

## Appendikset

### Om appendikset

Appendikset indeholder overordnet fire afsnit:

- Teoretisk udledning og definition af modeller,
- Supplerende statistiske resultater
- Deskriptiv statistik
- Udeladte undermodeller

De teoretiske afsnit i appendikset er særdeles vigtige. Enhver læser med kendskab til statistik bør, efter at have læst rapporten, læse disse teoriafsnit. De lægger et fundament for de antagelser rapporten bygger på og retfærdiggør meget af det senere arbejde. Dertil viser de også hvordan man kan håndtere en række statistisk teoretiske problemer, samt hvordan den endelige model bør tolkes

De supplerende statistiske resultater er vigtige for forskere og eksperter der arbejder med fattigdom. Med dem kan man trække langt flere konklusioner end den oprindelige rapport ligger op til. Desuden får man også et indblik i hvad man ikke kan udtale sig om, altså hvor det statistiske materiale tyder på insignifikans. Det tilskyndes derfor alle fattigdomsfaglige mennesker, at betragte graferne hertil, og se om de finder vigtigt materiale.

Den deskriptive statistik er anført for, at de statistiske eksperimenter kan gentages. Hertil er det muligt også at få adgang til SAS koden fra projektet - man skal blot sende en email til julianserban@gmail.com.

Der er to udeladte SSM undermodeller. Den første er modellen for dødsfald, da den giver et definitionsmæssigt absurd syn på fattigdom. F.eks. hvis blot flere fattige mennesker døde, ville fattigdommen i Danmark være faldende, hvilket er uacceptabelt. Den anden model er for pardannelses og separation. Den er udeladt af to lige vigtige årsager, at den tilfører en enorm kompleksitet til den overordnede model, og at dens indflydelse ville være relativ lille. Begge undermodeller er dog i sig selv udmarkede, og læsere som med særlig interesse for SSM modellen, bør også kigge på disse.

## Indhold

Appendikset .....	1
Om appendikset.....	1
Teori.....	3
Logistisk regression.....	10
Markov model.....	3
SSM modellen.....	11
SSM: Arbejdsløshedsmodel .....	11
SSM: Handikap.....	13
SSM: Uddannelse .....	14
SSM: Fødsler .....	15
PSM modellen.....	17
PSM: Variabel estimator .....	17
PSM: Variabel estimator .....	19
Gennemsnitlig vægtning .....	21
Resultater for data .....	22
ESS: Logistisk Regression .....	22
ESS: Resultater .....	32
LIS: Resultater .....	38
ESS: Fremskrivning .....	44
Deskriptiv statistik.....	51
ESS Data .....	51
ESS Uddannelsesniveauer .....	58
SSM - Udeladte undermodeller .....	59
Pardannelse og separation .....	60
Dødsfald .....	63

## Teori

### Markov model

Igennem vores arbejde, særligt SSM modellen, bruger vi de såkaldte endelige stationære tidskontinuerte markov kæder. At de er endelige antages herfra og diskuteres ikke videre. Modellen defineres formelt matematisk som en mængde  $S$  af  $N$  stadier:

$$\{1, 2, 3, \dots, N\} \in S$$

Mellem to stadier  $i, j \in S, i \neq j$  hører en transitionsintensitet,  $q_{i,j} \geq 0$ . Dertil defineres  $q_{i,i} = -\sum q_{i,j} \forall i \in S$ . Disse intensiteter opskrives typisk som en matrix benævnt intensitetsmatricen (eller infinitesimal generatoren) for en given stationær tidskontinuerte markov kæde:

$$Q = \begin{pmatrix} q_{1,1} & \cdots & q_{1,N} \\ \vdots & \ddots & \vdots \\ q_{N,1} & \cdots & q_{N,N} \end{pmatrix}$$

Til et givent tidspunkt  $t \geq 0$  vil vores markov kæde befinde sig i nøjagtigt ét af vores  $N$  stadier. Fra dette stade vil den med bevæge sig til et nyt stade  $j \neq i$ , med eksponentiel sandsynlighedsfordeling  $\exp(-q_{i,j})$ . Med andre ord bestemmer intensiteterne hvilken sandsynligheden markov kæden har til at hoppe til et andet stade.

Givet at kæden i tiden  $t + s$  har foretaget nøjagtigt et hop, kan vi udregne sandsynligheden for at den befinder sig i stadiet  $j$  som værende:

$$P(X(t+s) = j | X(t) = i) = \frac{q_{i,j}}{-q_{i,i}} \quad \forall j \neq i$$

Hvor  $X(t)$  og  $X(t+s)$  definerer stadiet kæden befinder sig i til tiden  $t$  og  $t+s$ . Endvidere kan vi udregne følgende matrice:

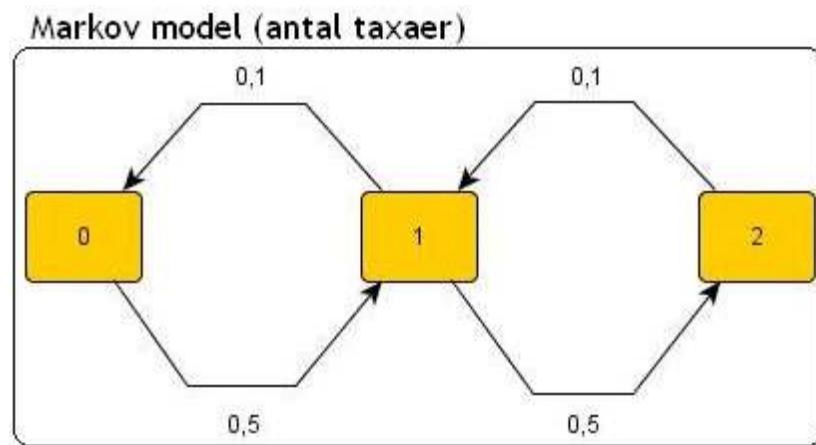
$$P(t) = \exp(Qt)$$

Hvor indgangen  $P(t+s)_{i,j}$  er sandsynligheden for kæden befinder sig i stadiet  $j$  til tiden  $t+s$ , givet at kæden befandt sig i stadiet  $i$  til tiden  $t$ .

Lad os illustrere modellen med et konkret eksempel: antallet af taxaer holdende ved hovedbanegården. Lad os antage at vi kun kigger på tidsrummet 12 - 15 på hverdage, og at der er så meget konstruktionsarbejde at der maksimalt kan holde 2 taxaer. Vi har således tre stadier i vores markov kæde. Lad os gå ud fra at alle taxaer i gennemsnit venter 10 min før de får en passager, men at der samtidig er så meget konkurrence om de få taxapladser, at der i gennemsnit kun går 2 min før en tom taxaplads bliver optaget. Da vi arbejder med eksponentiel fordelinger, og disse har middelværdi  $1/\lambda$ , hvor  $\lambda$  er fordelingens parameteren for eksponentiel fordelingen kan vi udregne intensiteterne således:

$$q_{0,1} = q_{1,2} = \frac{1}{2} = 0,5, \quad q_{2,1} = q_{1,0} = \frac{1}{10} = 0,1, \quad q_{0,0} = -0,5, \quad q_{2,2} = -0,1, \quad q_{1,1} = -(0,5 + 0,1) = -0,6$$

Hvor alle andre intensiteter er nul. Modellen kan illustreres som nedenstående diagram:



Vi kan desuden udregne vores matrice over sandsynligheder som:

$$Q = \begin{pmatrix} -0,5 & 0,5 & 0 \\ 0,1 & -0,6 & 0,5 \\ 0 & ... & 0,1 \end{pmatrix}$$

$$P(t) = \exp(Qt)$$

Hvis der f.eks. kl. 15.00 er 2 taxaer, så er sandsynligheden for at der kl 15:15 stadig er 2 taxaer:

$$P(15)_{2,2} = 0,8069 = 80,69\%$$

Det behøver dog ikke at være de samme taxaer. Eksempelvis kunne en af de to taxaer have samlet en passager op, hvorefter en ny taxa har optaget dens plads.

Givet at der kun er en taxa på taxaholdepladsen, kan vi også udregne sandsynligheden for at denne ene taxa skal samle en passager op før en ny taxa ankommer. Det svarer til at vi befinner os i stadiet 1 og går til stadiet 0 før vi går til stadiet 2, og kan udregnes som:

$$\frac{0,1}{0,6} = 0,1667 = 16,67\%$$

Vores intention er at benytte disse markov modeller til at modellere sociologiske processer. Endvidere er det eneste data vi har til rådighed aggregeret data til SSM modellen. Det vil sige at vi prøver at udtales om individers adfærd ud fra kollektiv adfærd. For at dette skal kunne lade sig gøre arbejder vi under en række antagelser:

1. Diskret udfaldsrum
2. Markov-egenskaben
3. Stationær proces (Tidshomogenitet)
4. Processen foregår i kontinuert tid
5. Eksponentielt fordelte spring mellem stadier
6. Homogenitet af befolkningen

Disse antagelser er meget almindelige og udbredte i litteraturen, om end de sjældent bliver så udpegslet. Nedenfor diskuterer vi disse antagelser i kontekst om social mobilitet generelt. Når vi i analysen udarbejder de enkelte modeller bliver de mest væsentlige aspekter diskuteret videre.

### Diskret udfaldsrum

Vores første antagelse, som alt vores senere arbejde bygger på, er at vi kan partitionere vores udfaldsrum ind i et endeligt antal stadier. Vi arbejder således med kategorier i vores data, som eksempelvis:

- Indkomstgrupper (0 – 9.999, 10.000 – 19.990 ... )
- Aldersgrupper (10-14, 15-19 ... )
- Fattig / Ikke fattig
- Uddannelsesniveauer
- osv..

Det er klart at man altid vil miste nogle detaljer på denne måde. Et individ kan f.eks. godt være 14,99, i gang med en uddannelse eller ligge i en gråzone mellem to indkomstgrupper. Det er i virkeligheden ikke så væsentligt, da vi arbejder med et stort antal mennesker og derfor vil eventuelle fejl blive fuldstændigt udjævnet. Dertil skal det tilføjes at en overvældende mængde psykologisk teori peger på, at mennesker selv forstår verden i kategorier, eller ”kasser” om man vil. Som konsekvens af dette bør måling af fattigdom også konsekvent arbejde med brede kategorier og ikke præcise tal som den rene statistik helst så.

### Markov-egenskaben

Markov-egenskaben er antagelsen om, at sandsynligheden for at bevæge sig til et nyt stadie kun afhænger af det nuværende stadie. I taxa-eksemplet fra før, betyder dette at antallet af taxaer på taxaholdepladsen kun afhænger af hvornår den sidste taxa ankom eller afgik. Vi arbejder i kontinuert tid, så det er umuligt at udtales sig om præcist hvilket tidspunkt der er afgørende. Hvis vi ved at der kl. 12.00 var to taxaer, men kl

13:00 kun en taxa, og vi gerne vil udtales os om hvordan situationen er kl 13:15, så afhænger dette udelukkende af den ene taxa kl 13.00.

Dette giver fin mening i taxa-eksemplet, men når vi skal til at kigge på sociologiske processer kan det blive mere omstændigt og usikkert. Hvis vi f.eks. gerne vil udtales os om hvorvidt en lavuddannet mand med arbejde i 2008, er blevet arbejdsløs i 2009, vil dette højst sandsynligt også afhænge af hans anciennitet på arbejdspladsen. Hvis han for eksempel i årene 2000-2007 har været arbejdsløs, og netop i 2008 har fået arbejde, er der en væsentlig større sandsynlighed for at han igen i 2009 er arbejdsløs, end hvis han i stedet havde haft dette arbejde igennem 5 år.

McFarland argumenterer for at markov-egenskaben i sociologiske processer, og særligt social mobilitet, har en forvrængede effekt på resultaterne fra modellen. Sociologiske teorier viser at tiden et individ befinder sig i et givent stadie (eksempelvis på en arbejdsplads eller i et ægteskab) påvirker individets billede af verden og derfor hans/hendes valg for at ændre stadie (som eksempelvis at skifte arbejde eller blive separeret). McFarland argumenterer og referer til empiriske studier, som til gengæld viser, at markov-egenskaben bliver mere og mere pålidelig jo mindre tidsrum man arbejder i. Det er f.eks. fuldstændig absurd hvis man kigger på social mobilitet over generationer, men kan godt være acceptabel over en kort årrække.

Vores håb er, at vi ved at anvende disse markov modeller til SSM modellen over en 5-årig periode, har reduceret problemerne ved markov-egenskaben. Vi vil blandt andet argumentere ud fra de resultater vi opnår og andre empiriske studier for at godtgøre denne antagelse. Det kan dog ikke lade sig gøre at måle usikkerheden direkte på denne antagelse – dette ville kræve longitudinal registerdata.

### Stationær proces

Antagelsen om at vi har en stationær proces kan ses som en parallel til markov-egenskaben. Stationær proces betyder at, sandsynlighederne at skifte stadier ikke ændrer sig over tid. Det vil sige, at to identiske individer, hvoraf den første blev arbejdsløs i 2004 og den anden i 2008, begge har samme sandsynlighed for at opnå finde arbejde indenfor en 2-årig periode. I en vis forstand antager vi at resten af verden står stille – dog ikke hele verden, kun den del af verden som kan påvirke vores lille system.

Ligesom markov-egenskaben, betyder det, at vi ved at arbejde med en 5-årig periode har begrænset eventuelle fejl på denne antagelse væsentligt. Overordnet set er denne antagelse ikke væsentlig, da samtlige aggregeret data vi anvender selv stærkt udstråler stationaritet. Eventuelle t-test kan udføres for at eftervise hvor lille en fejl denne antagelse udøver på vores markov kæde. I afsnittet om arbejde og arbejdsløshed fører vid og en kortere diskussion, for at vi i perioden 2002-2008 har et stationært arbejdsmarkedet – da det er klart at en økonomisk krise kan påvirke denne væsentligt. Heldigvis er konsekvenserne af den økonomiske krise ikke begyndt før 2009, og vores antagelse er derfor acceptabel.

### Processen foregår i kontinuert tid

Denne antagelse betyder kort sagt, at spring mellem forskellige stadier kan ske i kontinuert tid. Fra taxaexemplbet betyder det, at vi kan være i to forskellige stadier for tiderne 12:01 og 12:02, og måske i et helt tredje stadie i 12:01:30. For mange sociologiske processer er dette en fornuftig antagelse, da der ingen begrænsninger er på hvornår de formes. Et ægteskab kan således indgås 1 minut efter to mennesker har mødt hinanden, såvel som 10 år efter at de har mødt hinanden. Det vil naturligvis være højst usandsynligt at to mennesker ville indgå et ægteskab efter 1 minut, men dette bliver korrigert for af vores eksponentiel fordeling, som givet data vil tvinge denne sandsynlighed til næsten nul.

