktorene som påvirker lønnsnivået, da en grundig forståelse av disse faktorene også vil bidra til en dypere forståelse av de mekanismene som er involvert i fordelingen av økonomiske ressurser i arbeidsforhold på samfunnsnivå.

En anstendig lønn øker vanligvis økonomisk sikkerhet og gir muligheter for å anskaffe seg nødvendige varer og tjenester. Det påvirker levestandarden og kan ha direkte innvirkning på økonomisk vekst. Å undersøke individuelle egenskaper som påvirker lønnsnivået er av betydning for å forstå dens innvirkning på en persons livskvalitet, økonomiske trygghet, levestandard og sosiale mobilitet. Dette er viktig både for samfunnsøkonomi og sosialvitenskap.

Til tross for at det eksisterer betydelig kunnskap om hvordan ulike faktorer påvirker lønnsnivået, gjenstår det fremdeles mye å utforske når det gjelder samspillet mellom disse faktorene. Med det formål å oppnå en mer innsiktsfull forståelse av forholdene som påvirker lønn, er problemstillingen valgt for denne oppgaven:

## ***"I hvilken grad kan forskjeller i lønn forklares av variablene kjønn, alder og utdanning?"***

For å søke å besvare denne problemstillingen vil det bli gjennomført en omfattende diskusjon rundt hvordan de tre enkelte variablene påvirker lønn.

### **Hvordan påvirker kjønn lønn?**

Å utforske i hvilken grad kjønn påvirker lønnsnivået i Norge, som anses som et av verdens mest likestilte land, representerer et interessant tema. Dette skyldes i stor grad landets historiske beslutninger og tiltak for å sikre like muligheter og rettigheter for både kvinner og menn. Norge var blant annet et av de første landene som innførte stemmerett for kvinner på begynnelsen av 1900-tallet (Bech, 2013). I tillegg fikk kvinner i Norge rett til høyere utdanning på slutten av 1800-tallet (Steinkellner, 2010), og i dag utgjør kvinner flertallet av studentene i høyere utdanning, med 60% kvinnelige studenter i 2021 (Kifinfo, u.å.). Kvinner utgjør også en betydelig andel av fullførte utdanninger på universiteter og høyskoler, der 58,9% fullfører sin utdanning (Statistisk Sentralbyrå, 2023). En rekke lover og tiltak er iverksatt for å sikre likebehandling uavhengig av kjønn, deriblant Likestillingsloven som ble vedtatt i 1978 som fastslo at forskjellsbehandling var forbudt, og at det skulle være lik lønn for likt arbeid (St.meld. nr. 44 (2012-2013)). Videre pålegger FNs bærekraftsmål nr. 5, som fokuserer på likestilling og styrking av kvinners og jenters stilling, bedrifter og organisasjoner aktivt arbeide for å fremme likestilling (FN, 2023). Basert på de betydelige fremskrittene, lovene og tiltakene som er implementert for å fremme likestilling i Norge, kan en hypotese formuleres om at kjønn ikke vil ha en signifikant innvirkning på nivået av lønn.

### **Hvordan påvirker alder lønn?**

Alder kan påvirke en persons lønn på ulike måter, og det er flere faktorer som spiller inn, som arbeidserfaring, bransje, jobbansiennitet og konkurranse i arbeidsmarkedet. Generelt sett viser statistikk fra SSB at lønnen øker med alderen, der eldre arbeidstakere vanligvis har høyere lønn (Askvik, 2020).

Til tross for dette, viser undersøkelser at flertallet av norske ledere foretrekker å ansette en 20-åring fremfor en 60-åring. Det antas at eldre arbeidstakere kan møte utfordringer knyttet til helse, sykefravær, digitale ferdigheter, innovative evner og arbeidsprestasjoner,



noe som kan påvirke deres muligheter på arbeidsmarkedet (Hauge & Moen, 2023). Dette er i strid med anti-diskrimineringsloven, som forbyr direkte og indirekte diskriminering basert på blant annet alder (Arbeidsmiljøloven, 2022, §13-1 ). Alder kan muligens påvirke både tilgangen til arbeidsplasser og lønnsnivået til en arbeidstaker, til tross for eksisterende lovverk som tar sikte på å forhindre aldersdiskriminering. Dette indikerer behovet for videre studier for å undersøke de positive og negative aspektene av alder, som nevnt ovenfor, for å vurdere om de har en tendens til å påvirke lønnsnivået i en positiv eller negativ retning.

### **Hvordan påvirker utdanning lønn?**

Utdanningens økende betydning i samfunnet blir påpekt i en artikkel fra SSB som presenterer prognoser for arbeidsstyrken og sysselsettingen etter utdanningsnivå frem til 2040. Ifølge artikkelen vil etterspørselen etter arbeidstakere med høyere utdanning øke fra 36% i 2019 til 44% i 2040, mens behovet for arbeidstakere med videregående opplæring derimot vil synke fra 42% til 36% i samme periode (Statistisk Sentralbyrå, 2020). Derimot vil andelen arbeidstakere med grunnskole eller lavere utdanning være relativt stabil. Dette vil føre til en forventet økning i gjennomsnittlig utdanningsnivå blant arbeidsstyrken i fremtiden, og impliserer behovet for at arbeidsmarkedet tilrettelegger for denne utviklingen. På bakgrunn av de fremsatte prognosene angående arbeidsstyrken og sysselsettingen etter utdanningsnivå, kan det antas at utdanningsnivået vil få en økende betydning for lønnsnivået i fremtiden. Dette gir grunnlag for en nærmere undersøkelse av utdanning som en variabel i analysen.

Oppgaven vil først foreta en systematisk undersøkelse av hvordan disse variablene individuelt påvirker lønnsnivået til arbeidstakere, samt om det eksisterer signifikante forskjeller i lønn mellom ulike grupper av arbeidstakere. Dette vil bli gjort gjennom en grundig analyse av eksisterende data som omhandler lønn og arbeidstakere i Norge. Dataene som er benyttet er samlet inn fra eksterne kilder, og analysen vil inneholde både deskriptiv statistikk og mer avanserte analyseverktøy som multippel regresjonsanalyse. For å oppnå dette, har vi anvendt kvantitative metoder, herunder tall, grafer og illustrasjoner, for å nøye analysere og presentere resultatene som avdekker i hvilken grad kjønn, alder og utdanningsnivå påvirker lønnsnivået.

## Oppgavens oppbygning

Kapittel 2 omhandler en litteraturgjennomgang av tidligere forskning på faktorer som påvirker kjønnsforskjeller mellom kjønn, avkastning på utdanning og teorien om humankapital. Videre velges tre forklaringsfaktorer og andre mulige faktorer som påvirker lønn. Metodene som brukes i analysen, beskrivende statistikk, kvantitativ metode og økonometri, blir presentert i kapittel 3. I kapittel 4 gis en presentasjon av datamaterialet og avgrensningene som er gjort, samt definisjoner av variablene som er brukt i regresjonsanalysen. Kapittel 5 inneholder beskrivende statistikk som presenterer data på en systematisk og oversiktlig måte. Her blir også grafer og illustrasjoner presentert. I kapittel 6 blir først forutsetninger ved en regresjonsanalyse presentert, deretter kommer en presentasjon og drøfting av resultatene av regresjonsanalysen på påvirkningen av kjønn, alder og utdanning først separat, etterfulgt av en samkjøring av variablene i en multipel regresjon. Oppgaven avsluttes i kapittel 7 med en oppsummering av funnene i oppgaven.

## 2. Teori

Dette kapitlet har som formål å presentere flere teoretiske perspektiver og konsepter som kan bidra til en dypere forståelse av de observerte lønnsforskjellene basert på variabler som kjønn, alder og utdanningsnivå. Relevant tidligere forskning og noe teori vil bli gjennomgått. Videre identifiserer kapitlet muligens motstridende funn eller utfordringer knyttet til dette emnet. Det presenteres også de valgte variablene, med begrunnelse for deres inkludering. I tillegg blir andre variabler nevnt, som antas å påvirke lønn, men som ikke blir inkludert i regresjonsanalysen. Disse variablene kan være relevante for fremtidig forskning da de kan bidra til å forbedre validiteten og reliabiliteten av analysen når det gjelder å vurdere de ulike faktorene som kan påvirke lønn.

### 2.1 Tidligere forskning

Gjennom en analyse av eksisterende teoretisk rammeverk for å forstå sammenhengen mellom variablene, kan det bidra til en økt forståelse av mekanismene bak lønnsforskjeller og identifisere eventuelle kunnskapshull eller kontroverser i litteraturen. Dette kan også danne et solid grunnlag for utforming av egen empirisk forskning eller

analyse for å utforske dette temaet i denne bacheloroppgaven. Ved å basere analysen på tidligere teoretiske perspektiver og forskning, kan det bygges videre på eksisterende kunnskap og bidra til å fylle eventuelle gap eller uklarheter i litteraturen knyttet til sammenhengen mellom kjønn, alder, utdanning og lønn.

Studien “Children and Gender Inequality: Evidence from Denmark” ble publisert i *American Economic Journal* i 2019 (Kleven et al., 2019). Det baserer seg på administrative data fra Danmark og viser at barnefødsler spiller en viktig rolle for vedvarende kjønnsulikheter i arbeidsmarkedet til tross for økt utdanningsnivå blant kvinner og antidiskrimineringspolitikk. Resultater fra denne studien viser at kvinner begynner å henge etter karrieremessig kort tid etter at de blir foreldre, og det skapes en langsiktig lønnsforskjell på rundt 20% mellom kvinner og menn. Kvinner vil også ha en betydelig lavere sannsynlighet for å stige til leder-nivå etter å ha fått barn. Studien viser at kvinner som har barn er mer sannsynlig å jobbe i den offentlige sektoren, som tilbyr fleksible arbeidstimer, permisjonsdager når man har syke barn, og en gunstig holdning til foreldrepermisjon. Kvinner er derfor 12% mer sannsynlig enn menn til å jobbe i den offentlige sektoren 10 år etter fødselen. (Kleven et al., 2019). Disse funnene viser at til tross for antidiskrimineringspolitikk og utallige tiltak som er gjort for å fremme likestilling, kan kjønn ha en betydelig innvirkning på lønnen til en person og deres karrieremuligheter. Årsakene til lønnsforskjellene mellom menn og kvinner er komplekse og kan omfatte en rekke faktorer som nevnes i “Children and Gender Inequality”, inkludert fordeling av omsorgsarbeidet i hjemmet, kjønnsbasert diskriminering, ulikhet i karrieremuligheter, utdanningsvalg og yrkesvalg.

Til tross for at kvinner i dag har oppnådd høyere nivå av utdanning og yrkesdeltakelse sammenlignet med tidligere, fremkommer det fremdeles en kjønnsdisparitet i sektortilhørighet i henhold til statistikk fra Statistisk sentralbyrå (SSB) (Lekve, 2023). Kvinnelige ansatte er overrepresentert i offentlig sektor, mens menn dominerer i privat sektor. På den ene siden er andelen kvinner i offentlig sektor anslått til å være 70 prosent, mens menn utgjør kun 30 prosent. På den annen side er kvinner i mindretall i privat sektor, med en andel på 37 prosent, mens menn utgjør 63 prosent (Lekve, 2023). Det er også viktig å merke seg at lønningene generelt sett er lavere i offentlig sektor sammenlignet med privat sektor.

I tillegg viser data fra SSB at flere kvinner enn menn arbeider deltid, der 32 prosent av kvinnelige sysselsatte i alderen 20-66 år jobber deltid, sammenlignet med 15 prosent blant menn. Selv om det har vært en økning i andelen menn som jobber deltid de siste ti årene, er dette fortsatt en sjeldenhet blant mannlige arbeidstakere (Lekve, 2023). Disse faktorene kan bidra til å øke lønnsforskjellen mellom kjønnene og ha en negativ innvirkning på samfunnets økonomi. Dersom andelen kvinner i høyt betalte yrker øker kan dette føre til økt mangfold og innovasjon, noe man ellers ville gått glipp av.

En potensiell tolkning av SSBs data som viser at flere kvinner arbeider deltid enn menn, kan relateres til en ulik fordeling av omsorgsoppgaver og ansvar i hjemmet (Lekve, 2023). Dette fenomenet kan ha betydelig innvirkning på kvinners deltakelse i arbeidslivet, lønnsnivå og karriereutvikling. Det er mulig å tenke seg at kvinner ofte tar på seg hovedansvaret for omsorgsoppgaver og husarbeid, noe som kan føre til redusert arbeidsdeltakelse og begrensede muligheter for karriereutvikling. Dette kan videre bidra til opprettholdelsen av lønnsforskjeller mellom kjønnene og hindre kvinner i å realisere sitt fulle potensial på arbeidsplassen. I tråd med forskningen "Children and Gender Inequality: Evidence from Denmark", har det blitt observert en reduksjon i kvinners karriereutvikling og lønn i perioden rundt fødselen av barn, og denne effekten kan ha varig innvirkning (Kleven et al., 2019). Videre viser studien også at kvinner kan ha kortere arbeidserfaring på grunn av fravær fra arbeidsplassen for å ta seg av familieforpliktelser.

De faktorene som er nevnt ovenfor kan gi oss et mer nyansert bilde av problematikken rundt lønnsgapet mellom menn og kvinner, og indikerer at det finnes flere årsaker til dette gapet som går utover ren kjønnsbasert diskriminering eller stereotyper knyttet til kjønnsroller. En underrepresentasjon av kvinner blant de som har de aller høyeste lønningene bidrar til lønnsgapet (Lekve, 2023). Videre er en overrepresentasjon av kvinner i offentlig sektor, der gjennomsnittlig lønnsnivå er lavere sammenlignet med privat sektor, også en faktor som påvirker lønnsgapet. I tillegg arbeider flere kvinner i deltidsstillinger, og deltidsjobber har en tendens til å ha lavere lønnsnivå sammenlignet med heltidsjobber. Til slutt er det vanlig at kvinner tar på seg hovedansvaret for omsorgsoppgaver og husarbeid i hjemmet, noe som også kan bidra til lønnsgapet mellom kjønnene (Lekve, 2023).

Det kan postuleres at ulike valg- og atferdsmønstre mellom kjønnene kan ha en innvirkning på deres lønnsnivå. Studien "The College Payoff" som presenteres under, understreker at kjønn kan være en faktor som er mer avgjørende enn utdanning i å påvirke lønnsnivået (Carnevale et al., 2011), og det er derfor interessant å inkludere kjønn som en variabel som kan bli ytterligere undersøkt.

Humankapitalteorien har vært en betydelig teori innenfor forskning på lønnsforskjeller, og den amerikanske professoren innen økonomi og sosiologi, Gary S. Becker, har vært en pioner innenfor dette feltet. Innenfor økonomisk teori anses utgifter til utdanning, opplæring og lignende som investeringer i menneskelig kapital. Menneskelig kapital referer til kunnskap og ferdigheter som individer besitter, og som ikke kan skilles fra dem slik som finansielle eller fysiske eiendeler. Ifølge humankapitalteorien er utdanning, opplæring og helse blant de viktigste investeringene i menneskelig kapital. Studier i USA har vist at både videregående og høyere utdanning i stor grad bidrar til økt inntekt. Teorien antyder at årsaken til lønnsforskjeller ligger i arbeidstakernes individuelle egenskaper, og at en persons utdanning, erfaring og opplæring øker deres produktivitet og dermed deres potensial for økt inntekt (Becker, u.å.).

Ifølge den amerikanske rapporten "The College Payoff" fra Georgetown University Center on Education and the Workforce, er avkastningen på utdanning betydelig høyere enn avkastningen på andre investeringer, som for eksempel obligasjoner og aksjer (Carnevale et al., 2011). Rapporten fremhever at utdanning er en avgjørende faktor, og generelt vil personer med høyere utdanning tjene mer enn de med lavere utdanning. Likevel kan valg av yrke overstyre utdanningsnivået, spesielt i høytlønnede yrker, der en person med lavere utdanning kan tjene mer enn en annen person med høyere utdanning i et lavtlønnet yrke. På tross av at yrkesvalg kan overgå utdanning, vil utdanning i de fleste tilfeller være en avgjørende faktor innenfor yrkesgruppene. Rapporten peker også på at etnisitet og kjønn er mer avgjørende enn utdanning og yrke når det gjelder inntektsnivå. På lang sikt kan inntektsforskjellene mellom personer med og uten høyere utdanning være betydelige, og en person med en bachelorgrad kan forvente å tjene omtrent 1 million dollar mer i løpet av livet sammenlignet med en person med kun videregående skoleutdanning. Det er imidlertid viktig å merke seg at avkastningen på utdanning kan variere avhengig av faktorer som fagfelt, karrierevei og individuelle egenskaper og evner (Carnevale et al., 2011).

Utdanning kan gi arbeidstakere nødvendige kvalifikasjoner for spesialiserte jobber slik at man kan anta at høyere utdanningsnivå gir høyere mulighet for lønn. Dette kan også være gjeldende for spesialisert utdanning innen et fagområde, da spesifikke ferdigheter og kunnskap kan være ettertraktet av arbeidsgivere. Arbeidsgivere bruker ofte utdanning som vurderingsgrunnlag for kvalifikasjoner og erfaring, og høyere utdanning kan dermed gjøre en person mer konkurransedyktig i arbeidsmarkedet. Dette kan føre til muligheter til å forhandle om høyere lønn og åpne opp for karriereutvikling med stillinger som innehar større ansvar.

Alder kan påvirke en individs lønn gjennom komplekse mekanismer som kan ha både negative og positive utfall, med flere faktorer som bidrar til å forme denne sammenhengen. Fra et positivt perspektiv viser statistikk fra SSB at det generelt sett er en økning i lønn i takt med alderen, noe som kan reflektere økt arbeidserfaring, jobbansiennitet og konkurransesituasjonen i arbeidsmarkedet (Askvik, 2020). Det kan også antas at lengre ansettelsestid i en organisasjon kan bidra til en større kjennskap til organisasjonen og økt jobberfaring, noe som kan være en fordel i forhandlinger om lønn. Økt erfaring og jobbansiennitet kan fungere som sterke argumenter for å oppnå en høyere lønn. Videre kan eldre arbeidstakere i bransjer der erfaring og kunnskap er høyt verdsatt, også oppnå en høyere lønn sammenlignet med yngre arbeidstakere. Dette kan skyldes oppfatningen om at eldre arbeidstakere har opparbeidet seg større ferdigheter og kunnskap innenfor sitt arbeidsområde. Imidlertid har arbeidsmarkedet også gjennomgått endringer som følge av digitalisering og globalisering, noe som har ført til økte krav til arbeidstakere når det gjelder evnen til å tilpasse seg raskt til endringer og anvende teknologi. En studie har indikert at majoriteten av norske ledere har en tendens til å foretrekke å ansette personer i 20-årene framfor personer i 60-årene, basert på en antakelse om at eldre arbeidstakere har begrensede digitale ferdigheter som kan ha en negativ innvirkning på deres evne til innovasjon og produktivitet (Hauge & Moen, 2023). Videre viser studien at disse lederne også forbinder eldre arbeidstakere med høye sykefravær knyttet til helse. Slike antakelser kan begrense mulighetene for eldre arbeidstakere til å realisere sitt fulle potensial når det gjelder karrieremuligheter eller forhandling av høyere lønninger. Denne antakelsen kan resultere i aldersdiskriminering, som kan utgjøre et hinder for eldre arbeidstakere med hensyn til å få tilgang til jobbmuligheter og forhandle om høyere lønn. En slik praksis er i strid med prinsippene nedfelt i anti-diskrimineringsloven, som har som mål å forby enhver form for diskriminering basert på blant annet alder (Arbeidsmiljøloven, 2022, §13-1). Det

er også av interesse å undersøke sammenhenger mellom alder og andre faktorer som kan påvirke lønn. Statistisk sentralbyrå har påpekt en tendens til mindre lønnsforskjeller basert på kjønn blant yngre arbeidstakere, mens lønnsforskjellene øker med alderen (Askvik, 2020). Dette indikerer en mulig sammenheng mellom alder og en annen forklaringsfaktor for lønn i analysen. Videre forskning er nødvendig for å forstå hvordan alder kan påvirke lønnsnivået, og om ulike aspekter av alder, som aldersdiskriminering, har en tendens til å påvirke lønnsnivået i en positiv eller negativ retning.

Teoriene og rapportene utgjør grunnlaget for forventede resultater samt mulige begrensninger som kan oppstå under utførelsen av analysen. Det forventes å finne en sammenheng mellom kjønn og lønn, men det erkjennes at resultatene i denne studien ikke vil gi en helhetlig forståelse av de komplekse faktorene som påvirker kjønnsvariabler. Videre forventes det også å finne en sammenheng mellom utdanning og lønn, men det vil ikke bli demonstrert hvordan variasjonen i avkastning på utdanning kan skyldes individuelle valg av karriere og personlige egenskaper i våre resultater. Til slutt kan det antas at det vil være en sammenheng mellom alder og lønn, men det tas ikke hensyn til eventuelle tidsperioder utenfor arbeidslivet, eksempelvis fødselspermisjon, arbeidsledighet og langtidssykdom.

## 2.2 Begrunnelse for valg av variabler

Lønn er et omfattende tema som har en stor innvirkning på individets økonomiske muligheter og samfunnets velstand. Kjønn, alder og utdanning er noen av faktorene som spiller en rolle i et individs økonomiske situasjon. Forskjeller i lønn mellom ulike grupper av arbeidstakere har vært et betydningsfullt tema innen samfunnspolitikken og arbeidsmarkedet, og forskere har gjennomført flere studier for å undersøke hvordan ulike faktorer påvirker lønnsnivået. Kjønn, alder og utdanning kan tenkes å påvirke lønn og det kan derfor være verdt å undersøke nærmere hvordan disse variablene påvirker lønn i denne oppgaven, og diskutere i hvilken grad de påvirker lønn.

### 2.2.1 Kjønn

Et av formålene med denne oppgaven er å undersøke og forsøke å besvare om det foreligger en sammenheng mellom kjønn og lønn. På bakgrunn av forskningen "Children and Gender Inequality" antas det at det eksisterer en slik sammenheng, ettersom det

konstateres at menn i gjennomsnitt har høyere lønninger enn kvinner (Kleven et al., 2019). Studien argumenterer for at disse lønnsforskjellene kan forklares med ulike valg og beslutninger som tas av individer basert på kjønn, og som påvirker deres karrieremuligheter og lønnsnivå. Disse valgene kan bidra til opprettholdelsen av lønnsforskjellen mellom menn og kvinner.

