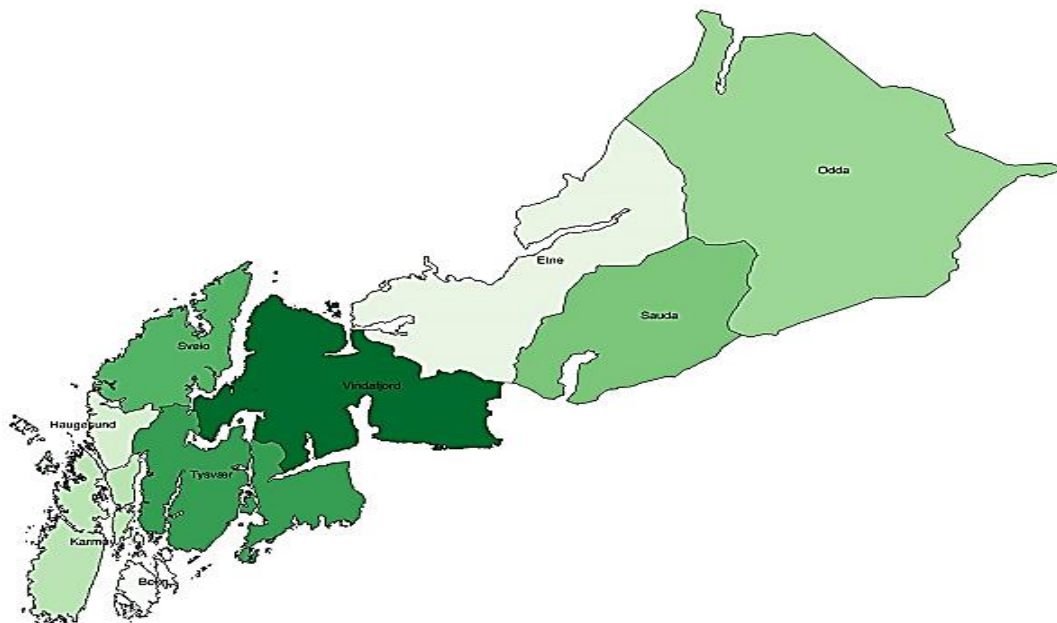




HØGSKOLEN STORD/HAUGESUND

## Sosioøkonomiske nabolagstrekk og boligpriser på Haugalandet



Bacheloroppgave utført ved

Høgskolen Stord/Haugesund – Økonomisk- administrativ utdanning

---

Av: Linn Jeanett Økland, kandidat nr. 13

Dette arbeidet er gjennomført som ledd i bachelorprogrammet i økonomi og administrasjon ved Høgskolen Stord/Haugesund og er godkjent som sådan. Godkjennelsen innebærer ikke at HSH inntår for metodene som er anvendt, resultatene som er fremkommet og konklusjoner og vurderinger i arbeidet.

---

Bacheloroppgavens tittel: *“Sosioøkonomiske nabolagstrekk og boligpriser på Haugalandet”*

---

Student: Linn Jeanett Økland

---

Veileder: Liv Aileen Osland

---

Gradering: Offentlig

---

## Forord

Som en del av det 3-årige studiet økonomi og administrasjon ved HSH, er dette en avsluttende oppgave før fullført bachelorgrad. Oppgaven er utført av Linn Jeanett Økland, med fordypning innen finans- og økonomistyring, våren 2014.

Samfunnsøkonomiske fag er spennende og jeg valgte derfor å rette oppgaven min inn mot dette fagområdet. Temaet for oppgaven er segregering og boligpriser på Haugalandet. Jeg ønsket å se på ulike sosioøkonomiske nabolagstrekk og dets innvirkning på grad av segregering. Deretter se hvilke av disse variablene som har størst betydning for variasjonen i boligpriser i distriktet.

Jeg vil benytte anledningen til å takke min engasjerende veileder Liv Aileen Osland for god støtte i arbeidet med oppgaven. Hun har gitt meg grundige tilbakemeldinger, gode råd og tips, og alltid stilt opp når jeg har hatt behov for veiledning. Hun har også vært behjelpelig med datamateriale til oppgaven. Jeg vil også takke min gode venninne, Christine Olsen Schei, som har hjulpet meg med regresjonsanalyser i STATA.

Haugesund 13. mai 2014

---

Linn Jeanett Økland

## Sammendrag

Denne bacheloroppgaven er basert på samfunnsøkonomiske fag, og omhandler sosioøkonomiske nabolagstrekk og boligpriser. Problemstillingen tar utgangspunkt i mål for segregering og boligpriser på Haugalandet. Jeg ønsket å få svar på i hvor stor grad sosioøkonomiske variabler påvirker boligpriser, og hvilke av disse variablene som har størst betydning for variasjonen i distriktets boligpriser. For å beskrive segregeringssituasjonen laget jeg indekser på region- og kommunenivå for ulike sosioøkonomiske variabler. Deretter så jeg på boligpriser i distriktet, for å undersøke om de variablene jeg hadde studert ville ha noen innvirkning på boligpriser. Til dette arbeidet benyttet jeg regresjonsanalyse.

Resultatene viste at de sosioøkonomiske nabolagstrekkene hadde lave segregeringstendenser i regionen, og at disse ikke hadde store betydninger for boligpriser i regresjonsanalysene. Flere forklaringsvariabler ble signifikante i analysene, men den økonomiske virkningen på boligpriser var lav. Det resulterte i at nabolagstrekkene hadde liten påvirkning på boligpriser. Den forklaringsvariabelen som i alle analysene ble statistisk signifikant og som hadde størst effekt på boligpriser var høy utdanning. Trekk ved det lokale nærmiljøet er av betydning, siden flere av variablene ble signifikante i regresjonsanalysene.

## INNHold

<b>Forord</b> .....	<b>iii</b>
<b>Sammendrag</b> .....	<b>iv</b>
<b>Liste over figurer</b> .....	<b>vii</b>
<b>Liste over tabeller</b> .....	<b>viii</b>
<b>1 Innledning</b> .....	<b>1</b>
1.1 Bakgrunn og problemstilling .....	1
1.2 Oppgavens struktur .....	2
1.3 Begrensninger .....	3
<b>2 Teori</b> .....	<b>4</b>
2.1 Nabolag .....	4
2.2 Segregering.....	5
2.3 Hedonsk metode .....	5
<b>3 Segregeringsmål for beskrivende statistikk</b> .....	<b>7</b>
3.1 Beskrivende statistikk av sosioøkonomiske trekk.....	9
3.2 Datagrunnlag for utdanning og innvandring.....	9
3.2.1 Ulikhets- og isolasjonsindeks for utdanning og innvandring på regionnivå	10
3.2.2 Ulikhetsindeks for utdanning og innvandring på kommunenivå .....	13
3.2.3 Isolasjonsindeks for utdanning og innvandring på kommunenivå .....	16
3.4 Sammenligning av ulikhets- og isolasjonsindeks med andre geografier .....	20
<b>4 Metode</b> .....	<b>22</b>
4.1 Valg av metode .....	22
4.2 Data/datainnsamling.....	23
4.3 Regresjon .....	24
4.3.1 Grunnlagsmodellen .....	26
<b>5 Resultater fra regresjonsanalysene</b> .....	<b>30</b>
<b>6 Konklusjon</b> .....	<b>39</b>

<b>7 Litteraturliste</b> .....	<b>41</b>
<b>8 Appendix</b> .....	<b>43</b>
Vedlegg 1: Beregning av ulikhetsindeks.....	43
Vedlegg 2: Beregning av isolasjonsindeks .....	46
Vedlegg 3: Populasjonsstørrelser.....	47
Vedlegg 4: Indeksverdier på regionnivå .....	50

## Liste over figurer

- Figur 3.1 *Ulikhetsindeks på regionnivå, andel lav utdanning*
- Figur 3.2 *Isolasjonsindeks på regionnivå, andel lav utdanning*
- Figur 3.3 *Ulikhetsindeks på regionnivå, andel videregående utdanning*
- Figur 3.4 *Isolasjonsindeks på regionnivå, andel videregående utdanning*
- Figur 3.5 *Ulikhetsindeks på regionnivå, andel høy utdanning*
- Figur 3.6 *Isolasjonsindeks på regionnivå, andel høy utdanning*
- Figur 3.7 *Ulikhetsindeks på regionnivå, andel total innvandring*
- Figur 3.8 *Isolasjonsindeks på regionnivå, andel total innvandring*
- Figur 3.9 *Ulikhetsindeks på regionnivå, andel ikke-vestlig innvandring*
- Figur 3.10 *Isolasjonsindeks på regionnivå, andel ikke-vestlig innvandring*
- Figur 3.11 *Ulikhetsindeks på kommunenivå, andel lav utdanning*
- Figur 3.12 *Ulikhetsindeks på kommunenivå, andel videregående utdanning*
- Figur 3.13 *Ulikhetsindeks på kommunenivå, andel høy utdanning*
- Figur 3.14 *Ulikhetsindeks på kommunenivå, andel total innvandring*
- Figur 3.15 *Ulikhetsindeks på kommunenivå, andel ikke-vestlig innvandring*
- Figur 3.16 *Isolasjonsindeks på kommunenivå, andel lav utdanning*
- Figur 3.17 *Isolasjonsindeks på kommunenivå, andel videregående utdanning*
- Figur 3.18 *Isolasjonsindeks på kommunenivå, andel høy utdanning*
- Figur 3.19 *Isolasjonsindeks på kommunenivå, andel total innvandring*
- Figur 3.20 *Isolasjonsindeks på kommunenivå, andel ikke-vestlig innvandring*

## Liste over tabeller

Tabell 3.1 *Viser total befolkning til den utvalgte aldersgruppen innenfor hver kommune i Haugaland, samt total befolkning i regionen, år 1995.*

Tabell 4.1 *Beskrivende statistikk*

Tabell 4.2 *Regresjonsformel*

Tabell 4.3 *Definisjoner av variabler i grunnmodellen*

Tabell 4.4 *Resultater fra Oslands grunnmodell*

Tabell 5.1 *Resultat av regresjonsanalyse som inkluderer grunnmodell og utdanningskategorier*

Tabell 5.2 *Resultat av regresjonsanalyse for innvandringskategorier*

Tabell 5.3 *Resultat av regresjonsanalyse for grunnmodell, utdanning- og innvandringskategorier*

Tabell 5.4 *Resultat av den siste regresjonsanalysen, en modell bestående av grunnmodell variabler, andel ikke-vestlig innvandring, videregående og høy utdanning*



# 1 Innledning

Segregering og nabolagseffekter er et tema som media stadig belyser. Vi lever i en globalisert verden, noe som har medført at innvandring fra andre land i stor grad har økt de siste tyve årene. Flerkulturelle samfunn har derfor utviklet seg i mange land. I Norden så man de første tendensene til segregering på 1980-tallet i svenske storbyer (Barne-, likestillings- og inkluderingsdepartementet, 2010). Senere har man også sett tendenser til segregering i Norges hovedstad, Oslo. Områder i Oslo hvor det bor flere innvandrere enn etnisk norske oppleves å være preget av dårlige levekår og lav inntekt. Segregering av ikke-vestlige innvandrere bosatt i norske byer varierer etter byenes befolkningsstørrelse (Statistisk Sentralbyrå, 2005).

Det er utført flere studier i Europa på nabolagseffekter og dets innvirkning på barn, men man ser at disse effektene er mer oppstykket og uklare enn studier gjort i USA, som har dokumentert nabolagseffekter på blant annet kriminalitet, rus, og utdanningsforløp (Wessel, 2013). I Europa har det vist seg at slike nabolagseffekter har en tendens til å forsvinne ved kontroll for demografi, sosial bakgrunn og livshendelser. Allikevel ser man her små til moderate effekter på kriminalitet, rus, utdanning, og inntekt (Wessel, 2013).

De fleste studier av segregering er knyttet til storbyområder. Andelen av innvandrere er trolig lavere i mindre norske byer. Segregering er likevel noe som også bør studeres i disse områdene. Et annet tema som også er lite studert i Norge, er i hvilken grad ulikheter i sosioøkonomiske trekk i nabolag påvirker boligpriser. Det er dette som er temaet for denne bacheloroppgaven.

## **1.1 Bakgrunn og problemstilling**

Ved valg av tema og problemstilling for denne oppgaven ble min store interesse for samfunnsøkonomiske fag tatt i betraktning. By- og regionaløkonomi er et fag innenfor dette fagfeltet, som omhandler segregering og nabolagseffekter. Ved å studere et geografisk område kan man oppdage ulike mønstre og sammenhenger i for eksempel bosettingsmønstre, miljø, utdanning, inntekt, og etniske kjennetegn. Emner som dette fenger meg, og jeg ville derfor studere nabolag og ulike sosioøkonomiske variabelers innvirkning på boligpriser.

Det er mange områder i Norge som hadde vært interessant å studere. Tilgjengelig data for kommunene i region Haugaland gjorde at jeg valgte å se på dette geografiske området. Med utgangspunkt i mål for segregering og boligpriser på Haugalandet ønsker jeg å finne svar på i hvor stor grad sosioøkonomiske variabler påvirker boligpriser og hvilke av disse variablene som har størst betydning for variasjonen i distriktets boligpriser. Jeg vil studere sosioøkonomiske nabolagstrekk som utdanningsnivå og innvandring. Det er ulike måter å gjøre dette på. I denne oppgaven vil jeg først presentere beskrivende statistikk via indekser – som er vanlig å bruke for å måle graden av segregering. Deretter vil jeg ta for meg boligpriser i Haugalandet, og se om de variablene jeg har studert vil ha noen innvirkning på boligprisene. Til denne analysen er regresjonsanalyse et nyttig verktøy å bruke.

## **1.2 Oppgavens struktur**

Kapittel 1, Innledning, omhandler bakgrunnen for valg av oppgave og tema, samt problemstilling, strukturen til oppgaven, og hvilke avgrensninger som ble gjort. I kapittel 2, Teori, går jeg nærmere inn på teorien som legger grunnlaget for oppgaven. Her presenteres grunnleggende teori om nabolag, forklaring på hva et nabolag er og dets spesielle kjennetegn, samt ulike sosioøkonomiske trekk som kan prege et nabolag. Segregering, som presenteres etter nabolag, definerer segregering, og hvordan dette kan knyttes opp mot boligpriser. Den hedoniske metoden er det siste underkapittelet i teoridelen, og viser hvilke faktorer som kan påvirke boligpriser, samt en modell som kan brukes til å se på nabolagets betydning for boligpriser. I kapittel 3 introduserer jeg to segregeringsmål, ulikhets- og isolasjonsindeks, som jeg har brukt til å presentere beskrivende statistikk av sosioøkonomiske trekk. Videre vises ulikhets- og isolasjonsindeks figurer innenfor forskningsområdet. Til slutt i dette kapittelet sammenlignes indeksverdier med andre geografier. I kapittel 4, Metodekapittelet, forklarer jeg hvilke metoder jeg har benyttet, informasjon om tallmaterialet, samt beskrivende statistikk for forskningsområdet. Videre introduseres teori om regresjon, som forklarer hva regresjon er, og jeg går nærmere inn på multippel regresjon og definerer ulike størrelser. Til slutt presenteres grunnlagsmodellen, som i kapittel 5 brukes til regresjonsanalysene. Resultater fra regresjonsanalysene oppgis i kapittel 5. De to første regresjonsanalysene viser resultater for hver enkel andelsvariabel. Deretter fremstilles en ny modell som inneholder alle variablene i samme modell. Den siste regresjonsmodellen inneholder de andelsvariablene som ble signifikante i den foregående modellen, og her presenteres de endelige resultatene. I

konklusjonen av oppgaven, kapittel 6, blir resultatene fra kapittel 5 diskutert. Som en avslutning på oppgaven, foreslås tips til videre forskning.