Man kan dog diskutere hvorvidt kontinuitet altid er fornuftigt i sig selv. Eksempelvis vil vi også gerne modellere individer som forsøger at opnå en højere uddannelse, men uddannelsesinstitutionerne arbejder i hovedreglen udelukkende med faste datoer for indgang og afgang. Ligeledes er det (næsten) umuligt at finde arbejde på en søndag, da samtlige virksomheder holder lukket.

Antagelsen er dog meget mild, da modellen vil blive korrigert af de data vi benytter. Data på uddannelser, parforhold, dødsfald etc... vil alle sammen afspejle de underliggende brud i den kontinuerte tid. Når vi opstiller modellerne for 5-årige perioder opnår vi derfor minimale fejl, som alle sammen er bør være jævnt fordelt omkring vores resultater.

Grunden til at vi vælger at arbejde med kontinuert tid, og ikke diskret tid, er først og fremmest fordi en opdeling af tiden i diskrete mængder vil medføre væsentlige fejl på aggregeret niveau, og desuden komplikere modellen unødvendigt.

### Eksponentielt fordelte spring mellem stadier

Eksponentielfordelingen er en sandsynlighedsfordeling som har en konstant rate. Det vil sige, at der er lige sandsynligheden er konstant for alle intervaller af en bestemt længde. Denne konstante rate medfører den særlige egenskab at den er hukommelsesløs. Netop derfor anvendes markov modeller hovedsagligt med denne fordeling. At den er hukommelsesløs er nemlig ensbetydende med at den overholder markov-egenskaben. Derudover er eksponentielfordeling bredt anvendt til at modellere en række sociologiske processer, og kan opfange mange af de underliggende processer som ikke kan observeres direkte.

Da vi stadig arbejder i en 5-årige periode, og desuden kontrollerer for en række forhold som køn, alder og uddannelse, virker det som en rimelig antagelse at vi har en konstant rate af sandsynlighed for hop i mellem stadier.

### Homogenitet af befolkningen

At vi antager homogenitet af befolkningen dækker over to ting. Først og fremmest antager vi, at aggregeret data er proportionelt til individuelt data. Det vil sige, at den gennemsnitlige intensitet for at hoppe mellem to stadier for er den samme for hele befolkningen såvel for individet. Sekundært antager vi at de individer

som ikke handler proportionelt til befolkningen er normalfordelte omkring gennemsnittet på tværs af de parametre vi ikke kontrollere.

Der er muligvis mange ting man ikke opfanger igennem aggregeret data, især når vi ikke kan kontrollere for indkomstniveauet i husstanden som et individ lever i. Det er klart, at der foregår to forskellige sociologiske processer for individer med samme køn, alder og uddannelse. Men hvis det ene individ hører til en økonomisk stærk husstand og den anden til en økonomisk svag husstand, vil deres handlinger afspejle dette. Dette kan vi ikke differentiere i mellem og vil derfor forvente at begge individer opfører sig ens igennem vores analyse.

Dette er således måske den mest omstridte antagelse vi foretager. Derfor vil vi løbende igennem vores arbejde benytte andre empiriske analyser til at begrænse fejl og forskubbe eventuelle fejl, således at hvis vi har fejl i vores resultater, da vil de fremstille et mere positivt mål på social mobilitet end det der reelt eksisterer. Denne tilgang retfærdiggøres af vores hypotese: at social mobilitet er langt ringere for husstande i de laveste indkomstgrupper end for alle andre. Ved at fremstille den sociale mobilitet mere positivt vil eventuelle resultater som bekræfter vores hypotese, ikke kun være tilstrækkelige for at bekræfte vores hypotese, men også skabe stærk formodning om at den sociale mobilitet er endnu værre. Vi vælger så at sige, kun at se toppen af isbjerget. Men det vil uden tvivl være det rigtige isbjerg vi kigger på.

Kilder:

INTRAGENERATIONAL SOCIAL MOBILITY AS A MARKOV  
PROCESS: INCLUDING A TIME-STATIONARY MARKOVIAN  
MODEL THAT EXPLAINS OBSERVED  
DECLINES IN MOBILITY RATES  
OVER TIME, David McFarland, 1970

On Estimating Transition Intensities of a Markov

Process with Aggregate Data of a Certain Type:

"Occurrences but no Exposures", 1986 Richard D. Gill

## Logistisk regression

Vi benytter logistisk regression til vores PSM model. Formelt definer vi en række binomial fordelte variable:

$$Y_i \sim Bin(n_i, p_i) \quad i \in \{1, 2, 3, \dots, N\}$$

Hvor antalsparametren  $n_i$  er velkendt, mens  $p_i$  antages ukendt. Vi opstiller nu:

$$p_i = E\left(\frac{Y_i}{n_i} \mid X_i\right)$$

$$logit(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_N x_{N,i}$$

Hvor  $X_i$  er en stokastisk variabel vektor, som indeholder de beskrivende kategoriske variabler, og  $x_{j,i}$  er den j'te indgang til  $X_i$ . Her er  $\beta_0, \dots, \beta_N$  parametre til modellen som estimeres gennem maximum likelihood estimation.

For en observeret stokastisk variabel vektor  $Z$ , i tilknytning til en uobserveret stokastisk bernoulli-fordelt variabel  $W$ , kan vi udregne sandsynligheden:

$$P(W = 1) = \frac{1}{1 + \exp(\beta_0 + \beta_1 z_1 + \dots + \beta_N z_N)}$$

Hvor  $z_j$  er den j'te indgang til  $Z$ .

Da vores  $Y_i$ 'er er binomialt fordelte, vil fejlen på vores logistiske regression også være binomialt fordelte. Det giver således ikke mening at uddybe fordelingen af vores fejl teoretisk. I stedet skal denne betragtes for den konkrete model.

Vores intention er at benytte logistisk regression til at modellere sociologiske processer, som beskrevet tidligere. For at dette skal kunne lade sig gøre arbejder vi under en række antagelser, hovedsagligt:

1. Logistisk (odds-lineær) sammenhæng mellem beskrivende variable og indkomst-mobilitet
2. Stationær proces (Tidshomogenitet)
3. Homogenitet af befolkningen (Se afsnittet for markov modeller)

Disse antagelser er alle sammen beskrevet i afsnittet for markov modeller, på nær den første om den logistiske sammenhæng mellem vores variable. Denne antagelse bygger som beskrevet i ligningen ovenfor på, en eksponentiel lineær sammenhæng af vores beskrivende variable. Men da vores variable i både definitionen og praksis er kategoriske, så reduceres antagelsen betragteligt. Vi får nemlig tildelt en konstant til hvert eneste udfald i hver eneste beskrivende variabel, for derefter at maksimere vores likelihood. Antagelsen bliver derfor en antagelse på tværs af de beskrivende variable, og ikke imellem dem. Da vi stadigvæk arbejder med spatielt data, bør modellens antagelser være relativt milde. Modellen er derfor, næsten udelukkende, kun begrænset af de beskrivende variable vi har til rådighed.

## SSM modellen

### SSM: Arbejdsløshedsmodel

Vores statistikker udregnet ud fra Laborta Data og OECD Statistics:

Uddannelsesniveau (EU Standard)			Level 0-2	Level 3-4	Level 5-6
Sandsynlighed for at have et job 5 år senere, givet at individet ikke har job på nuværende tidspunkt	Mænd	18 til 24	0,660410653	0,662508037	0,668728841
		25 til 29	0,670406395	0,683673419	0,718053177
		25 til 54	0,781171494	0,79225	0,819765784
		30 til 34	0,739878626	0,75187909	0,782228502
		35 til 39	0,772348602	0,783634133	0,81178118
		40 til 44	0,815269181	0,825498349	0,850448349
		45 til 49	0,838830313	0,848428447	0,871492858
		50 til 54	0,862221046	0,871158507	0,89226406
		55 til 59	0,777597353	0,81500439	0,880483842
		55 til 64	0,792264943	0,829182887	0,892593372
		60 til 64	0,821114528	0,857005323	0,916197684
		65 til 69	0,778728347	0,816098489	0,881420277
		70 til 74	0,90152219	0,934129246	0,980613021
		70+	0,90152219	0,934129246	0,980613021
		75+	0,90152219	0,934129246	0,980613021
		alle	0,757682031	0,770051343	0,800745471
Kvinder	Mænd	20 til 24	0,614715318	0,60862797	0,620019079
		25 til 29	0,663691984	0,63801438	0,683003878
		25 til 54	0,782436246	0,761170498	0,797823888
		30 til 34	0,753961157	0,731426563	0,770447549
		35 til 39	0,772783478	0,751073333	0,788554181
		40 til 44	0,814466452	0,794777627	0,828507316
		45 til 49	0,83683202	0,818335095	0,84986555
		50 til 54	0,861134612	0,844014706	0,87301345
		55 til 59	0,832886304	0,73589756	0,876047377
		55 til 64	0,844310131	0,748137802	0,886299542
		60 til 64	0,878526197	0,785025579	0,91685373
		65 til 69	0,874439191	0,780601979	0,913215854
		70 til 74	0,945695619	0,858371104	0,976212561
		70+	0,945695619	0,858371104	0,976212561
		75+	0,945695619	0,858371104	0,976212561
		alle	0,755921205	0,730072082	0,774400436

Uddannelsesniveau (EU Standard)			Level 0-2	Level 3-4	Level 5-6
Sandsynlighed for at have et job 5 år senere, givet at individet HAR job på nuværende tidspunkt	Mænd	18 til 24	0,661217676	0,663280414	0,669404721
		25 til 29	0,67539426	0,687699613	0,720151998
		25 til 54	0,792035845	0,801019679	0,824337345
		30 til 34	0,748217315	0,758610061	0,785737302
		35 til 39	0,782638586	0,791940184	0,816111056
		40 til 44	0,828534366	0,836205976	0,856030146
		45 til 49	0,853925055	0,860612889	0,877844506
		50 til 54	0,879272073	0,884922058	0,899438885
		55 til 59	0,838436237	0,855698869	0,892460979
		55 til 64	0,857092259	0,872545186	0,905355697
		60 til 64	0,89415755	0,905863019	0,930577406
		65 til 69	0,839870217	0,856995632	0,893457062
		70 til 74	1	1	1
		70+	1	1	1
		75+	1	1	1
		alle	0,767882777	0,778209823	0,804880858
	Kvinder	20 til 24	0,615164665	0,609134277	0,62021167
		25 til 29	0,665879531	0,641361631	0,684520533
		25 til 54	0,788078644	0,769804153	0,801735838
		30 til 34	0,758575719	0,738487486	0,773646887
		35 til 39	0,7780624	0,75915082	0,792214128
		40 til 44	0,821428844	0,805431059	0,833334436
		45 til 49	0,844822782	0,830562077	0,855405653
		50 til 54	0,870345188	0,858108173	0,879399267
		55 til 59	0,866989063	0,824839486	0,890985717
		55 til 64	0,880232789	0,841826126	0,902035069
		60 til 64	0,920205767	0,893728255	0,935111009
		65 til 69	0,915407046	0,88744846	0,931161373
		70 til 74	1	1	1
		70+	1	1	1
		75+	1	1	1
		alle	0,761246416	0,738585203	0,77795583

## SSM: Handikap

De statistiske resultater er givet i nedenstående tabeller:

Sandsynligheden for at blive alvorligt fysisk hämmet efter 5-årig periode			
Uddannelsesniveau (EU Standard)	Alder	Mænd	Kvinder
Level 0-2	15-24	0,010217	0,014832
Level 0-2	25-34	0,034599	0,058494
Level 0-2	35-44	0,026284	0,03933
Level 0-2	45-54	0,04006	0,025564
Level 0-2	55-64	0,040791	0,058122
Level 0-2	65-74	0,042986	0,020183
Level 0-2	75-84	0,018753	0,043719
Level 3-4	15-24	0,004215	0,024845
Level 3-4	25-34	0,010217	0,017682
Level 3-4	35-44	0,020183	0,031336
Level 3-4	45-54	0,023767	0,031336
Level 3-4	55-64	0,016969	0,055892
Level 3-4	65-74	0,024485	0,023049
Level 3-4	75-84	0,026284	0,054779
Level 5-6	15-24	0,007388	0,009509
Level 5-6	25-34	0,023767	0,014476
Level 5-6	35-44	0,012344	0,019825
Level 5-6	45-54	0,043352	0,037143
Level 5-6	55-64	0,031698	0,037143
Level 5-6	65-74	0,039695	0,031698

## SSM: Uddannelse

Der foreligger ingen pålidelige data for årene 18-24 blandt tilgang af elver til Level 5-6 uddannelser.

Årsagen til dette er formentlig at der er høj tilgang, høj frafald og et stort antal forkerte registreringer. For at sikre os at vi ikke skaber et negativt bias, vil vi igen foretage en antagelse. Vi antager at alle mellem 18-24, som har en Level 3-4 uddannelse, påbegynder en Level 5-6 uddannelse.

De statistiske resultater er givet i nedenstående tabel:

Sandsynlighed for at opnå højere uddannelsesniveau

ISCED97	Alder	Mænd	Kvinder
Level 3-4	18	0,10677693	0,092898615
Level 3-4	19	0,026442231	0,029656838
Level 3-4	20 - 24	0,019476158	0,024620518
Level 3-4	25 - 29	0,012815545	0,021013027
Level 3-4	30 - 34	0,006582434	0,019426392
Level 3-4	35 - 39	0,004195033	0,014449773
Level 3-4	40+	0,001611137	0,002557356
Level 5-6	18	1	1
Level 5-6	19	1	1
Level 5-6	20 - 24	1	1
Level 5-6	25 - 29	0,032509261	0,039962625
Level 5-6	30 - 34	0,012097579	0,017124927
Level 5-6	35 - 39	0,00485257	0,009438437
Level 5-6	40+	0,001091841	0,002520304

De 3 rækker markeret med rød er estimeret til 100% her, da der kun foreligger fejlfyldt data om disses sandsynlighed for at starte på en uddannelse. Antagelsen bør udelukkende skabe et mere positivt syn på den sociale mobilitet, og vi opnår således stadigvæk solide resultater ved at bruge disse data senere hen.

## SSM: Fødsler

Sandsynlighed for at føde levende antal børn efter 5-årig periode, givet at moderen IKKE har nogen børn, afhængig af moderens alder og højeste uddannelsesniveau (EU Standard).