I et forsøk på å undersøke om tidligere forskning som indikerer en eksisterende lønnsforskjell mellom menn og kvinner er relevant i Norge, som er kjent for å være et av verdens mest likestilte land, blir kjønn inkludert som en variabel i analysen.

### 2.2.2 Alder

Gitt tidligere forskning som indikerer både positive og negative sammenhenger mellom alder og lønn, ble alder valgt som en forklaringsvariabel for lønn i denne analysen.

Formålet med å inkludere alder som en variabel i analysen er å undersøke om det foreligger en negativ eller positiv korrelasjon mellom alder og lønnsnivå, samt å undersøke samspillet mellom alder og de øvrige forklaringsvariablene i analysen.

Inkluderingen av alder som en variabel kan bidra til en grundigere forståelse av relasjonen mellom alder og lønnsnivå, samtidig som det kan gi økt innsikt i dynamikken bak lønnsforskjellen mellom menn og kvinner, med særlig fokus på eventuelle variasjoner knyttet til aldersforskjeller.

### 2.2.3 Utdanning

Utdanning er en av de faktorene som kan ha en betydelig innvirkning på lønnen til en person, da arbeidsgivere gjerne legger vekt på utdanningsnivået når de fastsetter lønn. Det er flere måter utdanningen kan påvirke lønn på. Ifølge humankapitalteorien er lønnsforskjellene grunnet i forskjeller i egenskapene til en person, noe som kan bedres ved å investere i utdanning (Becker, u.å ). «The college payoff» rapporten viser til akkurat dette, da utdanning vil øke gjennomsnittslønnen til en person (Carnevale et al., 2011). Utdanning kan derfor være en hensiktsmessig faktor å undersøke videre.

Ved å inkludere utdanningsnivå i vår analyse, ønsker vi å øke forståelsen av i hvilken grad høyere utdanning påvirker lønnsnivået. Det rapporteres av SSB at ansatte med mastergrad



eller doktorgrad har en gjennomsnittlig lønn som er 25 prosent høyere enn de med kun bachelorgrad (Bye, 2018). På bakgrunn av dette er det valgt å sette søkelys på høyere utdanning på master- og doktorgradsnivå.

#### 2.2.4 Andre forklaringsfaktorer til lønn

Denne oppgaven setter søkelys på tre variabler som antas å ha en betydelig innvirkning på lønnsnivået til enkeltpersoner, nemlig utdanning, kjønn og alder. Det kan være viktig å merke seg at det er flere andre faktorer som også kan påvirke lønnsnivået, men som ikke vil bli hensyntatt i regresjonsanalysen. Disse faktorene kan likevel være av interesse for videre forskning på emnet.

En slik faktor kan være arbeidserfaring, da det er en vanlig antagelse om at personer med lengre erfaring innen en jobb eller bransje vil ha høyere lønnsnivå enn de med mindre erfaring. En annen faktor kan være geografisk plassering av arbeidsplassen, da det er kjent at lønnsnivået kan variere betydelig mellom områder eller regioner. Andre faktorer som kan ha betydning for lønnsnivået inkluderer sektoren eller bransjen som en person jobber i, stillingsnivået og ansiennitet. I rapporten «The College Payoff» fremheves betydningen av yrkesvalg dets innvirkning på individuell lønn (Carnevale et al., 2011). Valg av yrke og sektor som en person utdanner seg innen, anses derfor som en viktig faktor med hensyn til fremtidig personlig lønnsutvikling. I tillegg til kjønn så vektlegger også rapporten etnisitet. Ettersom denne rapporten er fra USA, kan det dog diskuteres hvor overførbart en vektlegging av etnisitet er til norske forhold. Tilgang til arbeid kan også påvirke lønn gjennom tilbud- og etterspørselsprinsippet i arbeidsmarkedet. Personer som har bedre tilgang til jobber med høy etterspørsel vil vanligvis ha høyere lønn, mens personer med mindre tilgang kan være mer villige til å godta en lavere lønn. Sett at en person innehar en spisskompetanse som er ettertraktet på arbeidsmarkedet så kan dette også ha en innvirkning på lønn. Er det mangel på disse arbeidstakerne er det sannsynlig å tro at lønnen deres vil øke da arbeidsgiverne gjerne må tilby høyere lønn for å tiltrekke seg og beholde arbeidstakere med den nødvendige kompetansen.

Alle disse faktorene kan også ha en samspillseffekt med de tre variablene som er inkludert i denne analysen og kunne derfor vært av interesse å inkludere i en videre og mer omfattende forskning. Samtidig kan det være viktig å merke seg at noen av disse faktorenes påvirkning ikke nødvendigvis er kvantifiserbart og kan da gjerne ikke

inkluderes i en regresjonsmodell selv om de kan ha en påvirkning på lønnsnivået. I tillegg kan det være begrensinger med tanke på tilgjengelighet av pålitelig data som kunne påvirket muligheten for å inkludere disse faktorene i regresjonsanalysen. Man kan dermed komme frem til at det kan være viktig å være oppmerksom på andre faktorer som kan påvirke lønnsnivået, selv om de ikke blir hensyntatt i regresjonsanalysen.

### 3. Metode

Statistikk som disiplin har som grunnprinsipp å anta at det eksisterer et problem eller et spørsmål som krever en løsning eller et svar. For å oppnå en best mulig besvarelse av spørsmålet, er det essensielt å inneha betydelig kunnskap om den aktuelle virkeligheten, og denne nødvendige kunnskapen opparbeides ved å samle inn og analysere data (Løvås, 2018). I dette kapittelet skal det gjennomgås de ulike metodene som anvendes i analysen for å oppnå en best mulig besvarelse av den aktuelle problemstillingen. Beskrivende statistikk, kvantitativ metode, økonometri, enkel og multippel regresjon, forklaringskraft og hypotesetesting er metoder som brukes for å analysere og forstå data i samfunnsvitenskap og økonomi. Ved hjelp av disse metodene har man mulighet til å beskrive sammenhenger mellom variabler, teste hypoteser, trekke konklusjoner og gi anbefalinger basert på dataene.

#### 3.1 Beskrivende statistikk

Beskrivende statistikk er en analysemetode som har til formål å gi en forståelse av utvalgte data gjennom å presentere nøkkeltall og grafer på en beskrivende måte (Foldnes et al., 2018). Videre kan beskrivende statistikk gi en fremstilling av gjennomsnitt, spredning, antall observasjoner, standardavvik, minimums- og maksimumsverdier og sammenheng mellom variablene som skal brukes i analysen kan inkluderes. Beskrivende statistikk kan gi innsikt og være nyttig for å forstå observasjonene i en analyse, men det er imidlertid viktig å erkjenne begrensningene ved beskrivende statistikk da den ikke kan brukes til å trekke konklusjoner om årsakssammenheng eller til å generalisere funn til større populasjoner.

### 3.2 Kvantitativ metode

Oppgavens formål er å undersøke noen av de faktorene som påvirker lønn ved hjelp av en multipel regresjonsanalyse, som vil vise i hvilken grad enkelte variabler påvirker. For å sikre at analysen er realiserbar og presis, vil det være hensiktsmessig å bruke en kvantitativ metode for innsamling av data. Kvantitativ metode refererer til forskningsmetoder som anvendes for innsamling og analyse av data som er kvantitative, det vil si data som kan representeres som tall eller andre mengdetemer (Grønmo, 2023). Denne metoden gir mulighet for generalisering av funn til en større populasjon, men gir ikke alltid dyp innsikt i årsakssammenhenger og kan overse viktige kvalitative aspekter ved fenomenet som studeres.

Eksempler på kvantitative metoder inkluderer spørreskjemaundersøkelser, observasjoner og eksperimenter for å samle inn kvantitative data, som deretter analyseres ved hjelp av statistiske verktøy.

Oppgaven baserer seg på data som er hentet fra det statistiske analyseverktøyet Microdata.no. Her har også dataen blitt avgrenset. Microdata.no er en innovativ og unik metode for deling av data internasjonalt. Tjenesten gir brukerne tilgang til å analysere uforandret registerdata fra SSB og andre kilder uten behov for søknader eller godkjente prosjekter. Denne tilnærmingen muliggjør en mer effektiv og fleksibel analyse av data, og eliminerer kravene til tidligere formalisert samarbeid med dataeier eller innsending av søknader om data adgang (Microdata, u.å.). Microdata.no tilbyr detaljert analyse av data, støtter utviklingen av nøyaktige analyser og har høy grad av sikkerhet og personvern ved å anonymisere dataene som deles. Tjenesten har vært en verdifull ressurs for oppgaven for å kunne gjennomføre analysen og ved å behandle store og komplekse datasett på en effektiv og detaljert måte. En betydelig andel av datamengden som er tilgjengelig på Microdata.no er tilknyttet registerdata fra SSB. Statistikken inneholder informasjon om de fleste relevante faktorer som er nødvendige for å analysere hva som påvirker lønn, som er fokuset i denne oppgaven. Disse registerdataene utgjør en sentral del av SSBs statistikkproduksjon, og er derfor av stor betydning for produksjon av pålitelige og nøyaktige statistiske analyser. Ved å ha direkte tilgang til SSBs registerdata via Microdata.no, kan det utføres mer detaljerte og sofistikerte analyser som kan gi verdifull innsikt i samfunnsfenomener og -trender.

Registerdataene fra SSB er basert på en kombinasjon av administrative registre, spørreundersøkelser og elektronisk rapportering. Ved hjelp av disse metodene kan SSB innhente omfattende og pålitelig statistikk på en rekke områder. For å samle inn informasjon som ikke er tilgjengelig i registrene, bruker SSB elektronisk rapportering eller intervjuer i form av telefonsamtaler eller besøk i hjemmet til de som blir spurt (Statistisk Sentralbyrå, u.å.). Dette bidrar til å sikre at relevant informasjon blir samlet inn på en effektiv og pålitelig måte. SSB har en viktig rolle, ikke bare i å produsere statistikk, men også i å utføre forskning og analysevirksomhet samt samarbeide internasjonalt for å hjelpe utviklingsland og deres statistikkbyrå. Deres rolle i å produsere pålitelig statistikk på både nasjonalt og internasjonalt nivå er dermed sentralt og avgjørende for både forskning og samfunnsutvikling (Statistisk Sentralbyrå, u.å.). De nyeste dataene tilgjengelig da denne oppgaven ble påbegynt var tall fra 2021, og derfor har ikke nyere tall blitt tatt med.

### 3.3 Økonometri

I økonometrien brukes matematiske og statistiske metoder for å modellere og analysere økonomiske sammenhenger (Idsø, 2023). Blant annet utvikles det hypoteser som senere tester hvordan ulike variabler kan påvirke hverandre. Denne oppgaven undersøker hvordan variablene kjønn, alder og utdanning påvirker lønn. I denne delen av oppgaven gis en forklaring på enkel regresjon og multippel regresjon, med spesiell vekt på semi-logaritmisk regresjon. Videre vil  $R^2$  bli forklart, samt hypotesetesting vil bli gjennomgått.

#### 3.3.1 Enkel regresjon

Formålet med enkel regresjon er å undersøke om det eksisterer en lineær sammenheng mellom to variabler (Løvås, 2018). Denne sammenhengen gjelder den avhengige variabelen (det vil si variabelen som påvirkes) og kun én enkelt uavhengig variabel (det vil si variabelen som påvirker). Den matematiske ligningen for denne sammenhengen er som følger:

$$Y = \alpha + b_1X$$

*Formel 1 Enkel regresjon (Ubøe, 2017)*

Verdien av den avhengige variabelen og verdien av den uavhengige variabelen er representert ved henholdsvis  $Y$  og  $X$ . Krysningspunktet mellom linjen og  $Y$ -aksen, det vil si verdien av  $Y$  når  $X = 0$ , er representert ved bokstaven  $\alpha$ . Det er verdien av  $Y$  når  $X$  ikke har noen effekt på  $Y$ -verdien. Stigningstallet, som forteller hvor mye  $Y$ -verdien endrer seg for hver enhet økning i  $X$ , er representert ved bokstaven  $b$ . Dette gir en ide om hvor bratt linjen er og gir muligheten til å forutsi  $Y$ -verdier for forskjellige  $X$ -verdier.

### 3.3.2 Multippel regresjon

Multippel regresjon er en annen form for regresjonsanalyse som involverer inkludering av to eller flere uavhengige variabler i modellen (Ubøe, 2017). Ved å ta i bruk flere uavhengige variabler samtidig, kan man evaluere den samlede effekten disse variablene har på den avhengige variabelen. Den matematiske ligningen for en multippel regresjon er som følger:

$$Y = \alpha + b_1X + b_2X + b_3X$$

*Formel 2 Multippel regresjon (Ubøe, 2017)*

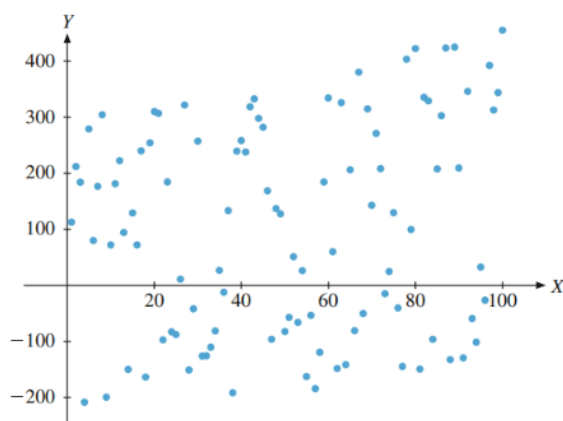
Funksjonsformen som er valgt i denne oppgaven er en semi-logaritmisk regresjon

$$\ln(Y) = \alpha + b_1X + b_2X + b_3X$$

hvor  $\ln(Y)$  er logaritmen til  $Y$ . Grunnen til at det er valgt en semi-logaritmisk funksjonsform er at denne formen ofte er mer hensiktsmessig å bruke når man analyserer sammenheng mellom lønn og andre variabler. Lønn har ofte en log-normal fordeling fordi lønninger kan variere fra å ha svært lave verdier til svært høye verdier. Å transformere til naturlig logaritme vil denne skjevheten transformeres til en mer balansert fordeling av verdiene, jevne ut spredningen av datamaterialet slik at man lettere kan tolke resultatene av regresjonsanalysen. En annen grunn til å bruke denne formen når den avhengige variabelen er lønn er at lønn har en tendens til å øke eksponentielt med eksempelvis utdanningsnivå og arbeidserfaring (økning i lønn er større for høyere nivåer av utdanning/erfaring). Ved å velge denne funksjonsformen kan sammenhengen mellom lønn og de andre variablene beskrives mer nøyaktig (Ubøe, 2017).

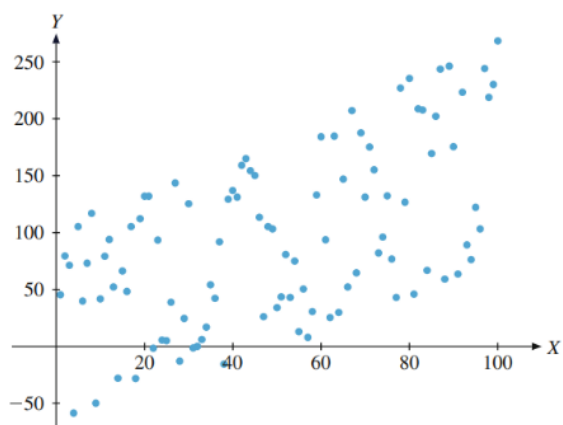
### 3.3.3 $R^2$ – Forklaringskraft

Forklaringskraft er en statistisk måling av modellens evne til å forklare variasjonene i Y-variabelen (Foldnes et al., 2018). Med andre ord, det angir modellens egnethet til å forklare variasjonen i Y-variabelen.  $R^2$ -verdien vil alltid variere mellom 0 og 1, hvor en verdi på 0 indikerer at ingen variasjoner i Y-variabelen kan forklares av X-variablene, mens en verdi på 1 angir at all variasjon i Y-variabelen kan forklares av X-variablene (Foldnes et al., 2018). For eksempel, hvis  $R^2$ -verdien er 0,4, indikerer dette at 40% av variasjonene i den avhengige variabelen (Y) kan forklares av endringer i en eller flere av de uavhengige variablene (X) (Ubøe, 2017).



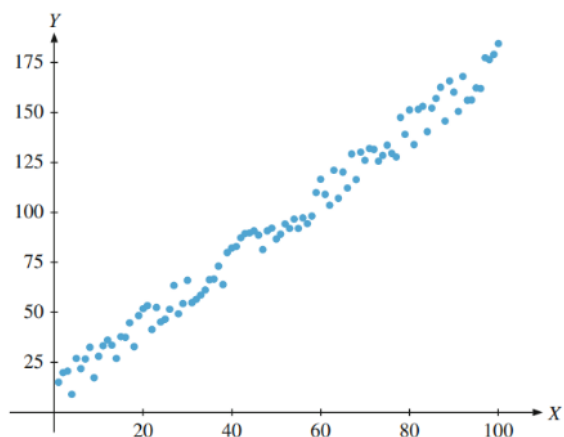
Figur 1 Liten forklaringskraft

(Ubøe, 2017, s.252).



Figur 2 Litt forklaringskraft

(Ubøe, 2017, s.253)



Figur 3 Sterk forklaringskraft

(Ubøe, 2017, s.253)

Figur 1, 2 og 3 illustrerer hvordan spredningen i modellen kan variere avhengig av graden av forklaringskraft i regresjonsmodellen, angående den avhengige variabelen Y basert på de uavhengige variablene X. I figur 1 viser det seg å være en minimal sammenheng mellom variablene X og Y, med en  $R^2$ -verdi på kun 2,3% for det aktuelle datasettet. I figur 2 derimot, er det tegn på en betydelig sammenheng mellom X og Y, da  $R^2$ -verdien øker til 30,7% og bekrefter en moderat sammenheng. I figur 3 derimot, observeres en sterk sammenheng mellom X og Y, med en betydelig høy  $R^2$ -verdi på 98,3%. (Ubøe, 2017).

### 3.3.4 Hypotesetesting

Hypotesetesting er en velkjent metode som brukes for å vurdere påstanden om sannheten til en antagelse eller påstand om en gruppe eller populasjon (Foldnes et al., 2018). Metoden innebærer formulering av en nullhypotese ( $H_0$ ) og en alternativ hypotese ( $H_A$ ).

Nullhypotesen er vanligvis en påstand om at det ikke finnes noen signifikant forskjell mellom variabler eller grupper som er valgt for forskningen, mens den alternative hypotesen er at det finnes en signifikant sammenheng mellom de valgte variablene (Løvås, 2018; Ubøe, 2017). Det er hovedsakelig nullhypotesen som testes, siden det antas at det ikke kan bekreftes en påstand med 100% sikkerhet. Hvis nullhypotesen blir avvist, gir det indirekte støtte til alternativhypotesen. Dette betyr at det kan trekkes en slutning om at det er en signifikant forskjell mellom de valgte variablene eller gruppene (Løvås, 2018; Ubøe, 2017).

Tosidig hypotesetesting, også kjent som toveis hypotesetesting, innebærer å teste for forskjeller både i positiv og negativ retning (Foldnes et al., 2018). Dette betyr at

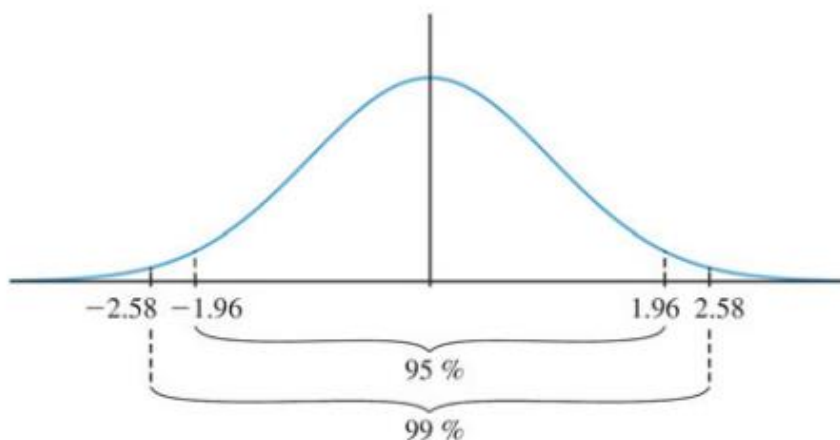
nullhypotesen formuleres som en påstand om at det ikke er noen signifikant forskjell mellom variablene eller gruppene, hverken positivt eller negativt:

$$H_0: \beta_i = 0 \text{ vs. } H_1: \beta_i \neq 0$$

Formel 3 Tosidig hypotesetesting (Ubøe, 2017)

Etter formulering av nullhypotesen og den alternative hypotesen, utføres en teststatistikk for å vurdere om nullhypotesen bør avvises eller ikke.

I denne oppgaven benyttes et signifikansnivå (alfa) på 5%, som tilsvarer et konfidensintervall på 95%. Signifikansnivået angir sannsynligheten for å avvise nullhypotesen når den faktisk er sann. Teststatistikken som anvendes er T-test, en hypotesetest som sammenligner gjennomsnittet av to grupper for å undersøke om forskjellene mellom gruppene er signifikante eller ikke (Ubøe, 2017). T-verdien indikerer størrelsen på forskjellen mellom gruppene, mens P-verdien angir sannsynligheten for å observere en forskjell som er større eller like stor som den observerte forskjellen. Dersom P-verdien er mindre enn det angitte signifikansnivået (5%), indikerer T-testen at forskjellene mellom gruppene er signifikante og ikke skyldes tilfeldigheter, og nullhypotesen forkastes til fordel for den alternative hypotesen. Dersom P-verdien er større enn signifikansnivået, indikerer dette at nullhypotesen ikke kan avvises, da det ikke er tilstrekkelig bevis for å støtte den alternative hypotesen (Løvås, 2018)



Figur 4 Standard normalfordeling med konfidensintervaller

(Ubøe, 2017, s.165)



I denne oppgaven utledes det tre hypoteser som testes ved å benytte t-test.