### **1.3 Begrensninger**

De sosioøkonomiske variablene som er benyttet i oppgaven er på grunnkrets nivå. Ved beregning av indekser hadde det vært ideelt å beregne disse på grunnkrets nivå. Da hadde jeg antageligvis fått mer konkrete resultater og segregeringstendensene hadde muligens vært annerledes. På grunn av datasettets omfang og oppgavens tidsperspektiv, har jeg benyttet data på region- og kommunenivå.

Boligprisdataene er på individnivå. I regresjonsdelen hadde jeg tilgang på to sett av boligprisdata, for årene 1997-2002 og årene 2004-2007. Jeg har analysert det første settet, men hadde jeg hatt mulighet ville jeg også laget regresjonsanalyser på det siste datasettet. Man ville da fått et bedre bilde av hvordan boligprisene forspiller seg over lengre tid i sammenheng med segregeringsandelene.

## 2 Teori

Teori betyr å ``granske``, og handler om kunnskaper og antakelser vedrørende virkeligheten (Johannessen, Kristoffersen & Tufte, 2011). Teorigrunnlaget tar utgangspunkt i relevant teori for det emnet jeg studerer, sosioøkonomiske variabler i relasjon til boligpriser. Dette er knyttet til sentrale trekk ved boområder og i hvor stor grad disse fanges opp i priser.

### 2.1 Nabolag

Et nabolag er et lite sammenhengende geografisk område hvor det som regel er gåavstand mellom innbyggernes boliger (Wessel, 2013). Et nabolag inneholder flere viktige institusjoner som skole, barnehage, møteplasser og lokale organisasjoner. Felles stedsidentitet, samhandling og tilhørighet kan nabolaget være preget av. Innenfor bolig- og by forskning er nabolag et viktig tema (Wessel, 2013).

Europeisk forskning gjennomført på nabolag og nabolageffekter som utdanningsnivå og inntekt viser sprik i resultatene (Wessel, 2013). Britiske og nederlandske studier har avfeid nabolageffekter som forklaring på variasjoner i utdanning, yrke og inntekt, mens Sverige og andre land peker på at slike effekter er av betydning (Wessel, 2013).

Et nabolag kan ha flere sosioøkonomiske trekk ved seg. Dette kan både være positive og negative egenskaper som for eksempel variasjoner i adferd, ulike holdninger, grad av kriminalitet, innvandring, fattigdom, bostabilitet, og utdanningsnivå. Slike trekk kan ha positiv/negativ effekt på barn og unge. Barn som bor i et nabolag hvor det eksisterer mye kriminalitet kan bli påvirket av dette, få dårlige forbilder, og etterligne eldre barn (Wessel, 2013). Negative egenskaper kan også føre til mangel på felles normer og verdier i et nabolag. Motsatt kan et velfungerende nabolag med gode skoler, fellesverdier og sosialisering ha positiv innvirkning på barn.

Vi vet at fysiske aspekter ved nabolag som for eksempel trafikkstøy kan ha betydning for boligpriser. Det er uklart hvilken betydning de nevnte sosioøkonomiske nabolagstrekkene har for boligpriser i Norge, spesielt i den typen regioner som jeg skal studere. Sammenlignet med storbyområder er det å forvente at den kvantitative effekten ikke er veldig stor. Gitt at boligen og variasjoner i boligpriser er av stor betydning for den enkelte husholdning, så er det likevel

viktig å studere i hvor stor grad variasjoner i boligpriser blir påvirket av variasjoner i nabolagstrekk også her. De nabolagstrekkene jeg fokuserer på er noen sosioøkonomiske faktorer.

## **2.2 Segregering**

Segregering blir omtalt som en motsetning til integrering, og er grupper av mennesker eller miljøer som lever adskilt eller som ikke deltar på de samme arenaene i samfunnet (Barne-, likestillings- og inkluderingsdepartementet, u.å). Dette vil si at segregering ser på hvor ulike grupper bosetter seg, og om det er noe mønster innenfor disse. Segregerte samfunn har eksistert lenge, helt tilbake til 1870-tallet i USA, der svarte og hvite mennesker ble skilt fra hverandre ved lov. Det ble opprettet egne skoler, transportmidler, fritidssentre, og fengsel for å skille mellom rasene (Library of congress, u.å). Selv om kulturforskjellene har blitt mindre i senere tid, ser man ofte en økt bevissthet om forskjellighet mellom grupper (Store Norske Leksikon, u.å). I Oslo er noen steder preget av segregering, der har man sett at bokonsentrasjonen til grupper med ulik opprinnelse (etnisk norske og innvandrere) dominerer ulike avgrensede områder.

Ulike typer segregering kan oppstå på sosiale steder som for eksempel på skolen, arbeidsplassen, og i bomiljøet (Wong, 2008). Innenfor disse områdene kan populasjonen deles inn i undergrupper, som for eksempel med hensyn på utdanning og innvandring. Segregering kan brukes til å se på forskjeller i bosetningsmønster innenfor ulike geografiske områder, som for eksempel i byer og mindre områder som nabolag. Ulike modeller kan benyttes for å finne grad av segregering i et gitt geografisk område. Ved å studere segregering i bomiljøer, kan man knytte dette opp mot boligpriser i det aktuelle området for å se om det har noen påvirkning på boligprisene.

## **2.3 Hedonsk metode**

Innenfor by- og regionaløkonomien ser man på bosetningsmønsteret i ulike deler av en geografi. Det vises at avstand fra sentrum er av betydning for boligpriser, dess lengre unna sentrum man beveger seg, dess lavere vil boligprisene være (McCann, 2001). Det er ikke bare avstand fra sentrum som påvirker boligprisen, men også andre forhold innenfor geografien

eller ved boligens lokalisering har en betydning. Dette kan være faktorer som nabolag, utsikt, solforhold, og nærhet til servicetilbud (eks. skole, aktiviteter, buss, og butikker) (Knudsen & Sødal, 2010). Disse attributtene vil bli tatt hensyn til i boligmarkedets prisvurdering av boliger.

Sosioøkonomiske variabler kan ha betydning for boligpriser innenfor et nabolag. En modell som kan benyttes til å se på nabolagets betydning for boligpriser er "Hedonic price model". Modellen identifiserer ulike faktorer i henhold til premisset om at boligprisen bestemmes av interne og eksterne kjennetegn (Investopedia, 2013). Dette vil med andre ord si at prisen på en bolig blir bestemt av boligens utseende, kvalitet, og størrelse (interne faktorer). Boligens beliggenhet, avstand fra skole og barnehage, pris på tilsvarende boliger i nabolaget (eksterne faktorer) kan også ha betydning. Den hedoniske modellen kan ta i bruk ulike sosioøkonomiske variabler og måler hvilke faktorer som har en innvirkning på boligprisen. Modellen kan anvende ulike segregeringsmål når man studerer sosioøkonomiske variabler og boligpris (Wong, 2008).

### 3 Segregeringsmål for beskrivende statistikk

Ulike mål kan benyttes til å finne grad av segregering i et gitt geografisk område. De målene jeg har konsentrert meg om er ulikhetsindeks og isolasjonsindeks. Ved beregning av indeksene har jeg brukt to sosioøkonomiske variabler; utdanning og innvandring.

Ulikhetsindeksen er et høyt anbefalt mål for segregering (Wong, 2008), og indeksen viser fordelingen mellom to grupper innenfor en populasjon. Man grupperer befolkningen i det gitte området inn i ulike variabler, som for eksempel inntekt, utdanningslengde, eller innvandring. Ser man for eksempel på variabelen innvandring innenfor et område, kan man ved hjelp av indeksen finne segregeringsmønsteret til denne gruppen.

Formel for ulikhetsindeks (Simpson, 2007):

$$ID = 0.5 \sum abs((N_{gi}/N_g) - (N_{hi}/N_h))$$

Forklaring av variabler:

$N_{gi}$  = Populasjon til gruppe  $g$  i området  $i$  (der området er en kommune).

$N_g$  = Total populasjon til gruppe  $g$  i regionen.

$N_{hi}$  = Populasjon til gruppe  $h$  i området  $i$  (der området er en kommune).

$N_h$  = Total populasjon til gruppe  $h$  i regionen.

Ulikhetsindeksen gir verdier mellom 0 og 1. Verdier under 0,30 betyr at området er lite segregert, verdier mellom 0,30-0,60 er moderate, mens verdier over 0,60 betyr at området er sterkt segregert (Statistisk Sentralbyrå, 2006).

Henviser til Appendix (vedlegg 1) for eksempel på utregning av ulikhetsindeks.

Et annet viktig segregeringsmål er isolasjonsindeksen. Denne finnes i flere varianter, og er et mål som kan brukes til å undersøke hvor segregert et område er. Indeksen viser sannsynligheten for at en tilfeldig valgt person i minoritetsgruppe X skal møte en annen person fra samme befolkningsgruppe i sitt underområde (Statistisk sentralbyrå, 2006). Det vil si at man ser på et geografisk område, der man benytter en faktor som deler populasjonen inn i grupper. På denne måten kan man se hvor isolert en gruppe er fra resten av befolkningen innenfor et område (Simpson, 2007).

Formel for isolasjonsindeks (Simpson, 2007):

$$P_g^* = \sum (N_{gi}/N_g)(N_{gi}/N_i)$$

Forklaring av variabler:

$N_{gi}$  = Populasjon til gruppe  $g$  i området  $i$  (der området er en kommune).

$N_g$  = Total populasjon til gruppe  $g$  i regionen.

$N_i$  = Total populasjon i det gitte området (kommune).

Isolasjonsverdiene avhenger av populasjonsstørrelsen, det vil si at man får høyere verdier jo større populasjon. Isolasjonsindeksen gir verdier mellom 0 og 1. Lave verdier tilsvarer lav segregering, mens høye verdier tilsvarer segregerte områder.

Siden indeksen ikke sammenligner grupper, men ser på en bestemt andel innenfor en geografi, bør man ved studering av isolasjonsverdier over tid være varsom med tolkninger, på grunn av at indeksen kan stige selv om segregering ikke stiger (Simpson, 2007). Dette vil si at isolasjonsverdien til den andelen man studerer vil være sensitiv.

Henviser til Appendix (vedlegg 2) for eksempel på utregning av isolasjonsindeks.



### **3.1 Beskrivende statistikk av sosioøkonomiske trekk**

De sosioøkonomiske variablene jeg har konsentrert meg om er utdanning og innvandring. I dagens samfunn er det nødvendig å ha en god utdanning for å kunne komme inn på arbeidsmarkedet, og flere velger dermed å ta høyere utdanning. Utdanning vil derfor være en viktig variabel å studere. Det vil også være interessant å se på innvandring, da antall innvandrere har økt i Norge de siste tiårene, og siden det trolig er slik at andel innvandrere er større i storbyer enn i mindre byer. Ved å beregne indekser på disse variablene, kan jeg se om det er noen tydelige endringer i segregeringsmønsteret. Det hadde vært interessant å studere flere nabolagstrekk, men da jeg ikke har data på andre variabler fokuserer jeg på utdanning og innvandring.

### **3.2 Datagrunnlag for utdanning og innvandring**

Datasettene er hentet fra folkeregisteret og SSB (fått tildelt av veileder). Dataene er fra år 1995 til år 2011. Jeg har regnet indekser for annethvert år innenfor denne tidsperioden. Aldersgruppen til den utvalgte populasjonen er 30-49 år. Det geografiske området jeg har studert er region Haugaland, som er delt inn i følgende kommuner: Haugesund, Sauda, Bokn, Tysvær, Karmøy, Vindafjord, Etne, Sveio og Odda (Utsira er utelatt). Haugesund er bykommune, mens de andre er nærliggende landkommuner. Haugesund og Karmøy har flest innbyggere, mens de resterende har lave befolkningstall. Tabellen nedenfor illustrerer populasjonsforskjellen mellom kommunene for et gitt år.

<b>Kommune</b>	<b>Populasjon</b>
Haugesund	7948
Sauda	1342
Bokn	194
Tysvær	2238
Karmøy	9862
Vindafjord	1252
Etne	952
Sveio	1265
Odda	2130
Total befolkning i regionen	27183

*Tabell 3.1. Viser total befolkning til den utvalgte aldersgruppen innenfor hver kommune i Haugaland, samt total befolkning i regionen, år 1995.*

Ved analysering av faktoren utdanning hadde jeg åtte ulike utdanningsnivåer som jeg valgte å dele inn i tre mindre nivåer for å samle kategoriene; grunnskoleutdanning, videregående utdanning, og universitet/høyskole utdanning. Kategorien grunnskole består av personer som har fullført ungdomskolen. Videregående består av påbegynt videregående til fullført, samt påbygg. Kategorien universitet/høyskole består av universitet, høyskole og forsker utdanning. Datasettet har oppgitte tall på grunnkrets nivå, hvor det er nærmere 360 grunnkretser i regionen. Jeg har valgt å se på regionnivå og kommunenivå, dette for å klare å illustrere utviklingen i regionen mer samlet. Se Appendix (vedlegg 3) for oversikt over populasjonsstørrelsene til hver kommune.

Ved analysering av faktorene total innvandring og ikke-vestlig innvandring har jeg brukt andelene til å se om det har vært noe endring i segregering i kommunene. Andelene har jeg sett i forhold til totalt antall personer som er norske (født i Norge) i regionen.

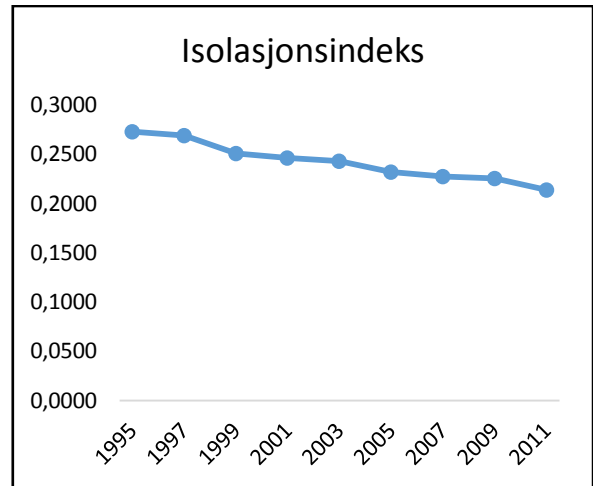
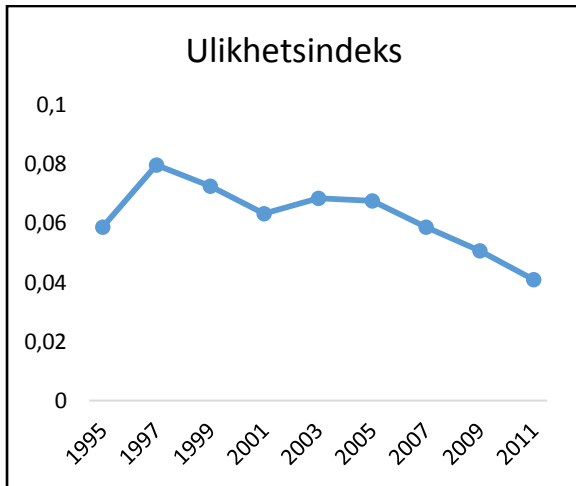
Offentlig statistikk definerer innvandrere som (Statistisk Sentralbyrå, 2013):

Innvandrere er personer som er født i utlandet av to utenlandsfødte foreldre og fire utenlandsfødte besteforeldre. Innvandrere har på et tidspunkt innvandret til Norge.