Level 0-2 (ISCED 1997)	Moderens alder	Et barn	To børn
	20-24	0,216257303	0,028383
	25-29	0,344414868	0,492621
	30-34	0,280109497	0,62157
	35-39	0,382376802	0,169704
	40-44	0,100024918	0,004927
	45-49	0,004023603	6,99E-06
	50+	1,51212E-05	9,83E-11
Level 3-4 (ISCED 1997)	20-24	0,077620808	0,00287
	25-29	0,087129778	0,003667
	30-34	0,058972609	0,001614
	35-39	0,028102696	0,000352
	40-44	0,005246676	1,19E-05
	45-49	0,000387455	6,46E-08
	50+	6,97184E-06	2,09E-11
Level 5-6 (ISCED 1997)	20-24	0,311326311	0,076239
	25-29	0,232796141	0,034128
	30-34	0,18917928	0,020543
	35-39	0,100463101	0,004974
	40-44	0,021501671	0,000204
	45-49	0,000913324	3,59E-07
	50+	1,18985E-05	6,09E-11

Sandsynlighed for at føde levende antal børn efter 5 årig periode, givet at individet KUN har et barn, afhængig af moderens alder og højeste uddannelsesniveau (EU Standard).

Level 0-2 (ISCED 1997)	Antal Børn:	Et barn	To børn
	20-24	0,203778122	0,009766
	25-29	0,568052408	0,212425
	30-34	0,567024735	0,286593
	35-39	0,431309997	0,06325
	40-44	0,089126793	0,001651
	45-49	0,003463338	2,3E-06
	50+	1,29989E-05	3,24E-11
Level 3-4 (ISCED 1997)	20-24	0,068559535	0,000958
	25-29	0,077240941	0,001226
	30-34	0,051727677	0,000537
	35-39	0,024383699	0,000116
	40-44	0,004517896	3,93E-06
	45-49	0,000333113	2,13E-08
	50+	5,99328E-06	6,88E-12
Level 5-6 (ISCED 1997)	20-24	0,314887277	0,027095
	25-29	0,221566219	0,011797
	30-34	0,175606501	0,007019
	35-39	0,089533045	0,001667
	40-44	0,018614643	6,75E-05
	45-49	0,00078536	1,18E-07
	50+	1,02284E-05	2E-11

Sandsynlighed for at føde levende antal børn efter 5 årig periode, givet at individet HAR to børn, afhængig af moderens alder og højeste uddannelsesniveau (EU Standard).

Level 0-2 (ISCED 1997)	Antal Børn:	Et barn	To børn
	20-24	0,085658873	0,00103
	25-29	0,381902741	0,02835
	30-34	0,438169295	0,040814
	35-39	0,214348353	0,007276
	40-44	0,03543	0,000169
	45-49	0,001328731	2,31E-07
	50+	4,98041E-06	3,25E-12
Level 3-4 (ISCED 1997)	20-24	0,027011576	9,76E-05
	25-29	0,030545893	0,000125
	30-34	0,020235656	5,45E-05
	35-39	0,00943269	1,17E-05
	40-44	0,001734032	3,94E-07
	45-49	0,000127645	2,13E-09
	50+	2,29627E-06	6,91E-13
Level 5-6 (ISCED 1997)	20-24	0,141685443	0,002961
	25-29	0,094051766	0,001251
	30-34	0,072732059	0,000734
	35-39	0,035597879	0,000171
	40-44	0,007184381	6,8E-06
	45-49	0,000300994	1,19E-08
	50+	3,91893E-06	2,01E-12

## PSM modellen

### PSM: Variabel estimerer

Nedenstående er vores maximum likelihood parametre med deres værdier og tilhørende 90%-konfidensintervaller for den primære PSM model:

**Profile Likelihood Confidence Interval for Parameters**  
Primær PSM, husstande uden partner

Parameter		Estimate	90% Confidence Limits
Intercept		5.0963	0.2834 .
GNDR	1	-0.00364	-0.0923 0.0852
BLGETMG	-1	8.0934	-0.7834 .
BLGETMG	0	-3.8298	. 0.6690
EDULVL	0	0.3059	0.1640 0.4482
EDULVL	1	-0.2258	-0.3499 -0.1024
EDULVLM	-1	-2.4209	-5.8171 -0.8199
EDULVLM	0	0.7676	0.2137 1.9048
EDULVLM	1	0.6407	0.0837 1.7790
EDULVLF	-1	0.7282	-0.0247 1.6383
EDULVLF	0	-0.1022	-0.4351 0.1897
EDULVLF	1	-0.2974	-0.6219 -0.0164
EDCTN	0	-1.0408	-1.1941 -0.8920
Head_Unemployed	0	-0.0323	-0.1971 0.1309
DSBLD	0	0.1344	-0.0559 0.3239
RTRD	0	-0.1011	-0.2171 0.0149
Number_Of_Children_S	0	-0.4353	-0.8565 -0.0156
Number_Of_Children_S	1	0.5575	0.2321 0.8936
HHMMB	1	0.3304	-0.0131 0.6685
HHMMB	2	0.0126	-0.2879 0.3063

**Profile Likelihood Confidence Interval for Parameters**  
**Primær PSM, husstande med partner**

Parameter		Estimate		90% Confidence Limits
Intercept		-4.0219	.	1.7710
GNDR	1	-0.2205	-0.4221	-0.0216
BLGETMG	-1	0.8291	-0.3633	2.6197
BLGETMG	0	-1.0550	-2.1189	-0.1691
EDULVL	0	-0.0640	-0.4333	0.2973
EDULVL	1	-0.2930	-0.6072	0.0110
EDULVLP	-1	-0.8552	-10.4814	3.1506
EDULVLP	0	0.7821	-0.5792	3.9987
EDULVLP	1	0.3219	-1.0348	3.5367
EDULVLM	-1	-0.8377	-2.5651	0.7723
EDULVLM	0	0.3898	-0.2348	1.0505
EDULVLM	1	0.7689	0.0973	1.4801
EDULVLF	-1	0.5917	-0.6382	2.1776
EDULVLF	0	-0.4036	-0.9988	0.1044
EDULVLF	1	-0.3283	-0.9346	0.1941
EDCTN	0	-0.0242	-0.6716	0.5731
Head_Unemployed	0	0.1620	-0.3985	0.6969
DSBLD	0	0.2617	-0.2635	0.7588
RTRD	0	-0.0543	-0.4128	0.3135
Number_Of_Children_S	0	-0.6832	-2.0593	0.6275
Number_Of_Children_S	1	0.2816	-0.5756	1.1745
Spouse_Unemployed	0	-0.4125	-1.0216	0.1268
DSBLDP	0	-0.2317	-1.0798	0.4670
RTRDP	0	-0.0547	-0.4294	0.3273
EDCTNP	0	0.4380	-0.1659	1.0574
HHMMB	1	-8.6673	.	6.8719
HHMMB	2	4.6067	-3.1879	.

## PSM: Variabel estimerer

Vi præsenterer nu estimererne for den sekundære PSM model, altså estimerne for LIS datasættet:

### Analysis of Maximum Likelihood Estimates

Sekundær PSM, husstande uden partner

Parameter	Estimate	Standard Error	Chi- Square	Pr > ChiSq
Intercept	1	0,0846	73,38	<,0001
D4	1 0,3516	0,0483	52,94	<,0001
	2 -0,5914	0,0468	159,79	<,0001
D27	0 -0,3116	0,077	16,38	<,0001
	1 0,8011	0,074	117,26	<,0001
	2 0,0979	0,0804	1,48	0,2232
MARRIED	0 ,	,	,	,
hpeduc	0 -0,5428	0,0198	749,2	<,0001
	1 0,1605	0,0215	55,63	<,0001
hpdisabl	0 -0,9083	0,0286	1006,09	<,0001
hpumas	0 -0,2555	0,0224	130,47	<,0001
hpethnat	0 0,4219	0,0732	33,23	<,0001
	1 -0,1449	0,0792	3,35	0,0673

### Maximum Likelihood Analysis of Variance

Sekundær PSM, husstande uden partner

Source	DF	Chi-Square	Pr > ChiSq
Intercept	1	73,38	<,0001
D4	2	374,4	<,0001
D27	3	290,31	<,0001
MARRIED	0*	,	,
Hpeduc	2	806,93	<,0001
Hpdisabl	1	1006,09	<,0001
hpumas	1	130,47	<,0001
hpethnat	2	110,98	<,0001
Likelihood Ratio	147	1637,02	<,0001

### Analysis of Maximum Likelihood Estimates

Sekundær PSM, husstande med partner

Parameter	Estimate	Standard Error	Chi- Square	Pr > ChiSq
Intercept	1	0,1925	47,2	<,0001
D4	2 -0,4962	0,12	17,11	<,0001
D27	0 0,5542	0,1825	9,22	0,0024
	1 0,066	0,0897	0,54	0,4619

	2	-0,1445	0,0858	2,83	0,0923
MARRIED	2	-0,101	0,0319	10,05	0,0015
hpeduc	0	-0,2731	0,0384	50,55	<,0001
	1	0,2504	0,039	41,29	<,0001
hpdisabl	0	-0,2643	0,0606	19,04	<,0001
hpumas	0	-0,2358	0,041	33,01	<,0001
hpethnat	0	0,1336	0,1529	0,76	0,3822
	1	-0,4758	0,1554	9,38	0,0022
speduc	0	-0,1566	0,04	15,29	<,0001
	1	0,1077	0,0414	6,77	0,0093
spdisabl	0	-0,6267	0,0584	115,11	<,0001
spumas	0	-0,2616	0,0366	50,99	<,0001
spethnat	0	0,2808	0,1171	5,75	0,0165
	1	-0,2069	0,114	3,29	0,0696

**Maximum Likelihood Analysis of Variance**  
Sekundær PSM, husstande med partner

Source	DF	Chi-Square	Pr > ChiSq
Intercept	1	47,2	<,0001
D4	1	17,11	<,0001
D27	3	33,04	<,0001
MARRIED	1	10,05	0,0015
Hpeduc	2	93,21	<,0001
Hpdisabl	1	19,04	<,0001
Hpumas	1	33,01	<,0001
Hpethnat	2	35,93	<,0001
Speduc	2	22,22	<,0001
Spdisabl	1	115,11	<,0001
Spumas	1	50,99	<,0001
Spethnat	2	24,15	<,0001
Likelihood Ratio	1,00E+03	1719,27	<,0001

## Gennemsnitlig vægtning

Datasæt	Kategori / År	Gennemsnitlig vægtning for alle husstande							
		2009	2008	2007	2006	2005	2004	2003	2002
2008	Under 70%-medianen	2,58	2,55	2,47	2,48	2,47	2,33	2,48	2,37
	Over 70%-medianen	0,82	0,83	0,84	0,84	0,84	0,85	0,84	0,85
2006	Under 70%-medianen	2,46	2,42	2,35	2,37	2,36	2,23	2,37	2,28
	Over 70%-medianen	0,76	0,77	0,78	0,78	0,78	0,79	0,78	0,79
2004	Under 70%-medianen	2,43	2,39	2,32	2,34	2,33	2,20	2,34	2,25
	Over 70%-medianen	0,76	0,76	0,77	0,77	0,77	0,78	0,77	0,78
2002	Under 70%-medianen	2,46	2,42	2,35	2,37	2,36	2,23	2,37	2,28
	Over 70%-medianen	0,76	0,77	0,78	0,78	0,78	0,79	0,78	0,79

For at vurdere konsekvensen af vores vægtning er ovenstående skema af vægtenes gennemsnit angivet.

Der er et klart system i, at husstande under 70% medianen skal vægtes 2,5, og husstande over 70% medianen kun skal vægtes omkring 0,8. Det indikerer, at vores samples er pænt fordelte, og at der ikke sker uigennemskuelige beregninger.

## Resultater for data

### ESS: Logistisk Regression

#### Logistic Regression Results (Husstande uden partner)

#### The LOGISTIC Procedure

##### Model Information

Data Set	SASUSER.FILTER_FOR_CON_DATA_FORMATT_0005
Response Variable	In_Poverty
Number of Response Levels	2
Weight Variable	Weight_Household_2009
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	849
Number of Observations Used	849
Sum of Weights Read	1956.889
Sum of Weights Used	1956.889

#### Response Profile

Ordered Value	In_Poverty	Total Frequency	Total Weight
1	0	425	641.2671
2	1	424	1315.6223

Probability modeled is In\_Poverty='1'.

#### Class Level Information

Class	Value Design Variables			
GNDR	1	1		
	2	-1		
BLGETMG	-1	1	0	
	0	0	1	
	1	-1	-1	
EDULVL	0	1	0	
	1	0	1	
	2	-1	-1	
EDULVLM	-1	1	0	0
	0	0	1	0
	1	0	0	1
	2	-1	-1	-1
EDULVLF	-1	1	0	0

	0	0	1	0
	1	0	0	1
	2	-1	-1	-1
EDCTN	0	1		
	1	-1		
Head_Unemployed	0	1		
	1	-1		
DSBLD	0	1		
	1	-1		
RTRD	0	1		
	1	-1		
Number_Of_Children_SEC0		1	0	
	1	0	1	
	2	-1	-1	
HHMMB	1	1	0	
	2	0	1	
	3	-1	-1	

**Model Convergence Status**

Convergence criterion (GCONV=1E-8) satisfied.