Hypotese 1: Følgende nullhypotese testes;  $H_0$  = det finnes ikke noen forskjell i lønn mellom kvinner og menn.  $H_A$  = det finnes en signifikant forskjell i lønn mellom kjønnene.

$$H_0: \beta_1 = 0, \quad H_A: \beta_1 \neq 0$$

Hypotese 2: Følgende nullhypotese testes;  $H_0$  = det finnes ikke noe forskjell i lønn mellom personer av ulik alder i aldersgruppen 25-67 år.  $H_A$  = det finnes en signifikant forskjell i lønn mellom personer av ulik alder.

$$H_0: \beta_2 = 0, \quad H_A: \beta_2 \neq 0$$

Hypotese 3: Følgende nullhypotese testes;  $H_0$  = det finnes ikke noe forskjell i lønn mellom personer med høy utdanning (master og doktorgrad) og personer med lavere til ingen utdanning (utdanning lavere enn masternivå).  $H_A$  = Det finnes en signifikant forskjell i lønn mellom de to utdanningsgruppene.

$$H_0: \beta_3 = 0, \quad H_A: \beta_3 \neq 0$$

## 4. Datamateriale

I dette kapitlet presenteres en grundig definisjon av hver type variabel, og deres funksjoner samt bruksområder i denne forskning og analysen forklares. Videre gis en detaljert beskrivelse av flere datasett som er benyttet med formålet å undersøke faktorer som påvirker lønnsnivået ved hjelp av de utvalgte variablene. Disse datasettene er nøye undersøkt og behandlet for å sikre at resultatene som kommer frem er pålitelige og informative, samt er relevante for problemstillingen i oppgaven. For å sikre både presisjon

og relevans i resultatene, har det blitt gjort flere avgrensninger i det presenterte datamaterialet. Disse avgrensningene inkluderer begrensning av antall år dataene er hentet fra, demografiske faktorer og utdanningsnivå. Videre har det blitt gjort en nøye utvelgelse av relevante variabler, samt ekskludering av irrelevante eller unøyaktige variabler. Det vil bli gitt en detaljert beskrivelse av disse datasettene og de gjennomførte avgrensningene.

#### 4.1 Definisjon av type variabler

En variabel refererer til en faktor eller et trekk som kan differensiere fra én person til en annen (Foldnes et al., 2018). En avhengig variabel er den som studeres og påvirkes av andre variabler, mens den uavhengige variabelen er den som kan kontrolleres og antas å påvirke den avhengige variabelen.

I denne oppgaven er lønn valgt som den avhengige variabelen. For at analysen skal være mer pålitelig og sammenlignbar har vi forklart i kapittel 6.1.3 at vi har valgt å omgjøre lønn til  $\ln$ lønn, som er logaritmen av lønn. Dette vil gi oss en mer normalfordelt variabel å jobbe med slik at det blir lettere å tolke effekten av de uavhengige variablene på lønnen.

Variablene kjønn, alder og utdanning er valgt som de uavhengige variablene i denne oppgaven på grunn av deres mulige innvirkning på lønn. Kjønn kan muligens påvirke lønnen i form av diskriminering, valg av utdanning, valg av yrke. Alder kan ha en innvirkning på lønn gjennom ansiennitet og erfaring, mens ettertraktede kvalifikasjoner og kompetanse kan resultere i at utdanning kan ha innvirkning på lønn.

En indikatorvariabel, også kjent som en dummy variabel, blir ofte benyttet når det er behov for å transformere en kategorisk variabel i et datasett til en numerisk variabel, slik at en spesifikk gruppe kan inkluderes i en regresjonsanalyse (Ubøe, 2017). I denne analysen er kjønn og utdanning kodet som dummy variabler, ettersom de representerer kategorier eller grupper som ikke kan kvantifiseres eller måles på en kontinuerlig skala. Ved å konvertere kjønn og utdanning til dummy variabler, blir det mulig å systematisk inkludere kategorisk informasjon i analysen, noe som forenkler tolkningen av resultater og konklusjonsdannelsen basert på disse.

## 4.2 Beskrivelse av datasett med avgrensning

Som tidligere forklart i kapittel 3.2 ble all data som ble benyttet i denne analysen importert og behandlet gjennom Mikrodata.no (microdata.no, 2023). Her beskrives detaljert alle prosedyrer som ble utført på plattformen for å generere de dataene som dannet grunnlaget for analysen. En omfattende gjennomgang av aktivitetene som utføres på Mikrodata er inkludert som vedlegg 1.

Først ble det importert et datasett med navnet "BEFOLKNING\_FOEDSELS\_AAR\_MND" som vi tildelte navnet «befolkning». Datasettet inneholder informasjon om alle som er registrert med et fødselsnummer i Norge. Formålet med å importere et datasett som inkluderte fødselsnumre var å muliggjøre en demografisk avgrensning basert på alder, noe som vil bli grundig forklart senere.

Videre ble et annet datasett med navnet "BEFOLKNING\_STATUSKODE" importert, og dette ble tildelt navnet "regstatus". Dette datasettet muliggjorde begrensning av utvalget til å kun inkludere personer som var bosatt i Norge per 01.01.2021. Opprinnelig hadde "regstatus"-datasettet en populasjon på 11 089 163 enheter, der kategori 1 representerte de som var bosatte i Norge per 01.01.2021, kategori 3 representerte de som hadde utvandret fra Norge, og kategori 5 representerte de som var registrert døde. Da denne problemstillingen omhandler forskjeller i lønnsnivået blant arbeidstakere i Norge er det hensiktsmessig at personer som har utvandret fra Norge og personer som er registrert døde blir fjernet fra datasettet. Begrensningen til kategori 1, bosatt i Norge per 01.01.2021, resulterte i en reduksjon av populasjonen til 5 389 955 enheter.

Gitt at oppgavens fokusgruppe var begrenset til personer som var både høyt utdannet og yrkesaktive, og som tilhørte enten det mannlige eller kvinnelige kjønn, var det nødvendig å begrense utvalget til personer som matchet denne profilen. I datasettet som ble kalt «befolkning» ble det derfor gjort en demografisk avgrensning til personer i aldersgruppen 25 til 67 år. Dette ble den første variabelen i datasettet og ble tildelt betegnelsen "alder". Denne avgrensningen ble foretatt på bakgrunn av at fullførelsen av en mastergrad er omtrent i en alder av 25 år, og normal pensjonsalder i Norge er 67 år. Personer som ikke tilhørte dette aldersintervallet ble ekskludert, noe som resulterte i en reduksjon av populasjonen til 3 035 242 enheter.

For å kunne undersøke en mulig sammenheng mellom kjønn og lønnsnivå, ble datasettet "BEFOLKNING\_KJOENN" importert og tildelt betegnelsen "kjønn". Dette datasettet inneholdt informasjon om fordelingen mellom kvinner og menn og ble importert med den hensikt å inkludere en dummyvariabel i analysen som kunne kategorisere kvinner og menn separat. Dummyvariabelen, som ble kalt «man\_first», ble gitt verdiene 1 for mann og 0 for kvinner.

<b>man_first</b>	<b>Prosent</b>	<b>Antall</b>
1 - Mann	51 %	1 551 498
0 - Kvinne	49 %	1 483 750
Total	100 %	3 035 242

Tabell 1 Oversikt av dummyvariabelen man\_first.

Tall er hentet fra microdata.no (microdata.no, 2023), og tabellen ble opprettet av oss ved hjelp av Excel.

For å analysere effekten av høyere utdanning på lønn ble datasettet "NUDB\_BU" importert og navngitt "utdanning". Datasettet inneholdt informasjon om personers høyeste fullførte utdanningsnivå per 01.01.2021. For at dette datasettet skulle kunne bidra til å undersøke effekten av høyere utdanning, spesifikt mastergrad og doktorgrad, på lønnsnivået, måtte det gjennomgå noen modifikasjoner. Det opprinnelige datasettet inkluderte 10 forskjellige kategorier, som de 3 035 242 observasjonene ble fordelt på.

<b>Utdanningsnivå</b>	<b>Observasjoner</b>
0 - Ingen	6 940
1 - Barneskole	16 950
2 - Ungdomsskole	538 513
3 - Videregående	143 370
4 - Videregående avsluttende utdanning	855 783
5 - Påbygg til videregående	109 235
6 - UH-utdanning - lavere nivå	851 042
7 - UH-utdanning - høyere nivå	358 693
8 - Doktorgrad	37 503
9 - Uoppgitt	117 213
<b>Total</b>	<b>3 035 242</b>

Tabell 2 Oversikt over ulike utdanningsnivåer og observasjoner

Tall er hentet fra microdata.no (Holseter, 2019; microdata.no, 2023). Tabellen ble opprettet av oss ved hjelp av Excel.

Av de 10 kategorier av utdanningsnivå ble kategori 0 til 6 definert som ingen/lav utdanning av SSB, mens kategori 7 og 8 ble definert som høy utdanning. For å isolere effekten av høyere utdanning, ble dummyvariabelen "høy\_utd" opprettet. Kategori 1 representerer utdanning på master- og doktorgradsnivå og 0 representerer ingen, uoppgitt og utdanning lavere enn masternivå.

<b>høy_utd</b>	<b>Observasjoner</b>	<b>Prosent</b>
0 - Under masternivå	2 639 046	86,95 %
1 - Master & doktorgrad	396 196	13,05 %
<b>Total</b>	<b>3 035 242</b>	<b>100 %</b>

Tabell 3 Oversikt over dummyvariabel høy\_utd

Tall er hentet fra microdata.no (microdata.no, 2023). Tabellen ble opprettet av oss ved hjelp av Excel.

Fra tabell 3 ser man at personer med mastergrad eller doktorgrad utgjør omtrent 13% av utvalget, mens personer med ingen, uoppgitt eller utdanning lavere enn mastergradsnivå utgjør omtrent 87%.

Datasettet, "INNTEKT\_WLONN", inneholdt variabler som omfattet kontantlønn, skattepliktige naturalytelser, samt syke- og fødselspenger, utbetalt innenfor et kalenderår. Med hensikt å analysere lønnsutviklingen over tid, ble variabler for hver lønns-år fra 2015 til 2021 opprettet. Disse variablene ble betegnet som lønn15, lønn16, lønn17, lønn18, lønn19, lønn20 og lønn21. Variablene ble transformert til naturlige logaritmer og de blir referert til som lnlønn15, lnlønn16, lnlønn17, lnlønn18, lnlønn19, lnlønn20 og lnlønn21 i analysen. Det er ikke spesifisert i datamaterialet om lønn inkluderer deltidsansatte eller om lønnen er basert på en heltidsekvivalent. Dette kan føre til skjevheter i datamaterialet, da sammenligning mellom heltids- og deltidsinntekter kan bli feilaktig.

Variabel	Gj.snitt	Std.avvik	Antall
lønn 15	kr 435 412	kr 286 471	2 490 580
lønn 16	kr 449 729	kr 288 158	2 502 252
lønn 17	kr 467 610	kr 293 387	2 515 985
lønn 18	kr 491 461	kr 302 766	2 530 718
lønn 19	kr 516 590	kr 314 184	2 549 916
lønn 20	kr 530 576	kr 321 817	2 523 805
lønn 21	kr 564 050	kr 335 788	2 488 391

Tabell 4 Oversikt over gjennomsnittslønn fra 2015 – 2021

Tall er hentet fra microdata.no (microdata.no, 2023). Tabellen ble opprettet av oss ved hjelp av Excel.

Variabel	Gj.snitt	Std.avvik	Antall
lnlønn15	12,62	1,11594	2 490 572
lnlønn16	12,66	1,10156	2 502 252
lnlønn17	12,72	1,08034	2 515 982
lnlønn18	12,78	1,06140	2 530 718
lnlønn19	12,82	1,10366	2 549 901
lnlønn20	12,86	1,06153	2 523 784
lnlønn21	12,93	1,05012	2 488 362

Tabell 5 Oversikt over gjennomsnittslønn konvertert til naturlig logaritme

Tall er hentet fra microdata.no (microdata.no, 2023). Tabellen ble opprettet av oss ved hjelp av Excel.

## 5. Beskrivende statistikk

Beskrivende statistikk refererer til en analysemetode som har til formål å gi en forståelse av de utvalgte dataene ved å presentere nøkkeltall og grafer på en treffsikker og beskrivende måte (Foldnes et al., 2018). Dette kapitlet tar sikte på å gi en systematisk og forståelig beskrivelse samt presentasjon av data ved hjelp av beskrivende statistikk.

Variabel	Gj.snitt	Std.avvik	Observasjoner	25% Kvartil	75% Kvartil
man_first	51,12 %	49,99 %	3 035 242	0	1
høy_utd	13,05 %	33,69 %	3 035 242	0	0
alder	45,40	12,19	3 035 242	35	56
lønn15	kr 435 412	kr 286 471	2 490 580	kr 231 740	kr 571 298
lønn16	kr 449 729	kr 288 158	2 502 252	kr 252 266	kr 585 855
lønn17	kr 467 610	kr 293 387	2 515 985	kr 273 562	kr 604 909
lønn18	kr 491 461	kr 302 766	2 530 718	kr 299 724	kr 630 414
lønn19	kr 516 590	kr 314 184	2 549 916	kr 325 083	kr 659 147
lønn20	kr 530 576	kr 321 817	2 523 805	kr 332 278	kr 675 691
lønn21	kr 564 050	kr 335 788	2 488 391	kr 367 741	kr 712 357

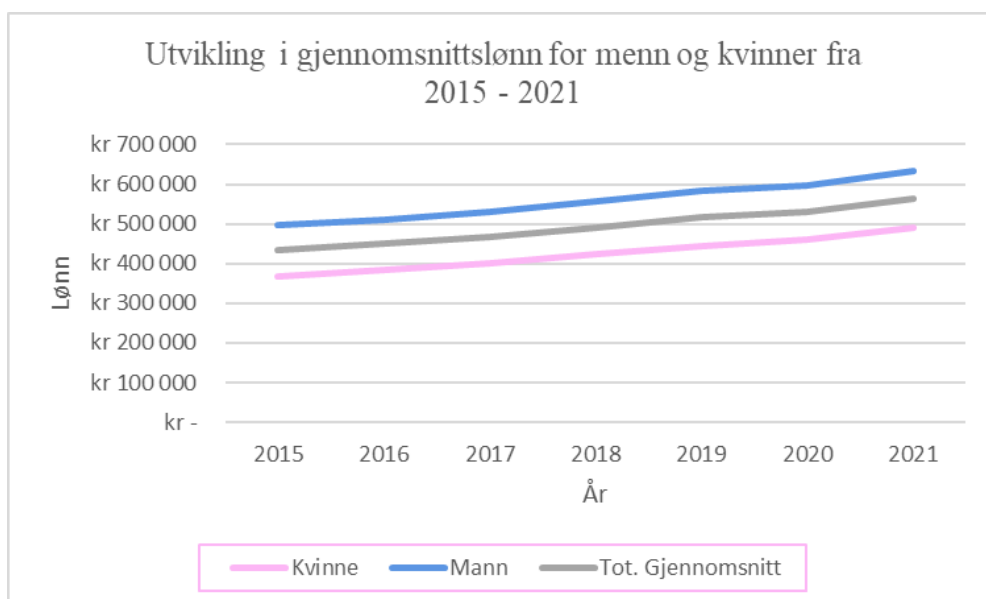
Tabell 6 Oversikt over variablene kjønn, alder og utdanning

Tall er hentet fra *microdata.no* (*microdata.no*, 2023). Tabellen ble opprettet av oss ved hjelp av Excel.

I tabell 6 ovenfor presenteres det ulike egenskaper ved populasjonen. Ifølge tabellen utgjør menn 51,12% av den totale populasjonen på 3 035 242. Videre viser tabellen at 13,05% av populasjonen har fullført en master- eller doktorgrad. Gjennomsnittsalderen i populasjonen er omtrent 45 år, og tabellen gir en oversikt over lønns gjennomsnittet fra 2015-2021 for en variert populasjon.

### 5.1 Kjønn

En grundig undersøkelse av deskriptiv statistikk, som inkluderer metrikker som gjennomsnittlig lønn for begge kjønn over tid, samt ulike mål for spredning og variasjon, som standardavvik, kan bidra til en mer grundig forståelse av hvordan kjønn påvirker lønn. Gjennomsnittslønn er en vanlig indikator på sentral tendens, som gir innsikt i ulikheter mellom gjennomsnittlig lønn for menn og kvinner. Ved å sammenligne gjennomsnittslønnen for menn og kvinner, er det mulig å bestemme om det eksisterer lønnsforskjeller basert på kjønn.



Figur 5 Viser utvikling i gjennomsnittslønn for menn og kvinner fra 2015-2021

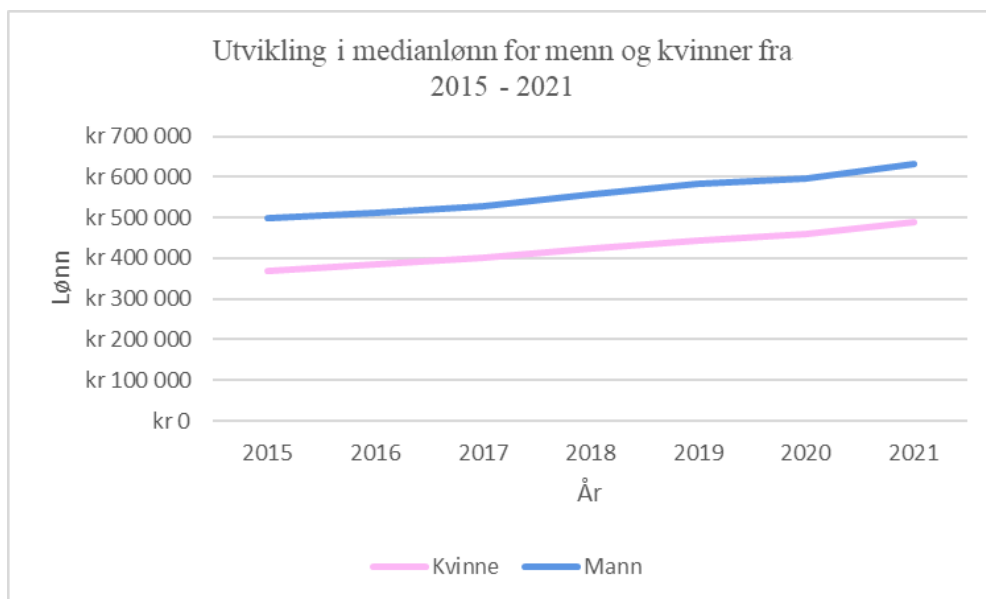
Tall er hentet fra *microdata.no* (*microdata.no*, 2023). Figuren ble opprettet av oss ved hjelp av Excel.

Som vist i figur 5, er det åpenbart at menn har en høyere gjennomsnittslønn enn kvinner, og at kvinner tjener under det generelle gjennomsnittet for begge kjønn. Dette gir grunnlag for å anta at kjønn spiller en betydelig rolle i bestemmelsen av lønn, ettersom det eksisterer en ulikhet i lønn mellom kjønnene.

En alternativ tilnærming for å analysere lønnsforskjeller er å vurdere medianlønn.

Medianlønn representerer den midterste verdien i et datasett og kan være et mer relevant mål på sentral tendens hvis det finnes noen få ekstremt høye eller ekstremt lave lønninger som påvirker gjennomsnittet (Løvås, 2018).





Figur 6 Viser utvikling i medianlønn for menn og kvinner fra 2015-2021

Tall er hentet fra microdata.no (microdata.no, 2023). Figuren ble opprettet av oss ved hjelp av Excel.

Figur 6 viser utviklingen i medianlønn over tid, inndelt etter kjønn. Som grafen viser, øker lønnen jevnt over for begge kjønn, men det er ikke tydelig om det forekommer noen reduksjon i lønnsforskjellene. For å undersøke dette mer nøyaktig kan det være hensiktsmessig å konvertere økningen til prosentandeler.

Årstall	Gjennomsnittlig årslønn		Lønnsøkning i prosent	
	Kvinner	Menn	Kvinner	Menn
2015	kr 368 550	kr 498 696		
2016	kr 384 265	kr 511 390	4,26 %	2,55 %
2017	kr 401 688	kr 529 372	4,53 %	3,52 %
2018	kr 422 511	kr 555 678	5,18 %	4,97 %
2019	kr 444 829	kr 583 135	5,28 %	4,94 %
2020	kr 459 700	kr 595 969	3,34 %	2,20 %
2021	kr 489 538	kr 632 902	6,49 %	6,20 %

Tabell 7 Viser prosentvis lønnsøkning for kvinner og menn fra 2015 – 2021

Tall er hentet fra microdata.no (microdata.no, 2023). Tabellen ble opprettet av oss ved hjelp av Excel.

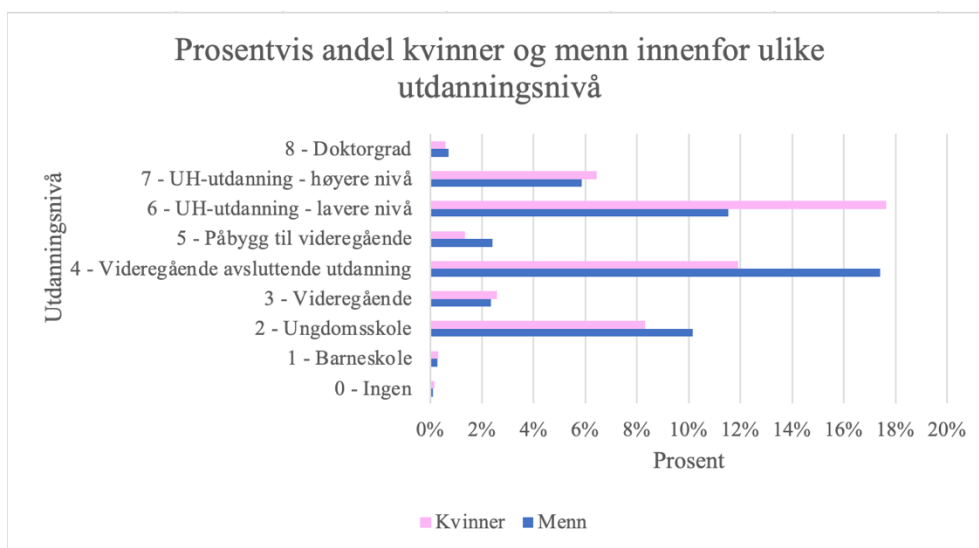
Figur 5 og 6 viser en jevn økning i henholdsvis gjennomsnitts- og medianlønn for både kvinner og menn i perioden 2015-2021. Tabell 7 viser gjennomsnittslønn, og det kommer frem at kvinner har en høyere prosentvis økning i lønnen sammenlignet med menn. Til

tross for dette viser tabellen at lønnsgapet mellom kvinner og menn fortsatt er betydelig, og selv om kvinner har en høyere gjennomsnittlig lønnsøkning enn menn, kan det fortsatt ta tid å tette lønnsgapet.