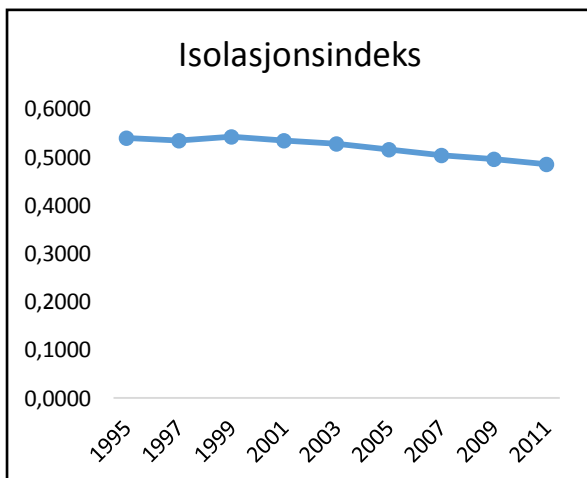
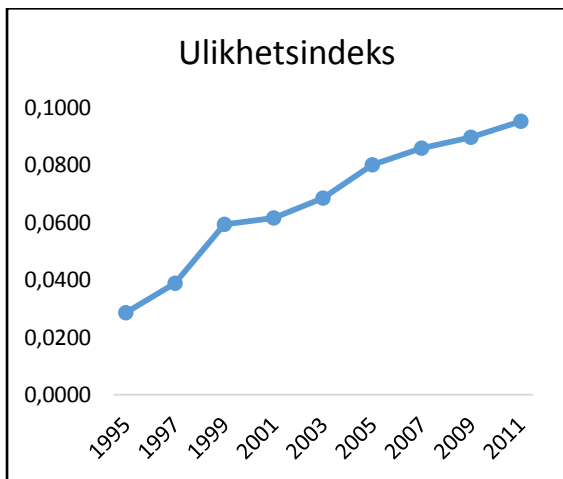
Faktor total innvandring innebærer all innvandring til regionen. Ikke-vestlig innvandring er innvandring fra; Jugoslavia (tidligere), Tyrkia, Afrika, Asia, Mellom-Amerika, Chile, Sør-Amerika for øvrig, nye EU-land, og Europa for øvrig. Ikke-vestlige innvandrere er egentlig en kategori som ikke brukes lenger i offisiell statistikk, men på grunn av at jeg bruker data som går litt langt tilbake i tid så brukes kategorien likevel. Datasettet har oppgitte tall på grunnkrets nivå, og er for tidsperioden 1995 til 2011. Jeg har også her valgt å se på regionnivå og kommunenivå.

### **3.2.1 Ulikhets- og isolasjonsindeks for utdanning og innvandring på regionnivå**

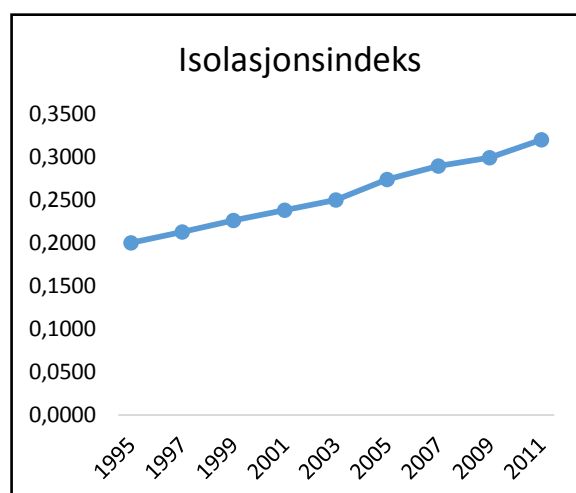
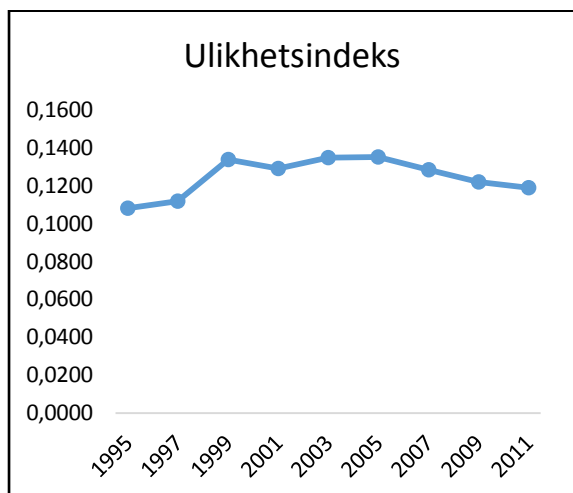
Nedenfor vises ulikhets- og isolasjonskurve for alle kategoriene innenfor utdanning og innvandring. Disse viser utviklingen til regionen i tidsperioden 1995 – 2011. Henviser til Appendix (vedlegg 4) for oversikt over ulikhets- og isolasjonsverdier til figurene.



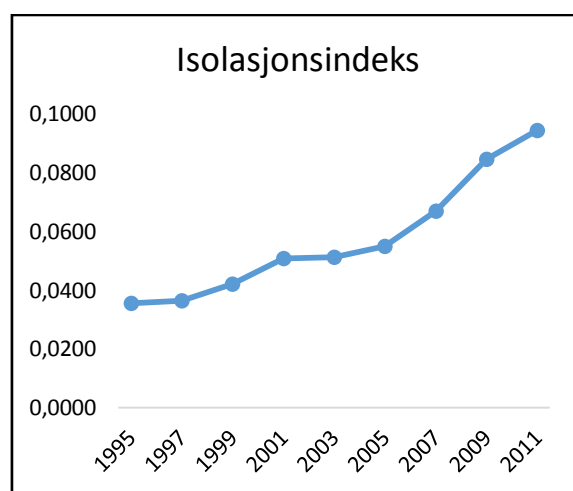
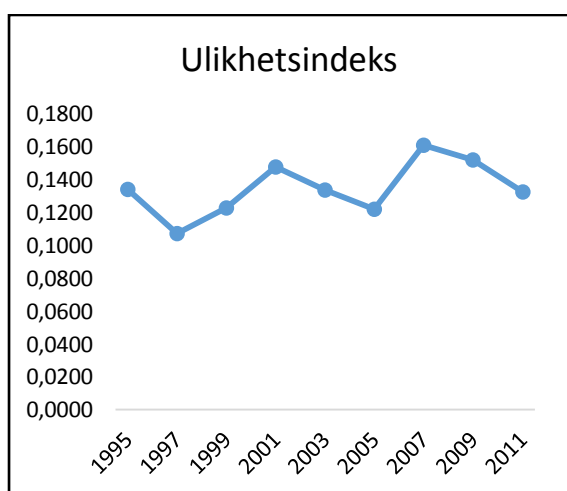
Figur 3.1. og figur 3.2. Viser ulikhets- og isolasjonskurve for andel lav utdanning i region Haugaland.



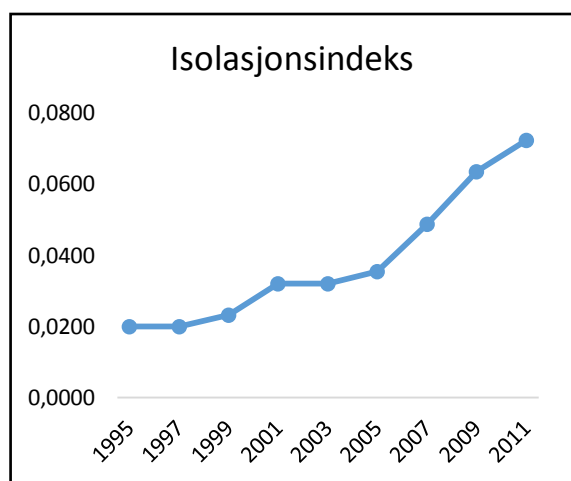
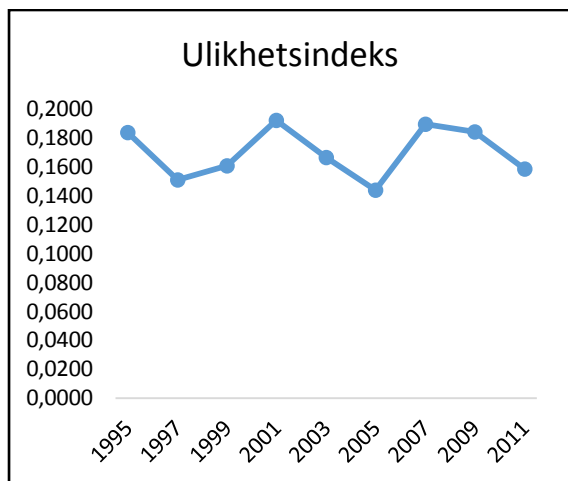
Figur 3.3. og figur 3.4. Viser ulikhets- og isolasjonskurve for andel videregående utdanning i region Haugaland.



Figur 3.5. og figur 3.6. Viser ulikhets- og isolasjonskurve for andel høy utdanning i region Haugaland.



Figur 3.7. og figur 3.8. Viser ulikhets- og isolasjonskurve for andel total innvandring i region Haugaland.



Figur 3.9. og figur 3.10. Viser ulikhets- og isolasjonskurve for andel ikke-vestlig innvandring i region Haugaland.

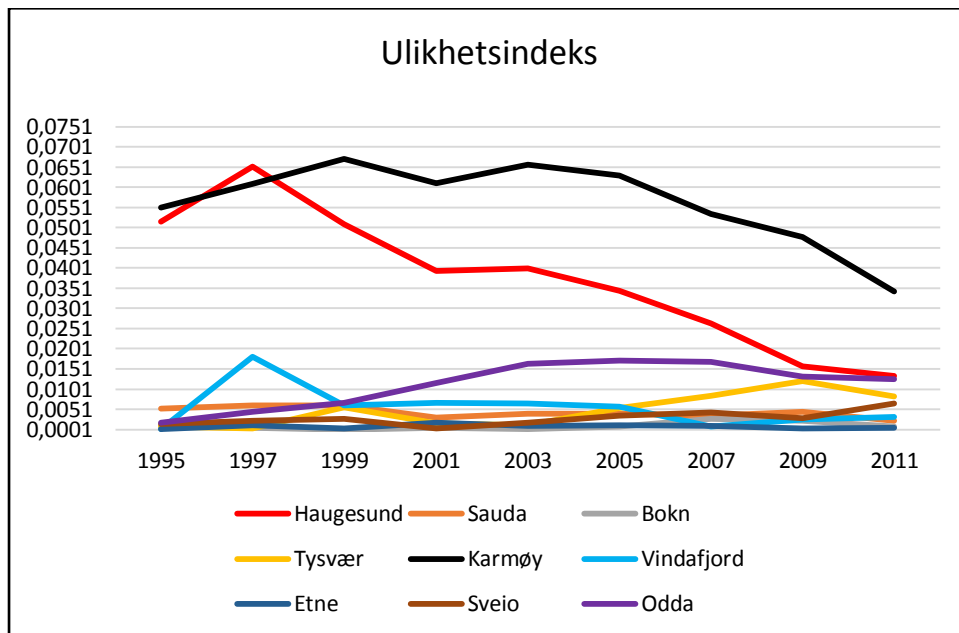
Ulikhetsindekser ser på fordelingen mellom grupper innenfor en gitt populasjon. Ved å studere ulikhetsindeksene på regionnivå ser jeg at andel lav utdanning faller og har lite variasjon i indeksverdiene. Andel videregående og høy utdanning har en liten økning, der videregående har den største økningen. Ulikhetsindeksen for andel total innvandrere og ikke-vestlige innvandrere har hatt svingninger i perioden, men totalt sett har disse gått litt ned. Alle ulikhetsverdiene ligger under 0,30, noe som tilsier at regionen er lite segregert. Segregeringsmønstrene til variablene tyder på små endringer.

I isolasjonsindeksene på regionnivå har andel lav og videregående utdanning blitt redusert, mens andelen for høyere utdanning har økt. Høyere utdanning har hatt størst endring i tidsperioden. En isolasjonsverdi på eksempel 0,2315 (år 2005 i figur 3.2) vil si at det er 23,15% sannsynlighet for at en person med bare grunnskole utdanning møter en annen person innenfor regionen med samme utdanning. Isolasjonskurven for total innvandrere og ikke-vestlige innvandrere er stigende i hele perioden, noe som kan tyde på at bokonsentrasjonen av disse har blitt større (men dette er små endringer).

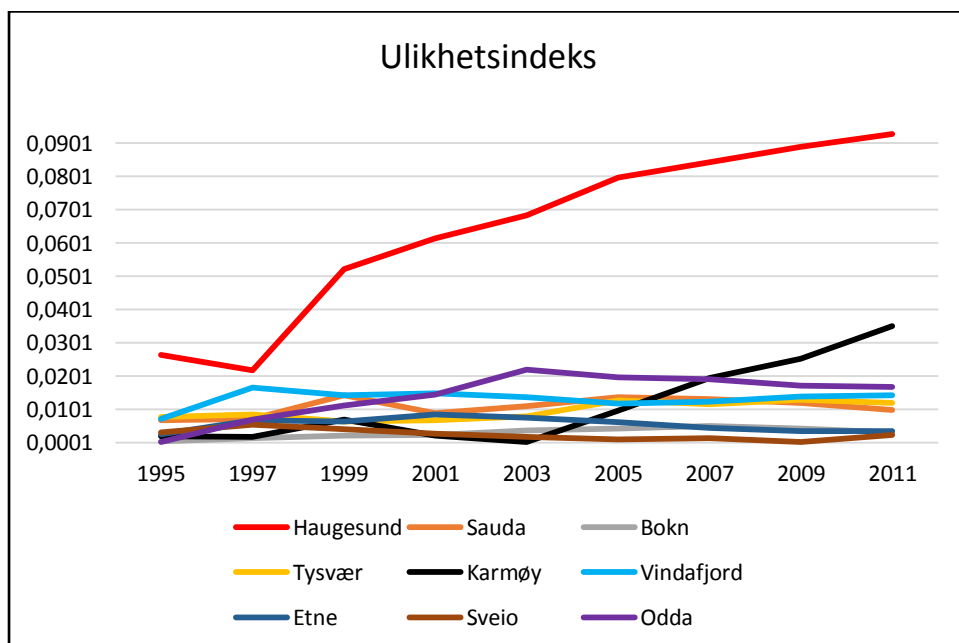
### 3.2.2 Ulikhetsindeks for utdanning og innvandring på kommunenivå

Nedenfor vises figurer for ulikhetsindeks. Disse viser indeksverdier for bykommunen Haugesund og landkommunene i år 1995 – 2011, annethvert år. Ved bruk av slike figurer er

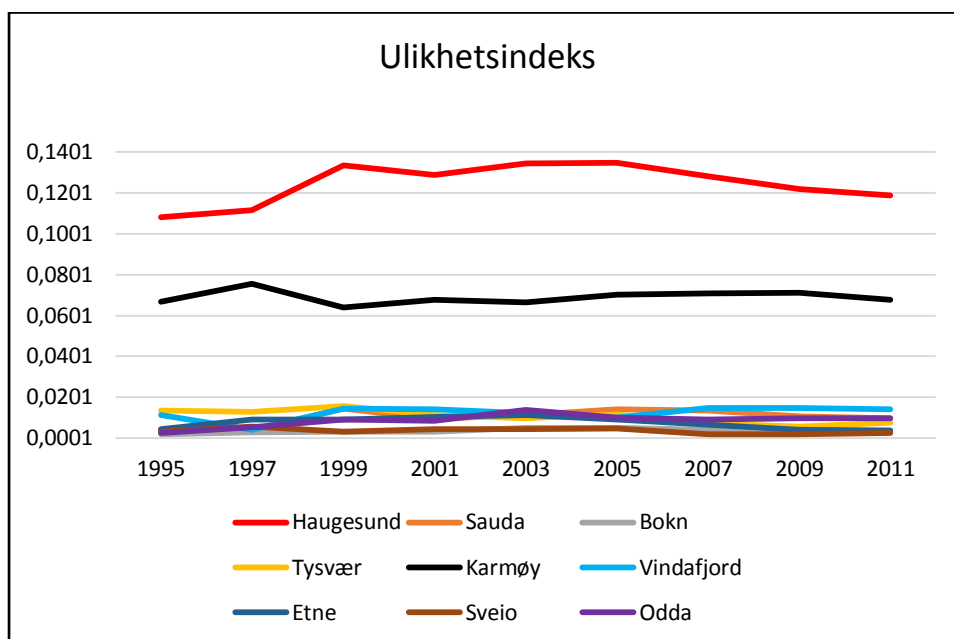
det enklere å sammenligne kommunene over tid. Indeksverdiene for kommunene er ikke oppgitt som vedlegg, men hvis det er ønskelig, kan man få dette tilsendt av forfatter.