**Model Fit Statistics**

	Intercept	Intercept and Covariates
Criterion	Only	Covariates
AIC	2477.605	2259.832
SC	2482.349	2354.713
-2 Log L	2475.605	2219.832

**Testing Global Null Hypothesis: BETA=0**

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	255.7732	19	<.0001
Score	230.4672	19	<.0001
Wald	193.4027	19	<.0001

**Type 3 Analysis of Effects****Wald**

<b>Effect</b>	<b>DF</b>	<b>Chi-Square</b>	<b>Pr &gt; ChiSq</b>
GNDR	1	0.0046	0.9462
BLGETMG	2	1.4570	0.4826
EDULVL	2	19.6619	<.0001
EDULVLM	3	7.6953	0.0527
EDULVLF	3	4.8543	0.1828
EDCTN	1	128.6529	<.0001
Head_Unemployed	1	0.1055	0.7454
DSBLD	1	1.3617	0.2433
RTRD	1	2.0562	0.1516
Number_Of_Children_S	2	10.7154	0.0047
HHMMB	2	2.9117	0.2332

**Analysis of Maximum Likelihood Estimates****Standard****Wald**

<b>Parameter</b>	<b>DF</b>	<b>Estimate</b>	<b>Error</b>	<b>Chi-Square</b>	<b>Pr &gt; ChiSq</b>
Intercept	1	5.0963	155.4	0.0011	0.9738
GNDR	1	-0.00364	0.0539	0.0046	0.9462
BLGETMG	-1	1	8.0934	310.8	0.0007
BLGETMG	0	1	-3.8298	155.4	0.0006
EDULVL	0	1	0.3059	0.0863	12.5473
EDULVL	1	1	-0.2258	0.0752	9.0153
EDULVLM	-1	1	-2.4209	1.2826	3.5626
EDULVLM	0	1	0.7676	0.4352	3.1112
EDULVLM	1	1	0.6407	0.4367	2.1524
EDULVLF	-1	1	0.7282	0.4948	2.1658
EDULVLF	0	1	-0.1022	0.1870	0.2986
EDULVLF	1	1	-0.2974	0.1809	2.7020
EDCTN	0	1	-1.0408	0.0918	128.6529
Head_Unemployed	0	1	-0.0323	0.0996	0.1055
DSBLD	0	1	0.1344	0.1152	1.3617
RTRD	0	1	-0.1011	0.0705	2.0562
Number_Of_Children_S	0	1	-0.4353	0.2551	2.9124
Number_Of_Children_S	1	1	0.5575	0.2005	7.7348
HHMMB	1	1	0.3304	0.2066	2.5573
HHMMB	2	1	0.0126	0.1799	0.0049

**Odds Ratio Estimates**

<b>Effect</b>	<b>90% Wald</b>		
	<b>Point Estimate</b>	<b>Confidence Limits</b>	
GNDR 1 vs 2	0.993	0.831	1.185
BLGETMG -1 vs 1	>999.999	<0.001	>999.999
BLGETMG 0 vs 1	1.543	0.854	2.788
EDULVL 0 vs 2	1.471	1.091	1.983
EDULVL 1 vs 2	0.864	0.658	1.136
EDULVLM -1 vs 2	0.032	0.002	0.543
EDULVLM 0 vs 2	0.783	0.575	1.065
EDULVLM 1 vs 2	0.689	0.511	0.930
EDULVLF -1 vs 2	2.877	0.959	8.631
EDULVLF 0 vs 2	1.254	0.940	1.673
EDULVLF 1 vs 2	1.032	0.799	1.333
EDCTN 0 vs 1	0.125	0.092	0.169
Head_Unemployed 0 vs 1	0.937	0.675	1.301
DSBLD 0 vs 1	1.309	0.896	1.912
RTRD 0 vs 1	0.817	0.648	1.030
Number_Of_Children_S 0 vs 2	0.731	0.296	1.805
Number_Of_Children_S 1 vs 2	1.973	0.901	4.323
HHMMB 1 vs 3	1.961	0.847	4.542
HHMMB 2 vs 3	1.427	0.649	3.139

**Association of Predicted Probabilities and Observed Responses**

Percent Concordant	64.5	Somers' D	0.310
Percent Discordant	33.5	Gamma	0.317
Percent Tied	2.0	Tau-a	0.155
Pairs	180200	c	0.655

**Profile Likelihood Confidence Interval for Parameters**

<b>Parameter</b>		<b>Estimate</b>	<b>90% Confidence Limits</b>
Intercept		5.0963	0.2834
GNDR	1	-0.00364	-0.0923 0.0852
BLGETMG	-1	8.0934	-0.7834
BLGETMG	0	-3.8298	. 0.6690
EDULVL	0	0.3059	0.1640 0.4482
EDULVL	1	-0.2258	-0.3499 -0.1024
EDULVLM	-1	-2.4209	-5.8171 -0.8199
EDULVLM	0	0.7676	0.2137 1.9048
EDULVLM	1	0.6407	0.0837 1.7790
EDULVLF	-1	0.7282	-0.0247 1.6383
EDULVLF	0	-0.1022	-0.4351 0.1897
EDULVLF	1	-0.2974	-0.6219 -0.0164
EDCTN	0	-1.0408	-1.1941 -0.8920
Head_Unemployed	0	-0.0323	-0.1971 0.1309
DSBLD	0	0.1344	-0.0559 0.3239
RTRD	0	-0.1011	-0.2171 0.0149
Number_Of_Children_S	0	-0.4353	-0.8565 -0.0156
Number_Of_Children_S	1	0.5575	0.2321 0.8936
HHMMB	1	0.3304	-0.0131 0.6685
HHMMB	2	0.0126	-0.2879 0.3063

Year: 2009.

**Logistic Regression Results (Husstande med partner)**  
**The LOGISTIC Procedure**

**Model Information**

Data Set	SASUSER.FILTER_FOR_CON_DATA_FORMATT_0004
Response Variable	In_Poverty
Number of Response Levels	2
Weight Variable	Weight_Household_2009
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	500
Number of Observations Used	500
Sum of Weights Read	399.3754
Sum of Weights Used	399.3754

**Response Profile**

Ordered Value	Total Frequency	Total Weight
1	247	125.52755
2	253	273.84783

Probability modeled is In\_Poverty='1'.

**Class Level Information**

Class	Value Design Variables			
GNDR	1	1		
	2	-1		
BLGETMG	-1	1	0	
	0	0	1	
	1	-1	-1	
EDULVL	0	1	0	
	1	0	1	
	2	-1	-1	
EDULVLP	-1	1	0	0
	0	0	1	0
	1	0	0	1
	2	-1	-1	-1
EDULVLM	-1	1	0	0
	0	0	1	0

	1	0	0	1
	2	-1	-1	-1
EDULVLF	-1	1	0	0
	0	0	1	0
	1	0	0	1
	2	-1	-1	-1
EDCTN	0	1		
	1	-1		
Head_Unemployed	0	1		
	1	-1		
DSBLD	0	1		
	1	-1		
RTRD	0	1		
	1	-1		
Number_Of_Children_SEC0	0	1	0	
	1	0	1	
	2	-1	-1	
Spouse_Unemployed	0	1		
	1	-1		
DSBLDP	0	1		
	1	-1		
RTRDP	0	1		
	1	-1		
EDCTNP	0	1		
	1	-1		
HHMMB	1	1	0	
	2	0	1	
	3	-1	-1	

**Model Convergence Status**

Convergence criterion (GCONV=1E-8) satisfied.

**Model Fit Statistics**

Criterion	Intercept	
	Intercept	and Only Covariates
AIC	499.227	519.149
SC	503.441	632.944
-2 Log L	497.227	465.149

**Testing Global Null Hypothesis: BETA=0**

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	32.0774	26	0.1906
Score	31.5260	26	0.2093
Wald	23.4957	26	0.6048

**Type 3 Analysis of Effects**

Effect	Wald	DF	Chi-Square	Pr > ChiSq
GNDR	1	3.2908	0.0697	
BLGETMG	2	4.3781	0.1120	
EDULVL	2	2.5102	0.2851	
EDULVLP	3	6.1129	0.1062	
EDULVLM	3	5.2042	0.1574	
EDULVLF	3	1.9856	0.5754	
EDCTN	1	0.0042	0.9484	
Head_Unemployed	1	0.2421	0.6227	
DSBLD	1	0.7269	0.3939	
RTRD	1	0.0609	0.8050	
Number_Of_Children_S	2	0.8970	0.6386	
Spouse_Unemployed	1	1.4288	0.2320	
DSBLDP	1	0.2595	0.6105	
RTRDP	1	0.0571	0.8112	
EDCTNP	1	1.4266	0.2323	
HHMMB	2	0.2240	0.8940	

**Analysis of Maximum Likelihood Estimates**

Parameter	DF	Estimate	Standard	Wald		
			Error	Chi-Square	Pr > ChiSq	
Intercept	1	-4.0219	295.3	0.0002	0.9891	
GNDR	1	-0.2205	0.1216	3.2908	0.0697	
BLGETMG	-1	1	0.8291	0.8355	0.9847	0.3211
BLGETMG	0	1	-1.0550	0.5693	3.4339	0.0639
EDULVL	0	1	-0.0640	0.2216	0.0834	0.7727
EDULVL	1	1	-0.2930	0.1874	2.4445	0.1179
EDULVLP	-1	1	-0.8552	2.5604	0.1116	0.7384
EDULVLP	0	1	0.7821	0.8728	0.8030	0.3702
EDULVLP	1	1	0.3219	0.8686	0.1374	0.7109
EDULVLM	-1	1	-0.8377	0.9566	0.7670	0.3812
EDULVLM	0	1	0.3898	0.3776	1.0658	0.3019
EDULVLM	1	1	0.7689	0.4096	3.5243	0.0605
EDULVLF	-1	1	0.5917	0.8189	0.5221	0.4700

EDULVLF	0	1	-0.4036	0.3266	1.5271	0.2165
EDULVLF	1	1	-0.3283	0.3345	0.9637	0.3262
EDCTN	0	1	-0.0242	0.3731	0.0042	0.9484
Head_Unemployed	0	1	0.1620	0.3293	0.2421	0.6227
DSBLD	0	1	0.2617	0.3069	0.7269	0.3939
RTRD	0	1	-0.0543	0.2198	0.0609	0.8050
Number_Of_Children_S 0	1	-0.6832	0.7777	0.7718	0.3797	
Number_Of_Children_S 1	1	0.2816	0.5224	0.2906	0.5899	
Spouse_Unemployed	0	1	-0.4125	0.3451	1.4288	0.2320
DSBLDP	0	1	-0.2317	0.4547	0.2595	0.6105
RTRDP	0	1	-0.0547	0.2289	0.0571	0.8112
EDCTNP	0	1	0.4380	0.3667	1.4266	0.2323
-HHMMB	1	1	-8.6673	590.7	0.0002	0.9883
HHMMB	2	1	4.6067	295.3	0.0002	0.9876

**Odds Ratio Estimates**

<b>Effect</b>	<b>90% Wald</b>		
	<b>Point Estimate</b>	<b>Confidence Limits</b>	
GNDR 1 vs 2	0.643	0.431	0.960
BLGETMG -1 vs 1	1.828	0.237	14.072
BLGETMG 0 vs 1	0.278	0.096	0.804
EDULVL 0 vs 2	0.656	0.296	1.457
EDULVL 1 vs 2	0.522	0.253	1.075
EDULVLP -1 vs 2	0.545	0.002	154.375
EDULVLP 0 vs 2	2.804	1.361	5.778
EDULVLP 1 vs 2	1.770	0.916	3.419
EDULVLM -1 vs 2	0.596	0.069	5.193
EDULVLM 0 vs 2	2.036	0.913	4.538
EDULVLM 1 vs 2	2.974	1.273	6.948
EDULVLF -1 vs 2	1.571	0.234	10.560
EDULVLF 0 vs 2	0.581	0.266	1.267
EDULVLF 1 vs 2	0.626	0.281	1.394
EDCTN 0 vs 1	0.953	0.279	3.251
Head_Unemployed 0 vs 1	1.383	0.468	4.085
DSBLD 0 vs 1	1.688	0.615	4.632
RTRD 0 vs 1	0.897	0.435	1.849
Number_Of_Children_S 0 vs 2	0.338	0.051	2.261
Number_Of_Children_S 1 vs 2	0.887	0.340	2.312
Spouse_Unemployed 0 vs 1	0.438	0.141	1.364
DSBLDP 0 vs 1	0.629	0.141	2.809
RTRDP 0 vs 1	0.896	0.422	1.903
EDCTNP 0 vs 1	2.401	0.719	8.023

HHMMB 1 vs 3	<0.001	<0.001	>999.999
HHMMB 2 vs 3	1.727	0.259	11.531

**Association of Predicted Probabilities and  
Observed Responses**

Percent Concordant	58.1	Somers' D	0.180
Percent Discordant	40.1	Gamma	0.183
Percent Tied	1.8	Tau-a	0.090
Pairs	62491	c	0.590

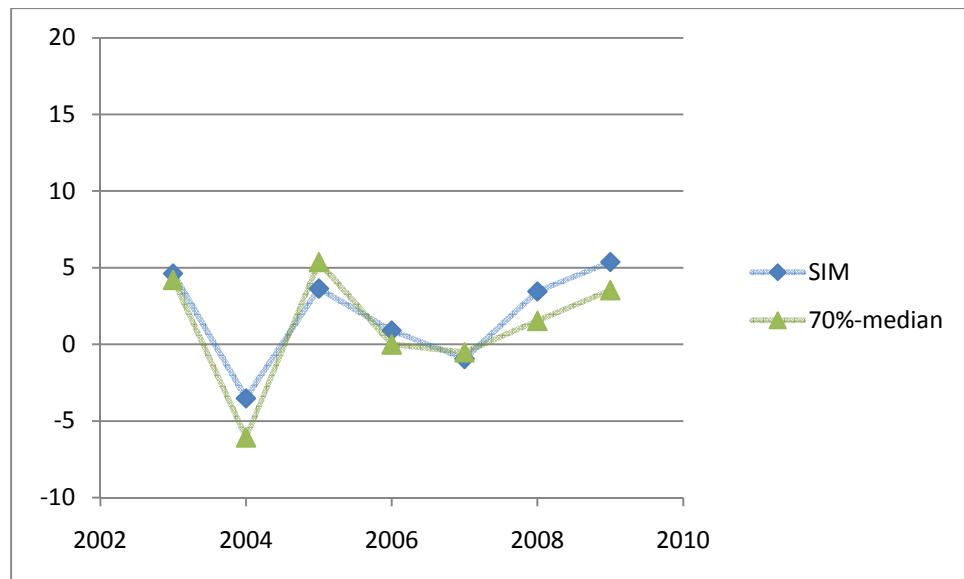
**Profile Likelihood Confidence Interval for Parameters**

Parameter		Estimate	90% Confidence Limits
Intercept		-4.0219	. 1.7710
GNDR	1	-0.2205	-0.4221 -0.0216
BLGETMG	-1	0.8291	-0.3633 2.6197
BLGETMG	0	-1.0550	-2.1189 -0.1691
EDULVL	0	-0.0640	-0.4333 0.2973
EDULVL	1	-0.2930	-0.6072 0.0110
EDULVLP	-1	-0.8552	-10.4814 3.1506
EDULVLP	0	0.7821	-0.5792 3.9987
EDULVLP	1	0.3219	-1.0348 3.5367
EDULVLM	-1	-0.8377	-2.5651 0.7723
EDULVLM	0	0.3898	-0.2348 1.0505
EDULVLM	1	0.7689	0.0973 1.4801
EDULVLF	-1	0.5917	-0.6382 2.1776
EDULVLF	0	-0.4036	-0.9988 0.1044
EDULVLF	1	-0.3283	-0.9346 0.1941
EDCTN	0	-0.0242	-0.6716 0.5731
Head_Unemployed	0	0.1620	-0.3985 0.6969
DSBLD	0	0.2617	-0.2635 0.7588
RTRD	0	-0.0543	-0.4128 0.3135
Number_Of_Children_S	0	-0.6832	-2.0593 0.6275
Number_Of_Children_S	1	0.2816	-0.5756 1.1745
Spouse_Unemployed	0	-0.4125	-1.0216 0.1268
DSBLDP	0	-0.2317	-1.0798 0.4670
RTRDP	0	-0.0547	-0.4294 0.3273
EDCTNP	0	0.4380	-0.1659 1.0574
HHMMB	1	-8.6673	. 6.8719
HHMMB	2	4.6067	-3.1879 .