Videre undersøkes det om andre faktorer kan forklare det observerte lønnsgapet mellom kvinner og menn, da det ville være overfladisk å konkludere med at lønnsgapet utelukkende skyldes kjønn. Det vil utføres en undersøkelse av om ulikt utdanningsnivå, yrkesvalg eller fysiologiske forhold kan være avgjørende for lønnsgapet mellom kjønnene. Dette vil bidra til en grundigere analyse av lønnsgapet og identifisere eventuelle andre faktorer som kan bidra til forskjellene mellom kvinner og menn.

- Kan utdanning forklare lønnsgapet mellom menn og kvinner?

Senere, i kapittel 5.3 blir det presentert statistiske data som viser at flere kvinner har høyere utdanning i forhold til menn.



Figur 7 Prosentvis fordeling av kvinner og menn over de ulike utdanningsnivåene

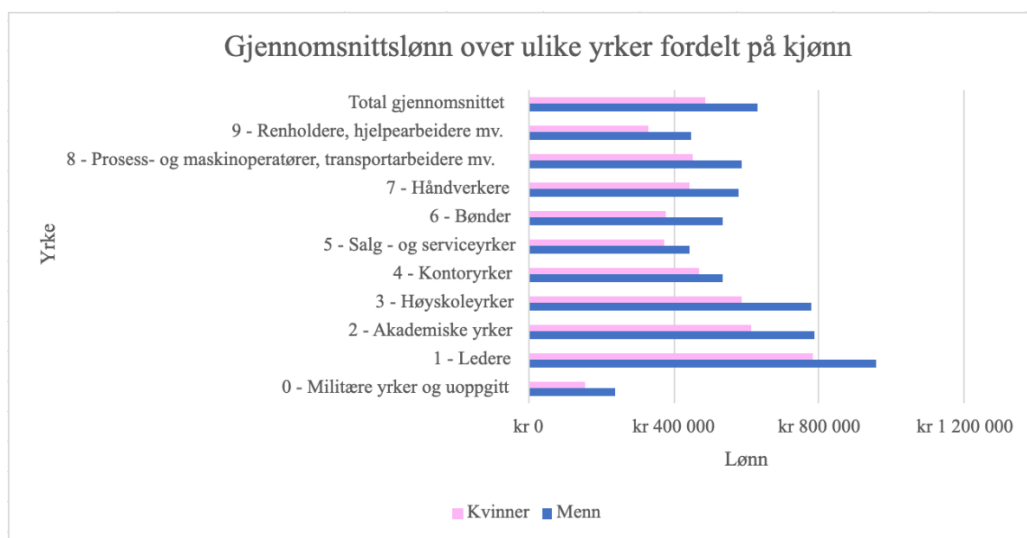
Tall er hentet fra microdata.no (Holseter, 2019; microdata.no, 2023). Figuren ble opprettet av oss ved hjelp av Excel.

Figur 7 viser at kvinner er overrepresentert på nivå 6 og 7, noe som i utgangspunktet skulle tilsi positive konsekvenser for lønnsnivået, da det er en antakelse om at høyere utdanning øker sjansen for å få jobber med høyere lønn.

Ifølge «The College Payoff»-rapporten er utdanning en avgjørende faktor for inntekt, da personer med høyere utdanning generelt sett tjener mer enn de med lavere utdanning (Carnevale et al., 2011). Imidlertid sier rapporten at valg av yrke kan overstyre utdanningsnivået, spesielt i høytlønnede yrker. Dette betyr at selv om en person har lavere utdanning, kan de likevel tjene mer enn en person med høyere utdanning i et lavtlønnet yrke, avhengig av yrkesvalget de har tatt.

Valg av yrke kan også påvirke en persons karrierevei, arbeidserfaring og muligheter for forfremmelser og lønnsøkninger over tid. Noen yrker kan ha større karrieremuligheter og potensiale for økonomisk vekst, mens andre yrker muligens har begrensede muligheter for lønnsutvikling. Ved å vurdere yrkesvalg kan man få et mer komplett bilde av hvordan ulike yrker kan påvirke en persons inntektsmuligheter over tid.

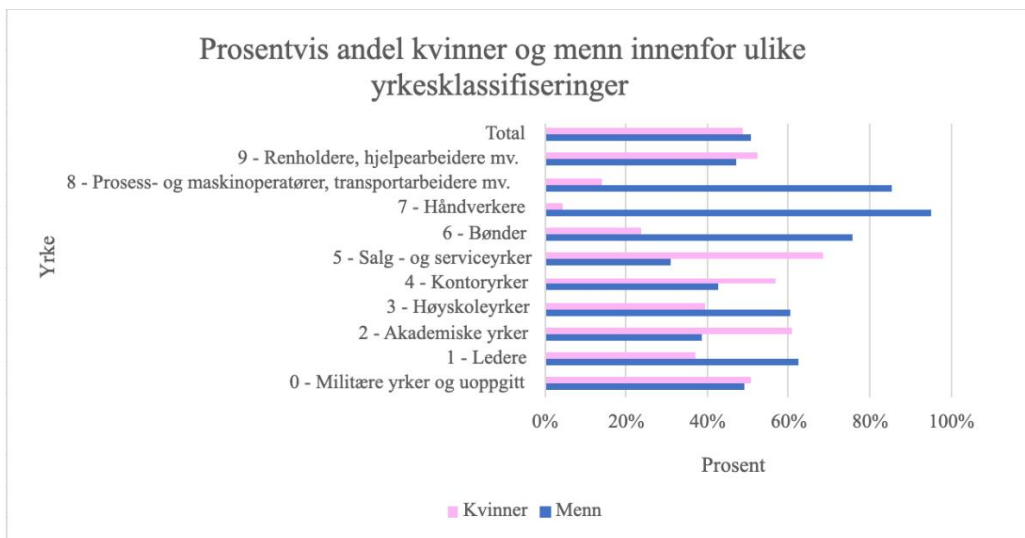
Det er dermed avgjørende å analysere begge kjønns valg av yrke for å identifisere eventuelle sammenhenger mellom yrkesvalg og lønnsgapet mellom kvinner og menn.



Figur 8 Gjennomsnittslønn i de ulike yrkesgruppene fordelt på kjønn

Tall er hentet fra [microdata.no](https://microdata.no) (microdata.no, 2023). Figuren ble opprettet av oss ved hjelp av Excel.

Figur 8 viser at yrker som ledere, akademikere og yrker innenfor høyskolesektoren ligger høyere på lønsskalaen, og dermed har et høyere inntektsnivå. På den annen side er yrker som militær/uoppgitt yrke, salg- og service samt renholdere/hjelpemedarbeidere plassert lengre nede på lønsskalaen, og har dermed lavere lønnsnivå.

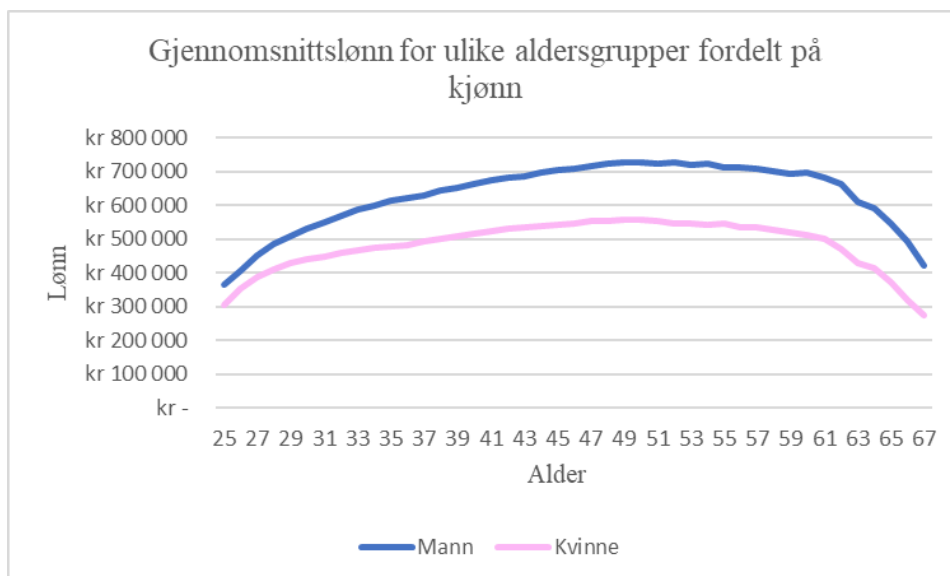


Figur 9 Prosentvise andelen kvinner og menn innenfor ulike yrkesklassifiseringer

Tall er hentet fra microdata.no (microdata.no, 2023). Figuren ble opprettet av oss ved hjelp av Excel.

Videre viser figur 9 at kun ett av yrkene med høyest lønn har en større prosentandel kvinner, mens flere av yrkene lengre nede på lønnskalaen er overrepresentert av kvinner. Disse funnene kan peke på mulige lønnsforskjeller mellom kjønn knyttet til yrkesvalg og kan indikere eksisterende kjønnsrelaterte skjevheter i arbeidsmarkedet som kan påvirke inntektsulikhetene mellom kjønnene.

Studien "Children and Gender inequality" viser at barnefødsler fører til kjønnsulikheter i arbeidsmarkedet og en lønnsforskjell på 20% mellom kvinner og menn (Kleven et al., 2019). Det viser også at kvinner har lavere sannsynlighet for å oppnå lederstillinger etter å ha fått barn, men er mer tilbøyelige til å jobbe i den offentlige sektoren som tilbyr fleksibilitet og positive holdninger til foreldrepermisjon. Dermed er det relevant å undersøke lønnsutviklingen for både menn og kvinner for å se om det oppstår en større og vedvarende lønnsforskjell etter gjennomsnittsalderen for førstegangsfødende. Ifølge Statistisk sentralbyrå har gjennomsnittsalderen for førstegangsfødende økt og var 30,1 år i 2021, opp fra 29,9 år i 2020. Dette er første gang snittalderen har kommet over 30 år. For fedre var gjennomsnittsalderen 32,1 år. I løpet av de siste ti årene har gjennomsnittsalderen for mødres første fødsel økt med to år (Andersen, 2022).



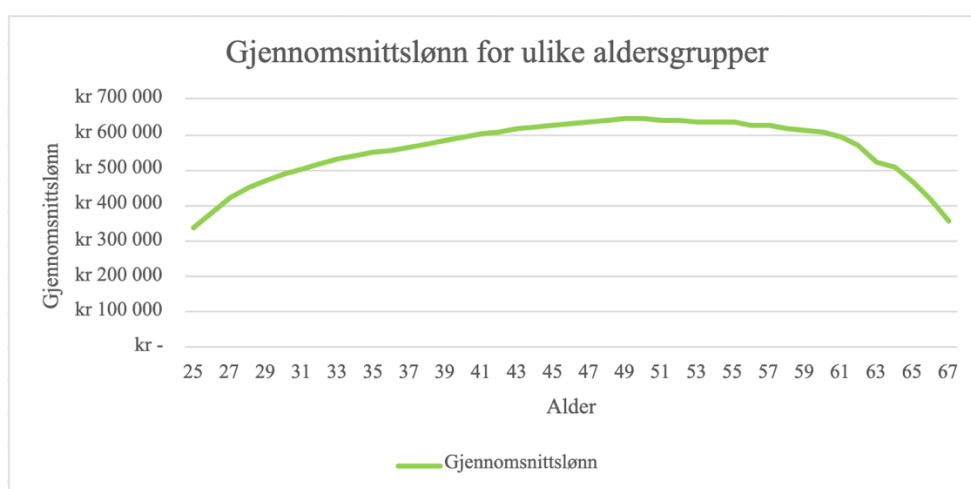
Figur 10 Gjennomsnittslønnen for ulike aldersgrupper fordelt på kjønn

Tall er hentet fra microdata.no (microdata.no, 2023). Figuren ble opprettet av oss ved hjelp av Excel.

Figur 10 illustrerer lønnsutviklingen for menn og kvinner, og det kommer frem at lønnsforskjellen mellom kjønnene er relativt liten før de når 30 år, men øker betydelig etter denne alderen. Dette funnet er i samsvar med funnene presentert i studien "Children and Gender inequality" (Kleven et al., 2019).

## 5.2 Alder

I denne oppgaven er alder identifisert som en av de utvalgte variablene som skal analyseres med formål om å undersøke faktorer som påvirker lønn. Ved hjelp av deskriptiv statistikk er det mulig å undersøke og beskrive forholdet mellom alder og lønn.



Figur 11 Gjennomsnittslønn for ulike aldersgrupper fra 25-67 år

Tall er hentet fra microdata.no (microdata.no, 2023). Figuren ble opprettet av oss ved hjelp av Excel.

I figur 11 presenteres en graf som viser gjennomsnittlig lønn for ulike aldersgrupper. Grafen illustrerer at sammenhengen mellom alder og lønn er positiv, men ikke lineær. Observasjonen av knekker i lønnskurven for ulike aldersgrupper kan antyde tilstedeværelsen av ulike faktorer som påvirker lønnsutviklingen. Dette indikerer at alder ikke følger en lineær sammenheng med lønn, og kompleksiteten i denne relasjonen blir fremhevet.

Figur 11 presenterer gjennomsnittlig lønn for ulike aldersgrupper og viser at det først er en positiv men ikke lineær sammenheng mellom alder og lønn. Kurven viser at det er en knekk i gjennomsnittslønn rundt 50 årene. Fra denne alderen kan man se at alder påvirker lønnsnivået negativt, noe som kan skyldes ulike faktorer.

Kurven viser som nevnt en markant stigning i starten, noe som kan forklares av at lønnen øker raskt i begynnelsen av karrieren når nyutdannede opparbeider seg erfaring og avanserer i karrierestigen. Etter en viss alder kan imidlertid lønnsveksten avta eller flate ut som følge av ulike faktorer som redusert mobilitet i organisasjonen eller bransjen, samt muligheten for å nå et lønnstak. Det er også viktige variasjoner i denne sammenhengen avhengig av yrke, bransje, geografisk plassering og andre faktorer. For eksempel kan noen yrker og bransjer ha klare "topp"-år der lønnen når et høydepunkt på et bestemt tidspunkt i karrieren, mens andre yrker kan ha jevnere lønnsvekst over tid.

Det er også verdt å merke seg at noen arbeidsgivere kan ha en negativ oppfatning av eldre arbeidstakers innovative evner og produktivitet (Hauge & Moen, 2023), noe som kan føre til aldersdiskriminering og spille en rolle i en ikke-lineær sammenheng mellom alder og lønn.

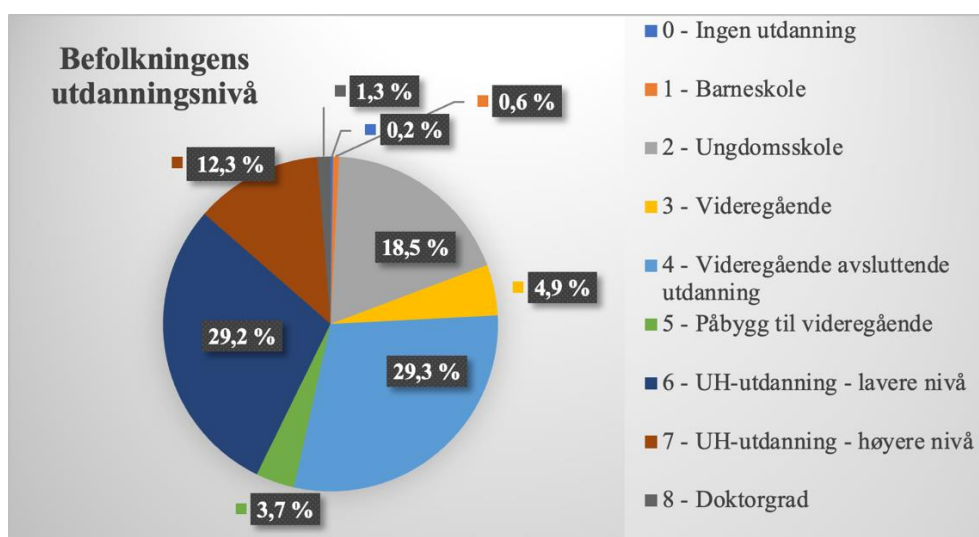
Kurven avtar en god del i nærheten av pensjonsalder, hvor det antas at personer enten reduserer arbeidstiden eller pensjonerer seg tidligere enn andre. Ved utgangen av desember 2021 var antallet personer som mottok alderspensjon 1 001 500 (NAV, 2022).

Gjennomsnittsalderen ved første uttak av alderspensjon i 2021 var 65,5 år, med en gjennomsnittlig alder på 66,1 år for kvinner og 64,9 år for menn. Det finnes ulike alderspensjonsordninger som tilbys i forskjellige yrker, AFP-avtaler og private avtaler som kan føre til at enkelte har muligheten til å velge å gå av med pensjon tidligere enn ved fylte 67 år (Forbrukerrådet, u.å.; KLP, u.å.). Andre velger gjerne å jobbe mindre fra en viss alder og gradvis trappe ned arbeidsmengden før de pensjonerer seg fullstendig. Dette kan

muligens forklare knekken i kurven ved 57 år, da både arbeidsstyrken og arbeidstimer reduseres.

### 5.3 Utdanning

Befolkningens utdanningsnivå er høyt i Norge sammenlignet med mange andre land, ifølge statistikk fra Statistisk sentralbyrå. I 2020 hadde 37,5% av befolkningen mellom 25 og 64 år høyere utdanning, mens 5,3% hadde grunnskole eller lavere utdanning, og 37,2% hadde videregående utdanning. Utdanningsnivået var høyest blant kvinner, hvor 41,5% hadde høyere utdanning, mens andelen var 33,4% for menn (Statistisk Sentralbyrå, 2022).

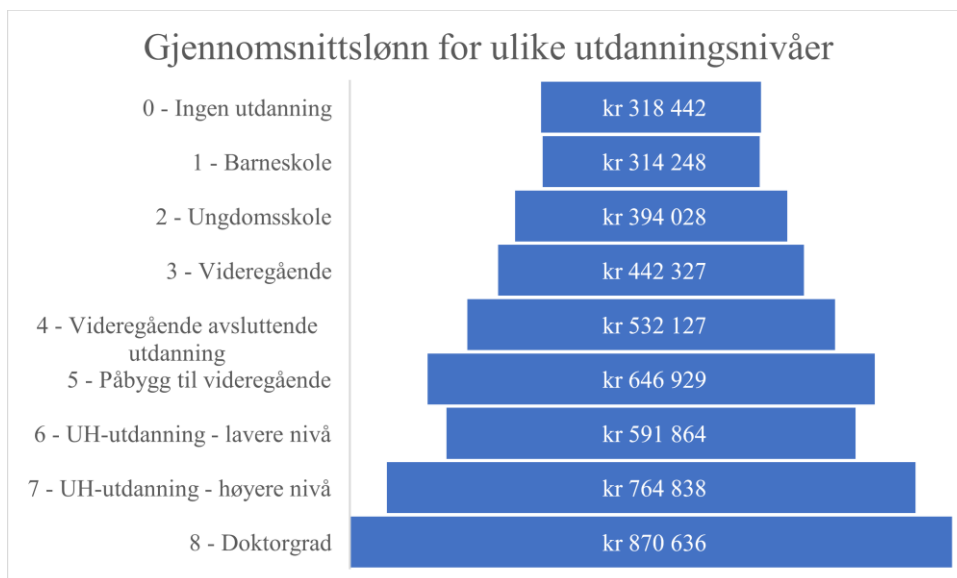


Figur 12 Befolkningens prosentvis fordeling mellom de ulike utdanningsnivåene

Tall er hentet fra microdata.no (Holseter, 2019; microdata.no, 2023). Figuren ble opprettet av oss ved hjelp av Excel.

Disse funnene stemmer godt overens med data fra mikrodata for 2021 som danner grunnlaget for denne analysen. Befolkningens utdanningsnivå er illustrert i figur 12.

Ved å utføre en analyse av befolkningens høyeste fullførte utdanningsnivå i forhold til deres lønn kan man få innsikt i en eventuell sammenheng mellom utdanning og inntekt.



Figur 13 Gjennomsnittslønn fordelt på ulike utdanningsnivåer

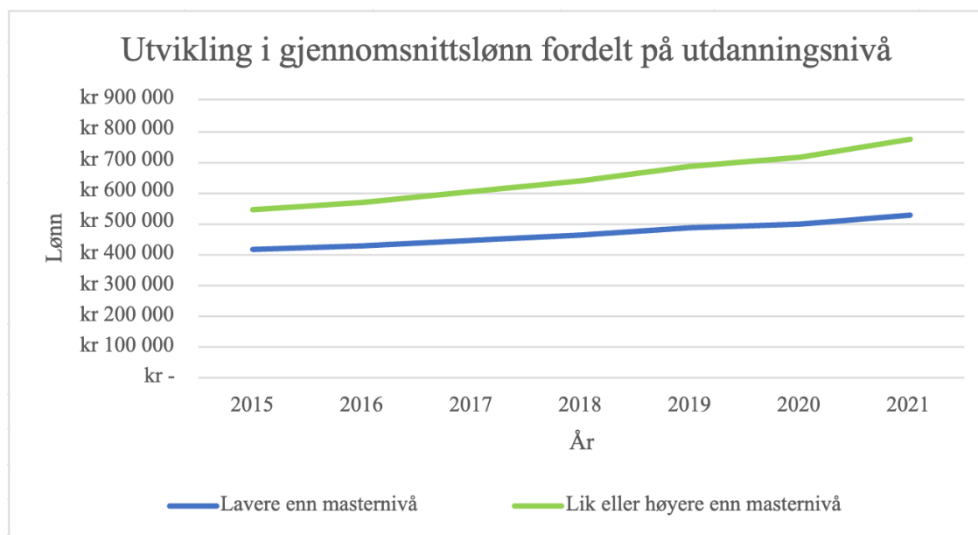
Tall er hentet fra microdata.no (Holseter, 2019; microdata.no, 2023). Figuren ble opprettet av oss ved hjelp av Excel.

Figur 13 viser gjennomsnittslønnen for individer med ulike høyeste fullførte utdanningsnivåer, og tydeliggjør en klar tendens til at lønnen øker i tråd med økt utdanningsnivå. Det er imidlertid viktig å bemerke at det finnes en avvikende trend for utdanningsnivå 5 - yrkesfaglig utdanning, som viser en høyere gjennomsnittslønn sammenlignet med nivå 6 - høyere utdanning. Dette fenomenet blir påpekt i rapporten fra "The College Payoff", der det understrekes at yrkesvalg kan overstyre utdanningsnivået, spesielt innen høytlønnede yrker. Dette innebærer at en person med lavere utdanning kan potensielt tjene mer enn en annen person med høyere utdanning i et lavtlønnet yrke. (Carnevale et al., 2011). Dette kan utgjøre en forklaring på hvorfor personer med utdanningsnivå 5 har en gjennomsnittlig høyere inntekt enn de med utdanningsnivå 4, til tross for at sistnevnte har en høyere grad av formell utdanning.