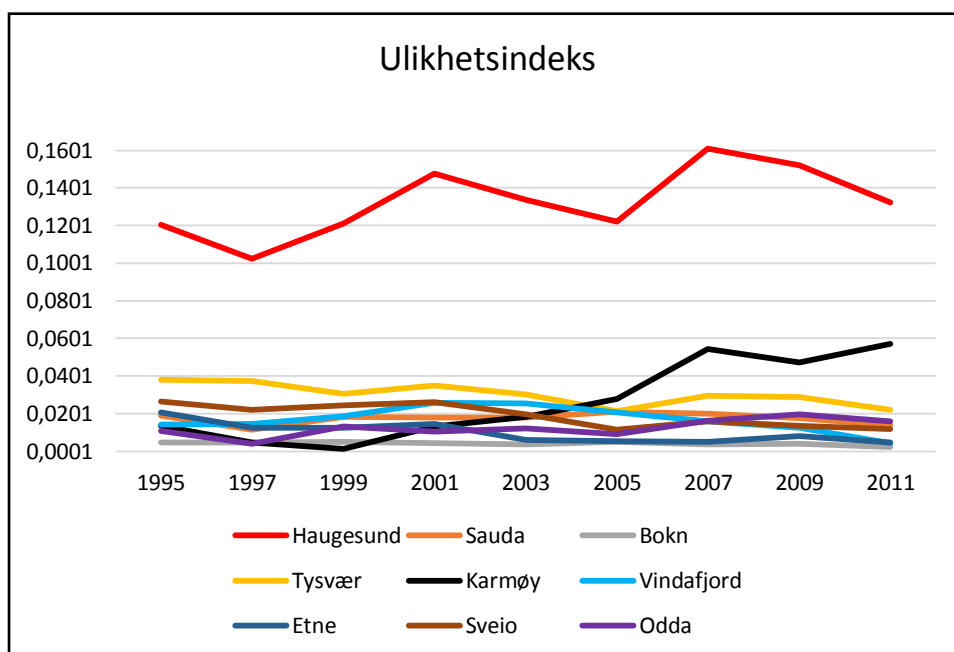
Figur 3.11. Viser ulikhetskurve for kommunene i region Haugaland, andel lav utdanning.



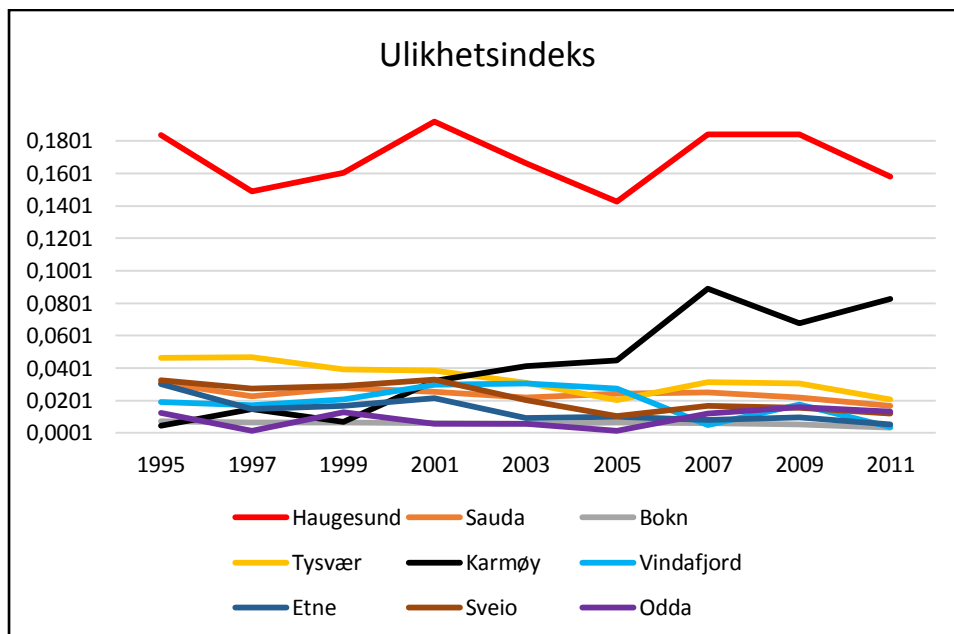
Figur 3.12. Viser ulikhetskurve for kommunene i region Haugaland, andel videregående utdanning.



Figur 3.13. Viser ulikhetskurve for kommunene i region Haugaland, andel høy utdanning.



Figur 3.14. Viser ulikhetskurve for kommunene i region Haugaland, andel total innvandring.



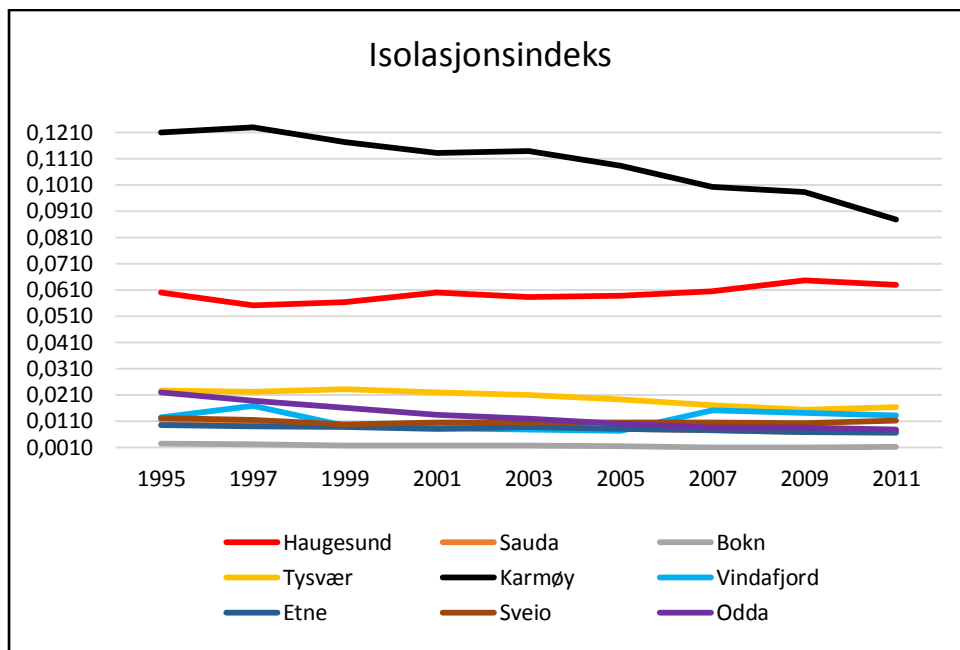
Figur 3.15. Viser ulikhetskurve for kommunene i region Haugaland, andel ikke-vestlig innvandring.

Mønsteret i ulikhetsindeksene på kommunenivå viser at andel videregående og høyere utdanning har økt, mens lav utdanning har redusert. Dette peker på at videregående og høyere utdanning har blitt mer segregert i perioden, det vil si at antall personer som tar videregående og høyere utdanning har økt. I samtlige figurer har bykommunen Haugesund og landkommunen Karmøy de høyeste verdiene. De resterende landkommunene har lave verdier. Populasjonsstørrelsen til disse kommunene er en del lavere enn for Haugesund og Karmøy, som kan være årsaken til at de ligger lavt i figurene (henviser til tabell 3.1). Forskjellen i indeksverdier mellom de minste kommunene er så å si null, dette kan sees tydelig ut i fra figurene.

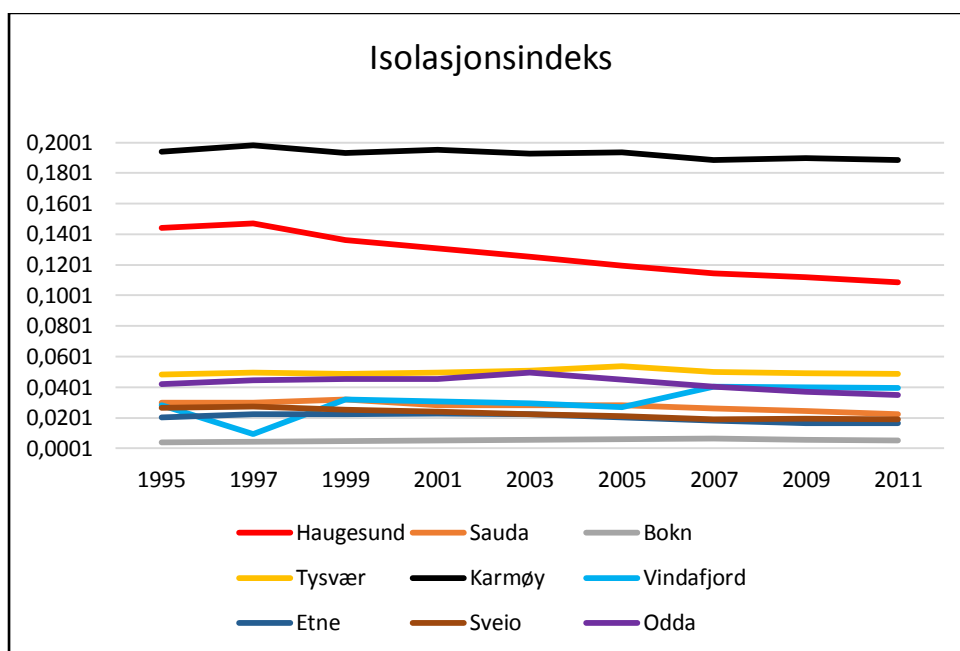
### 3.2.3 Isolasjonsindeks for utdanning og innvandring på kommunenivå

Figurene nedenfor viser isolasjonsverdier for bykommunen Haugesund og landkommunene i år 1995 – 2011, annethvert år.

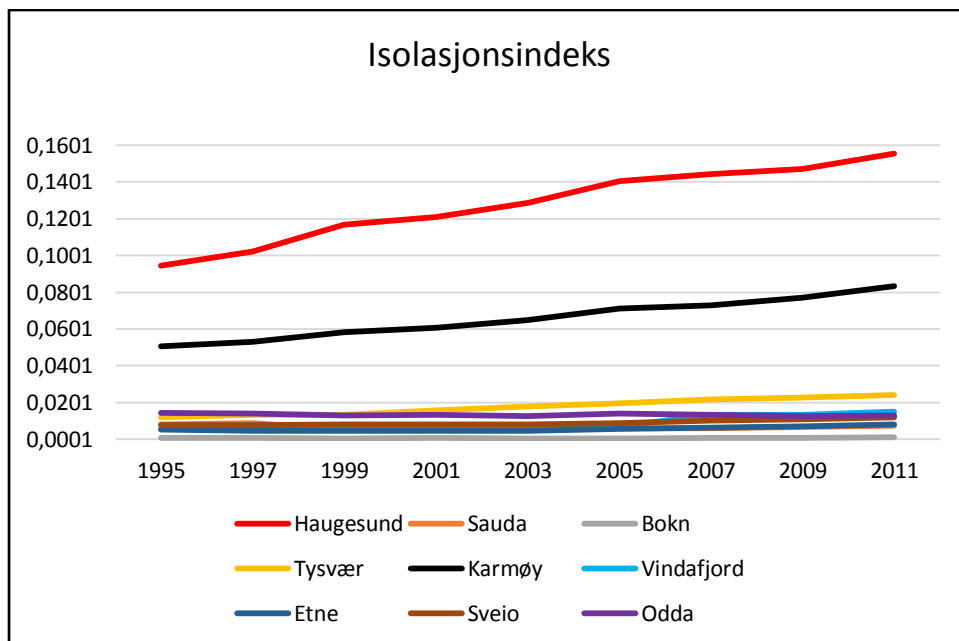




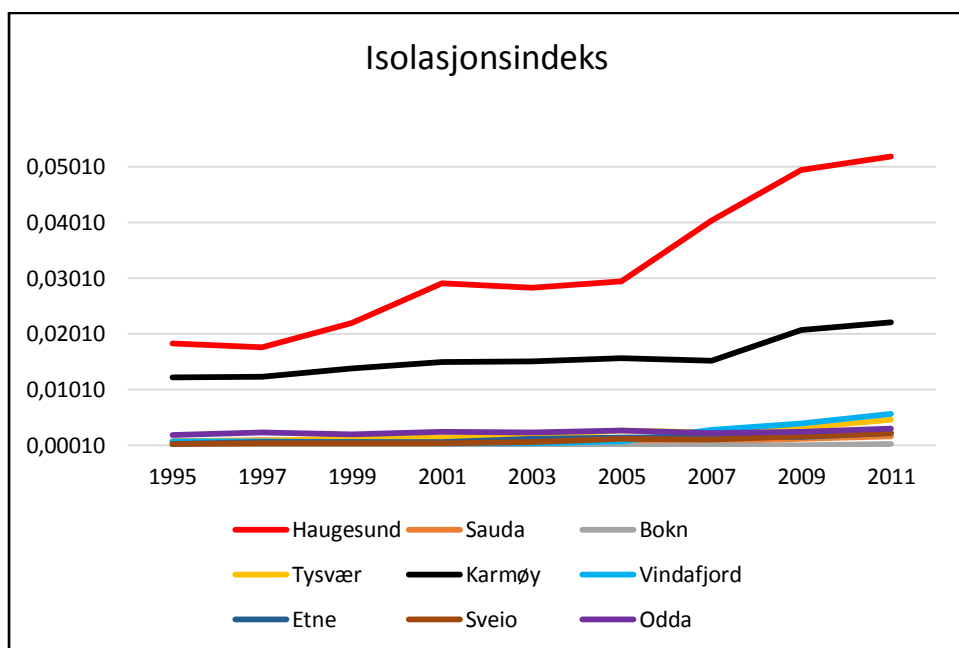
Figur 3.16. Viser isolasjonsindeks for kommunene i region Haugaland, andel lav utdanning.



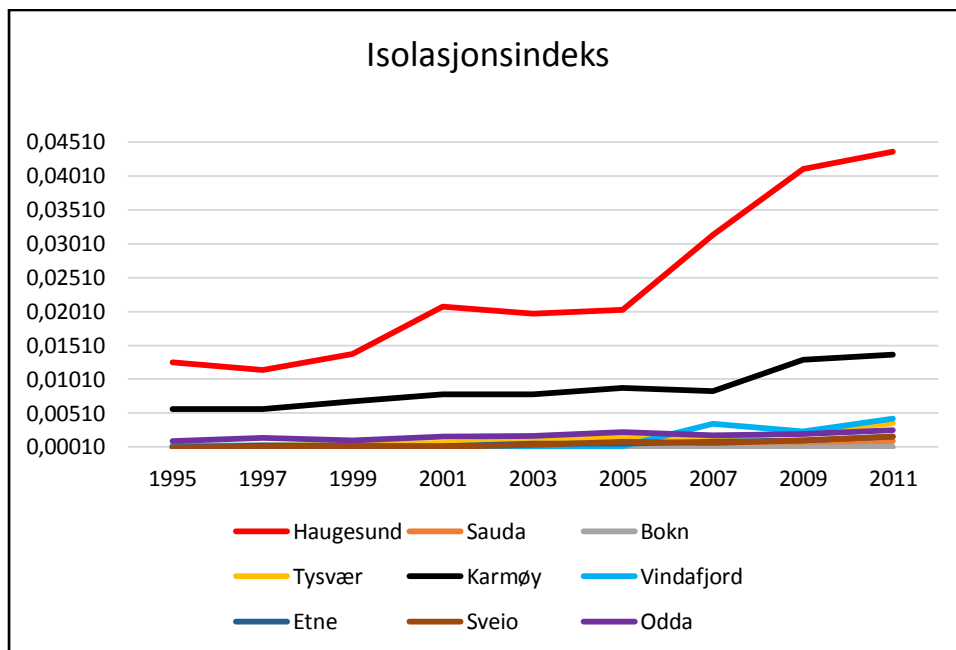
Figur 3.17. Viser isolasjonsindeks for kommunene i region Haugaland, andel videregående utdanning.