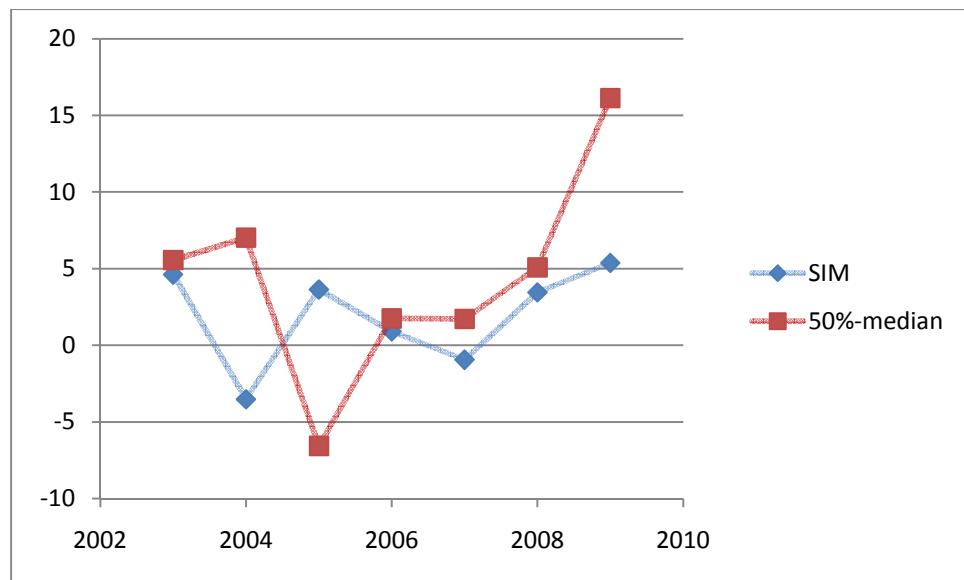
Year: 2009.

**ESS: Resultater**

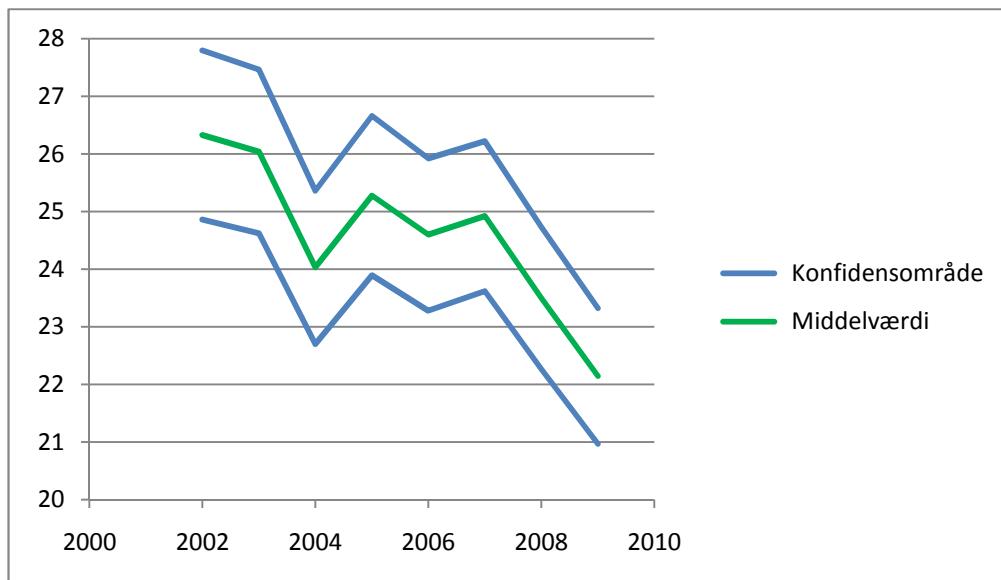
Nedenstående diagram illustrerer de procentvise stigninger for SIM og 70%-medianen fra 2003 til 2009.



Nedenstående diagram illustrerer de procentvise stigninger for SIM og 50%-medianen fra 2003 til 2009.

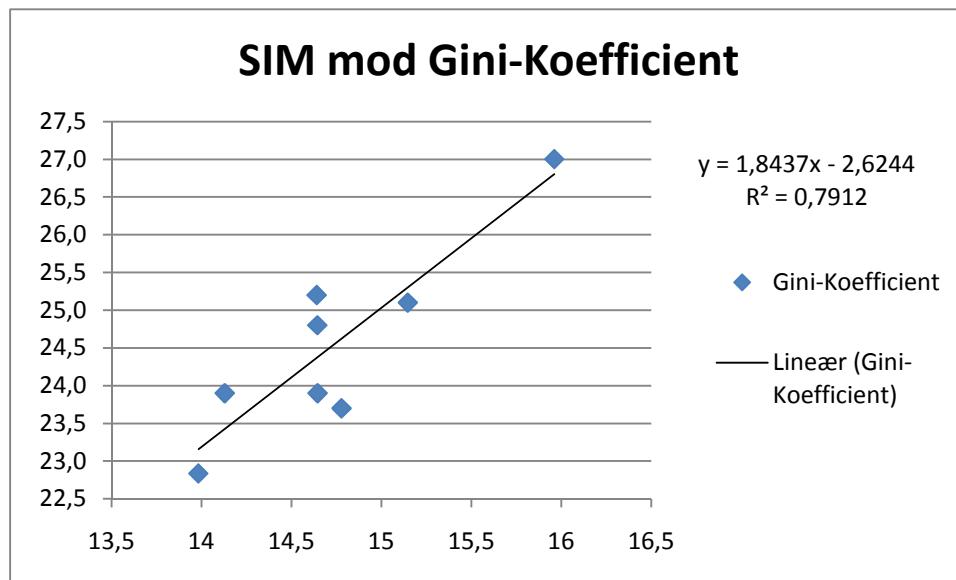


Nedenstående diagram viser procentdelen af fattige med de tilhørende 80% profil-konfidensintervaller for T-fordelingen. Denne skal anvendes til at forstå udviklingen af fattigdom over det sidste årti.

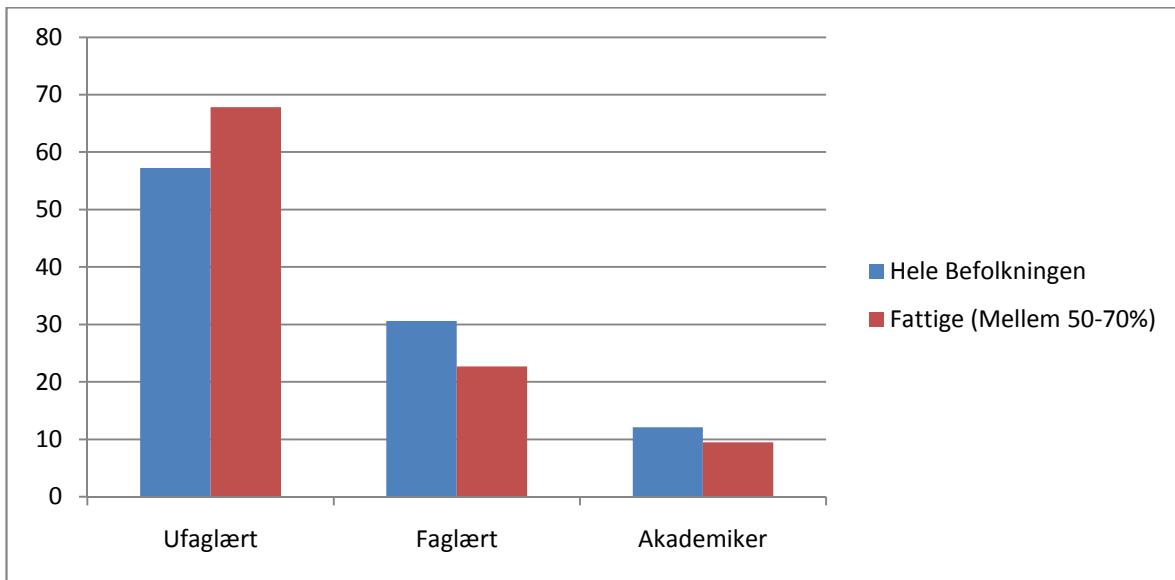


Nedenstående illustrerer sammenhængen mellem SIM og Gini-koefficienten:

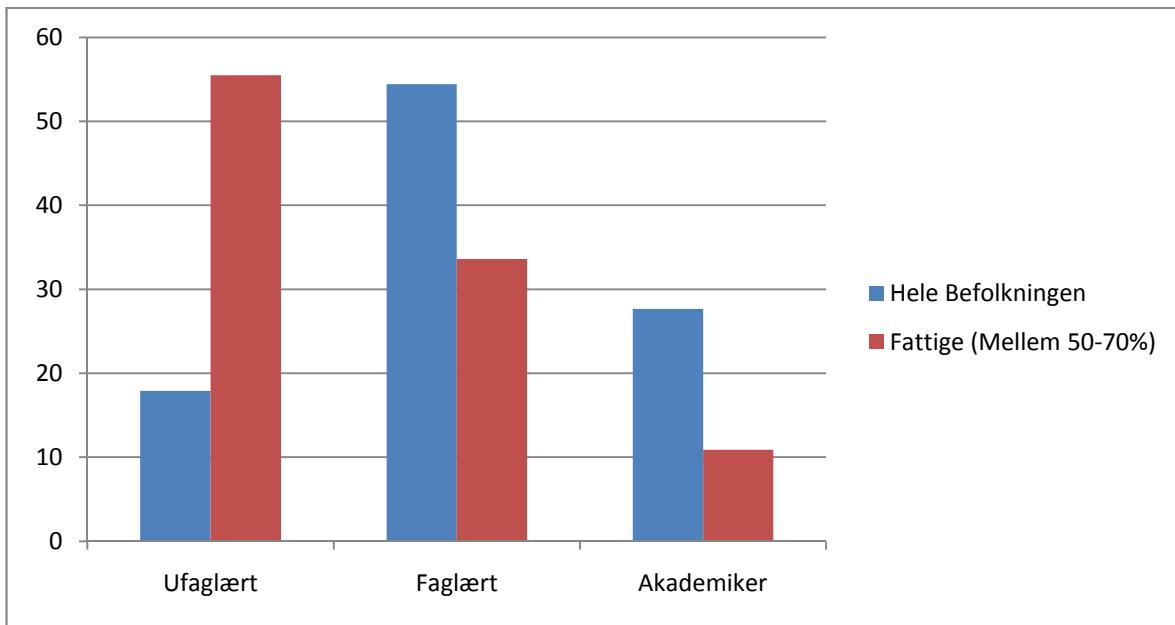
År	Social-mobilitets fattigdomsgrænse	Gini-Koefficient
2009	15,9606	27,0
2008	15,14696	25,1
2007	14,64096	25,2
2006	14,77888	23,7
2005	14,64579	23,9
2004	14,13051	23,9
2003	14,64446	24,8
2002	13,98392	22,8



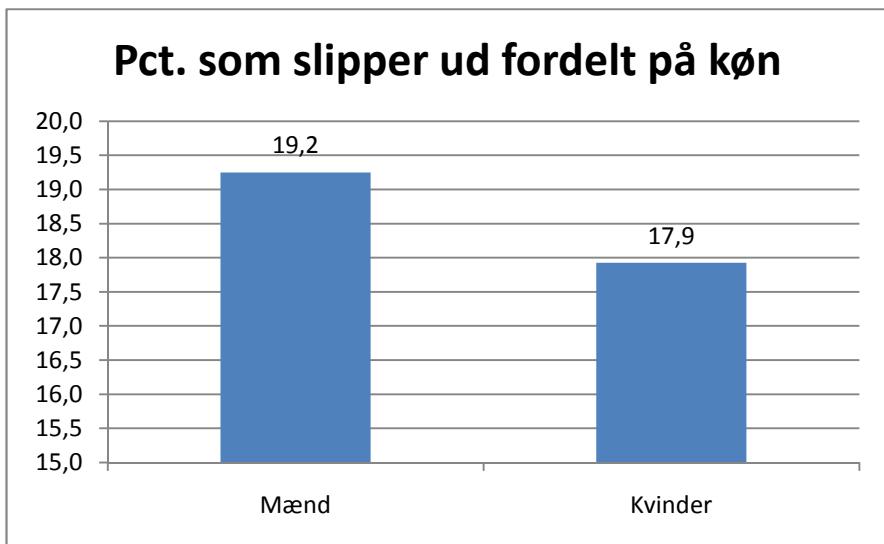
Nedenstående illustrerer sammensætningen af befolkningen og de ifølge SIM fattige mellem 50-70% af indkomstmedianen:



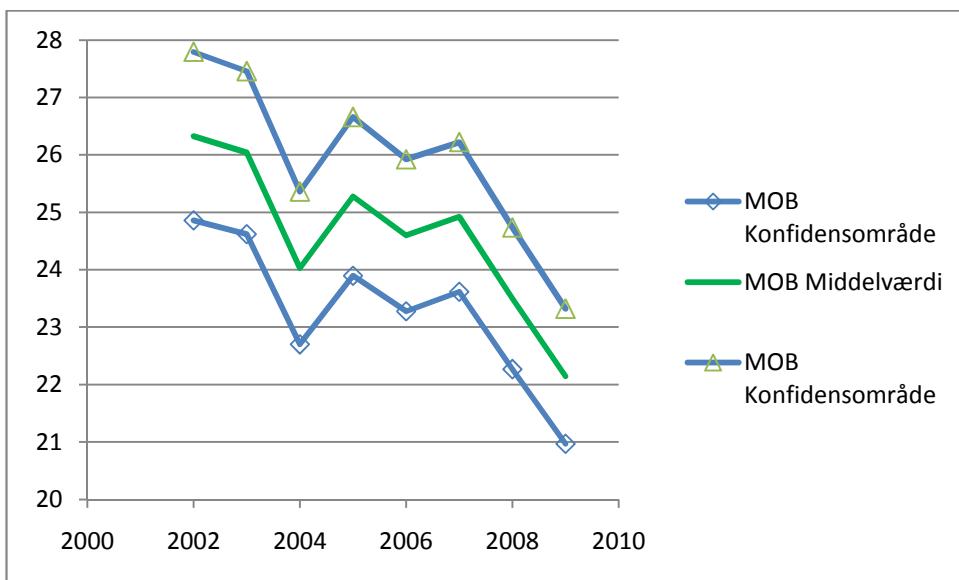
Næste graf illustrerer det samme, dog ud fra partnerens uddannelse:



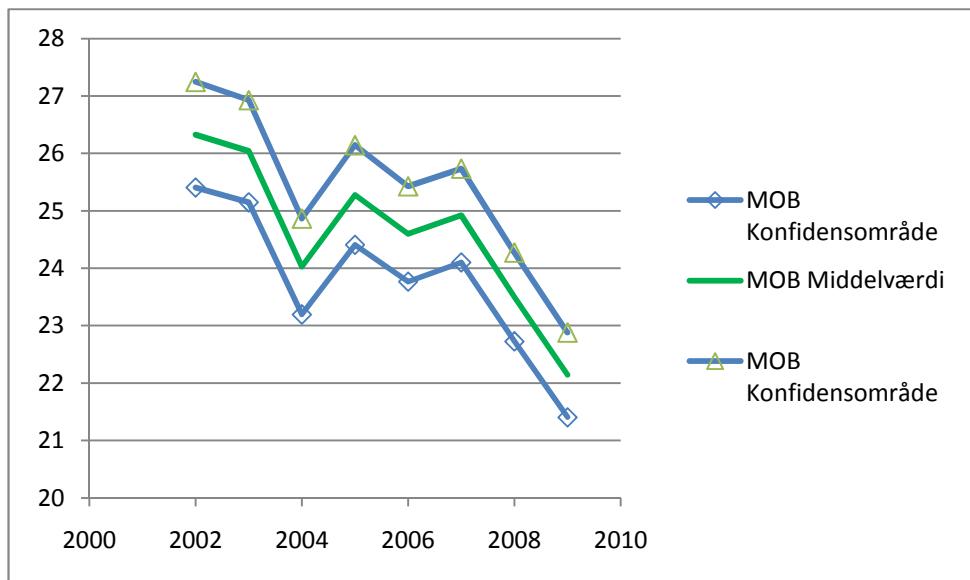
Nedenstående viser procentdelen af henholdsvis mænd og kvinder der slipper fri. Resultatet er ikke signifikant.



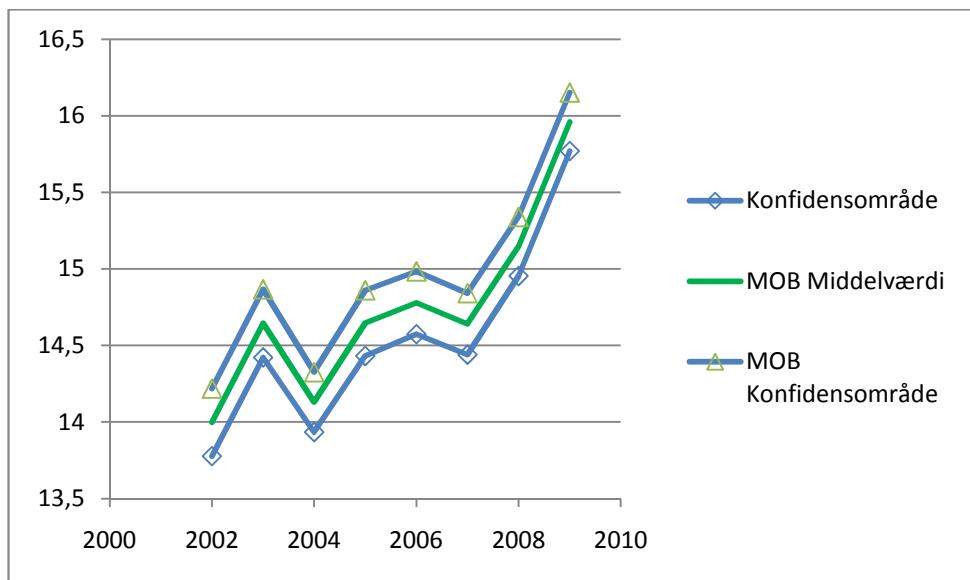
Nedenstående diagram viser procentdelen af fattige som undslipper fattigdom med de tilhørende 80% profil-konfidensintervaller for T-fordelingen. Denne skal anvendes til at forstå udviklingen af fattigdom over det sidste årti.



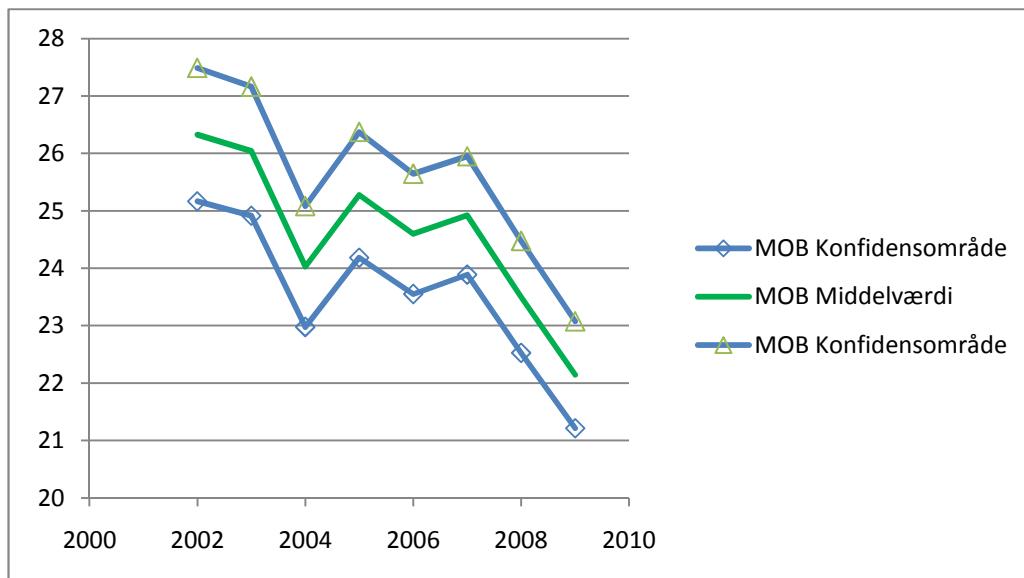
Nedenstående diagram viser procentdelen af fattige med de tilhørende 80% de simultane-konfidensintervaller for T-fordelingen. Denne skal anvendes til at sammenligne årene.



Vi viser yderligere 2 resultater, under normalfordelingsantagelsen, for at understrege robustheden af vores data. Nedenstående diagram viser SIM med tilhørende 80% simultane-konfidensintervaller for normalfordelingen. Denne kan anvendes til at sammenligne enkelte år.

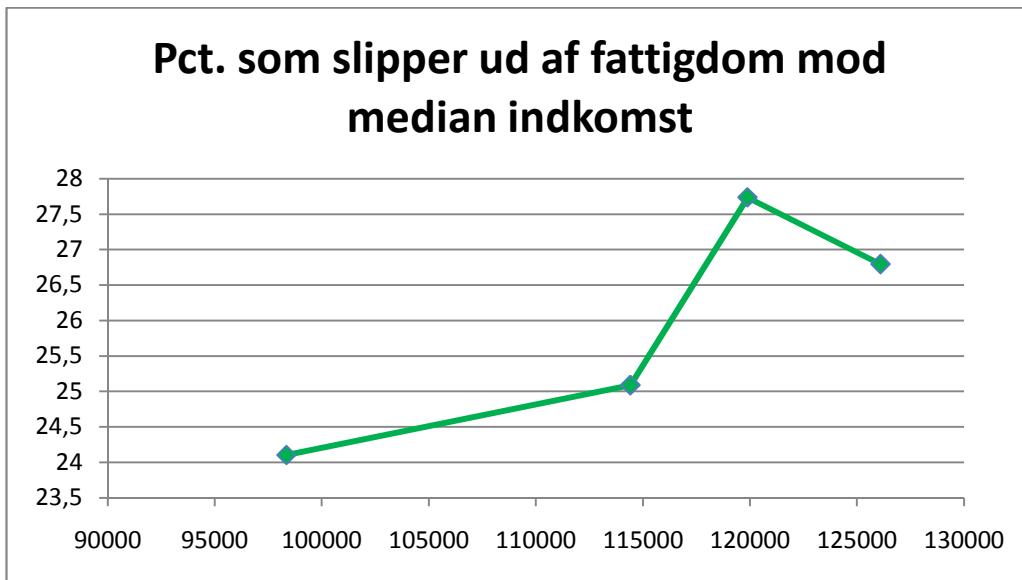


ØNedenstående diagram viser procentdelen af fattige med de tilhørende 80% simultane-konfidensintervaller for normal-fordelingen. Denne kan anvendes til at sammenligne enkelte år.

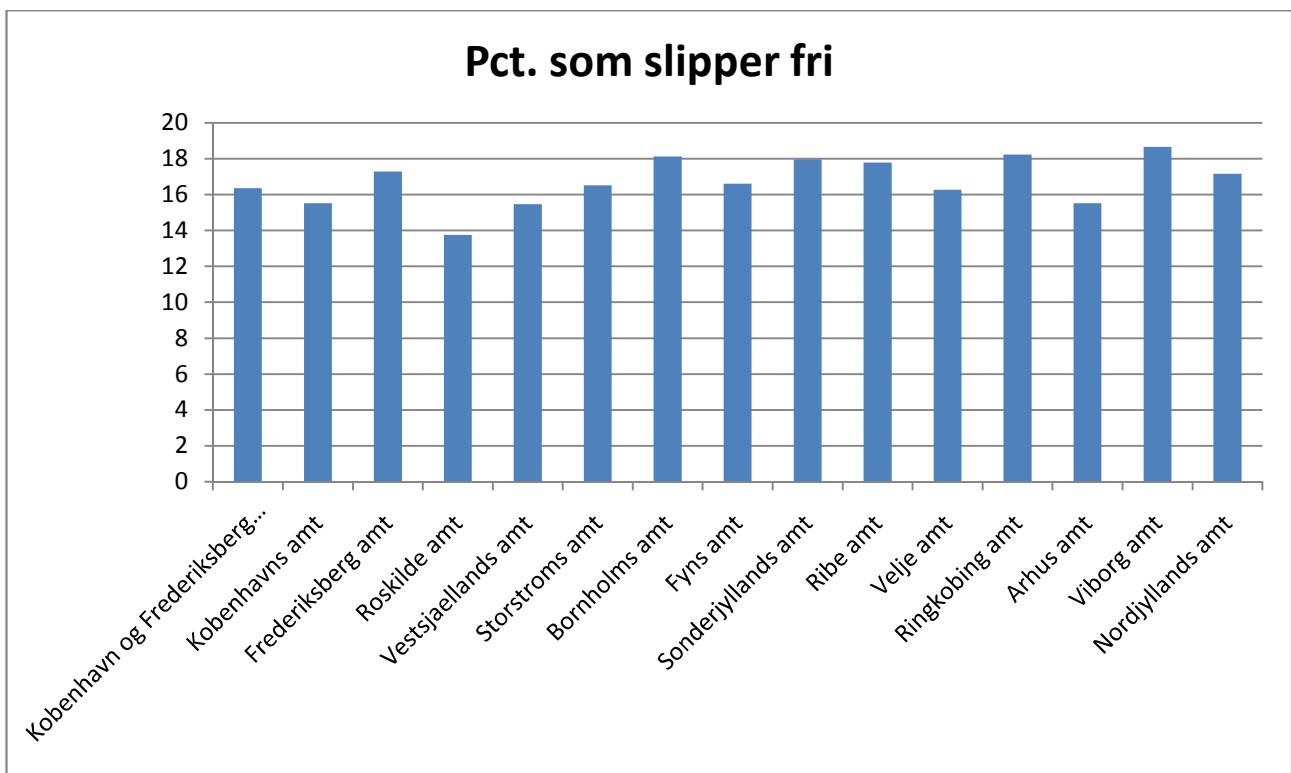


**LIS: Resultater**

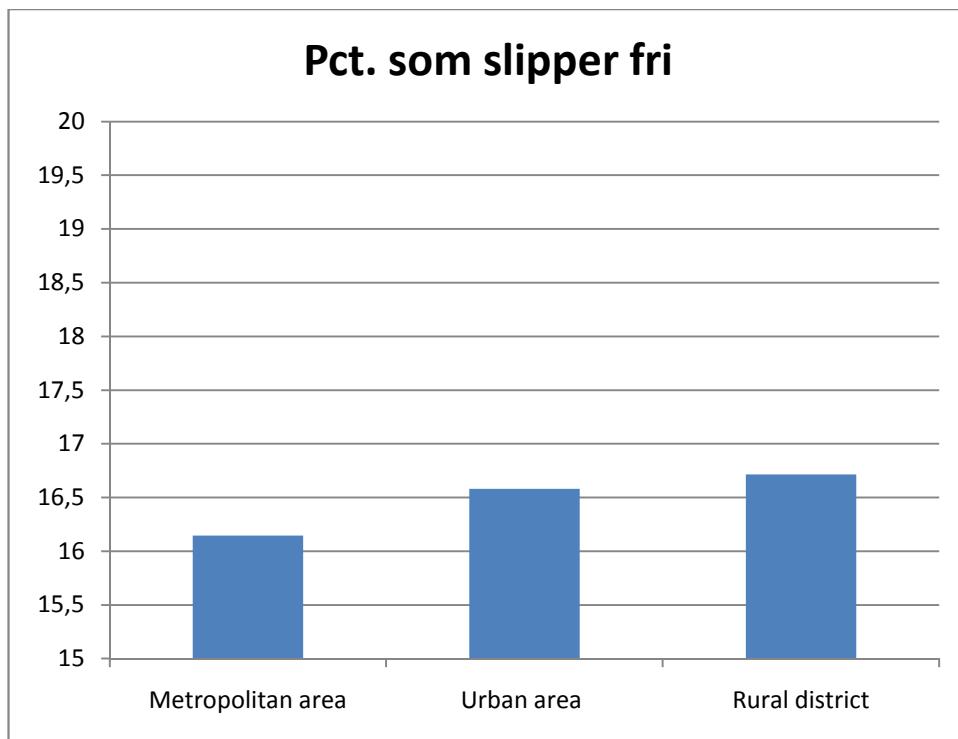
Nedenstående graf viser kvartilerne for alle fattige under 70% medianen, deres chance for at undslippe fattigdom mod deres årlige medianindkomst.