Personer med høyere utdanning har en konkurransefordel på arbeidsmarkedet på grunn av deres spesifikke ferdigheter og kunnskaper som gjør dem i stand til å anvende teknologi og konkurrere internasjonalt. Prognoser fra SSB viser at gjennomsnittlig utdanningsnivå i arbeidsstyrken forventes å øke frem mot 2040 (Statistisk Sentralbyrå, 2020). Dette skyldes økte krav til utdanning i samfunnet som følge av faktorer som teknologi og automatisering på arbeidsplassen, globalisering og økt internasjonal konkurranse, samt endringer i samfunnsstruktur og behov. Som et resultat av dette er arbeidsgivere villige til å tilby høyere lønn til personer med høyere utdanning under forutsetning for at etterspørselen øker



mer enn tilbudet, da de har nødvendige ferdigheter og kunnskaper for å konkurrere i dagens og fremtidige arbeidsmarked. Det er derfor relevant å undersøke effekten av ulike utdanningsnivåer, spesielt mellom de med utdanning under mastergrad og de med utdanning over mastergrad, da disse gruppene kan ha ulike muligheter for lønnsforfremmelse.



Figur 14 Utvikling i gjennomsnittslønn fordelt på utdanningsnivå.

Tall er hentet fra [microdata.no](https://microdata.no) (microdata.no, 2023). Figuren ble opprettet av oss ved hjelp av Excel.

Figur 14 viser at personer med en mastergrad eller doktorgrad oppnår en mer positiv effekt på lønnen, sammenlignet med de med lavere utdanningsnivå. Dette funnet indikerer en positiv sammenheng mellom høyere utdanning og lønnsnivå, og støtter dermed hypotesen om at høyere utdanning er assosiert med økt lønnsnivå.

## 6. Regresjonsanalyse med drøfting

I kapittel 3 ble det presentert teorien bak multippel regresjon. Regresjonsmodellene i denne oppgaven brukes til å analysere i hvilken grad kjønn, alder og utdanning påvirker lønn.

Dette forutsetter at vi i utgangspunktet mistenker at det kan være en forbindelse mellom de variablene vi studerer. Responsvariabelen,  $Y$ , og forklaringsvariablene,  $X$ , må ha en form for relasjon. Målet med oppgaven er å undersøke om denne sammenhengen virkelig eksisterer.

Analysen baserer seg på  $n$  observasjoner, og den skal se om det er en sammenheng mellom lønn og henholdsvis kjønn, alder og utdanning (Løvås, 2018). Utgangspunktet er en mann med høyere utdanning i alderen 25-67 år. Analyse baserer seg på  $n$  observasjoner, og den skal se om det er en sammenheng mellom lønn og henholdsvis kjønn, alder og utdanning (Løvås, 2018). Utgangspunktet er en mann med høyere utdanning i alderen 25-67 år.

Analysen starter med å se på hver variabel individuelt for å undersøke hvor mye de ulike variablene påvirker lønnen. Til slutt utføres en multippel regresjon for å vurdere den samlede effekten av variablene.

Hensikten med å utføre individuelle regresjoner som undersøker sammenhengen mellom lønn og hver enkelt av de uavhengige variablene er å vurdere hvilken effekt kjønn, alder og utdanning har på den avhengige variabelen uavhengig av hverandre. Ved å holde de andre uavhengige variablene konstante mens man undersøker effekten for hver enkel variabel, kan man bedre forstå hvilke variabler som har størst innvirkning på gjennomsnittlig lønnsnivå.

Når det deretter utføres en multippel regresjonsanalyse med alle de uavhengige variablene inkludert, kan man evaluere den totale effekten av alle variablene i samspill med hverandre og hvilken effekt de har på lønnen.

Som angitt i kapittel 4.1, har det blitt utført en transformasjon av den avhengige variabelen, lønn, ved å ta den naturlige logaritmen av lønn.

## 6.1 Forutsetninger ved Regresjonsanalyse

For å sikre pålitelighet og validitet av resultatene ved utførelse av en regresjonsanalyse, er det nødvendig å ta hensyn til visse forutsetninger. Disse inkluderer: lineær sammenheng, uavhengige observasjoner, normalfordeling, multikollinearitet og store datasett (Løvås, 2018). Å ta hensyn til disse forutsetningene vil bidra til å sikre at resultatene av regresjonsanalysen er pålitelige og kan generaliseres til hele populasjonen. Det er derfor viktig å nøye vurdere disse forutsetningene før man utfører en regresjonsanalyse og ta de nødvendige tiltakene for å adressere eventuelle brudd på disse forutsetningene.

### 6.1.1 Lineær sammenheng

En nødvendig forutsetning for å utføre en regresjonsanalyse er at det eksisterer en lineær sammenheng mellom den avhengige variabelen og de uavhengige variablene (Løvås, 2018). Denne lineære sammenhengen innebærer at en endring i de uavhengige variablene skal være proporsjonal med endringen i den avhengige variabelen. Med andre ord antas det at det finnes en konstant rate av endring i den avhengige variabelen per enhet av endring i hver av de uavhengige variablene, som kan representere ulike faktorer som kan påvirke den avhengige variabelen. Slik linearitet muliggjør bruk av lineære regresjonsmodeller for å utføre prediksjon og forståelse av den avhengige variabelens verdier basert på verdiene av de uavhengige variablene. Det er derfor viktig å undersøke og validere denne forutsetningen før regresjonsanalysen utføres for å sikre nøyaktige og pålitelige resultater. En enkel måte å undersøke den lineære sammenhengen på kan være å se om det er en lineær sammenheng når det plottes data inn i et diagram.

I akademisk sammenheng har det tradisjonelt blitt postulert at lønn øker i en lineær forbindelse med alderen, hvor erfaring og ansiennitet anses som primære faktorer. Imidlertid blir det i kapittel 5.2 forklart hvordan forholdet mellom alder og lønn ikke nødvendigvis følger en lineær trend. Dette framkommer klart i figur 11, der det observeres knekkpunkter i grafen ved ulike aldersgrupper. Dette fenomenet representerer en betydelig forutsetning som må tas i betraktning ved tolkningen av resultatene fra regresjonsanalysen.

### 6.1.2 Uavhengige observasjoner

I statistisk analyse er det en forutsetning at observasjonene er uavhengige av hverandre, noe som betyr at resultatet av én observasjon ikke påvirker resultatene av andre observasjoner i datasettet (Foldnes et al., 2018). Uavhengighet refererer til tilfeller der

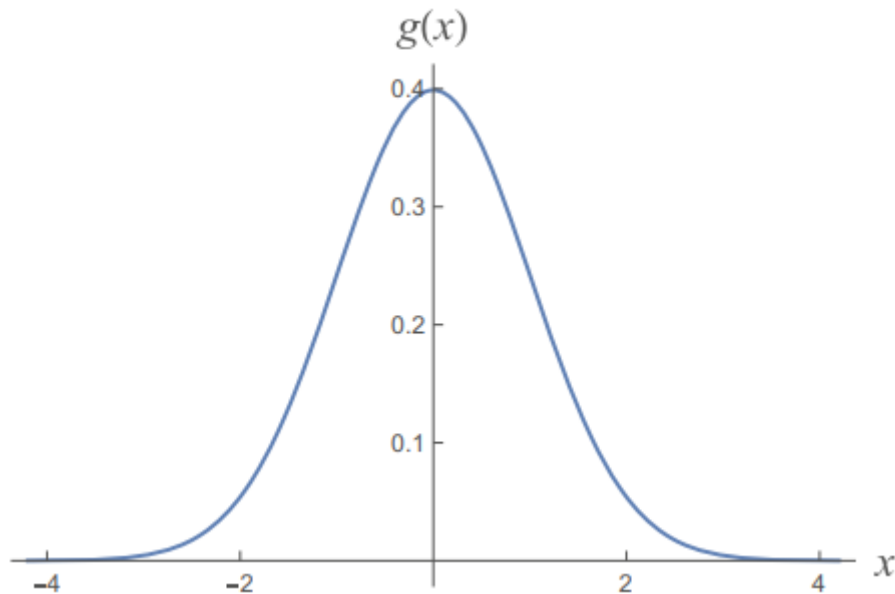
kunnskap om eller kjennskap til én hendelse, A eller B, ikke har noen innvirkning på sannsynligheten for den andre hendelsen (Foldnes et al., 2018). Dette konseptet kan uttrykkes matematisk som  $P(A|B) = P(A)$ , der  $P(A|B)$  er sannsynligheten for hendelse A gitt hendelse B, og  $P(A)$  er sannsynligheten for hendelse A uavhengig av hendelse B (Foldnes et al., 2018).

Ved å studere effekten av variabler som kjønn, utdanningsnivå eller alder på lønn, vil hver person som deltar i studien representere en uavhengig observasjon. Resultatet av lønnen til én person vil ikke påvirke resultatet av lønnen til en annen person.

Det er av betydning å ha uavhengige observasjoner i statistisk analyse, fordi det gjør det mulig å anvende statistiske metoder som forutsetter uavhengighet, for eksempel t-tester som benyttes i denne analysen. Hvis observasjonene ikke er uavhengige, kan det føre til feilaktige konklusjoner når vi bruker denne metoden.

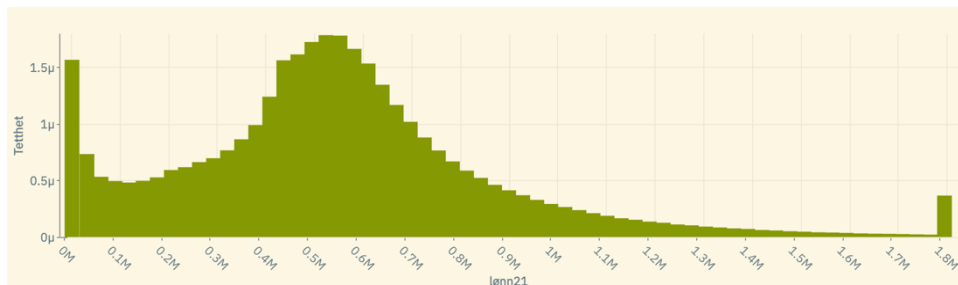
### 6.1.3 Normalfordeling

Dataene i en regresjonsanalyse bør være normalfordelt grunnet at mange statistiske metoder, inkludert regresjonsanalyse, baserer seg på antagelse om normalfordeling (Løvås, 2018). Normalfordeling karakteriseres av at den har en symmetrisk form, hvor de fleste verdiene ligger nær gjennomsnittet, mens dataens fordeling følger en bestemt matematisk formel (Foldnes et al., 2018). Det kan føre til feilaktige konklusjoner dersom dataene i regresjonsanalysen ikke er normalfordelt. Eksempelvis at den har en mer eller mindre prediktiv kraft enn det som er reelt. I tillegg kan det føre til at de estimerte koeffisientene i modellen er feil og ikke kan tolkes som representativ for det faktiske forholdet mellom variablene. Det er derfor viktig å undersøke om dataene er normalfordelt for å kunne bruke regresjonsanalyse på en pålitelig måte. I tilfeller hvor det er avvik fra normalfordelingen, kan det være nødvendig å benytte statistiske metoder som tar høyde for dette.



Figur 15 Normalfordeling (Ubøe, 2017, s.125)

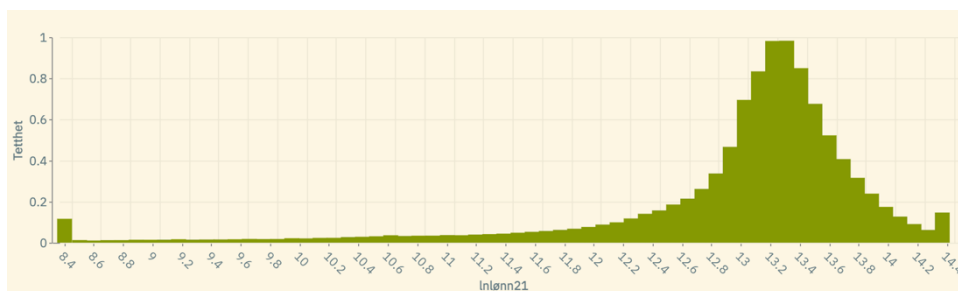
Som nevnt i avsnitt 6.1 så er en av forutsetningene for en regresjonsanalyse at datamaterialet er normalfordelt. Det ble konstruert et histogram for lønnsvariabelen med det formål å evaluere normalfordelingsantagelsen.



Figur 16 Fordeling av lønnsvariabelen før logaritmisk transformasjon

Tall er hentet fra [microdata.no](http://microdata.no) (microdata.no, 2023). Figuren ble opprettet av oss ved hjelp av [microdata](http://microdata.no).

Basert på figur 16 og informasjonen innledningsvis om normalfordeling kan det konkluderes med at lønnsvariabelen ikke er normalfordelt. Det ble utført en logaritmisk transformasjon av lønnsverdiene til den naturlige logaritmen av lønn for at analysen skal vær mer pålitelig og sammenlignbar, slik at det blir lettere å tolke effekten av de uavhengige variablene på lønnsvariabelen.



Figur 17 Fordeling av lønnsvariabelen etter logaritmisk transformasjon

Tall er hentet fra *microdata.no* (*microdata.no*, 2023). Figuren ble opprettet av oss ved hjelp av *microdata*.

Basert på figur 17 kan det observeres at lønnsvariabelen fortsatt har en viss skjevhet i fordelingen selv etter logaritmisk transformasjon, men den viser en mer tilnærmet normalfordeling sammenlignet med den opprinnelige lønnsvariabelen før logaritmisk transformasjon ble foretatt.

#### 6.1.4 Multikollinearitet

Multikollinearitet er en vanlig utfordring i statistisk analyse, der to eller flere forklaringsvariabler i en modell er sterkt innbyrdes korrelert (Løvås, 2018). Dette kan føre til at variablene i modellen i stor grad kan forklare det samme fenomenet, og det kan være unødvendig eller til og med ugunstig å inkludere begge variablene i modellen.

Multikollinearitet kan påvirke evnen til å trekke konklusjoner fra modellen, siden det kan være vanskelig å identifisere hvilken av variablene som faktisk har en betydelig effekt på responsvariabelen. Det kan også føre til overestimering eller underestimering av koeffisientene, og dermed føre til feilaktige estimater og feilslutninger.

En mulig løsning på multikollinearitetsproblemet er å fjerne en av de korrelerte forklaringsvariablene fra modellen. Imidlertid kan det være utfordrende å bestemme hvilken variabel som skal fjernes, og det kan avhenge av formålet med modellen og hvilke variabler som er mest relevante for forskningen.

I sum kan multikollinearitet være en utfordring i statistisk analyse som kan føre til unøyaktige estimater og feilaktige konklusjoner. Derfor er det viktig å være oppmerksom på dette fenomenet og anvende teknikker som kan bidra til å håndtere og minimere effekten av multikollinearitet i modellene våre (Løvås, 2018).

På microdata.no, som benyttes for datainnsamling og databehandling i denne studien, er det mulig å utføre testing av variablene for multikollinearitet ved hjelp av beregning av variansinflasjonsfaktor (VIF) og toleranseverdi. VIF beregner variasjonen i statistikk grunnet inter- korrelasjoner i datasettet. Jo høyere VIF, desto mindre pålitelig er statistikken på grunn av økt kollinearitet. VIF på 1 betyr ingen kollinearitet, mens mellom 1 og 5 indikerer moderat kollinearitet og over 5 betyr høy grad av kollinearitet (The Investopedia team, 2023).

<b>Variance inflation factor (VIF)</b>	
<b>Variabler</b>	<b>VIF - Multikollinearitet</b>
man_first	1.00093
alder	1.003131
høy_utd	1.004003
Gj.snitt	1.002688

Tabell 8 Resultater fra VIF-test for multikollinearitet mellom forklaringsvariablene

Tall er hentet fra microdata.no (microdata.no, 2023). Tabellen ble opprettet av oss ved hjelp av Excel.

Tabell 8 viser en lav grad av multikollinearitet mellom variablene i regresjonsanalysen. Altså, våre forklaringsvariabler er ikke sterkt innbyrdes korrelerte.

## 6.2 Hvor stor innvirkning har kjønn på lønn

<b>Antall Obs:</b> 2488363						
<b>F(1, 2488365):</b> 34898,602761						
<b>R<sup>2</sup>:</b> 0,013831						
<b>Justert R<sup>2</sup>:</b> 0,01383						
<b>Root MSE:</b> 1,084978						
<b>lnlønn21</b>	<b>Coef.</b>	<b>Std.feil</b>	<b>t</b>	<b>P&gt; t </b>	<b>[95% Konf.</b>	<b>intervall]</b>
man_first	0,257179	0,001377	186,811677	0	0,254481	0,259878
Konst	12,796363	0,000992	12893,21268	0	12,794418	12,798309

Tabell 9 Viser i hvilken grad lønn påvirkes av kjønn

Tall er hentet fra microdata.no (microdata.no, 2023). Tabellen er opprettet av oss ved hjelp av Microdata og Excel.

$$\lnlønn21 = \alpha + \beta_1 \text{mann\_first}$$

I tabell 9 undersøkes sammenhengen mellom gjennomsnittlig lønn og kjønn ved å bruke en dummyvariabel for kjønn (1 representerer mann og 0 representerer kvinne). Tidligere forskning indikerer at det vanligvis eksisterer et lønnsgap mellom kvinner og menn, der menn ofte tjener mer enn kvinner for samme type arbeid (Statistisk Sentralbyrå, 2023). Barnefødsler har en negativ innvirkning på kvinners lønnsnivå, og flere studier har dokumentert at denne lønnsulikheten ofte vedvarer over lang tid. Dette fører til at kvinner opplever en ufordelaktig posisjon i arbeidsmarkedet etter at de har fått barn (Kleven et al., 2019). I tillegg så påpeker «The College Payoff» at kjønn spiller en betydelig rolle når det gjelder lønn (Carnevale et al., 2011).

Det kommer fram i regresjonsanalysen at koeffisienten til "man\_first" er omtrent 0,2572. Dette indikerer at det finnes en positiv sammenheng mellom å være mann og lønnen man får. Gjennomsnittslønnen er  $((e^{0,2572}) - 1) * 100 = 29,33\%$  høyere for menn enn for kvinner i den aktuelle datamengden. Bruk av uttrykket  $((e^{\beta}) - 1) * 100$  kommer av at koeffisienten er basert på en naturlig logaritme. Formelen brukes til å omgjøre verdien til en prosentvis endring i responsvariabelen (i dette tilfellet lønn) for en økning i den uavhengige variabelen (i dette tilfellet kjønn) med en enhet. Altså, prosentvis endring i lønn når man beveger seg fra kvinner (0 på kjønnsvariabelen) til menn (1 på kjønnsvariabelen). Det bør bemerkes at denne konklusjonen er begrenset til den spesifikke datamengden, og det er ikke nødvendigvis sikkert at resultatene kan generaliseres til hele populasjonen.

R<sup>2</sup>-verdien (0,01383) antyder at kun 1,383% av variasjonene i lønn kan forklares av kjønn alene. Andre faktorer som ikke er inkludert i modellen kan derfor ha en betydelig innvirkning på lønnen. Det er derfor viktig å evaluere om koeffisienten for "man\_first" er statistisk signifikant og ikke tilfeldig.

For å evaluere om koeffisienten er signifikant forskjellig fra null, er det utført en t-test. Konfidensintervallet for koeffisienten, som er fra 0,254481 - 0,259878 med 95% sikkerhet, indikerer at den sanne verdien av koeffisienten vil ligge innenfor dette intervallet. Med andre ord, kan det med 95% sikkerhet sies at gjennomsnittslønnen for menn er fra 28,98% til 29,68% høyere enn gjennomsnittslønnen for kvinner.



T-testen (186,811677) og en p-verdi tilnærmet null, indikerer at det er svært lite sannsynlig at forskjellen i gjennomsnittslønn mellom kjønnene skyldes tilfeldigheter. Forskjellene er derfor statistisk signifikante.

I og med at det foreløpig bare er tatt hensyn til en variabel (kjønn) kan resultatene så langt muligens tyde på kjønnsdiskriminering mellom kvinner og menn, noen som teorien om barnestrafen også støtter opp om. Imidlertid har den avhengige variabelen, lønn, flere forklaringsfaktorer enn bare kjønn alene, derfor er en utvidelse av modellen verdt å undersøke.

### 6.3 Hvor stor innvirkning har alder på lønn

<b>Antall Obs:</b> 2488363						
<b>F(1, 2488365):</b> 2212,665292						
<b>R<sup>2</sup>:</b> 0,000888						
<b>Justert R<sup>2</sup>:</b> 0,000888						
<b>Root MSE:</b> 1,092074						
<b>lnlønn21</b>	<b>Coef.</b>	<b>Std.feil</b>	<b>t</b>	<b>P&gt; t </b>	<b>[95% Konf.</b>	<b>intervall]</b>
<b>alder</b>	0,002762	0,000059	47,038976	0	0,002646	0,002877
<b>Konst</b>	12,807503	0,002695	4751,924691	0	12,80222	12,812786

Tabell 10 Viser i hvilken grad lønn påvirkes av alder

Tall er hentet fra *microdata.no* (*microdata.no*, 2023). Tabellen er opprettet av oss ved hjelp av *Microdata* og *Excel*.

$$\ln\text{lønn}21 = \alpha + \beta_2 \text{alder}$$

Tabell 10 illustrerer hvordan alder påvirker lønnen til en arbeidstaker. Generelt sett kan det antas at eldre arbeidstakere har lengre arbeidserfaring og kompetanse, og dermed en høyere lønn. Statistisk Sentralbyrå peker også på sammenhengen mellom alder og lønn (Askvik, 2020).

Imidlertid kan den reelle utviklingen av lønn over tid være mer kompleks. Dette er belyst og illustrert i kapittel 5.2, der det fremheves ulike faktorer som bidrar til å anerkjenne at sammenhengen mellom alder og lønn ikke nødvendigvis er lineær. Det ikke-lineære forhold mellom alder og lønn må tas hensyn til ved tolkning av resultatene fra regresjonsanalysen da en lineær regresjon i dette tilfellet kan undervurdere kompleksiteten i forholdet mellom alder og lønn.