Figur 3.18. Viser isolasjonsindeks for kommunene i region Haugaland, andel høy utdanning.



Figur 3.19. Viser isolasjonsindeks for kommunene i region Haugaland, andel total innvandring.



Figur 3.20. Viser isolasjonsindeks for kommunene i region Haugaland, andel ikke-vestlig innvandring.

Isolasjonsindeksene på kommunenivå tyder på at andel lav utdanning har redusert, vært relativt jevn for videregående utdanning, og økt for høy utdanning. Dette peker på en reduksjon i bokkonsentrasjonen for lav og videregående utdanning, mens bokkonsentrasjonen til høyere utdanning tilsier økning. Total innvandring og ikke-vestlig innvandring viser en økning i perioden, noe som kan tyde på at disse har blitt mer segregert i kommunene.

Det må påpekes at hvis man hadde studert de samme andelene på grunnkrets nivå, ville trolig mønsteret vært annerledes. Som nevnt tidligere må man være forsiktig med tolkning av isolasjonsindeks dersom total populasjonsstørrelse endrer seg over tid, da indeksen vil være sensitiv ovenfor slike endringer. Populasjonsstørrelsene til samtlige kommuner har økt i perioden, derfor kan man ikke si med sikkerhet om segregeringsmønsteret har endret seg.

### **3.4 Sammenligning av ulikhets- og isolasjonsindeks med andre geografier**

Statistisk sentralbyrå har en oversikt over ulikhets- og isolasjonsindeks verdier for ikke-vestlige innvandrere bosatt i norske byer pr 01.01.2005 (Statistisk Sentralbyrå, u.å). Indeksene til SSB er beregnet på grunnkrets nivå, der de har regnet verdiene ut i fra at det gjennomsnittlig er 390 personer pr krets, og de har oppgitte verdier for 20 norske byer (Oslo er ikke med). Jeg har en grovere inndeling på indeksene, region- og kommunenivå. Det vil dermed være noe avvik mellom verdiene til SSB og mine. Verdiene er oppgitt i prosent. Ulikhetsindeksen til SSB viser at Drammen har den største verdien, samme gjelder for isolasjonsindeksen. Ulikhetsindeksen på grunnkrets nivå for Haugesund er 36,7, mens den er 14,26 på kommunenivå. Dette er en vesentlig forskjell. Isolasjonsindeksen på grunnkrets nivå for Haugesund er 10,0, mens den på kommunenivå er 2,039. Også her er det forskjeller, noe som vil være naturlig, da en finere inndeling av geografien vil gi høyere verdier enn en grovere inndeling. Beregning på grunnkrets nivå gir et bedre bilde av geografien, dette ser man tydelig ut i fra nivåforskjellene.

En tidligere bacheloroppgave som omhandler segregering i Stavanger har sett på andel total innvandrere på bydelsnivå (Bergtun & Schei, 2013). Studien deres benytter ulikhets- og isolasjonsindeks til å se om det har vært endringer i årene 2000 til 2012. Ulikhetsindeksen på bydelsnivå viser nedgang fra år 2000 til 2006, og fra år 2007 til 2012 ser man små tendenser til segregering. Ved sammenligning av byene Haugesund og Stavanger i samme tidsperiode ser jeg at Haugesund også har hatt en nedgang i ulikhetsindeksen fra år 2001 til 2005. I år 2007 stiger indeksen igjen, for å så gå litt ned igjen i år 2009 og år 2011. Jeg ser de samme tegnene i Haugesund og Stavanger for årene 2001 til 2007, begge geografiene viser tendenser til nedgang i segregering, for å så få en økning igjen. Isolasjonsindeksen for Stavanger viser at indeksen har steget fra år 2001 til år 2012. Haugesund har i samme periode (2001 – 2011) hatt en økning i indeksverdiene, med unntak i år 2003, der indeksen gikk noe ned. Dette tyder på at både Stavanger og Haugesund viser de samme tendensene, økning i segregering. Stavanger viser høyere segregeringstendenser enn Haugesund, noe som kan skyldes geografi inndelingen.

Ved beregning av indekser på ulike nivåer vil resultatene bli forskjellig. Ved beregning på regionnivå vil man typisk få svært lave indeksverdier, som er tilfellet i mine beregninger. På kommunenivå vil indeksverdiene også være preget av lave verdier, men disse vil vanligvis

være noe høyere enn på regionnivå, siden kommunenivå har en litt finere inndeling. Bydelsnivå vil typisk ha høyere segregeringsverdier enn kommune- og regionnivå, da denne inndelingen er mer nøyaktig enn de andre. Nivåforskjeller har betydning for tolkningen av indeks resultater, en inndeling på regionnivå kan gi et helt annet bilde av geografien enn på grunnkrets nivå. Når grad av segregering skal undersøkes vil data på grunnkrets nivå være mer optimalt, da disse verdiene gjenspeiler virkeligheten best.

Indeksberegningene gjort på region- og kommunenivå tilsier at graden av segregering i Haugaland er lav, det er små til moderate endringer i tidsperioden. På regionnivå kan det se ut som at grad av segregering blant lav utdanning har gått ned, mens det for videregående og høy utdanning har økt (ulikhetsindeks). Total innvandring og ikke-vestlige innvandring viser svingninger i perioden, men totalt sett kan det se ut som at grad av segregering har gått noe ned (ulikhetsindeks). Isolasjonsindeksene for variablene viser følgende tendenser; grad av segregering tilsier økning for høy utdanning, total innvandring og ikke-vestlige innvandring, mens grad av segregering er redusert for lav og videregående utdanning. På generelt kommunenivå har Haugesund og Karmøy de høyeste indeksverdiene, disse kommunene tyder også på størst økning i segregering. De resterende kommunene viser små endringer i segregering som er av liten betydning.

## 4 Metode

Metode stammer fra det greske ordet *methodos*, og betyr å følge en bestemt vei frem mot et mål (Johannessen et al., 2011). Samfunnsvitenskapelig metode brukes til å finne resultater ved å undersøke et sosialt fenomen. Med utgangspunkt i relevant teori og fakta for undersøkelsen kan man trekke konklusjoner og se om disse samsvarer med virkeligheten. En viktig del av den empiriske forskningen er å bruke innsamlet data til analysering og tolkning (Johannessen et al., 2011).

### 4.1 Valg av metode

Når et fenomen skal undersøkes er det i metodelæren vanlig å skille mellom kvalitativ og kvantitativ metode. I følge Johannessen et al (2011) sier kvalitativ metode noe om *''kvalitet eller spesielle kjennetegn/egenskaper ved det fenomenet som studeres''*. Denne metoden brukes på områder som har blitt forsket lite på, det vil si områder der man har lite informasjon om fenomenet, og når man ønsker å gå mer i dybden på fenomenet. Dybdeintervju er et eksempel på dette. Ved å benytte personlig intervju får man mer detaljert og spesifisert informasjon, enn om man hadde brukt et spørreskjema som sendes ut til mange. Spørreskjema brukes i kvantitativ metode. Kvantitativ tilnærming handler om å telle opp fenomener (Johannessen et al., 2011). Dette vil med andre ord si å kartlegge utbredelsen av et fenomen innenfor et større område. Denne metoden brukes i motsetning til kvalitativ, på områder hvor det eksisterer tidligere forskning og informasjon om fenomenet. Kvantitativ metode ser på ulike trekk og egenskaper ved fenomenet, men går ikke i dybden på disse.

I denne oppgaven har jeg brukt kvantitativ metode. Ulike metoder kan benyttes for å analysere datamaterialet innenfor kvantitativ metode. Jeg har sett på regresjonsanalyse og minste kvadraters metode.

## 4.2 Data/datainnsamling

Tallmaterialet for utdanning og innvandring er forklart tidligere i oppgaven, se avsnitt 3.2.

Tallmaterialet for boligene er hentet fra eiendomsregisteret og statistisk sentralbyrå (Osland, 2010) og er for tidsperioden 1997 - 2002. Dataene inneholder informasjon om følgende: salgspris, postnummer, tomteareal, alder, nettoareal, byggeår, boligtype, garasje, antall WC, avstand fra sentrum, tilgjengelighet til arbeidsplasser, og dummyvariabler. Tallmaterialet er for kommunene Haugesund, Karmøy, Bokn, Tysvær, Sveio, Ølen, Vindafjord, og Etne, som ligger i region Haugaland. Nedenfor vises beskrivende statistikk for forskningsområdet.

Variable	Whole Area	Haugesund
Price	1,101,467 (469,809)	1,212,030 (525,935)
Price per square meter	7,747 (2,587)	8,410 (2,619)
Lot size	813 (504)	719 (409)
Size of house	146 (52)	147 (53)
Number of toilets	2 (0.69)	2 (0.72)
Minutes to CBD by car	16 (14)	5 (2)
Garage	0.63 (0.48)	0.64 (0.48)
Age of house	32 (31)	39 (34)
<i>n</i>	1,691	766

Note: Mean values, standard deviation follow in parentheses.

Tabell 4.1. Beskrivende statistikk. Hentet fra Osland (2010):

<http://ideas.repec.org/a/jre/issued/v32n32010p289-320.html>. Gjengitt med tillatelse.

### 4.3 Regresjon

Ubøe, (2009, s. 243) forklarer regresjonsanalyse på følgende måte:

Når vi gjør observasjoner, finner vi ofte problemstillinger der observasjonene har en stigende eller fallende trend. Dette kan for eksempel skyldes at vi har gjort observasjonene over tid og at tidsaspektet har gitt systematiske endringer i tallmaterialet. For å avgjøre om en slik tendens er tilfeldig eller ikke, bruker vi regresjonsanalyse.

Regresjonsanalyse opererer med en variabel som er avhengig (Y) og en eller flere uavhengige variabler (X). De uavhengige variablene er forklaringsvariabler og brukes til å undersøke hvilke innvirkning de har på den avhengige variabelen, samt å se om det eksisterer noen sammenhenger mellom variablene. I følge Johannessen et al (2011) finnes det ulike metoder for å gjennomføre en regresjonsanalyse. Jeg benytter meg av multippel regresjonsanalyse og minste kvadraters metode. Denne modellen forklarer variasjonen i den avhengige variabelen ved hjelp av flere uavhengige variabler. Sagt med andre ord vil man kunne se hvordan prisen på boligen endres når en av boligens attributter endres, mens resten av attributtene holdes konstant.

Formel for multippel regresjon (Ubøe, 2009, s. 259):

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_r X_r + e$$

$\alpha$  er konstantleddet,  $\beta_1, \beta_2, \dots, \beta_r$  er regresjonskoeffisienter, og  $X_1, X_2, \dots, X_r$  er forklaringsvariabler. Leddet  $e$  står for error, det vil si de uobserverte faktorene, residualene, som er den resterende variasjonen i boligpriser som forklaringsvariablene ikke fanger opp. Når man beregner regresjonskoeffisientene vil regresjonsanalysen ta utgangspunkt i de kvadrerte residualene, og deretter kvadrere disse av alle observasjonene for å få kvadratsummen til residualene. Ved å anvende minste kvadraters metode beregner regresjonsanalysen regresjonskoeffisientene  $\beta_1, \beta_2, \dots, \beta_r$  for å gjøre det samlede kvadratiske avviket så lite som mulig. Regresjonskoeffisientene vil uttrykke den regresjonslinjen som passer best til observasjonene (Johannessen et al., 2011).



Når multippel regresjonsanalyse gjennomføres får man opp et sett av verdier som tolkes ut i fra om verdiene er av signifikans. De verdiene jeg anser som viktige når resultatene skal tolkes er  $R^2$ , justert  $R^2$ , regresjonskoeffisienter, T-verdier, og P-verdier.

$R^2$  er et mål som viser hvor god modellen er til å predikere den avhengige variabelen.  $R^2$  beregner andelen av variansen til Y som kan forklares av regresjonsmodellen. Sagt på en annen måte viser  $R^2$  hvor mye av spredningen i den avhengige variabelen (Y) som kan skyldes variasjon i de uavhengige variablene (X) (Johannessen et al., 2011).  $R^2$  har verdier som ligger mellom 0 til 1. En verdi på 0 betyr at punktene er spredt tilfeldig rundt regresjonslinjen, det vil si at ingenting av variasjonen i den avhengige variabelen vil være forklart av de uavhengige variablene. Er  $R^2$  derimot 1 vil alle punktene ligge på regresjonslinjen, som vil si at all variasjon i den avhengige variabelen vil være forklart av de uavhengige variablene.

Justert  $R^2$  har den samme tolkningen som den vanlige  $R^2$ , men denne korrigerer for antall variabler som inngår i modellen (Johannessen et al., 2011). Det vil si at når flere uavhengige variabler inkluderes i modellen vil  $R^2$  stige, men ikke justert  $R^2$ .

Regresjonskoeffisientene viser i gjennomsnitt hvor mye den avhengige variabelen endres når de uavhengige variablene øker med en verdienhet (Johannessen et al., 2011). I multippel regresjonsanalyse kan koeffisientene sammenlignes for å finne den uavhengige variabelen som har høyest effekt. For hver ny uavhengig variabel som blir lagt inn i modellen vil koeffisientene endre seg, og dermed også resultatene. Effekten av regresjonskoeffisientene uttrykkes med de opprinnelige målestokkene for variablene (Johannessen et al., 2011). Målestokken i denne oppgaven er antall kroner. Verdien til en regresjonskoeffisient vil vise hvor mye boligprisen endrer seg for de uavhengige variablene. Når en variabel endrer seg vil effekten av dette gjenspeiles i boligprisen, enten i positiv eller negativ retning.

T-verdi er en statistisk test som tar hensyn til antall observasjoner i utvalget. Signifikanstesten avdekker signifikante størrelser, det vil si at man kan se hvilke av de uavhengige variablene som har en t-verdi som er statistisk signifikant. For å vite når verdien er signifikant og at det ikke skyldes tilfeldigheter, må et nivå for t-verdien velges. Jeg har valgt et konfidensintervall på 90%, som vil si en t-verdi på 1,645. Hvis den observerte t-verdien er større enn  $\pm 1,645$  vil utslagene være av signifikans. For å kunne se om innvirkningen på avhengig variabel er av signifikant størrelse må en nullhypotese velges. I følge Ubøe (2009) tar man utgangspunkt i at

nullhypotesen er rett, og finner sannsynligheten for at observatoren avviker mer enn den observerte verdien. Dette kalles en P-verdi og blir målt ut i fra observert verdi (beregnes ut ifra t-verdien). Man ser altså om det er sammenheng mellom avhengig variabel og de uavhengige variablene ved å teste nullhypotesen. Nivået på 10% vil gi en P-verdi på 0,1. Når P-verdien er lavere enn signifikansnivået 0,1 vil nullhypotesen forkastes. Når nullhypoteser forkastes kan man påstå at resultatene er statistisk signifikant, og at det faktisk er sammenhenger mellom avhengig og uavhengig variabel.