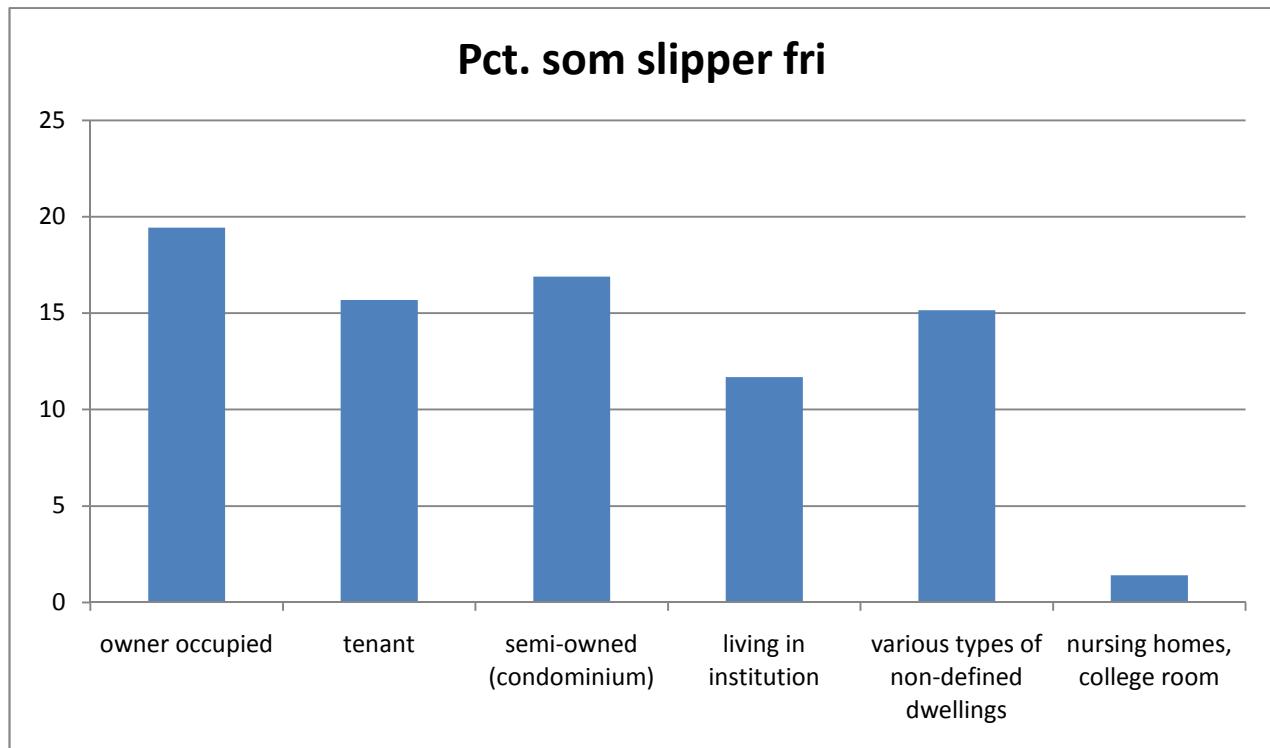
Nedenstående viser procentdelen af mennesker under 70% medianen som slipper fri fordelt på amter.



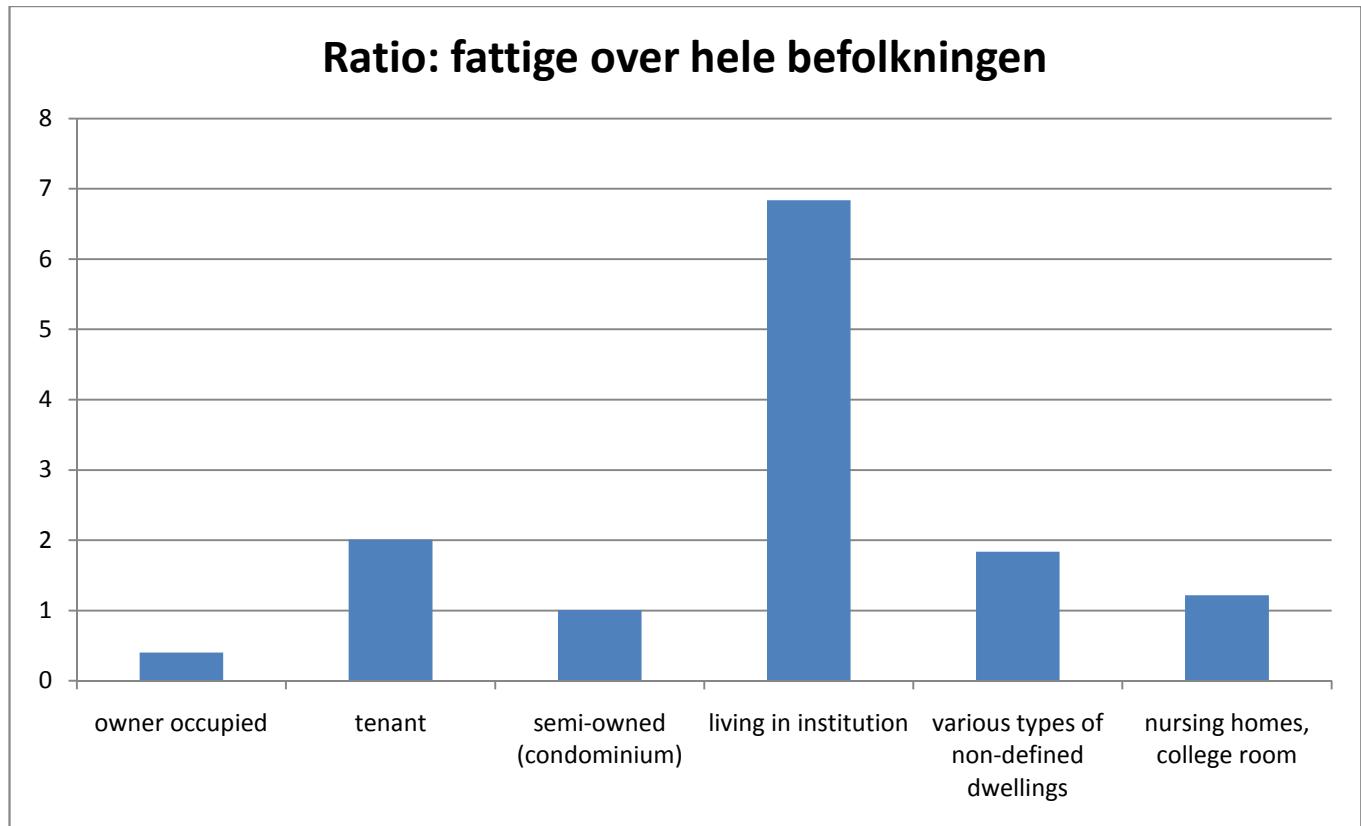
Nedenstående viser procentdelen af mennesker under 70% medianen som slipper fri fordelt på beboelsesområder.



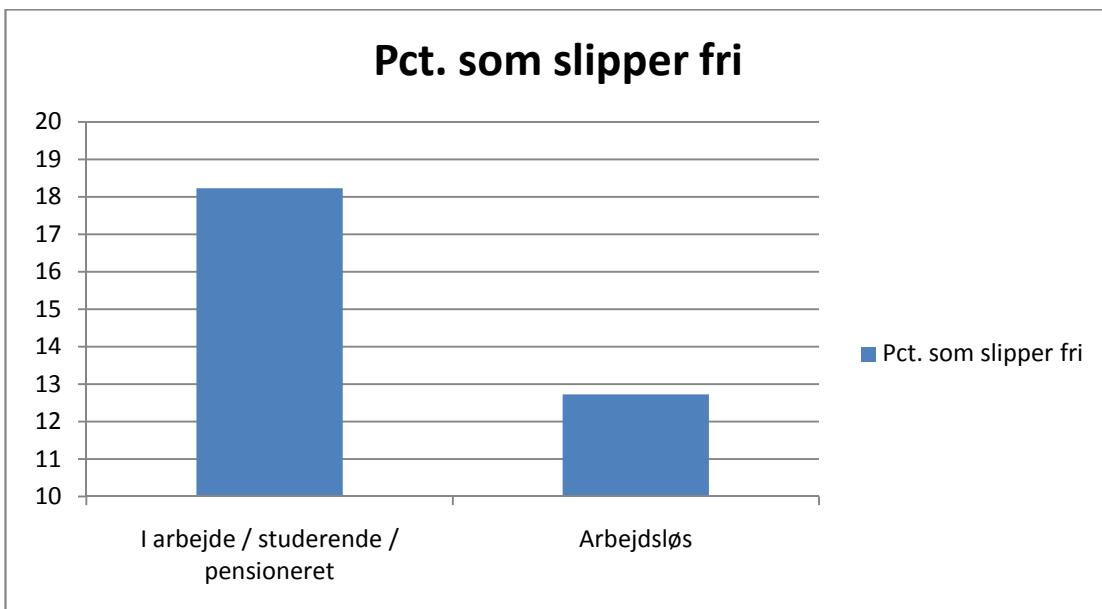
Nedenstående viser procentdelen af mennesker under 70% medianen som slipper fri fordelt på husstandstype.



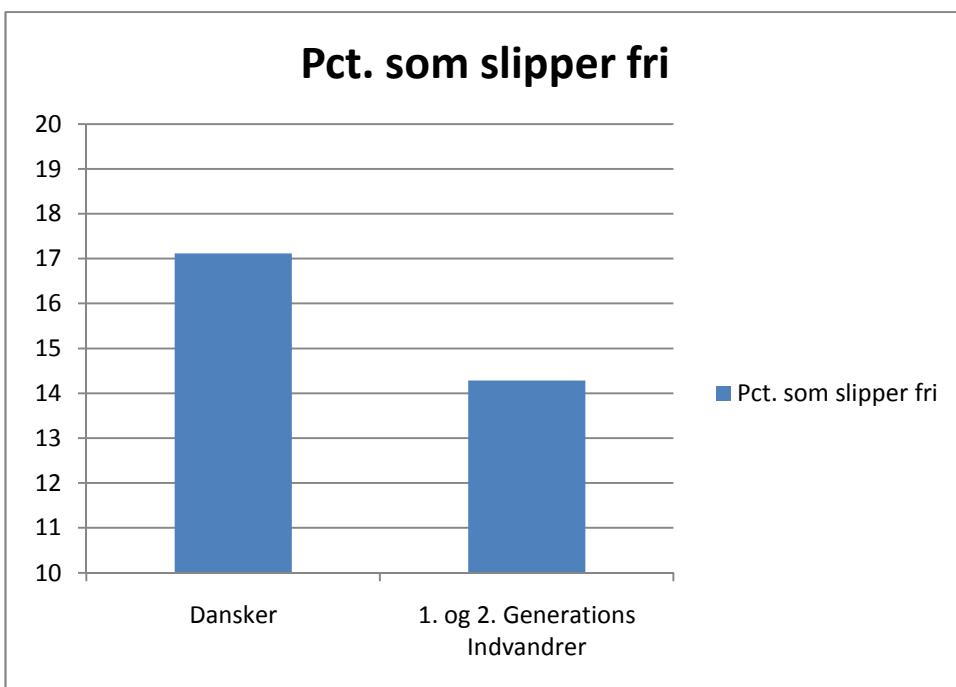
Nedenstående viser ratioen af fattige mennesker fordelt på husstandstype.