Dersom koeffisienten for alder var et negativt tall ville det ha betydd at desto eldre en person blir, desto lavere kan lønnen forventes å være. I denne regresjonsanalyse er

koeffisienten for alder  $((e^{0,002762}) - 1) * 100 = 0,2766\%$  noe som betyr at det er en positiv sammenheng mellom alder og lønn. For hver økning i alder med ett år (i intervallet 25-67år) forventes lønnen å øke med gjennomsnittlig 0,2766%.

En t-test ble utført for å evaluere om koeffisienten var signifikant forskjellig fra null. Konfidensintervallet for koeffisienten ble beregnet med en sikkerhet på 95% og viste en verdi på mellom 0,002646 – 0,002877. Resultatene indikerer at den sanne verdien av koeffisienten mest sannsynlig vil være innenfor dette intervallet. Med 95% sikkerhet kan man derfor konkludere med at gjennomsnittslønnen øker et sted mellom 0,2650% – 0,2881% for hvert år økningen i alder.

P- verdien i analysen er tilnærmet lik 0, noe som betyr at den observerte effekten av variablene i modellen mest sannsynlig ikke er et resultat av tilfeldigheter. Det er altså en veldig sterk indikasjon på at det er en signifikant og reell sammenheng mellom lønn og alder. Men man må være oppmerksom på at p-verdien alene ikke er tilstrekkelig til å fastslå kausalitet mellom variablene, og andre analytiske verktøy bør brukes for å bekrefte funnene.

$R^2$ -verdien i analysen er 0,000888, noe som tilsier at kun 0,0888% av variasjonen i den avhengige variabelen kan forklares av variasjonen i uavhengig variabel. Med andre ord så forklarer alder i seg selv ikke nok av lønnsnivået til en person. Det kan være andre viktigere faktorer som påvirker lønnen som ikke er inkludert i denne modellen. Det er dermed viktig å evaluere  $R^2$ -verdien sammen med andre statistiske resultater for å kunne gi en helhetlig vurdering av resultatene i regresjonsanalysen.

Basert på resultatene av p-verdien og  $R^2$  kan man konkludere at det er en statistisk signifikant, men svak sammenheng mellom lønn og alder. I og med at det kun er 0,0888% av variasjonen i lønn som kan forklares av variasjonen i alder kan det være flere faktorer som kan spille inn på lønn som ikke er inkludert i modellen.

## 6.4 Hvor stor innvirkning utdanning på lønn

<b>Antall Obs:</b> 2488363						
<b>F(1, 2488365):</b> 69774,490233						
<b>R<sup>2</sup>:</b> 0,027275						
<b>Justert R<sup>2</sup>:</b> 0,027275						
<b>Root MSE:</b> 1,077556						
<b>lnlønn21</b>	<b>Coef.</b>	<b>Std.feil</b>	<b>t</b>	<b>P&gt; t </b>	<b>[95% Konf.</b>	<b>intervall]</b>
<b>høy_utd</b>	0,510898	0,001934	264,148614	0	0,507107	0,514688
<b>Konst</b>	12,855401	0,000739	17390,52949	0	12,853952	12,85685

Tabell 11 Viser i hvilken grad lønn påvirkes av utdanning

Tall er hentet fra microdata.no (microdata.no, 2023). Tabellen er opprettet av oss ved hjelp av Microdata og Excel.

$$\ln\text{løn}n21 = \alpha + \beta_3 \text{høy\_utd}$$

Tabell 11 illustrerer hvordan høyere utdanning påvirker lønnsnivået til en arbeidstaker. Høyere utdanning har generelt sett en positiv påvirkning på lønn. Har man høyere utdanning kan det antas at man har mer spesialisert kunnskap og ferdigheter som er etterspurt på arbeidsmarkedet, og dermed større muligheter til å få høyere betalte jobber. Det kan likevel være individuelle forskjeller, samt det kan være avhengig av yrke og bransje. Er man høyt utdannet i en bransje med stor tilgang på arbeidskraft kan man oppleve å tjene mindre enn en person med lavere utdanning i en bransje som har høyere etterspørsel etter arbeidskraft. Den positive relasjonen mellom lønn og utdanning er understøttet av rapporten «The College Payoff» og humankapitalteorien som begge viser til at generelt sett vil en høyere utdanning gi høyere lønn (Becker, u.å.; Carnevale et al., 2011)

For å kunne undersøke sammenhengen mellom høy utdanning og lønn brukes en dummyvariabel, kalt høy\_utd, hvor 1 representerer høyere utdanning på master- eller doktorgradsnivå, og 0 representerer alle som ikke har master- eller doktorgradsutdanning.

Resultatene fra regresjonsanalysen indikerer at høyere utdanning (1) har en signifikant og positiv påvirkning på lønnsnivået. Koeffisienten for høy utdanning er 0,510898 noe som betyr at de med høyere utdanning har en lønn som er  $((e^{0,510898}) - 1) * 100 = 66,68\%$  høyere sammenlignet med de som ikke har høyere utdanning (0). Dette antyder at det er en direkte sammenheng mellom høyere utdanning og økt lønnsnivå.

For å vurdere signifikansen av koeffisienten har en t-test blitt gjennomført.

Konfidensintervallet for koeffisienten, som strekker seg fra 0,507107 til 0,514688 betyr at med en 95% sikkerhet vil den faktiske verdien av koeffisienten falle innenfor dette området.

P- verdien som ble observert var tilnærmet lik 0, noe som indikerer at effekten av variablene i modellen mest sannsynlig ikke skyldes tilfeldigheter. Dette gir en sterk indikasjon på at det eksisterer en signifikant og reell sammenheng mellom lønn og høy utdanning, noe som øker troverdigheten til resultatene og konklusjonen. Også her er det viktig å merke seg at p-verdien ikke alene er tilstrekkelige til å fastslå årsakssammenheng mellom variabelen; her lønn og høy utdanning, og andre analytiske verktøy bør anvendes for å validere funnene.

R<sup>2</sup> - verdien i analysen er 0,027275, noe som tilsier at omtrent 2,7% av variasjonen i den avhengige variabelen, lønn, kan forklares av variasjonen i uavhengig variabel, høy utdanning. Med andre ord så er den statistiske modellen ikke i stand til å forklare mer enn 2,7% av variasjonen i den avhengige i lønn. Dette betyr ikke nødvendigvis at modellen er irrelevant eller dårlig, men indikerer at andre faktorer kan spille en betydelig rolle i å forklare variasjonen i lønn.

Med bakgrunn i resultatene fra p-verdien og R<sup>2</sup>-verdien kan vi konkludere med at det er en statistisk signifikant, men likevel svak sammenheng mellom høy utdanning og lønn når det gjelder å forklare variasjonen i lønn.

## 6.5 Multipl regressjon med alle variablene

Antall Obs: 2488363						
F(3, 2488363): 37991,47982						
R <sup>2</sup> : 0,043797						
Justert R <sup>2</sup> : 0,043796						
Root MSE: 1,068367						
Inlønn21	Coef.	Std.feil	t	P> t	[95% Konf.	intervall]
man_first	0,267915	0,001356	197,543894	0	0,265257	0,270573
alder	0,003572	0,000058	62,099395	0	0,003459	0,003685
hoy_utd	0,528958	0,001921	275,288243	0	0,525192	0,532724
Konst	12,555024	0,002761	4547,331894	0	12,549613	12,560435

Tabell 12 Viser resultatene fra en multipl regressjonsanalyse

Tall er hentet fra [microdata.no](http://microdata.no) (microdata.no, 2023). Tabellen er opprettet av oss ved hjelp av Microdata og Excel.

Når det nå skal undersøkes sammenhengen mellom den avhengige variabelen (lønn) og alle de uavhengige variablene (kjønn, alder og utdanning) sammen, gjennomføres en multippel regresjonsanalyse. Formålet med analysen er som tidligere nevnt å evaluere den totale effekten av alle variablene i samspill med hverandre og undersøke hvilken effekt variablene har på lønnen. Resultatet av regresjonsanalysen vises av tabell 12.

#### Koeffisientene

Ved utførelse av en multippel regresjonsanalyse endres koeffisienten for kjønn (representert ved mann) i modellen seg til 0,267915. Dette indikerer at menn har en gjennomsnittslønn som er  $((e^{0,267915}) - 1) * 100 = 30,72\%$  høyere enn hva gjennomsnittslønnen er for kvinner når både alder og utdanningsnivå blir tatt i betraktning i datasettet. Dette representerer en økning fra 29,33% (med kjønn som eneste variabel) til 30,72% når alder og utdanningsnivå legges til i modellen. En mulig forklaring på dette funnet kan være at menn i større grad oppnår økt avkastning på høyere utdanning enn hva kvinner gjør, eller at menn i større grad har lønninger som øker i takt med deres alder.

For koeffisienten til alder viser den nå en verdi på 0,003572. Dette betyr at gjennomsnittslønnen vil øke med 0,3572% for hver økning på ett år i alder når både kjønn og utdanningsnivå blir tatt i betraktning. Denne koeffisienten har økt fra 0,2762% (som eneste variable) til 0,357% når kontrollvariablene legges til. Resultatet indikerer at alder har en positiv sammenheng med lønn når andre faktorer tas hensyn til, samvariasjon mellom alder og de andre kontrollvariablene.

Til slutt viser koeffisienten for høy utdanning en verdi på 0,528958 når kontrollvariablene kjønn og alder inkluderes. Dette betyr at når kjønn og alder tas i betraktning, øker betydningen av høyere utdanning som en prediktor for lønn. Gjennomsnittslønnen vil være omtrent  $((e^{0,528958}) - 1) * 100 = 69,72\%$  høyere for personer med høyere utdanning (master eller doktorgradsnivå), enn for personer med ingen, uoppgitt eller utdanning under masternivå i denne datamengden. Effekten av høyere utdanning når utdanning var den eneste variabelen var 66,68%. Altså 66,68% høyere gjennomsnittslønn for personer med høyere utdanning sammenlignet med personer uten høyere utdanning. At variablene høyere utdanning øker til 69,72% indikerer at både kjønn og alder har en signifikant effekt på

sammenhengen mellom lønn og høy utdanning. Men andre ord, effekten av utdanning på lønn avhenger av kjønn og alder.

#### Modelltilpasning

Man ser at det er en positiv endring i  $R^2$  når variablene først kjøres separat som eneste prediktor for lønn, og når alle variablene blir kjørt sammen.

I dette tilfellet ser vi at når variablene blir utført separat, bidrar hver av dem med en liten del til å forklare variasjonen i lønn, som indikeres av de lave  $R^2$ -verdiene. Når regresjonen blir utført med alle variablene i sammen modell, øker  $R^2$ -verdien betydelig, noe som indikerer at kombinasjonen av variablene gir mer forklaring på variasjonen i lønn enn hver enkelt variabel gjør på egenhånd.

En økning i  $R^2$ -verdi til 0,056 betyr at omtrent 5,6% av variasjonen i lønnen kan forklares av kombinasjonen av de tre uavhengige variablene som predikterer. Modellen er i stand til å forklare en liten, men likevel signifikant andel av variasjonene i lønnen.

Resultatene viser en P-verdi som er tilnærmet null, noe som indikerer en signifikant sammenheng mellom variablene i modellen og lønnen. Med et konfidensintervall på 95%, er det lite sannsynlig at koeffisientene er tilfeldige og at det ikke finnes en reell sammenheng mellom variablene. Det bør likevel bemerkes at selv om koeffisientene er signifikante, er det ikke nødvendigvis slik at modellen er en god forklaring på den avhengige variabelen lønn. Det kan eksistere andre uavhengige variabler som kan påvirker lønnen, men som ikke er inkludert i det aktuelle datasettet som blir analyseres. Disse variablene er tidligere gjennomgått i kapittel 2.2.4 og inkluderer faktorer som arbeidserfaring, geografisk plassering, yrkesvalg, sektor, bransje, stillingsnivå og etterspørselen i arbeidsmarkedet. Disse variablene kan alle ha en innvirkning på bestemmelsen av et individs lønnsnivå.

## 7. Avslutning/konklusjon

Målet med denne bacheloroppgaven har vært å utføre en grundig analyse og belyse faktorene som kan påvirke lønnsnivået. Spesifikt ble kjønn, alder og utdanning valgt som forklaringsvariabler. Det ble utforsket ulike aspekter ved forhold som påvirker lønn, og flere regresjonsanalyser ble gjennomført for å undersøke om det finnes en sammenheng mellom den avhengige variabelen lønn, og de uavhengige variablene kjønn, alder og utdanning. Det ble etablert en hypotese basert på tidligere forskning om potensielle påvirkninger av de uavhengige variablene på lønn. Imidlertid ble det gjennomført en regresjonsanalyse for å undersøke i hvilken grad disse faktorene faktisk kunne påvirke lønnsnivået. I dette avsluttende kapitlet vil de viktigste funnene bli nøye gjennomgått og oppsummert, og implikasjoner av resultatene for praksis og videre forskning vil bli drøftet.

Tidligere forskning avslører at kvinner ofte tjener mindre enn menn, og at barnefødsler kan ha en negativ innvirkning på kvinners lønn (Kleven et al., 2019; Lekve, 2023).

Regresjonsanalysen viser at det finnes en positiv sammenheng mellom å være mann og lønnen man får. Gjennomsnittslønnen er omtrent 29,33% høyere for menn enn for kvinner i den aktuelle datamengden.  $R^2$  indikere at kun 1,383% av variasjonen i lønn kan forklares av kjønn, og andre faktorer kan derfor ha en betydelig innvirkning på lønnen. T-testen som er gjennomført avslører at koeffisienten for kjønn er signifikant forskjellig fra null.

Konfidensintervallet og p-verdien indikerer at forskjellene i gjennomsnittslønn mellom kjønnene ikke skyldes tilfeldigheter og er derfor signifikante.

Det er nødvendig å understreke at de foreliggende resultatene ikke gir et helhetlig bilde av de underliggende faktorene som kan påvirke resultatene. Først og fremst er den lønnsvariabelen som er brukt i denne analysen ikke spesifisert som heltids-ekvivalent eller ikke. Dette er av avgjørende betydning for resultatene, da 32 prosent av kvinnelige sysselsatte i alderen 20-66 år jobber deltid, sammenlignet med 15 prosent blant menn (Lekve, 2023). Overrepresentasjonen av kvinner i deltidsstillinger kan til dels forklares av studien "Children and Gender Inequality", der resultatene viser at kvinner har 12 prosent høyere sannsynlighet enn menn for å jobbe i offentlig sektor 10 år etter fødselen (Kleven et al., 2019). Dette skyldes at offentlig sektor tilbyr fleksible arbeidstimer, permisjonsdager for syke barn og en positiv holdning til foreldrepermisjon. Det er av betydning å bemerke

at deltidsstillinger er vanlige innenfor offentlig sektor, og at lønnsnivået i denne sektoren er lavere sammenlignet med privat sektor. Videre er det en overrepresentasjon av menn, med 63 prosent menn og 37 prosent kvinner i denne sektoren. Basert på dette kan det konkluderes med at resultatene som viser en gjennomsnittlig lønnsforskjell på omtrent 29,33 prosent mellom menn og kvinner, ikke nødvendigvis er fullt ut representativt. Det kan derfor være verdifullt å utforske en utvidelse av modellen ved å inkludere flere forklaringsfaktorer som tar hensyn til de ulike valgene kjønnene tar med tanke på sektor, yrke og arbeidstimer. Dette vil bidra til å oppnå en mer helhetlig forståelse av sammenhengen mellom lønn og kjønn, og dermed kaste lys over eventuelle andre faktorer som kan bidra til de observerte forskjellene i lønn mellom menn og kvinner.

Når sammenhengen mellom alder og lønnsnivået blir undersøkt, viser resultatene at alder også har en positiv effekt på lønnen, noe som muligens kan forklares av mer arbeidserfaring og kompetanse. Resultatene fra analysen viser at gjennomsnittslønnen øker med 0,2762% for hver økning med ett år i alder. Analysen viser en signifikant, men svak sammenheng mellom alder og lønn, da kun 0,0888% av variasjonen i lønn kan forklares av variasjon i alder.

Det er nødvendig å vise nøye overveielse når man tolker resultatene som indikerer en korrelasjon mellom alder og lønn, ettersom slike funn er basert på en kompleks sammenveving av faktorer, inkludert akkumulert erfaring, ansiennitet og kompetanse som er utviklet over tid i arbeidslivet. Disse faktorene kan utgjøre en solid basis for forhandlinger om økt lønn. Likevel tar ikke variabelen hensyn til perioder med fravær fra arbeidsmarkedet, hvor kompetanse og ansiennitet ikke kan økes, og derfor kan den prosentvise økningen for hvert år stanse. Videre inkluderer ikke variabelen mulig aldersdiskriminering mot eldre arbeidstakere, som kan ha en svekket posisjon når de forhandler om lønnsøkning, da det er en antakelse blant flere norske ledere om at eldre arbeidstakere kan møte utfordringer knyttet til helse, sykefravær, digitale ferdigheter, innovasjonsevner og arbeidsprestasjoner (Hauge & Moen, 2023). De ovennevnte forholdene sår tvil om at det er en lineær sammenheng, noe som understøttes av datamaterialet i kapittel 5.2. Dette medfører begrensninger som ikke kan bli løst i denne analysen ved bruk av konvensjonelle statistiske metoder. For å korrekt estimere dette forholdet, vil det være nødvendig å benytte andre statistiske metoder som er bedre egnet for å håndtere den ikke-lineære natur av dette forholdet.



Analysen viser også at det er en positiv signifikant sammenheng mellom høyere utdanning og lønnsnivå. Regresjonsanalysen indikerer at personer med høyere utdanning har en lønn som er anslagsvis 66,68% høyere enn for de uten høyere utdanning. Denne sammenhengen kan forstås ut fra antagelsen om at utdanning gir en spesialisert kunnskap og ferdigheter som er etterspurt i arbeidsmarkedet, og dermed gir større muligheter for å få høyere betalte jobber. Disse funnene er konsistente med rapporten "The College Payoff" (Carnevale et al., 2011) som konkluderer med at utdanning er en avgjørende faktor for lønnsnivået, og at personer med høyere utdanning generelt tjener mer enn de med lavere utdanning. Videre analyse av konfidensintervallet og p-verdien indikerer at effekten av variabelen i modellen er sannsynligvis ikke tilfeldig. Imidlertid er  $R^2$ -verdien lav (2,7%), noe som tyder på at kun en liten del av variasjonen i lønn kan forklares av variasjonen i utdanningsnivå. Konklusjonen er at det er en statistisk signifikant, men likevel svak sammenheng mellom høyere utdanning og lønn når det gjelder å forklare variasjonen i lønn. I henhold til humankapitalteorien er det også andre individuelle egenskaper ved en arbeidstaker, slik som erfaring og opplæring, som kan forklare deres lønn. Teoretisk sett kan disse egenskapene bidra til å øke en persons produktivitet og dermed deres potensiale for økt inntekt, og dermed kan ikke utdanning alene forklare variasjonen i lønn, som påpekt av Becker (u.å.). I henhold til rapporten fra "The College Payoff" kan valg av yrker med høye inntektsnivåer ha en større innvirkning på inntekten enn den formelle utdannelsen, som illustrert og forklart i kapittel 5.3. (Carnevale et al., 2011). Dette forklares og illustreres i kapittel 5.3.

Selv om resultatene fra regresjonsanalysen indikerer at personer med høyere utdanning har en lønn som er omtrent 66,68% høyere enn de uten høyere utdanning, er det viktig å merke seg at avkastningen på utdanning kan variere betydelig avhengig av faktorer som fagfelt, karrierevei, individuelle egenskaper og evner, som ikke er inkludert som variabler i denne analysen (Carnevale et al., 2011). Dermed kan resultatene gi en begrenset forklaringskraft for den faktiske påvirkningen av høyere utdanning på lønnen.

Det ville derfor være interessant å utføre ytterligere forskning for å undersøke hvordan effekten av høyere utdanning på lønnen varierer innenfor ulike fagfelt, bransjer og forhåndsdefinerte egenskaper som kjønn og alder. Dette kunne bidra til å få en mer detaljert og nyansert forståelse av sammenhengen mellom høyere utdanning og lønn, og gi et grundigere innblikk i de komplekse faktorene som påvirker dette forholdet.

Resultatene fra de individuelle regresjonsanalysene antyder en positiv korrelasjon mellom de forklarende variablene og lønn, men modellene viser en begrenset forklaringskraft, da variablene alene har manglende egenskaper som kan forsterke variasjonen i lønn.

Resultatene fra den multiple regresjonsanalysen viser en økning i effekten av å være mann på lønnsnivået fra 29,33% til 30,72% når alder og utdanning blir inkludert som variabler. Dette indikerer at menn har en gjennomsnittslønn som er 30,72% høyere enn gjennomsnittslønnen for kvinner. Videre øker effekten av alder på lønn fra 0,2766% til 0,3572% når kjønn og høyere utdanning blir tatt i betraktning, noe som tyder på at alder har en økende innvirkning på lønn når man tar hensyn til kjønn og høyere utdanning, men i mindre grad enn de andre variablene. Koeffisienten for høyere utdanning øker også fra 66,68% til 69,72% når man tar hensyn til kjønn og alder. Dette indikerer at betydningen av høyere utdanning som en prediktor for lønn øker når kjønn og alder blir tatt i betraktning. Determinasjonskoeffisienten ( $R^2$ ) for modellen er beregnet til å være 5,6%. Denne indikerer at hver av de forklarende variablene bidrar med en begrenset andel i å forklare variasjonen i lønn når de betraktes hver for seg. Imidlertid øker  $R^2$  når alle variablene blir kombinert i én modell, noe som antyder at kombinasjonen av variablene gir en bedre forklaring på variasjonen i lønn enn hver enkelt variabel for seg selv.

Til tross for eksisterende begrensninger knyttet til datamengden i denne studien, kan de presenterte funnene bidra til en mer dyptgående forståelse av de ulike faktorene som kan påvirke lønnsnivå for arbeidstakere. Konklusjonen av studien indikerer derfor at det er flere komplekse faktorer som spiller en rolle i lønnsnivå, utover de som er inkludert i denne analysen, slik som individuelle valg og personlige egenskaper som overstiger betydningen av kjønn, alder og utdanningsnivå.

## Litteraturliste

Andersen, E. (2022). Økt fruktbarhet for første gang på 12 år. *SSB*.