### 4.3.1 Grunnlagsmodellen

Hedonisk metode kan benyttes til å undersøke hvordan ulike attributter påvirker boligprisen. Man benytter den hedoniske prisfunksjonen til å se hvordan boligpriser endres når man legger til ulike attributter som er relatert til boligen. Jeg har tatt utgangspunkt i en basemodell fra en artikkel som ser på spatial økonometri i forhold til hedoniske boligpriser (Osland, 2010). Denne artikkelen benytter følgende log-log modell for estimering av dataene:

$$\begin{aligned}
 \log P_{it} = & \beta_0 + \beta_1 \log LOTSIZE_i + \beta_2 (RUR \cdot \log LOT_i) \\
 & + \beta_3 \log AGE_i + \beta_4 \log sqAGE_i + \beta_5 (REBUILD \cdot \log AGE_i) \\
 & + \beta_6 GARAGE_i + \beta_7 \log LIVAREA_i + \beta_8 \log TOILETS_i \\
 & + \beta_9 \log MINCBD_i + \beta_{10} \log ACCESS_i \\
 & + \sum_{t=1997}^{02} \beta_t YEARDUM_{it} + \varepsilon_{it}.
 \end{aligned} \tag{8}$$

Tabell 4.2. Estimeringsformel for regresjonsanalyse. Hentet fra Osland (2010):

<http://ideas.repec.org/a/jre/issued/v32n32010p289-320.html>. Gjengitt med tillatelse.

Ønsker man for eksempel å se hvordan variabelen areal i formelen påvirker boligprisen,  $\beta_7 \log LIVAREA_i$ , der det antas at  $\beta$  er 0,20, vil dette bety at 1% endring i areal gir 0,20% endring i boligprisen. Sagt på en annen måte, for hver kvadratmeter økning vil prisen på

boligen stige med 0,20%. Legges det for eksempel til en segregeringsvariabel i tillegg, vil man kunne se om denne også virker inn på boligprisen.

Det er tidkrevende og vanskelig å lage en god regresjonsmodell, og siden basemodellen til Osland er en godt utarbeidet og testet modell, bruker jeg denne til regresjonsanalysene. Formelen ovenfor viser alle variablene som inngår i Oslands modell.

Når man skal utføre regresjon er det vanlig å velge en funksjonsform. Dette kan for eksempel være lin-lin, log-lin, eller en log-log modell, hvor sistnevnte er benyttet i denne oppgaven. Ved å anvende log-log som funksjonsform kan man predikere en verdi på den avhengige variabelen når den uavhengige variabelen har en bestemt verdi (Johannessen et al., 2011).

Forklaring av variabler i modellen:

Variable	Definition
<i>P</i>	Dependent variable. Selling price in NOK, deflated by the consumer price index; base year is 1998.
<i>LOTSIZE</i>	Lot size measured in square meters.
<i>RUR · LOT</i>	Lot size multiplied by a dummy variable, which takes the value of 1 if the house is located in a rural area, else 0. A rural area is defined on the basis of population density. All municipalities outside Haugesund are rural areas, except for houses located in the mainland area of Karmøy.
<i>AGE</i>	Age of building, measured in years.
<i>sqAGE</i>	Age of building squared.
<i>REBUILD AGE</i>	A dummy variable indicating if the building has been rebuilt / renovated or not multiplied by the age-variable.
<i>GARAGE</i>	A dummy variable indicating whether the house has a garage or not.
<i>LIVAREA</i>	Size of house measured in square meters.
<i>TOILETS</i>	Number of toilets in the house.
<i>MINCBD</i>	Distance from the center of the largest city, Haugesund. Distance is measured by minutes between postal codes, travelling by car, accounting for speed limits.
<i>ACCESS</i>	Access to work places is measured by $S_j = \sum_{k=1}^{55} D_k \gamma \exp(\sigma d_{jk})$ . $D_k$ represents number of jobs in postal zone $k$ , $d_{jk}$ represents minutes between zone $j$ and $k$ , traveling by car. $\sigma$ and $\gamma$ are parameters to be estimated by Maximum Likelihood estimation using the base model. The estimated values on the two parameters are $-1.172$ (4.31) and $0.1107$ (1.94), respectively, with robust z-values in parenthesis. The <i>ACCESS</i> variable is computed based on these imputed values.
<i>YEARUM</i>	A dummy variable indicating the year the house has been sold (1997–2002). The dummy variable for 1998 is excluded in order to avoid perfect multicollinearity.

Tabell 4.3. Definisjoner av variabler i basemodellen. Hentet fra Osland (2010):

<http://ideas.repec.org/a/jre/issued/v32n32010p289-320.html>. Gjengitt med tillatelse.

Som tabellen ovenfor viser er den avhengige variabelen pris, mens de andre variablene er uavhengige forklaringsvariabler. Når jeg har utført regresjonsanalyser har jeg brukt variablene i grunnlagsmodellen (estimeringsformel, tabell 4.2). Jeg har deretter lagt til nye variabler, for å kunne studere om disse nabolagstrekkene har noen innvirkning på boligpriser. De nye segregeringsvariablene er andeler, det vil si utdanningsandeler og innvandringsandeler. I de siste regresjonene vil alle de opprinnelige variablene pluss de nye variablene (andelene) være i en og samme modell. Da kan man se hvilke variabler som er signifikante, og igjen sette de signifikante variablene i en ny regresjon, for å studere ytterligere endringer.

Nedenfor vises resultatene til Oslands basemodell:

Variables	OLS Base Model (robust t-values)
Constant	11.016 (74.35)
logLOTSIZE	0.107 (7.34)
RUR · logLOT	-0.018 (-5.43)
logAGE	-0.061 (-3.80)
logsqAGE	-0.017 (-4.59)
REBUILD·logAGE	0.021 (4.79)
GARAGE	0.103 (7.43)
logLIVARE	0.436 (16.61)
logTOILETS	0.171 (8.43)
logMINCBD	-0.077 (-5.29)

Variables	OLS Base Model (robust t-values)
logACCESS	0.257 (4.91)
YEARDUM97	-0.101 (-4.95)
YEARDUM99	0.135 (6.15)
YEARDUM00	0.231 (10.71)
YEARDUM01	0.314 (15.38)
YEARDUM02	0.410 (19.77)

Tabell 4.4. Presentasjon av resultatene for basemodellen (grunnmodell). Hentet fra Osland (2010): <http://ideas.repec.org/a/jre/issued/v32n32010p289-320.html>. Gjengitt med tillatelse.

Tabellen viser alle variablene som inngår i modellen, verdiene til regresjonskoeffisientene og t-verdier i parentes. Variabelen logLIVARE har de høyeste verdiene, en koeffisient på 0,436, og en t-verdi på 16,61.  $R^2$  og justert  $R^2$  vises ikke i tabellen, men begge er relativ høye, henholdsvis 0,7052 og 0,7025 (Osland, 2010). De estimerte parameterne er signifikante.

Før en starter med regresjonsanalyse er det lurt å ha noen formeninger om variablene som inngår i modellen, det vil i dette tilfellet si hvilke forventninger jeg har til de ulike andelene som blir tilføyd den opprinnelige basemodellen. Som utgangspunkt beskriver andelsvariablene trekk ved et nabolag. Det er ikke slik at man ikke vil bo i nærheten av lavt utdannede folk, eller at man vil bo i nærheten av høyt utdannede folk. Variablene fanger opp noe ved nabolaget, men nøyaktig hva dette er, er usikkert. Antakelsene om variablene nedenfor er basert på et helt generelt perspektiv.

Innenfor utdanningskategoriene antar jeg at hvis andelen lav utdanning er signifikant, så forventes den muligens å ha negativ innvirkning på boligpriser. Motsatt er tilfellet for andelen høy utdanning, hvis variabelen er signifikant vil denne mest sannsynlig forventes å ha positiv effekt på boligpriser. Variabelen videregående utdanning er litt vanskeligere å predikere, men er denne signifikant vil andelen trolig ha en liten tendens til positiv innvirkning på boligpriser. I innvandringskategoriene vil antageligvis andel total innvandring ha lite eller ingen positiv innvirkning på boligpriser, hvis denne er av signifikans. Andelen til ikke-vestlige innvandrere er svært lav, og hvis denne variabelen er signifikant, vil muligens denne ha en negativ effekt på boligpriser. Generelt sett; når grad av segregering er stor, forventes variablene å gi utslag på boligpriser, mens når grad av segregering er liten, forventes variablene å ha liten innvirkning. Med tanke på at grad av segregering for andelene på region- og kommunenivå var lav, forventes andelene i regresjonsmodellene å ha lite effekt på boligprisene. Det vil si at den kvantitative effekten forventes ikke å være stor.

## 5 Resultater fra regresjonsanalysene

Regresjonsanalysene er gjennomført i Excel, med unntak fra de to siste regresjonsanalysene, som er utført i analyseprogrammet STATA. Tallmaterialet som er anvendt i regresjonene er Oslands grunnmodell. Modellen er basert på eneboliger, solgte eneboliger i tidsperioden 1997 – 2002. Antall observasjoner er 1691. Videre er det i tallmaterialet lagt inn grunnkretser, andel for grunnskole, videregående og høy utdanning, samt andeler på total innvandring og ikke-vestlig innvandring.

Regresjonsanalysene er gjennomført på hver enkelt segregeringsvariabel, det vil si at de 5 første analysene inneholder grunnmodellen pluss en andel hver. Tabell 5.1. og 5.2. viser koeffisientene og t-verdiene til de uavhengige utdanningsvariablene, samt  $R^2$  og justert  $R^2$ .

Tabell 5.1. Resultat av regresjonsanalyse, grunnmodell og utdanningskategorier.

Uavhengige variabler	Grunnmodell Koeffisient (t-verdi)	Lav utdanning Koeffisient (t-verdi)	Videregående utdanning Koeffisient (t-verdi)	Høy utdanning Koeffisient (t-verdi)
Skjæringspunkt	11,0700727 (88,01)	11,1052503 (86,415)	11,10259429 (86,956)	11,13389562 (86,753)
Intomt	0,09301535 (8,317)	0,09388192 (8,382)	0,093619764 (8,368)	0,092807229 (8,310)
InruralTomt	-0,022247449 (-5,909)	-0,022595373 (-5,988)	-0,022446131 (-5,960)	-0,021693016 (-5,759)
Inalder	-0,06864437 (-3,915)	-0,071046337 (-4,032)	-0,071625093 (-4,060)	-0,072152506 (-4,107)
SqInalder	-0,016494238 (-4,706)	-0,016014244 (-4,545)	-0,01581206 (-4,473)	-0,015701942 (-4,466)
Inalderomb	0,020917055 (5,335)	0,021088284 (5,377)	0,02109914 (5,381)	0,021111828 (5,392)
Innareal	0,441159796 (21,081)	0,441289751 (21,092)	0,441482396 (21,103)	0,441046921 (21,106)
Inwc	0,168261982 (9,692)	0,166999443 (9,607)	0,166909871 (9,604)	0,166711156 (9,610)
Garasjedum	0,106488388 (8,069)	0,106826761 (8,095)	0,106609608 (8,081)	0,106607559 (8,090)
dum1997 <sup>1</sup>	-0,073953646 (-3,844)	-0,072750029 (-3,778)	-0,072750454 (-3,780)	-0,072156411 (-3,754)

<sup>1</sup> "dum" står for dummyvariabel. Dummyvariablene må kodes om til å ha verdien 0 eller 1 for å kunne benyttes i regresjonsanalyser. I hver regresjonsmodell må én tidsdummy utelates for å unngå multikollinearitet, dum1998 er tatt ut av alle modellene.

dum1999	0,112330665 (5,391)	0,113360981 (5,438)	0,113505289 (5,445)	0,113264579 (5,443)
dum2000	0,177865234 (9,192)	0,178323383 (9,216)	0,178319707 (9,217)	0,177505272 (9,186)
dum2001	0,230841397 (11,788)	0,232054172 (11,840)	0,231845871 (11,836)	0,231625968 (11,844)
dum2002	0,311130029 (14,871)	0,31281502 (14,928)	0,312689281 (14,932)	0,311886756 (14,927)
InMINCBD	-0,073798682 (-5,338)	-0,077424618 (-5,495)	-0,077319739 (-5,512)	-0,078231389 (-5,617)
InACCESSIBILI TY	0,210697828 (5,925)	0,215175913 (6,025)	0,218311705 (6,077)	0,216476076 (6,083)
<b>Inandellavutd</b>		<b>0,00659945 (1,327)</b>		
<b>Inandelvidutd</b>			<b>0,007918523 (1,463)</b>	
<b>Inandelhøyutd</b>				<b>0,01075922 (2,417)</b>
R <sup>2</sup>	0,6828	0,6832	0,6832	0,6839
Justert R <sup>2</sup>	0,6800	0,6801	0,6802	0,6809

I grunnmodellen er alle variablene signifikante, der netto areal er mest signifikant (koeffisient og t-verdi). R<sup>2</sup> er 0,6828 som betyr at 68,28 % av variansen forklares av regresjonsmodellen, og når det justeres for antall variabler i modellen er R<sup>2</sup> 68%. Grunnmodellen viser at koeffisienten til alder er negativ, det vil si at boligprisen synker jo eldre boligen er. MINCBD har også negativ koeffisient som tyder på at jo lengre bort fra sentrum man beveger seg, jo lavere vil boligprisene være. En bolig som er tilknyttet en stor tomt vil vanligvis ha høyere boligpris enn boliger med liten tomt. Dette er gjenspeilt i grunnmodellen, der koeffisienten til tomt er positiv. Variabelen WC har også en positiv koeffisient, antall toaletter i en bolig har en innvirkning på boligpriser. Hvis en bolig har garasje vil dette ha en positiv effekt på boligprisen, boligprisen vil stige. I dummyvariablene mangler tidsdummiene for år 1998, dette gjelder alle regresjonsmodellene, og skyldes at år 1998 er referanseåret. Dette vil si at årsdummiene tolkes i forhold til referanseåret. Alle dummyvariablene er signifikante og har positive koeffisienter (unntatt dum1997). I modellen stiger tidsdummiene hvert år (koeffisient og t-verdi), dette skyldes trolig stigende boligpriser.

Nullhypotesen i regresjonsmodellene er at det ikke er noen sammenheng mellom boligpris og hver enkel variabel. Er t-verdi i absolutt verdi større enn 1,645 forkastes nullhypotesen og variablene er signifikante. I grunnmodellen er alle t-verdiene til de uavhengige variablene

signifikante. Dette gjelder også for regresjonsanalysene gjort på de tre utdanningskategoriene (variablene fra grunnmodellen).

Når andel lav utdanning blir inkludert i grunnmodellen (uavhengig variabel, tabell 5.1) vises det at koeffisienten til denne er liten, 0,00659, men positiv. Dette tilsier at variabelen har en liten positiv innvirkning på boligpris. T-verdien er 1,327, noe som ikke er signifikant og nullhypotesen beholdes.

I regresjonsanalysen som inkluderer grunnmodellen og forklaringsvariabel videregående utdanning, er koeffisienten til andelen noe høyere enn for lav utdanning, 0,00791. Videregående utdanning har trolig en høyere økonomisk innvirkning på boligpriser enn lav utdanning; en prosentvis økning i andelen tilsvarer 0,0079% økning i boligpriser. T-verdien er 1,463, denne er for liten til at nullhypotesen forkastes.

Den siste regresjonsanalysen innenfor utdanningskategoriene er høy utdanning. Koeffisienten til denne andelen er 0,01075, noe som er høyere enn lav og videregående utdanning. Høy utdanning antyder dermed en positiv økonomisk effekt på boligpriser. T-verdien er høyere enn signifikansnivået på  $\pm 1,645$ , nullhypotesen må dermed forkastes, som tilsier at det er en sammenheng mellom boligpris og høy utdanning.

Ser man på  $R^2$  verdiene er disse for utdanningsandelene litt høyere enn grunnmodellen, der høy utdanning har den største verdien, en forklaringskraft på 68,39%. Det samme gjelder for justert  $R^2$ . P-verdiene er ikke oppgitt i tabellene siden alle P-verdiene til de uavhengige variablene som opprinnelig inngår i grunnmodellen er 0 (signifikant). P-verdiene til lav, videregående og høy utdanning er henholdsvis 0,18, 0,14, og 0,02. P-verdien til høy utdanning viser at variabelen bidrar signifikant og dette stemmer med tilhørende t-verdi.