<https://www.ssb.no/befolkning/fodte-og-dode/statistikk/fodte/artikler/okt-fruktbarhet-for-forste-gang-pa-12-ar>

Arbeidsmiljøloven-aml. (2023). *Lov om arbeidsmiljø, arbeidstid og stillingsvern mv. (Arbeidsmiljøloven)—Kapittel 13. Vern mot diskriminering—Lovdata (LOV-2022-12-20-99)*. Arbeidsmiljøloven. [https://lovdata.no/dokument/NL/lov/2005-06-17-62?fbclid=IwAR32HlaoSTjfnVNZCNDvaJ-iG\\_gQuy8vvgdkEbdJVVgOAgTqoXgATePRmVxQ#KAPITTEL\\_14](https://lovdata.no/dokument/NL/lov/2005-06-17-62?fbclid=IwAR32HlaoSTjfnVNZCNDvaJ-iG_gQuy8vvgdkEbdJVVgOAgTqoXgATePRmVxQ#KAPITTEL_14)

Askvik, T. (2020, 16. november). *Alder og yrke påvirker lønnsgapet*. Statistisk Sentralbyrå.

<https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/alder-og-yrke-pavirker-lonnsgapet>

Bech, J. (2013, 24.juni). 100 år med allmenn stemmerett for kvinner-høyere utdanning. *Lovdata*.

[https://lovdata.no/artikkel/100\\_ar\\_med\\_allmenn\\_stemmerett\\_for\\_kvinner\\_hoyere\\_uttanning/1224?fbclid=IwAR3eP6fhtytBMsZU-mj5j8qEGyCwmCv-cDy8xHRgamCj3zCMjZCEPVM8WA](https://lovdata.no/artikkel/100_ar_med_allmenn_stemmerett_for_kvinner_hoyere_uttanning/1224?fbclid=IwAR3eP6fhtytBMsZU-mj5j8qEGyCwmCv-cDy8xHRgamCj3zCMjZCEPVM8WA)

Becker, G. S. (u.å.). *Human Capital*. Econlib. Hentet 12. april 2023, fra

<https://www.econlib.org/library/Enc/HumanCapital.html>

Bye, K. S. (2018, 21. november). *Mastergrad gir høyere lønn*. Statistisk Sentralbyrå.

<https://www.ssb.no/arbeid-og-lonn/lonn-og-arbeidskraftkostnader/artikler/mastergrad-gir-hoyere-lonn>

Carnevale, A. P., Rose, S. J., & Cheah, B. (2011). *The College Payoff: Education, Occupations, Lifetime Earnings*. Georgetown University Center on Education and the Workforce. <https://cew.georgetown.edu/cew-reports/the-college-payoff/>

- FN. (2023, 3.februar). *Likestilling mellom kjønnene*. FN-Sambandet.  
<https://www.fn.no/om-fn/fns-baerekraftsmaal/likestilling-mellom-kjoennene>
- Foldnes, N., Grønneberg, S., & Hermansen, G. H. (2018). *Statistikk og dataanalyse: En moderne innføring*. Cappelen Damm AS.
- Forbrukerrådet. (u.å.). *Hva er privat AFP?* Finansportalen. Hentet 17. april 2023, fra  
<https://www.finansportalen.no/pensjon/om-privat-afp>
- Grønmo, S. (2023). Kvantitativ metode. I *Store norske leksikon*.  
[https://snl.no/kvantitativ\\_metode](https://snl.no/kvantitativ_metode)
- Hauge, L., & Moen, R. (2023). Aldersdiskriminering i norsk arbeidsliv. *Revisjon og Regnskap. Utgave nr.2*. [https://www.revregn.no/journal/2023/2/m-934/Aldersdiskriminering\\_i\\_norsk\\_arbeidsliv](https://www.revregn.no/journal/2023/2/m-934/Aldersdiskriminering_i_norsk_arbeidsliv)
- Holseter, A. M. R. (2019, 12.september). *Hvordan klassifiseres en persons høyeste utdanningsnivå?* Statistisk Sentralbyrå. <https://www.ssb.no/utdanning/artikler-og-publikasjoner/hvordan-klassifiseres-en-persons-hoyeste-utdanningsniva>
- Idsø, J. (2023). Økonometri. I *Store norske leksikon*. <https://snl.no/%C3%B8konometri>
- Kifinfo. (u.å.). Statistikk om kjønnsbalanse og mangfold. *kifinfo*. Hentet 12. april 2023, fra  
<https://kifinfo.no/nb/content/statistikk-0>
- Kleven, H., Landais, C., & Sjøgaard, J. E. (2019). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4), 181–209. <https://doi.org/10.1257/app.20180010>
- KLP. (u.å.). *Pensjonsalder og ulike aldersgrenser i Norge—KLP.no*. Hentet 17. april 2023, fra  
<https://www.klp.no/pensjon/regler/aldersgrenser?fbclid=IwAR1m2PDi6nIn1wavxT0A4VBwm16YAbou-d77q5kPdeiNUSiXENyjIw90QpA>

- Lekve, M. (2023, 8.mars). *Hvor likestilt er vi?* Statistisk Sentralbyrå.  
<https://www.ssb.no/befolkning/likestilling/artikler/hvor-likestilt-er-vi>
- Løvås, G. G. (2018). *Statistikk for universiteter og høyskoler* (4.utgave).  
Universitetsforlaget AS.
- Microdata. (u.å.). *Om microdata.no*. microdata.no. Hentet 12. april 2023, fra  
<https://www.microdata.no/om-microdata-no/>
- microdata.no. (2023, 12.april). *Microdata.no-registerdata uten å søke*. microdata.no.  
<https://www.microdata.no/>
- NAV (2022, 3.mai). *Alderspensjon. Desember 2021*. NAV.  
[https://www.nav.no/no/nav-og-samfunn/statistikk/pensjon-statistikk/relatert-informasjon/arkiv-alderspensjon.desember-2021?fbclid=IwAR0LtYsBSxgqDyks\\_P8vyVsGo1gUNbgaIKyNGL2Kad7xR\\_LO35jjOXRMa9A](https://www.nav.no/no/nav-og-samfunn/statistikk/pensjon-statistikk/relatert-informasjon/arkiv-alderspensjon.desember-2021?fbclid=IwAR0LtYsBSxgqDyks_P8vyVsGo1gUNbgaIKyNGL2Kad7xR_LO35jjOXRMa9A)
- St.meld. nr. 44 (2012-2013). *Likestilling kommer ikke av seg selv*. Regjeringen.  
<https://www.regjeringen.no/no/dokumenter/meld-st-44-20122013/id731019/?fbclid=IwAR1O9-YrL6-9rTFIY8MtqiBnduB-StPbihRAUpubwipAOjuhw7U1Xwrg-uI>
- Statistisk Sentralbyrå. (u.å.). *Tall som forteller*. Statistisk Sentralbyrå. Hentet 12. april 2023, fra <https://www.ssb.no/omssb/ssbs-virksomhet/tall-som-forteller>
- Statistisk Sentralbyrå. (u.å.). *SSBs virksomhet*. Statistisk Sentralbyrå.  
<https://www.ssb.no/omssb/ssbs-virksomhet?fbclid=IwAR34wXEaYxp0bGXkzY50G9fVTY7WyeTxI79vIzsqaccOBNx5YfEu4FCRERU>
- Statistisk Sentralbyrå. (2020, 9.november). *Framskrivinger av arbeidsstyrken og sysselsettingen etter utdanning mot 2040*. Statistisk Sentralbyrå.  
<https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/framskrivinger-av->

[arbeidsstyrken-og-sysselsettingen-etter-utdanning-mot-2040](#)

Statistisk Sentralbyrå. (2022, 16.juni). *Befolkningens utdanningsnivå*. Statistisk Sentralbyrå.

<https://www.ssb.no/utdanning/utdanningsniva/statistikk/befolkningens-utdanningsniva>

Statistisk Sentralbyrå. (2023, 2.februar). *Lønn*. Statistisk Sentralbyrå.

<https://www.ssb.no/arbeid-og-lonn/lonn-og-arbeidskraftkostnader/statistikk/lonn>

Steinkellner, A. (2010, 1.mai). *Fra den første kvinnelige student*. Statistisk Sentralbyrå.

[https://www.ssb.no/befolkning/artikler-og-publikasjoner/fra-den-forste-kvinnelige-student?fbclid=IwAR0ytOCcjTWPlvYTC8zco87rsA\\_vgYdTfLCg5SwF1L83Mkbwd3rBQiwIQZI](https://www.ssb.no/befolkning/artikler-og-publikasjoner/fra-den-forste-kvinnelige-student?fbclid=IwAR0ytOCcjTWPlvYTC8zco87rsA_vgYdTfLCg5SwF1L83Mkbwd3rBQiwIQZI)

The Investopedia team. (2023, 12.februar). *Variance Inflation Factor (VIF)*. Investopedia.

<https://www.investopedia.com/terms/v/variance-inflation-factor.asp>

Ubøe, J. (2017). *Introductory Statistics for Business and Economics*. Springer International Publishing.

Utdanning.no (u.å). *Kjønnsdelt arbeidsmarked*. Utdanning.no.

[https://utdanning.no/utdanningsvalg\\_artikkel\\_kjonnsdelt\\_arbeidsmarked?fbclid=IwAR0qxb6fs8JOTssl-xDSZ1AGoEgBXeGftDEs0SG34\\_dJqhCEwU5\\_Ebp\\_sM](https://utdanning.no/utdanningsvalg_artikkel_kjonnsdelt_arbeidsmarked?fbclid=IwAR0qxb6fs8JOTssl-xDSZ1AGoEgBXeGftDEs0SG34_dJqhCEwU5_Ebp_sM)

Utenriksdepartementet. (2016, 2.september). *Frihet, makt og muligheter-handlingsplan for kvinners rettigheter og likestilling i utenriks- og utviklingspolitikken 2016-2020 (E-962 B)*. Utenriksdepartementet.

[https://www.regjeringen.no/globalassets/departementene/ud/vedlegg/fn/kvinner\\_handlingsplan.pdf?fbclid=IwAR1km-4hcVAiBU7L7ksBQG6K4gxkoogmpKWSXxBwjTT1ZO7XEzwLPHEbM\\_M](https://www.regjeringen.no/globalassets/departementene/ud/vedlegg/fn/kvinner_handlingsplan.pdf?fbclid=IwAR1km-4hcVAiBU7L7ksBQG6K4gxkoogmpKWSXxBwjTT1ZO7XEzwLPHEbM_M)

## Figurliste

Figur 1 Liten forklaringskraft .....	22
Figur 2 Litt forklaringskraft.....	22
Figur 3 Sterk forklaringskraft .....	23
Figur 4 Standard normalfordeling med konfidensintervaller .....	24
Figur 5 Viser utvikling i gjennomsnittslønn for menn og kvinner fra 2015-2021 .....	32
Figur 6 Viser utvikling i medianlønn for menn og kvinner fra 2015-2021 .....	33
Figur 7 Prosentvis fordeling av kvinner og menn over de ulike utdanningsnivåene .....	34
Figur 8 Gjennomsnittslønn i de ulike yrkesgruppene fordelt på kjønn .....	35
Figur 9 Prosentvise andelen kvinner og menn innenfor ulike yrkesklassifiseringer .....	36
Figur 10 Gjennomsnittslønnen for ulike aldersgrupper fordelt på kjønn .....	37
Figur 11 Gjennomsnittslønn for ulike aldersgrupper fra 25-67 år .....	37
Figur 12 Befolkningens prosentvis fordeling mellom de ulike utdanningsnivåene .....	39
Figur 13 Gjennomsnittslønn fordelt på ulike utdanningsnivåer .....	40
Figur 14 Utvikling i gjennomsnittslønn fordelt på utdanningsnivå.....	41
Figur 15 Normalfordeling (Ubøe, 2017, s.125).....	45
Figur 16 Fordeling av lønnsvariabelen før logaritmisk transformasjon .....	45
Figur 17 Fordeling av lønnsvariabelen etter logaritmisk transformasjon.....	46

## Tabelliste

Tabell 1 Oversikt av dummyvariabelen man_first. ....	28
Tabell 2 Oversikt over ulike utdanningsnivåer og observasjoner .....	29
Tabell 3 Oversikt over dummyvariabel høy_utd.....	29
Tabell 4 Oversikt over gjennomsnittslønn fra 2015 – 2021 .....	30
Tabell 5 Oversikt over gjennomsnittslønn konvertert til naturlig logaritme .....	30
Tabell 6 Oversikt over variablene kjønn, alder og utdanning .....	31
Tabell 7 Viser prosentvis lønnsøkning for kvinner og menn fra 2015 – 2021 .....	33
Tabell 8 Resultater fra VIF-test for multikollinearitet mellom forklaringsvariablene .....	47
Tabell 9 Viser i hvilken grad lønn påvirkes av kjønn.....	47
Tabell 10 Viser i hvilken grad lønn påvirkes av alder.....	49
Tabell 11 Viser i hvilken grad lønn påvirkes av utdanning.....	51
Tabell 12 Viser resultatene fra en multippel regresjonsanalyse .....	52

## Vedlegg 1



» **require no.ssb.fdb:20 as ds**

Opprettet en kobling fra *no.ssb.fdb:20* til *ds*

» **create-dataset lønnsgapet**

Et tomt dataset, *lønnsgapet* ble opprettet og valgt

**lønnsgapet» import ds/BEFOLKNING\_FOEDSELS\_AAR\_MND as befolkning**

Importerte *BEFOLKNING\_FOEDSELS\_AAR\_MND* som *befolkning* til *lønnsgapet* med 11 089 163 enheter

**lønnsgapet» import ds/BEFOLKNING\_STATUSKODE 2021-01-01 as regstatus**

Importerte *BEFOLKNING\_STATUSKODE* på datoen *2021-01-01* som *regstatus* til *lønnsgapet* med 11 089 163 enheter, hvorav 2 486 719 missingverdier

**lønnsgapet» keep if regstatus == "1"**

5 699 208 enheter ble fjernet fra datasettet.

**lønnsgapet» summarize regstatus**

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
regstatus	-	-	5389955	-	-	-	-	-

**lønnsgapet» generate alder = 2021 - int(befolkning /100)**

Genererte *alder*

med 5 389 955 enheter

**lønnsgapet» keep if alder >= 25 & alder <= 67**

2 354 713 enheter ble fjernet fra datasettet.

**lønnsgapet» summarize alder**

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
alder	45.3784	12.1874	3035242	25	35	45	56	67

**lønnsgapet» import ds/BEFOLKNING\_KJOENN as kjønn**

Importerte *BEFOLKNING\_KJOENN* som *kjønn* til *lønnsgapet* med 3 035 242 enheter

**lønnsgapet» tabulate kjønn, cellpct**

<i>kjønn</i>	1 - Mann	51.12
	2 - Kvinne	48.88
<i>Total</i>		100

**lønnsgapet» tabulate kjønn**

<i>kjønn</i>	1 - Mann	1551498
	2 - Kvinne	1483750
<i>Total</i>		3035242

**lønnsgapet» import ds/NUDB\_BU 2021-01-01 as utdanning**

Importerte *NUDB\_BU* på datoen *2021-01-01* som *utdanning* til *lønnsgapet* med 3 035 242 enheter, hvorav 117 209 missingverdier

**lønnsgapet**» summarize **utdanning**

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
utdanning	-	-	2918039	-	-	-	-	-

**lønnsgapet**» import ds/INNTEKT\_WLONN 2015-12-31 as lønn15

Importerte *INNTEKT\_WLONN* på datoen *2015-12-31* som *lønn15* til *lønnsgapet* med 3 035 242 enheter, hvorav 544 668 missingverdier

**lønnsgapet**» import ds/INNTEKT\_WLONN 2016-12-31 as lønn16

Importerte *INNTEKT\_WLONN* på datoen *2016-12-31* som *lønn16* til *lønnsgapet* med 3 035 242 enheter, hvorav 532 993 missingverdier

**lønnsgapet**» import ds/INNTEKT\_WLONN 2017-12-31 as lønn17

Importerte *INNTEKT\_WLONN* på datoen *2017-12-31* som *lønn17* til *lønnsgapet* med 3 035 242 enheter, hvorav 519 256 missingverdier

**lønnsgapet**» import ds/INNTEKT\_WLONN 2018-12-31 as lønn18

Importerte *INNTEKT\_WLONN* på datoen *2018-12-31* som *lønn18* til *lønnsgapet* med 3 035 242 enheter, hvorav 504 523 missingverdier

**lønnsgapet**» import ds/INNTEKT\_WLONN 2019-12-31 as lønn19

Importerte *INNTEKT\_WLONN* på datoen *2019-12-31* som *lønn19* til *lønnsgapet* med 3 035 242 enheter, hvorav 485 323 missingverdier

**lønnsgapet**» import ds/INNTEKT\_WLONN 2020-12-31 as lønn20

Importerte *INNTEKT\_WLONN* på datoen *2020-12-31* som *lønn20* til *lønnsgapet* med 3 035 242 enheter, hvorav 511 434 missingverdier

**lønnsgapet**» generate **lnlønn15 =ln(lønn15)**

Genererte *lnlønn15*  
med 3 035 242 enheter, hvorav 544 669 missingverdier

**lønnsgapet**» summarize **lnlønn15**

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
lnlønn15	12.6168	1.1159	2490572	8.1795	12.3534	12.9589	13.2557	14.2194

**lønnsgapet**» generate **lnlønn16 =ln(lønn16)**

Genererte *lnlønn16*  
med 3 035 242 enheter, hvorav 532 993 missingverdier

**lønnsgapet**» summarize **lnlønn16**

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
lnlønn16	12.6635	1.1016	2502252	8.1605	12.4382	12.9925	13.2808	14.2296

**lønnsgapet**» generate **lnlønn17 =ln(lønn17)**

Genererte *lnlønn17*  
med 3 035 242 enheter, hvorav 519 258 missingverdier

**lønnsgapet**» summarize **lnlønn17**

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
lnlønn17	12.7171	1.0803	2515982	8.2196	12.5193	13.0296	13.3128	14.2544

**lønnsgapet**» generate `lnlønn18 =ln(lønn18)`

Genererte *lnlønn18*

med 3 035 242 enheter, hvorav 504 523 missingverdier

**lønnsgapet**» summarize `lnlønn18`

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
lnlønn18	12.7803	1.0614	2530718	8.2774	12.6106	13.0736	13.3541	14.2943

**lønnsgapet**» generate `lnlønn19 =ln(lønn19)`

Genererte *lnlønn19*

med 3 035 242 enheter, hvorav 485 345 missingverdier

**lønnsgapet**» summarize `lnlønn19`

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
lnlønn19	12.823	1.1037	2549901	7.9153	12.6919	13.1216	13.3987	14.3343

**lønnsgapet**» generate `lnlønn20 =ln(lønn20)`

Genererte *lnlønn20*

med 3 035 242 enheter, hvorav 511 455 missingverdier

**lønnsgapet**» import `ds/INNTEKT_WLONN 2021-12-31 as lønn21`

Importerte *INNTEKT\_WLONN* på datoen *2021-12-31* som *lønn21* til *lønnsgapet* med 3 035 242 enheter, hvorav 546 847 missingverdier

**lønnsgapet**» generate `lnlønn21 =ln(lønn21)`

Genererte *lnlønn21*

med 3 035 242 enheter, hvorav 546 876 missingverdier

**lønnsgapet**» summarize `lønn20`

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
lønn20	530576.2661	321816.8594	2523805	4068	332278	513130	675691	1725846

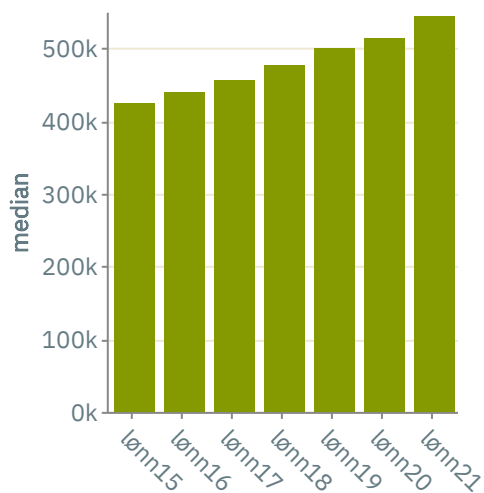
**lønnsgapet**» summarize `lønn21`

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
lønn21	564049.7897	335788.3113	2488391	4269	367741	543064.5	712357.25	1824432

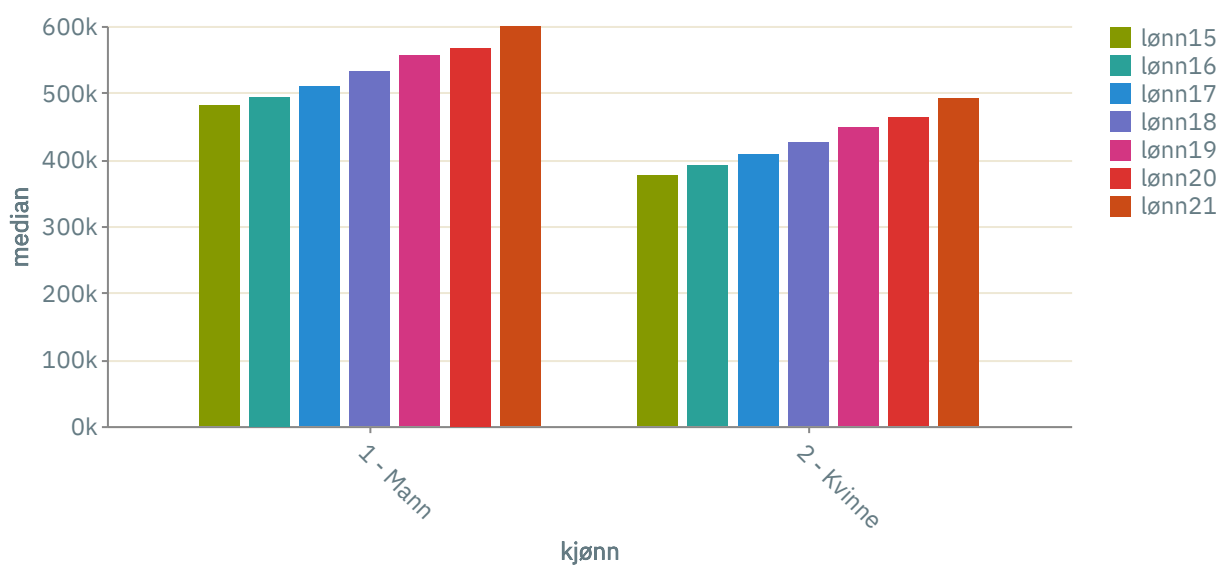
**lønnsgapet**» summarize `lønn15 lønn16 lønn17 lønn18 lønn19 lønn20 lønn21, gini`

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%	Gini
lønn15	435411.8047	286470.914	2490580	3567	231740	424604	571298	1497688	0.3569
lønn16	449728.6797	288158.0075	2502252	3500	252266	439102.5	585855	1513011	0.3477
lønn17	467610.0648	293387.0977	2515985	3713	273562	455690	604909	1550996	0.34
lønn18	491461.4653	302765.8293	2530718	3934	299724	476192	630414.25	1614053	0.333
lønn19	516590.0946	314183.6485	2549916	2734	325083	499615	659147	1679869	0.3282
lønn20	530576.2661	321816.8594	2523805	4068	332278	513130	675691	1725846	0.3272
lønn21	564049.7897	335788.3113	2488391	4269	367741	543064.5	712357.25	1824432	0.3199

lønnsgapet» barchart (median) lønn15 lønn16 lønn17 lønn18 lønn19 lønn20 lønn21



lønnsgapet» barchart (median) lønn15 lønn16 lønn17 lønn18 lønn19 lønn20 lønn21, over(kjønn)



lønnsgapet» tabulate kjønn, summarize(lønn15)

kjønn	1 - Mann	498695.65
	2 - Kvinne	368549.99
Total		435411.8

lønnsgapet» tabulate kjønn, summarize(lønn16)

kjønn	1 - Mann	511389.58
	2 - Kvinne	384265.1
Total		449728.68

**lønnsgapet**» tabulate kjønn, summarize(lønn17)

kjønn	1 - Mann	529372.37
	2 - Kvinne	401688
Total		467610.06

**lønnsgapet**» tabulate kjønn, summarize(lønn18)

kjønn	1 - Mann	555678.16
	2 - Kvinne	422510.62
Total		491461.47

**lønnsgapet**» tabulate kjønn, summarize(lønn19)

kjønn	1 - Mann	583135.18
	2 - Kvinne	444829.36
Total		516590.09

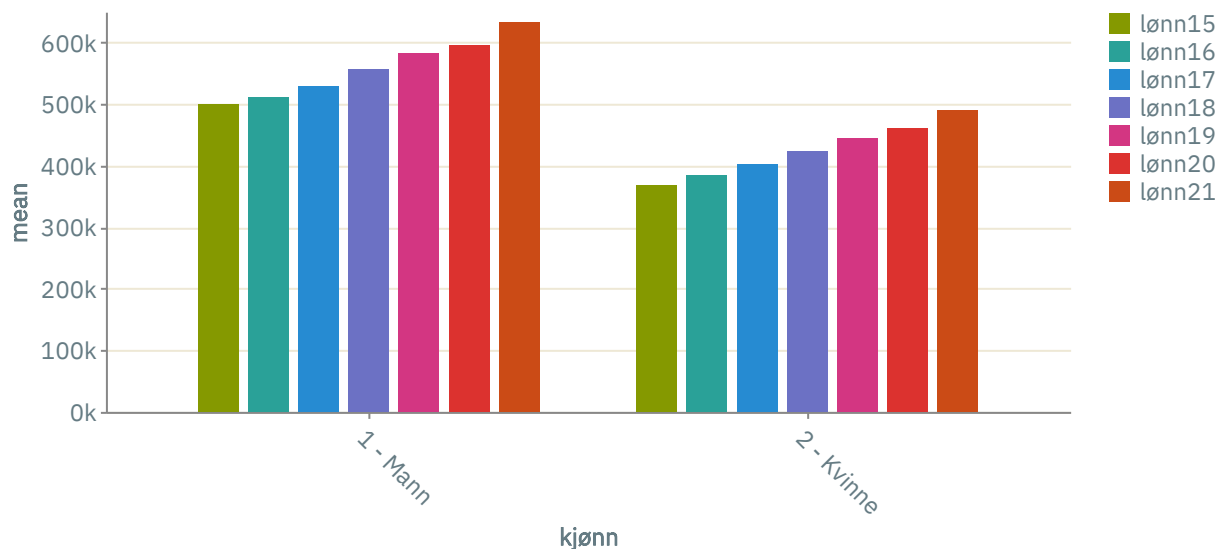
**lønnsgapet**» tabulate kjønn, summarize(lønn20)

kjønn	1 - Mann	595969.1
	2 - Kvinne	459699.91
Total		530576.27

**lønnsgapet**» tabulate kjønn, summarize(lønn21)

kjønn	1 - Mann	632901.76
	2 - Kvinne	489538.4
Total		564049.79

lønnsgapet» barchart (mean) lønn15 lønn16 lønn17 lønn18 lønn19 lønn20 lønn21, over(kjønn)



lønnsgapet» generate utdnivå=substr(utdanning, 1, 1)

Genererte *utdnivå*

med 3 035 242 enheter, hvorav 117 209 missingverdier

lønnsgapet» destring utdnivå

Konverterte *utdnivå* til tallverdier i ny variabel

lønnsgapet» define-labels utdlabel 0"Ingen" 1"Barneskole" 2"Ungdomsskole" 3"Videregående " 4"Videregående avsluttende utdanning" 5"Påbygg til videregående" 6"UH-utdanning - lavere nivå" 7"UH-utdanning - høyere nivå" 8"Doktorgrad" 9"Uoppgitt"

Opprettet kodelisten *utdlabel1* med 10 etiketter

lønnsgapet» assign-labels utdnivå utdlabel

Tilegnet kodelisten *utdlabel1* til variabelen *utdnivå*

lønnsgapet» tabulate utdnivå kjønn, cellpct

utdnivå	kjønn		Total
	1 - Mann	2 - Kvinne	
0 - Ingen	0.09	0.15	0.24
1 - Barneskole	0.28	0.3	0.58
2 - Ungdomsskole	10.14	8.32	18.45
3 - Videregående	2.34	2.57	4.91
4 - Videregående avsluttende utdanning	17.41	11.91	29.33
5 - Påbygg til videregående	2.41	1.33	3.74
6 - UH-utdanning - lavere nivå	11.52	17.64	29.16
7 - UH-utdanning - høyere nivå	5.86	6.43	12.29
8 - Doktorgrad	0.7	0.59	1.29
<b>Total</b>	<b>50.76</b>	<b>49.24</b>	<b>100</b>

**lønnsgapet** » summarize utdnivå

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
utdnivå	4.6063	1.7229	2918039	2	4	4	6	8

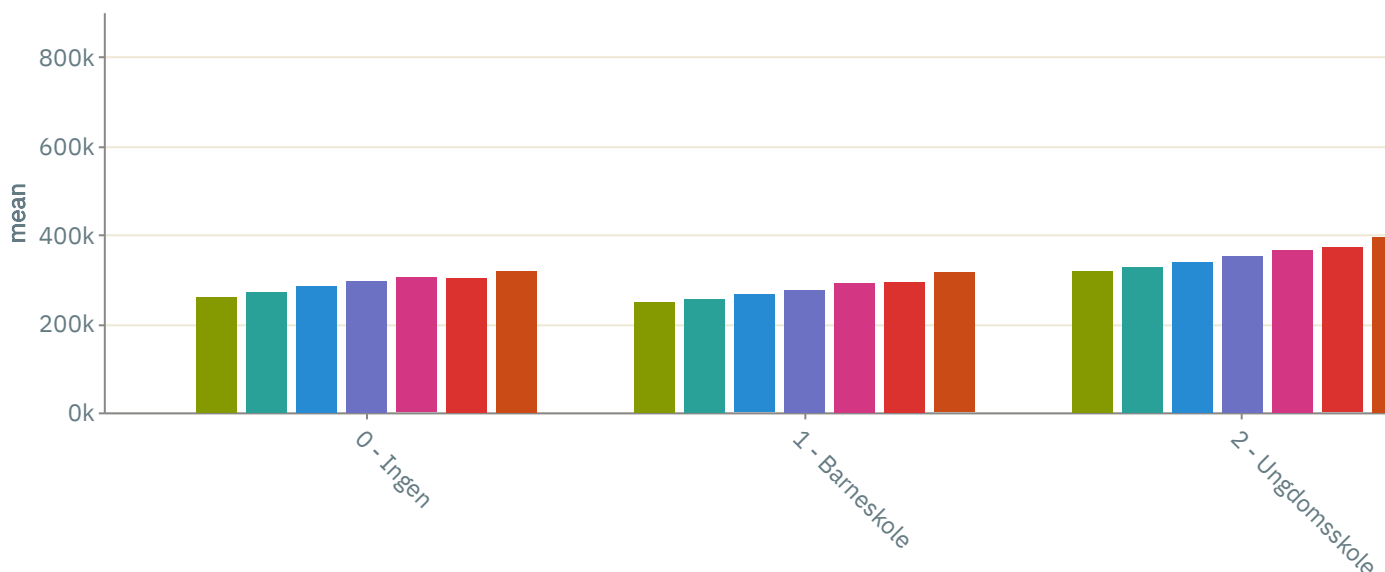
**lønnsgapet** » tabulate kjønn utdnivå, summarize(lønn15)

		utdnivå							
		0 - Ingen	1 - Barneskole	2 - Ungdomsskole	3 - Videregående	4 - Videregående avsluttende utdanning	5 - Påbygg til videregående	6 - UH-utdanning - lavere nivå	7 - UH-utdanning - høyere nivå
kjønn	1 - Mann	305610.35	276877.2	361423.96	477574.35	489190.19	591711.06	530904.89	621902.71
	2 - Kvinne	212196.61	215433.38	261821.95	351348.15	342389.58	381104.61	390267.58	452195.7
	Total	258949.14	247640.79	318488.15	411873.87	430523.43	518365.69	445223.14	533415.38

**lønnsgapet** » tabulate kjønn utdnivå, summarize(lønn21)

		utdnivå							
		0 - Ingen	1 - Barneskole	2 - Ungdomsskole	3 - Videregående	4 - Videregående avsluttende utdanning	5 - Påbygg til videregående	6 - UH-utdanning - lavere nivå	7 - UH-utdanning - høyere nivå
kjønn	1 - Mann	382346.47	351129.95	448843.07	513793.81	603118.34	737838.32	688977.21	859511.75
	2 - Kvinne	258615.08	270433.83	316772	372024.47	423331.12	473226.53	528192.75	678579.04
	Total	318441.69	314248.43	394027.67	442326.8	532126.81	646928.71	591863.86	764838.47

**lønnsgapet** » barchart (mean) lønn15 lønn16 lønn17 lønn18 lønn19 lønn20 lønn21, over(utdnivå)



## Lønnsgapet» tabulate utdnivå

utdnivå	0 - Ingen	6940
	1 - Barneskole	16950
	2 - Ungdomsskole	538513
	3 - Videregående	143370
	4 - Videregående avsluttende utdanning	855783
	5 - Påbygg til videregående	109235
	6 - UH-utdanning - lavere nivå	851042
	7 - UH-utdanning - høyere nivå	358693
	8 - Doktorgrad	37503
	<i>Total</i>	<i>2918039</i>

## Lønnsgapet» tabulate utdnivå, cellpct

utdnivå	0 - Ingen	0.24
	1 - Barneskole	0.58
	2 - Ungdomsskole	18.45
	3 - Videregående	4.91
	4 - Videregående avsluttende utdanning	29.33
	5 - Påbygg til videregående	3.74
	6 - UH-utdanning - lavere nivå	29.16
	7 - UH-utdanning - høyere nivå	12.29
	8 - Doktorgrad	1.29
	<i>Total</i>	<i>100</i>



**lønnsgapet**» tabulate utdnivå kjønn, summarize(lønn15)

	kjønn		Total
	1 - Mann	2 - Kvinne	
0 - Ingen	305610.35	212196.61	258949.14
1 - Barneskole	276877.2	215433.38	247640.79
2 - Ungdomsskole	361423.96	261821.95	318488.15
3 - Videregående	477574.35	351348.15	411873.87
4 - Videregående avsluttende utdanning	489190.19	342389.58	430523.43
5 - Påbygg til videregående	591711.06	381104.61	518365.69
6 - UH-utdanning - lavere nivå	530904.89	390267.58	445223.14
7 - UH-utdanning - høyere nivå	621902.71	452195.7	533415.38
8 - Doktorgrad	740243.2	587775.3	672029.78
<b>Total</b>	<b>501483.12</b>	<b>369383.14</b>	<b>436803.64</b>

**lønnsgapet**» tabulate utdnivå kjønn, summarize(lønn21)

	kjønn		Total
	1 - Mann	2 - Kvinne	
0 - Ingen	382346.47	258615.08	318441.69
1 - Barneskole	351129.95	270433.83	314248.43
2 - Ungdomsskole	448843.07	316772	394027.67
3 - Videregående	513793.81	372024.47	442326.8
4 - Videregående avsluttende utdanning	603118.34	423331.12	532126.81
5 - Påbygg til videregående	737838.32	473226.53	646928.71
6 - UH-utdanning - lavere nivå	688977.21	528192.75	591863.86
7 - UH-utdanning - høyere nivå	859511.75	678579.04	764838.47
8 - Doktorgrad	939995.25	786980.75	870636.06
<b>Total</b>	<b>638142.26</b>	<b>492350.34</b>	<b>567438.32</b>

**lønnsgapet**» generate høy\_utd = 0

Genererte høy\_utd  
med 3 035 242 enheter

**lønnsgapet**» replace høy\_utd =1 if utdnivå >= 7 & utdnivå <= 8

Byttet ut verdier i høy\_utd  
med 3 035 242 enheter

**lønnsgapet**» summarize høy\_utd

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
høy_utd	0.1305	0.3369	3035242	0	0	0	0	1

**lønnsgapet**» generate `man_first = 0`

Genererte `man_first`  
med 3 035 242 enheter

**lønnsgapet**» replace `man_first = 1 if kjønn == "1"`

Byttet ut verdier i `man_first`  
med 3 035 242 enheter

**lønnsgapet**» summarize `man_first`

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
man_first	0.5112	0.4999	3035242	0	0	1	1	1

**lønnsgapet**» regress `lnlønn21 man_first`

Kilde	SS	df	MS	Antall Obs:
Modell	41081.8	1	41081.8	2488363
Residual	2.9292451e+6	2.488365e+6	1.17717	F(1, 2488365): 34898.602761
Total	2.970327e+6	2.488366e+6	1.19368	R <sup>2</sup> : 0.01383
				Justert R <sup>2</sup> : 0.01383
				Root MSE: 1.08497

lnlønn21	Coef.	Std.feil	t	P> t	[95% Konf. intervall]
man_first	0.25717	0.00137	186.811	0	0.25448 0.25987
Konst	12.7963	0.00099	12893.2	0	12.7944 12.7983

**lønnsgapet**» regress `lnlønn21 man_first alder`

Kilde	SS	df	MS	Antall Obs:
Modell	43591.2	2	21795.6	2488363
Residual	2.9267357e+6	2.488364e+6	1.17616	F(2, 2488364): 18531.052037
Total	2.970327e+6	2.488366e+6	1.19368	R <sup>2</sup> : 0.01467
				Justert R <sup>2</sup> : 0.01467
				Root MSE: 1.08451

lnlønn21	Coef.	Std.feil	t	P> t	[95% Konf. intervall]
man_first	0.25677	0.00137	186.597	0	0.25408 0.25947
alder	0.00269	0.00005	46.1908	0	0.00257 0.0028
Konst	12.677	0.00276	4582.66	0	12.6716 12.6825

**lønnsgapet**» regress `lnlønn21 man_first alder høy_utd`

Kilde	SS	df	MS	Antall Obs:
Modell	1.3009126e+5	3	43363.7	2488363
Residual	2.8402357e+6	2.488363e+6	1.1414	F(3, 2488363): 37991.479821
Total	2.970327e+6	2.488366e+6	1.19368	R <sup>2</sup> : 0.04379
				Justert R <sup>2</sup> : 0.04379
				Root MSE: 1.06836

lnlønn21	Coef.	Std.feil	t	P> t	[95% Konf. intervall]
man_first	0.26791	0.00135	197.543	0	0.26525 0.27057
alder	0.00357	0.00005	62.0993	0	0.00345 0.00368
høy_utd	0.52895	0.00192	275.288	0	0.52519 0.53272
Konst	12.555	0.00276	4547.33	0	12.5496 12.5604

**lønnsgapet**» import `ds/REGSYS_ARB_YRKE_STYRK08 2021-11-16` as `styrk`

Importerte `REGSYS_ARB_YRKE_STYRK08` på datoen `2021-11-16` som `styrk` til `lønnsgapet` med 3 035 242 enheter, hvorav 52 163 missingverdier

**lønnsgapet**» generate `yrke = substr(styrk, 1, 1)`

Genererte `yrke`  
med 3 035 242 enheter, hvorav 52 163 missingverdier

**lønnsgapet» define-labels yrkerstring** 0" Militære yrker og uoppgitt" 1" Ledere" 2" Akademiske yrker" 3 "Høyskoleyrker" 4 "Kontoryrker" 5"Salg - og serviceyrker" 6"Bønder" 7" Håndverkere" 8" Prosess- og maskinoperatører, transportarbeidere mv." 9"Renholdere, hjelpearbeidere mv."

Opprettet kodelisten *yrkerstring* med 10 etiketter

**lønnsgapet» destring yrke**

Konverterte *yrke* til tallverdier i ny variabel

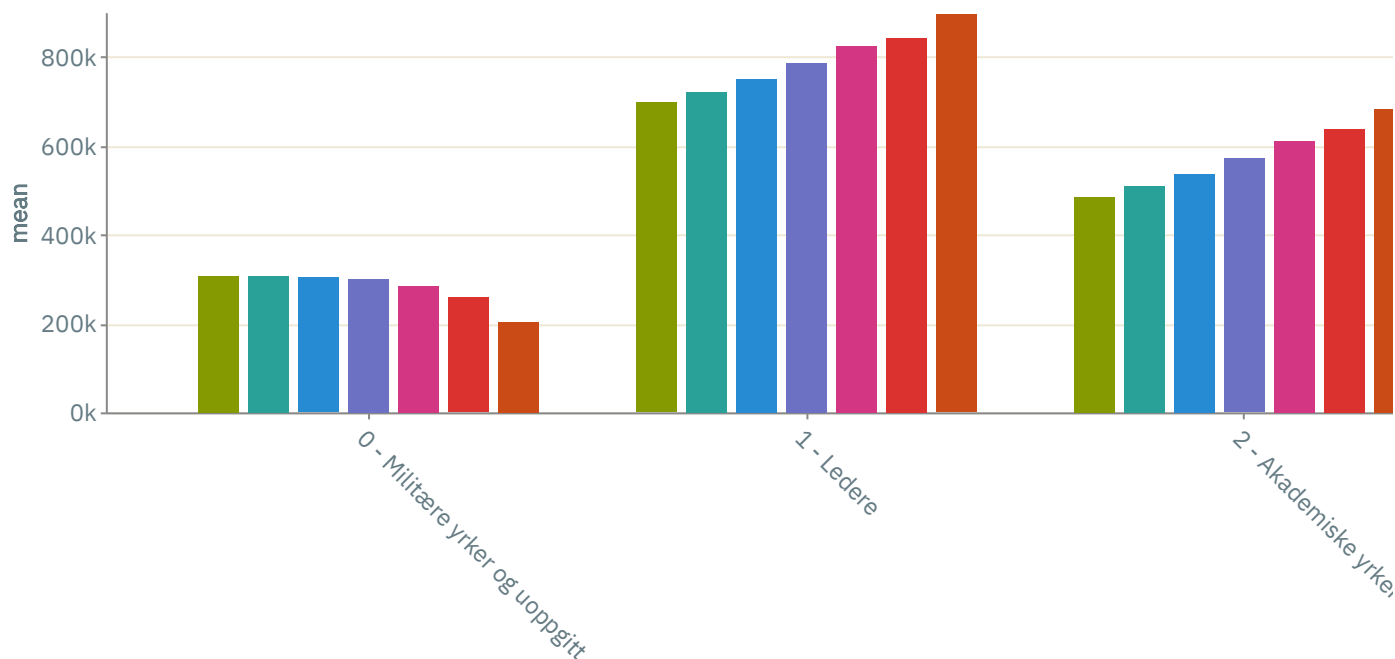
**lønnsgapet» assign-labels yrke yrkerstring**

Tilegnet kodelisten *yrkerstring* til variabelen *yrke*

**lønnsgapet» tabulate yrke kjønn, rowpct**

	kjønn		Total
	1 - Mann	2 - Kvinne	
0 - Militære yrker og uoppgitt	49.15	50.85	100
1 - Ledere	62.61	37.39	100
2 - Akademiske yrker	38.79	61.21	100
3 - Høyskoleyrker	60.52	39.48	100
4 - Kontoryrker	42.84	57.16	100
5 - Salg - og serviceyrker	31.36	68.64	100
6 - Bønder	75.97	24.05	100
7 - Håndverkere	95.44	4.56	100
8 - Prosess- og maskinoperatører, transportarbeidere mv.	85.77	14.23	100
9 - Renholdere, hjelpearbeidere mv.	47.39	52.61	100
<b>Total</b>	<b>51.01</b>	<b>48.99</b>	<b>100</b>

**lønnsgapet» barchart (mean) lønn15 lønn16 lønn17 lønn18 lønn19 lønn20 lønn21, over(yrke)**



## Lønnsgapet» tabulate yrke kjønn, summarize(lønn21)

	kjønn		Total
	1 - Mann	2 - Kvinne	
0 - Militære yrker og uoppgitt	239690.72	160438.9	203120.38
1 - Ledere	960356.12	785084.61	894819.21
2 - Akademiske yrker	790150.07	614753.74	682786.14
3 - Høyskoleyrker	779305.74	587408.01	703536.96
4 - Kontoryrker	537288.55	470148.45	498913.81
5 - Salg - og serviceyrker	446964.71	374838.98	397456.87
6 - Bønder	535460.79	380856.05	498369.35
7 - Håndverkere	581834.3	446539.96	575666.41
8 - Prosess- og maskinoperatører, transportarbeidere mv.	586998.61	453945.86	568072.94
9 - Renholdere, hjelpearbeidere mv.	449668.99	331849.53	387685.76
<b>Total</b>	<b>633169.19</b>	<b>489697.35</b>	<b>564246.45</b>

## Lønnsgapet» summarize høy\_utd

Variabel	Gj.snitt	Std.avvik	Antall	1%	25%	50%	75%	99%
høy_utd	0.1305	0.3369	3035242	0	0	0	0	1