Tabell 5.2. Resultat av regresjonsanalyse, innvandringskategorier.

Uavhengige variabler	Total innvandring Koeffisient (t-verdi)	Ikke-vestlig innvandring Koeffisient (t-verdi)
Skjæringspunkt	11,10791463 (85,637)	11,0636152 (86,749)
Intomt	0,093590317 (8,361)	0,092690339 (8,249)
InruralTomt	-0,022116342 (-5,873)	-0,022230237 (-5,902)
Inalder	-0,069972237 (-3,984)	-0,068433141 (-3,899)
SqInalder	-0,016343245 (-4,660)	-0,016514484 (-4,709)
Inalderomb	0,021135326 (5,386)	0,020813699 (5,288)
Innareal	0,44100307 (21,076)	0,441411571 (21,072)
Inwc	0,167585852 (9,649)	0,16821823 (9,687)
Garasjedum	0,106369157 (8,061)	0,106569748 (8,072)
dum1997	-0,073358461 (-3,813)	-0,073957447 (-3,844)
dum1999	0,112656506 (5,407)	0,112359386 (5,391)
dum2000	0,177040447 (9,144)	0,178401144 (9,180)
dum2001	0,23005663 (11,743)	0,231056478 (11,788)
dum2002	0,311398289 (14,885)	0,310923995 (14,850)
InMINCBD	-0,076568755 (-5,462)	-0,073541088 (-5,308)
InACCESSIBILITY	0,210916574 (5,932)	0,210504065 (5,917)
<b>Inandeltotinnv</b>	<b>0,004571142 (1,191)</b>	
<b>Inandelikkevestinnv</b>		<b>-0,000850919 (-0,308)</b>
R <sup>2</sup>	0,6831	0,6828
Justert R <sup>2</sup>	0,6801	0,6798

Tabell 5.2. viser resultatene fra de to regresjonsanalysene som inneholder andelene total innvandring og ikke-vestlig innvandring. Begge analysene er utført med utgangspunkt i grunnmodellen.

Koeffisienten til andel total innvandring er 0,00457, noe som tilsier at andelen har en liten positiv økonomisk påvirkning på boligpriser. En prosentvis økning i andelen tilsvarer 0,00457% økning i boligpriser. T-verdien er 1,191, dette bekrefter at andel total innvandring ikke har noe sammenheng med boligpriser, og dermed forkastes ikke denne nullhypotesen. I regresjonsanalysen for forklaringsvariabelen ikke-vestlig innvandring har andelen negativ regresjonskoeffisient og t-verdi. Siden koeffisienten er negativ, antyder dette at variabelen har en negativ effekt på boligpriser. T-verdien er ikke signifikant og nullhypotesen beholdes også i dette tilfellet.

P-verdiene til total og ikke-vestlig innvandring har verdier på henholdsvis 0,23 og 0,76. Justert R<sup>2</sup> for total innvandring har lik forklaringskraft som andel lav utdanning, og ikke-vestlig innvandring har den laveste variansen av modellene (tabell 5.1 og 5.2). Det bør nevnes at andelene for total og ikke-vestlig innvandring er lavere enn andelene for utdanningskategoriene, disse vil dermed ha lavere verdier i regresjonsanalysene.

Den neste regresjonsanalysen er utført i statistikk programmet STATA, og inneholder grunnmodellen samt alle utdannings- og innvandrings andelene. Jeg har satt alle andelene i en og samme modell for å se om dette endret de resultatene jeg fikk i de første regresjonsanalysene, og om eventuelt flere andeler ble signifikante. Tabell 5.3. viser resultatene av denne modellen.

*Tabell 5.3. Resultat av regresjonsanalyse, en samlet modell som inkluderer grunnmodell variabler, samt andeler for utdanning og innvandring.*

<b>Uavhengige variabler</b>	<b>Samlet modell Koeffisient (t-verdi)</b>
Skjæringspunkt	11,14289 (85,60)
Intomt	0,0855666 (7,52)
InruralTomt	-0,0177753 (-4,45)
Inalder	-0,0664591 (-3,76)
SqInalder	-0,0168437 (-4,74)
Inalderomb	0,0200586 (5,11)

Innareal	0,4406373 (21,09)
Inwc	0,1685725 (9,71)
Garasjedum	0,1064597 (8,08)
dum1997	-0,0726758 (-3,79)
dum1999	0,1098721 (5,28)
dum2000	0,1766259 (9,08)
dum2001	0,2287375 (11,62)
dum2002	0,3029762 (14,38)
InMINCBD	-0,072307 (-5,10)
InACCESSIBILITY	0,1923707 (5,19)
<b>Inandellavutd</b>	<b>-0,0105048</b> <b>(-0,67)</b>
<b>Inandelvidutd</b>	<b>-0,0393744</b> <b>(-1,85)</b>
<b>Inandelhøyutd</b>	<b>0,0521188</b> <b>(3,55)</b>
<b>Inandeltotinnv</b>	<b>0,0038419</b> <b>(0,51)</b>
<b>Inandelikkevestinnv</b>	<b>-0,0066476</b> <b>(-1,79)</b>
R <sup>2</sup>	0,6860
Justert R <sup>2</sup>	0,6823

Hvis man sammenligner den siste regresjonsanalysen med de modellene som har andeler i seg (tabell 5.1 og 5.2), har samlet modell de laveste koeffisient verdiene, som tilsier en lavere effekt enn de tidligere analysene. Sammenlignes t-verdiene, har de fleste verdiene også her gått noe ned. I tabell 5.3 ser man at når alle andelene inkluderes i samme modell, endres resultatene for disse. Koeffisienten til lav utdanning sammenlignet med den første regresjonsanalysen (tabell 5.1) har endret fortegn, noe som tyder på at denne variabelen nå har en negativ innvirkning på boligpriser. T-verdien har også endret fortegn (-0,67), men er ikke signifikant. Andel videregående utdanning har en regresjonskoeffisient på -0,03937, altså en negativ økonomisk effekt på boligpriser som tilsier at 1% endring i andelen gir -0,039% endring i boligpriser.

T-verdi er -1,85 og nullhypotesen forkastes, andelen er av signifikant størrelse. Forklaringsvariabel høy utdanning er fortsatt signifikant, og tyder dermed på at det er en positiv sammenheng mellom høyere utdanning og boligpriser. Total

innvandring har en koeffisient verdi på 0,00384, noe som er lavere enn for den tidligere regresjonsanalysen utført på andelen (tabell 5.2). T-verdien til total innvandring er ikke signifikant. Ikke-vestlig innvandring er i denne modellen høyere enn signifikansnivået på  $\pm 1,645$ , dermed forkastes nullhypotesen, og dette indikerer at det er en negativ sammenheng mellom ikke-vestlig innvandring og boligpriser.

P-verdiene til andelene er som følger: lav utdanning 0,506, videregående utdanning 0,065, høy utdanning 0,00, total innvandring 0,608, og ikke-vestlig innvandring 0,074. Disse verdiene bekrefter det t-verdiene antyder; at videregående utdanning, høy utdanning, og ikke-vestlig innvandring er signifikante med verdier under 0,1. Andelene lav utdanning og total innvandring har ikke verdier som er av signifikans. Justert  $R^2$  i samlet modell har den høyeste verdien av alle regresjonsanalysene, forklaringskraften til denne modellen er 68,23%.

Som vist i tabell 5.3, endret resultatene seg da alle variablene ble inkludert i en og samme modell. Hvis man nå lager en regresjonsanalyse som kun består av de variablene som ble signifikante i den samlede modellen (tabell 5.3), vil resultatene endre seg ytterligere? Jeg laget en siste modell for å få svar på dette, og resultatene for denne modellen vises nedenfor.

*Tabell 5.4. Resultat av den siste regresjonsanalysen, en modell bestående av grunnmodell variabler, andel ikke-vestlig innvandring, videregående og høy utdanning.*

<b>Uavhengige variabler</b>	<b>Endelig regresjonsmodell Koeffisient (t-verdi)</b>
Skjæringspunkt	11,14231 (86,63)
Intomt	0,08625 (7,62)
InruralTomt	-0,0183113 (-4,66)
Inalder	-0,0666625 (-3,78)
SqInalder	-0,0168043 (-4,74)
Inalderomb	0,0200767 (5,12)
Innareal	0,4404812 (21,10)
Inwc	0,1683902 (9,71)

Garasjedum	0,1069171 (8,13)
dum1997	-0,0723614 (-3,77)
dum1999	0,1101717 (5,30)
dum2000	0,1771983 (9,13)
dum2001	0,2302453 (11,77)
dum2002	0,304258 (14,49)
InMINCBD	-0,0727866 (-5,18)
InACCESSIBILITY	0,1927547 (5,25)
<b>Inandelvidutd</b>	<b>-0,0461025</b> <b>(-2,69)</b>
<b>Inandelhøyutd</b>	<b>0,0516258</b> <b>(3,58)</b>
<b>Inandelikkevestinnv</b>	<b>-0,0058848</b> <b>(-1,85)</b>
R <sup>2</sup>	0,6859
Justert R <sup>2</sup>	0,6826

Resultatene fra den endelige modellen viser at alle andelsvariablene er signifikante, og har endret seg lite fra foregående modell (tabell 5.3). T-verdien til forklaringsvariabelen videregående utdanning har i denne modellen økt, men er fortsatt negativ og signifikant. Tilsvarende koeffisient har en negativ økonomisk effekt på boligpriser. For høy utdanning tilsier regresjonskoeffisienten at en prosentvis økning i andelen tilsvarer 0,0516% økning i boligpriser. Andelens t-verdi er også i denne modellen større enn signifikansnivået på  $\pm 1,645$ . Nullhypotesen for andel ikke-vestlig innvandring forkastes, variabelen er signifikant. Koeffisienten tyder på en lav negativ innvirkning på boligpriser. I kroner har den altså kun en helt marginal betydning. Variabelen er statistisk signifikant, men ikke økonomisk signifikant. P-verdiene til andelsvariablene videregående utdanning, høy utdanning og ikke-vestlig innvandring er henholdsvis 0,007, 0,00, og 0,065. Justert R<sup>2</sup> for modellen er 0,6826, dette tilsier at modellens forklaringskraft er på 68,26%. I følge regresjonsmodellen fanger andelsvariablene opp noe ved nabolag, siden størrelsene er signifikante. Nøyaktig hva dette er, er vanskelig å si.



## 6 Konklusjon

I denne oppgaven ønsket jeg å finne svar på i hvor stor grad sosioøkonomiske variabler påvirker boligpriser på Haugalandet, og undersøke hvilke av disse nabolagstrekkene som har størst betydning for variasjonen i distriktets boligpriser. Ved hjelp av segregeringsindekser og regresjonsanalyse har jeg prøvd å besvare problemstillingen. I arbeidet med dette, har jeg benyttet to sosioøkonomiske variabler; utdanning og innvandring. Jeg har presentert beskrivende statistikk via indekser, for å lettere måle graden av segregering. Deretter studerte jeg boligpriser i regionen, for å se om de sosioøkonomiske variablene hadde noen innvirkning på boligprisene.

De to segregeringsmålene, ulikhets- og isolasjonsindeks for andelsvariablene antydte lave segregeringstendenser i regionen. I indeksene ble det likevel observert små forskjeller i andelene. Ulikhetsindeksene på regionnivå for andelene videregående og høy utdanning viste en antydning til økning i segregering, og motsatt for de resterende andelene. I isolasjonsindeksene på regionnivå ble det observert økning i total og ikke-vestlig innvandring, samt høy utdanning. De resterende variablene tydet på nedgang.

Indeksene på kommunenivå viser i mer detalj hvordan segregeringstendensene forspiller seg på de ulike andelsvariablene. Mønsteret i ulikhetsindeksene tydet på en økning i segregering for andelene videregående og høy utdanning, mens isolasjonsindeksene antydte økning i høy utdanning, samt total og ikke-vestlig innvandring. I samtlige indeksfigurer har bykommunen Haugesund og landkommunen Karmøy de høyeste indeksverdiene, dette skyldes trolig at disse kommunene har de høyeste innbyggertallene i regionen.

I regresjonsmodellene benyttet jeg en basemodell som utgangspunkt for analysene. I de første regresjonsanalysene ble følgende observert: forklaringsvariabelen høy utdanning var den eneste andelsvariabelen som ble signifikant. Det ble ikke påvist noen sammenhenger mellom de resterende variablene og boligpriser. Resultatene endret seg da jeg inkluderte alle andelene i samme modell (samlet modell, tabell 5.3). I denne regresjonsmodellen ble videregående utdanning, høy utdanning, og ikke-vestlig innvandring signifikante. I følge regresjonsteorien tyder dette på sammenhenger mellom disse forklaringsvariablene og boligpriser. Den siste regresjonsmodellen inneholdt alle de signifikante andelsvariablene. Resultatene fra denne hadde liten grad av endring og alle var fortsatt signifikante. Sammenlignes justert  $R^2$  i

modellene, er justert  $R^2$  høyest i den siste regresjonsmodellen, forklaringskraften til denne modellen er best. Det påpekes at andelen høy utdanning er den eneste andelen som ble signifikant i samtlige regresjonsmodeller. Når denne variabelen økte i et område (nabolag), steg også boligprisene. Økningen var statistisk signifikant, men hadde liten kronemessig betydning.

Ifølge teorien påvirkes boligpris av nabolagstrekk, og man vil kunne oppdage segregeringstendenser i ulike geografiske områder. Det ble i andelsvariablene observert lave segregeringstendenser i region Haugaland, og disse hadde heller ikke de store betydningene for boligpriser i regresjonsanalysene. Flere andeler ble signifikante, men den økonomiske virkningen på boligpriser var lav. Det resulterte i at andelene hadde liten påvirkning på boligpriser. Små segregeringstendenser i regionen antyder at målene på nabolagstrekk ikke er sentrale for boligpriser. I regresjonsanalysene ble allikevel flere av de sosioøkonomiske variablene signifikante, noe som viser at trekk ved det lokale nærmiljøet er av betydning.

For å få et bedre bilde av forklaringsvariablenes tilknytning til boligpriser kunne man studert utviklingen over en lengre tidsperiode. Dette kan være tips til videre forskning, da det eksisterer tallmateriale på boligpriser for tidsperioden 2004-2007. Det hadde vært interessant å se hvordan utviklingen er for disse årene, og om det er noen endringer i de andelene som ble signifikante i mine regresjonsmodeller.



## 7 Litteraturliste

- Barne-, likestillings- og inkluderingsdepartementet (2010). *Nordisk fagseminar om segregering*. Hentet 12. januar 2014 fra <http://www.regjeringen.no/nb/dep/bld/dep/styrer-rad-og-utvalg/inkluderingsutvalget/fagseminar-om-segregering.html?id=625471>
- Barne-, likestillings- og inkluderingsdepartementet (u.å). *Begreper og definisjoner*. Hentet 14. januar 2014 fra <http://www.regjeringen.no/nb/dep/bld/dok/nouer/2011/nou-2011-14/3/5.html?id=650770>
- Bergtun, A., Schei, C. (2013). *Segregering og sosioøkonomiske skilnader i Stavanger* (Bachelorgradsoppgave, Høgskolen Stord/Haugesund). Hentet fra [http://brage.bibsys.no/hsh/handle/URN:NBN:no-bibsys\\_brage\\_41149](http://brage.bibsys.no/hsh/handle/URN:NBN:no-bibsys_brage_41149)
- Investopedia. (2013). *Hedonic Pricing*. Hentet 14. januar 2014 fra <http://www.investopedia.com/terms/h/hedonicpricing.asp>
- Johannessen, A., Kristoffersen, L. & Tufte, P. A. (2011). *Forskningsmetode for økonomisk-administrative fag*. Oslo: Abstrakt.
- Kartverket. (u.å). *Fylkes- og kommuneoversikt*. Hentet 22. desember 2013 fra <http://www.kartverket.no/Kunnskap/Fakta-om-Norge/Fylker-og-kommuner/Tabell/>
- Knudsen, J.P., & Sødal, S. (2010). *Økonomi og tid*. Fagbokforlaget, Bergen.
- Library of congress. (u.å). *A Century of Racial Segregation*. Hentet 14. januar 2014 fra <http://www.loc.gov/exhibits/brown/brown-segregation.html>
- McCann, P. (2001). *Urban and Regional Economics*. Oxford: Oxford University Press.
- Osland, L. A. (2010). *An Application of Spatial Econometrics in Relation to Hedonic House Price Modeling*. Hentet fra [http://aux.zicklin.baruch.cuny.edu/jrer/papers/pdf/past/vol32n03/03.289\\_320.pdf](http://aux.zicklin.baruch.cuny.edu/jrer/papers/pdf/past/vol32n03/03.289_320.pdf)

- Rognsaa, A. (2010). *Prosjektoppgaven*. Oslo: Universitetsforlaget AS.
- Simpson, L. (2007). *Ghettos of the mind: the empirical behaviour of indices of segregation and diversity*. Journal of the Royal Statistical Society Series A-statistics in Society, 170(2), 405-424.  
doi: 10.1111/j.1467-90085X.2007.00465.x
- Statistisk Sentralbyrå. (u.å). *Ikke-vestlige innvandrere bosatt i norske og danske byer under 100 000 innbyggere per 1.1.2005 (Norge) og 1.1.2004 (Danmark)*. Hentet fra <https://www.ssb.no/a/samfunnsspeilet/utg/200604/05/tab-2006-10-10-04.html>
- Statistisk Sentralbyrå. (2006). *Innvandrere i Norge og Danmark*. Hentet fra <https://www.ssb.no/befolkning/artikler-og-publikasjoner/hoyest-innvandrerandel-i-oslo-men-storre-segregasjon-i-danske-byer>
- Statistisk Sentralbyrå. (2013). *Innvandrere og norskfødte med innvandrerforeldre, 1. januar 2013*. Hentet 25. januar 2014 fra <http://www.ssb.no/befolkning/statistikker/innvbef/aar/2013-04-25?fane=om>
- Store Norske Leksikon (u.å). *Flerkulturelle samfunn*. Hentet fra [http://snl.no/flerkulturelle\\_samfunn](http://snl.no/flerkulturelle_samfunn)
- Ubøe (2009). *Statistikk for økonomifag*. Oslo: Gyldendal.
- Wessel, T. (2013). *Hvilken betydning har nabolaget for barns utvikling og livssjanser?* (Artikkel nr. 12, s.247). BUFDIR, Oppvekstrapporten (2013). Hentet 22. desember 2013 fra [http://www.bufetat.no/PageFiles/9263/1/Web\\_Oppvekstrapporten\\_2013\\_es.pdf](http://www.bufetat.no/PageFiles/9263/1/Web_Oppvekstrapporten_2013_es.pdf)
- Wong, D.W.S. (2008). *Conceptual and Operational Issues in Incorporating Segregation Measurements in Hedonic Price Modeling*. Hedonic Methods in Housing Markets.  
doi: 10.1007/978-0-387-76815-1\_7

## 8 Appendix

### Vedlegg 1: Beregning av ulikhetsindeks

For å vise hvordan ulikhetsindeksen beregnes benytter jeg et eksempel, der faktor er utdanningsnivå og andelen jeg ser på er oppgitt for hver kommune i region Haugaland. Tabellen nedenfor viser fordelingen av variabel lav utdanning (bare grunnskole) for personer i aldersgruppen 30-49 år. Den resterende befolkningen (``Resten``) er personer med videregående og høyere utdanning.

Kommune	Lav utdanning	Resten	Total populasjon (utdanningskategorier)
Haugesund	1871	6077	7948
Sauda	334	1008	1342
Bokn	57	137	194
Tysvær	609	1629	2238
Karmøy	2960	6902	9862
Vindafjord	337	915	1252
Etne	258	694	952
Sveio	334	931	1265
Odda	585	1545	2130
Totalt i regionen	7345	19838	27183

Tabell 1. Viser total populasjon i Haugaland for utdanningsnivåene (alder 30-49 år) og fordelingen av lavt utdannede i år 1995.

Eksempel på utregning av ulikhetsindeks for hele regionen:

$$\text{ID} = 0.5 \left[ \left| \frac{\text{Lav utdanning Haugesund}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Resten utdanning Haugesund}}{\text{Total resten utdanning region Haugaland}} \right| + \left| \frac{\text{Lav utdanning Sauda}}{\text{Total lav utdanning Region Haugaland}} - \frac{\text{Resten utdanning Sauda}}{\text{Total resten utdanning region Haugaland}} \right| \right]$$

$$\begin{aligned}
& + \left| \frac{\text{Lav utdanning Bokn}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Resten utdanning Bokn}}{\text{Total resten utdanning region Haugaland}} \right| \\
& + \left| \frac{\text{Lav utdanning Tysvær}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Resten utdanning Tysvær}}{\text{Total resten utdanning region Haugaland}} \right| \\
& + \left| \frac{\text{Lav utdanning Karmøy}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Resten utdanning Karmøy}}{\text{Total resten utdanning region Haugaland}} \right| \\
& + \left| \frac{\text{Lav utdanning Vindafjord}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Resten utdanning Vindafjord}}{\text{Total resten utdanning region Haugaland}} \right| \\
& + \left| \frac{\text{Lav utdanning Etne}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Resten utdanning Etne}}{\text{Total resten utdanning region Haugaland}} \right| \\
& + \left| \frac{\text{Lav utdanning Sveio}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Resten utdanning Sveio}}{\text{Total resten utdanning region Haugaland}} \right| \\
& + \left| \frac{\text{Lav utdanning Odda}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Resten utdanning Odda}}{\text{Total resten utdanning region Haugaland}} \right| \Big]
\end{aligned}$$

$$\begin{aligned}
ID = 0.5 \Big[ & \left| \frac{1871}{7345} - \frac{6077}{19838} \right| + \left| \frac{334}{7345} - \frac{1008}{19838} \right| \\
& + \left| \frac{57}{7345} - \frac{137}{19838} \right| + \left| \frac{609}{7345} - \frac{1629}{19838} \right| \\
& + \left| \frac{2960}{7345} - \frac{6902}{19838} \right| + \left| \frac{337}{7345} - \frac{915}{19838} \right| \\
& + \left| \frac{258}{7345} - \frac{694}{19838} \right| + \left| \frac{334}{7345} - \frac{931}{19838} \right| \\
& + \left| \frac{585}{7345} - \frac{1545}{19838} \right| \Big]
\end{aligned}$$

$$ID = 0,0586377$$

Ulikhetsindeksen for region Haugaland i år 1995 har en verdi på under 0,30, noe som tilsier at området er lite segregert. Dette vil si at de med bare grunnskoleutdanning er svært lite segregert innenfor distriktet.



## Vedlegg 2: Beregning av isolasjonsindeks

Jeg benytter et eksempel for å vise hvordan isolasjonsindeksen kan beregnes. Variabel er utdanningsnivå, og geografisk område er kommunene i region Haugaland. Eksempelet viser variabel lav utdanning (bare grunnskole) for år 1995. Henviser til tabell 1 for populasjonsinndeling.

Eksempel på utregning av isolasjonsindeks for hele regionen:

$$\begin{aligned} P_g^* = & \left( \frac{\text{Lav utdanning Haugesund}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Lav utdanning Haugesund}}{\text{Total populasjon Haugesund}} \right) \\ & + \left( \frac{\text{Lav utdanning Sauda}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Lav utdanning Sauda}}{\text{Total populasjon Sauda}} \right) \\ & + \left( \frac{\text{Lav utdanning Bokn}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Lav utdanning Bokn}}{\text{Total populasjon Bokn}} \right) \\ & + \left( \frac{\text{Lav utdanning Tysvær}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Lav utdanning Tysvær}}{\text{Total populasjon Tysvær}} \right) \\ & + \left( \frac{\text{Lav utdanning Karmøy}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Lav utdanning Karmøy}}{\text{Total populasjon Karmøy}} \right) \\ & + \left( \frac{\text{Lav utdanning Vindafjord}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Lav utdanning Vindafjord}}{\text{Total populasjon Vindafjord}} \right) \\ & + \left( \frac{\text{Lav utdanning Etne}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Lav utdanning Etne}}{\text{Total populasjon Etne}} \right) \\ & + \left( \frac{\text{Lav utdanning Sveio}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Lav utdanning Sveio}}{\text{Total populasjon Sveio}} \right) \\ & + \left( \frac{\text{Lav utdanning Odda}}{\text{Total lav utdanning region Haugaland}} - \frac{\text{Lav utdanning Odda}}{\text{Total populasjon Odda}} \right) \end{aligned}$$

$$\begin{aligned} P_g^* = & \left( \frac{1871}{7345} - \frac{1871}{7948} \right) + \left( \frac{334}{7345} - \frac{334}{1342} \right) \\ & + \left( \frac{57}{7345} - \frac{57}{194} \right) + \left( \frac{609}{7345} - \frac{609}{2238} \right) \end{aligned}$$

$$\begin{aligned}
& + \left( \frac{2960}{7345} - \frac{2960}{9862} \right) + \left( \frac{337}{7345} - \frac{337}{1252} \right) \\
& + \left( \frac{258}{7345} - \frac{258}{952} \right) + \left( \frac{334}{7345} - \frac{334}{1265} \right) \\
& + \left( \frac{585}{7345} - \frac{585}{2130} \right)
\end{aligned}$$

$$P_g * = 0,268743859$$

Isolasjonsindeksen for region Haugaland i år 1995 har en verdi på 0,268743859. Dette vil si at det er 26,87% sannsynlighet for at en person med bare grunnskoleutdanning tilfeldig møter en annen person med samme utdanning innenfor distriktet.

### ***Vedlegg 3: Populasjonsstørrelser***

Utdanningene innenfor regionen er delt inn i 3 ulike kategorier (opprinnelig 8 kategorier) for å samle de ulike utdanningene. Kategoriene er som følger: Barne- og ungdom skole utdanning (1), Videregående utdanning (2), Universitet/Høyskoleutdanning (3).

Tabellene nedenfor viser populasjonsstørrelsene innenfor hver kategori i tidsperioden 1995 til 2011 (annenhver år), samt total populasjon for utdanningsnivåene.

#### **Haugesund**

År/Utdanningsnivå	1	2	3	Total populasjon
1995	1871	4094	1983	7948
1997	1773	4122	2130	8025
1999	1802	4132	2413	8347
2001	1884	4111	2582	8577
2003	1847	4001	2731	8579
2005	1840	3925	3041	8806
2007	1912	3919	3285	9116
2009	2051	3996	3516	9563
2011	2015	3976	3824	9815

### Sauda

År/Utdanningsnivå	1	2	3	Total populasjon
1995	334	768	240	1342
1997	319	757	255	1331
1999	283	783	203	1269
2001	294	738	241	1273
2003	273	719	233	1225
2005	250	702	231	1183
2007	237	663	235	1135
2009	229	645	258	1132
2011	224	606	278	1108

### Bokn

År/Utdanningsnivå	1	2	3	Total populasjon
1995	57	108	29	194
1997	53	111	26	190
1999	47	117	25	189
2001	48	127	31	206
2003	49	135	24	208
2005	43	137	27	207
2007	34	144	34	212
2009	34	134	38	206
2011	41	127	45	213

### Tysvær

År/Utdanningsnivå	1	2	3	Total populasjon
1995	609	1256	373	2238
1997	601	1275	411	2287
1999	615	1320	433	2368
2001	612	1361	504	2477
2003	603	1387	553	2543
2005	577	1441	624	2642
2007	541	1382	678	2601
2009	527	1394	727	2648
2011	542	1409	801	2752



### **Karmøy**

År/Utdanningsnivå	1	2	3	Total populasjon
1995	2960	5289	1613	9862
1997	2961	5343	1715	10019
1999	2875	5444	1883	10202
2001	2853	5524	2016	10393
2003	2853	5489	2148	10490
2005	2736	5470	2370	10576
2007	2654	5420	2513	10587
2009	2720	5599	2738	11057
2011	2524	5553	2969	11046

### **Vindafjord**

År/Utdanningsnivå	1	2	3	Total populasjon
1995	337	722	193	1252
1997	321	344	193	858
1999	283	783	203	1269
2001	261	749	205	1215
2003	252	723	220	1195
2005	235	674	246	1155
2007	447	1093	470	2010
2009	439	1099	491	2029
2011	423	1101	546	2070

### **Etne**

År/Utdanningsnivå	1	2	3	Total populasjon
1995	258	530	164	952
1997	249	561	157	967
1999	242	574	167	983
2001	233	587	174	994
2003	243	574	177	994
2005	224	533	199	956
2007	214	498	219	931
2009	204	478	239	921
2011	201	486	271	958

## Sveio

År/Utdanningsnivå	1	2	3	Total populasjon
1995	334	702	229	1265
1997	318	703	231	1252
1999	286	681	245	1212
2001	296	659	251	1206
2003	288	623	254	1165
2005	281	596	274	1151
2007	284	570	311	1165
2009	293	605	345	1243
2011	310	603	386	1299

## Odda

År/Utdanningsnivå	1	2	3	Total populasjon
1995	585	1147	398	2130
1997	526	1156	402	2084
1999	477	1191	399	2067
2001	435	1182	418	2035
2003	409	1235	418	2062
2005	356	1129	449	1934
2007	316	1035	442	1793
2009	315	975	436	1726
2011	294	939	459	1692

## Vedlegg 4: Indeksverdier på regionnivå

Nedenfor vises indeksverdier på regionnivå, for alle utdannings- og innvandrings kategoriene.

### Lav utdanning

År	Ulikhetsindeks
1995	0,0586
1997	0,0796
1999	0,0726
2001	0,0632
2003	0,0684
2005	0,0675
2007	0,0587
2009	0,0506
2011	0,0409

År	Isolasjonsindeks
1995	0,2728
1997	0,2687
1999	0,2509
2001	0,2463
2003	0,2426
2005	0,2315
2007	0,2270
2009	0,2250
2011	0,2136

### Videregående utdanning

År	Ulikhetsindeks
1995	0,0286
1997	0,0387
1999	0,0593
2001	0,0615
2003	0,0685
2005	0,0799
2007	0,0858
2009	0,0895
2011	0,0952

År	Isolasjonsindeks
1995	0,5383
1997	0,5338
1999	0,5410
2001	0,5330
2003	0,5271
2005	0,5155
2007	0,5037
2009	0,4946
2011	0,4841

### Høy utdanning

År	Ulikhetsindeks
1995	0,1081
1997	0,1116
1999	0,1336
2001	0,1290
2003	0,1346
2005	0,1349
2007	0,1282
2009	0,1219
2011	0,1189

År	Isolasjonsindeks
1995	0,1995
1997	0,2127
1999	0,2257
2001	0,2374
2003	0,2499
2005	0,2735
2007	0,2888
2009	0,2985
2011	0,3195

### Total innvandring

År	Ulikhetsindeks
1995	0,1341
1997	0,1070
1999	0,1227
2001	0,1477
2003	0,1336
2005	0,1221
2007	0,1611
2009	0,1520
2011	0,1323

År	Isolasjonsindeks
1995	0,0356
1997	0,0363
1999	0,0420
2001	0,0508
2003	0,0512
2005	0,0550
2007	0,0668
2009	0,0846
2011	0,0943

### **Ikke-vestlig innvandring**

År	Ulikhetsindeks
1995	0,1838
1997	0,1506
1999	0,1606
2001	0,1922
2003	0,1664
2005	0,1439
2007	0,1895
2009	0,1842
2011	0,1583

År	Isolasjonsindeks
1995	0,0200
1997	0,0199
1999	0,0232
2001	0,0320
2003	0,0321
2005	0,0355
2007	0,0487
2009	0,0636
2011	0,0723