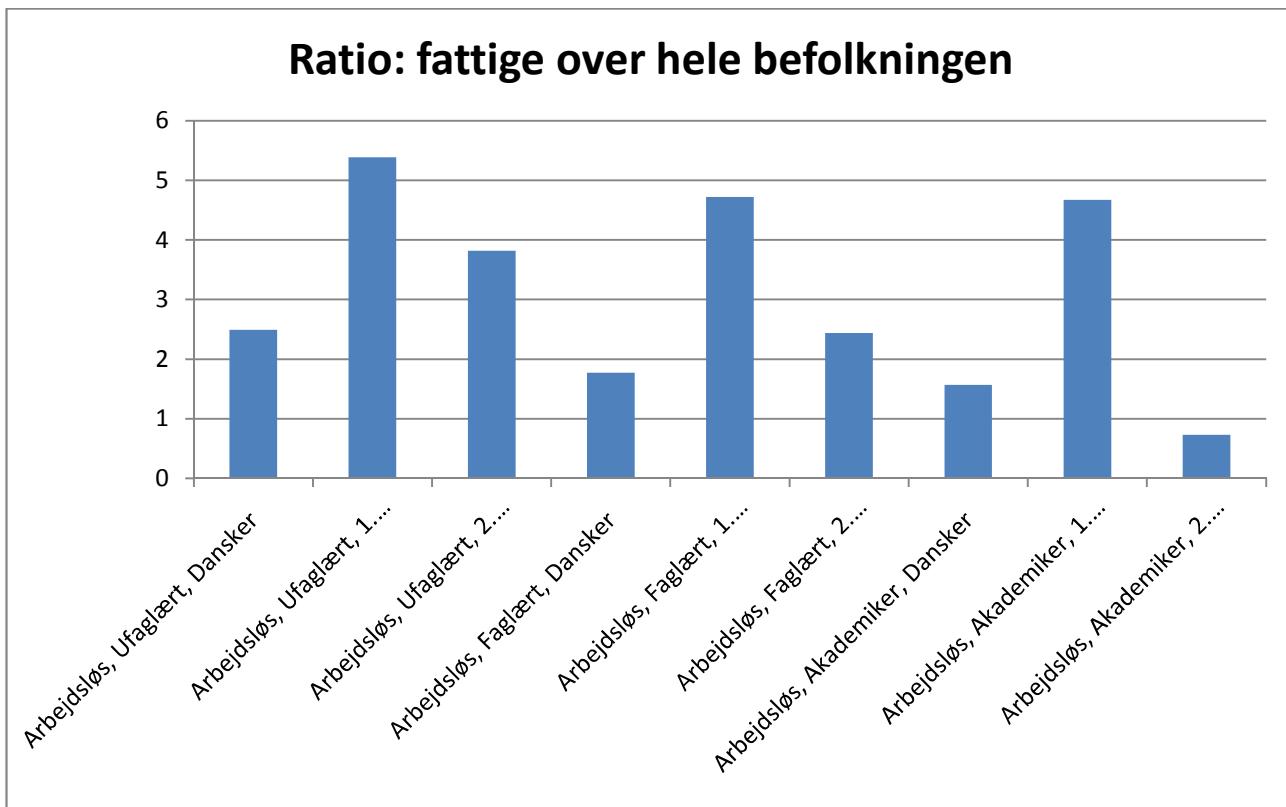
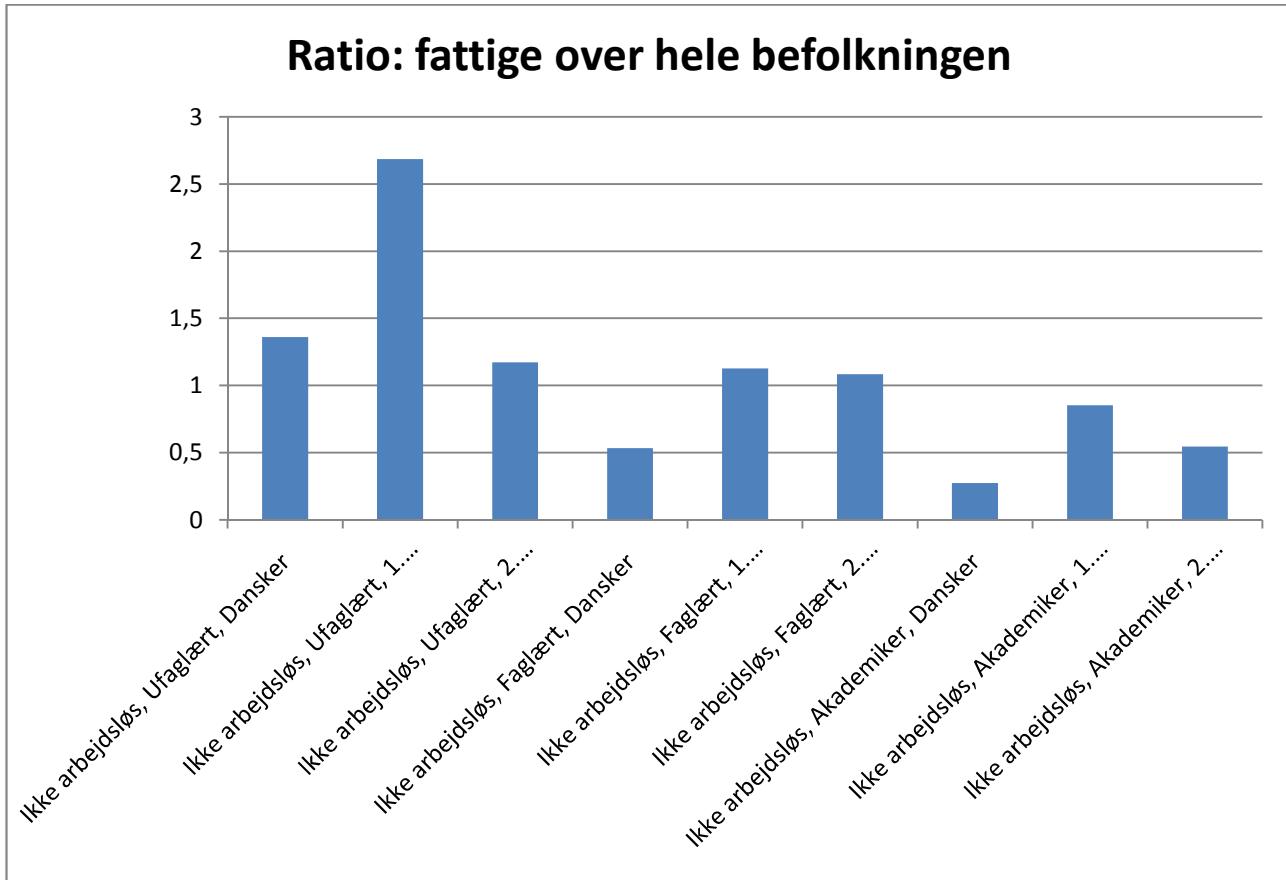
Nedenstående viser procentdelen af mennesker under 70% medianen som slipper fri fordelt på aktivitetstype.



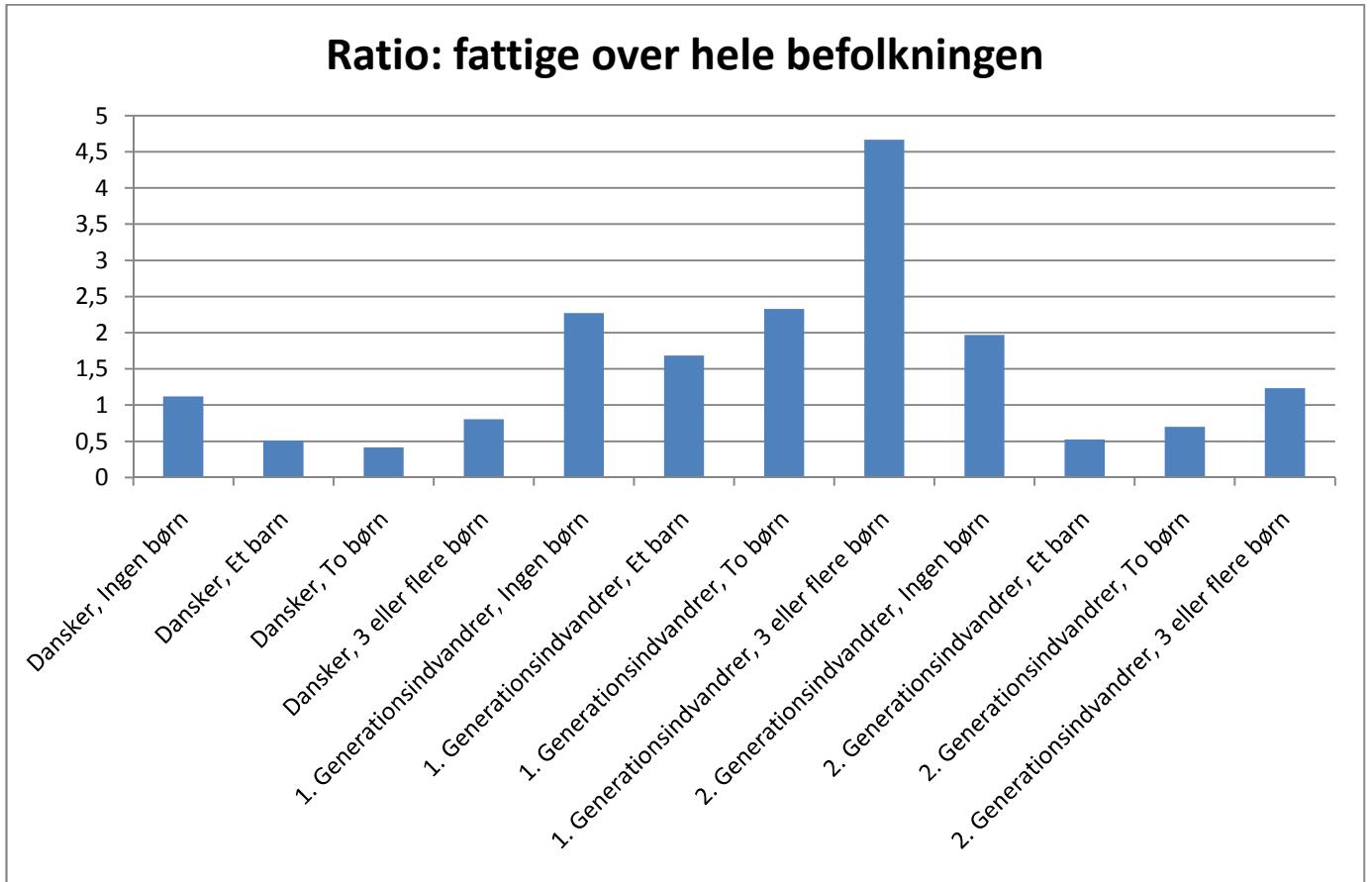
Nedenstående viser procentdelen af mennesker under 70% medianen som slipper fri fordelt på etnicitet.



Nedenstående viser ratioen af fattige fordelt på etnicitet, uddannelsesniveau og arbejdsløshed.



Nedenstående viser ratioen af fattige fordelt på etnicitet og antal børn.



## ESS: Fremskrivning

Vi starter med at fremskrive SIM, udelukkende ved brug af variablen år. Nedenstående viser de statistiske resultater for dette.

```

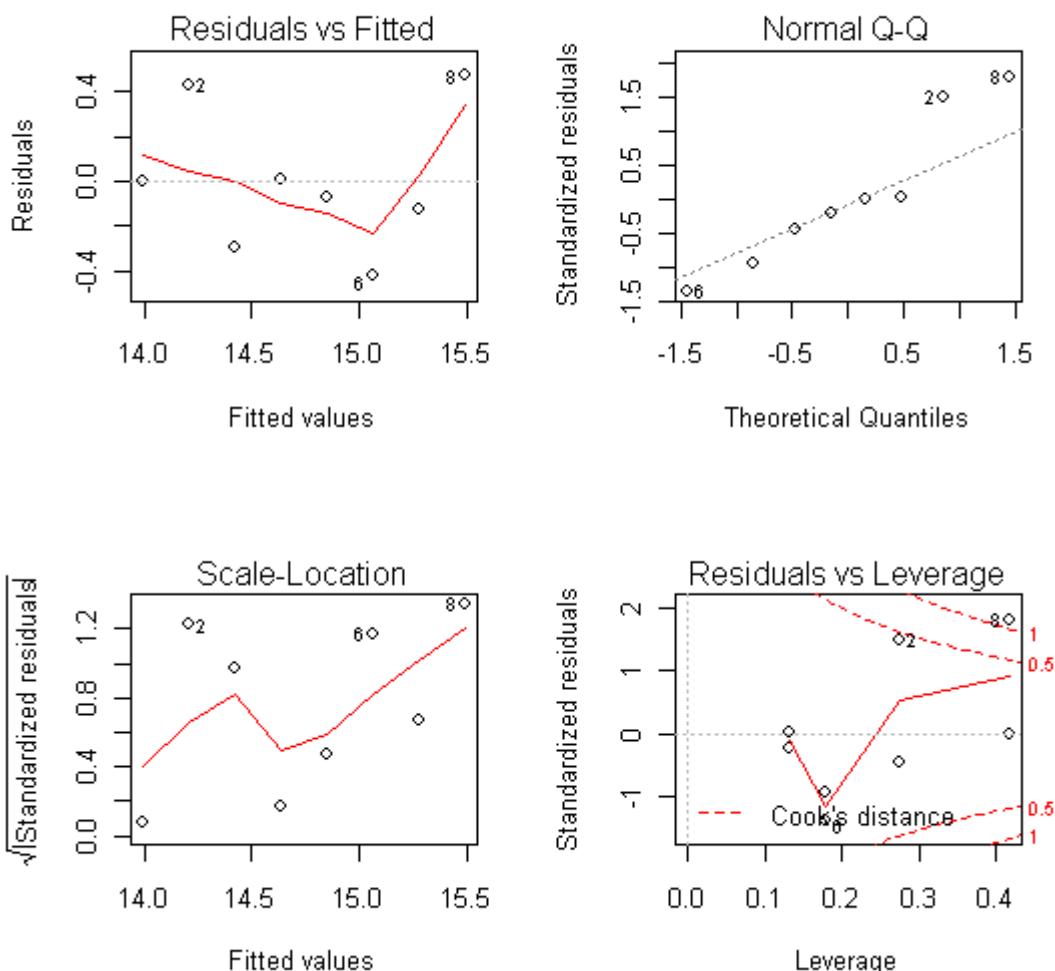
Call:
lm(formula = SIM ~ ÅR, data = mydata)

Residuals:
    Min      1Q Median      3Q     Max 
-0.42222 -0.17034 -0.03487  0.11542  0.47090 

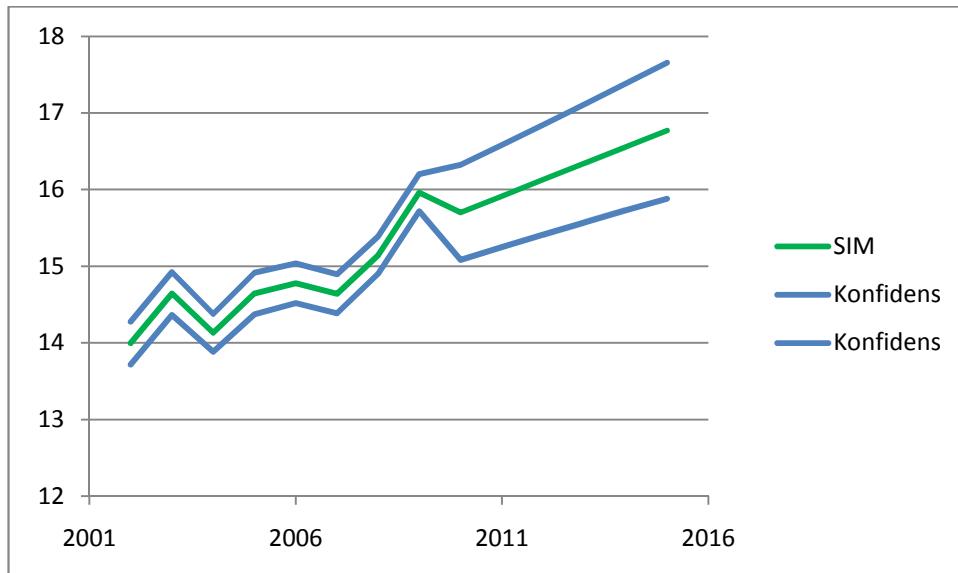
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -412.94642 105.42214 -3.917 0.00783 ** 
ÅR           0.21326  0.05257  4.057 0.00667 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3407 on 6 degrees of freedom
Multiple R-squared: 0.7328,   Adjusted R-squared: 0.6883 
F-statistic: 16.46 on 1 and 6 DF, p-value: 0.006674

```



5 %      95 %  
 (Intercept) -617.8006475 -208.0921904  
 ÅR        0.11111123 0.3154045



Herefter fremskrives SIM ud fra både ÅR og procent af BNP som anvendes til indkomstoverførslер.  
 Nedenstående viser de statistisk resultater for dette.

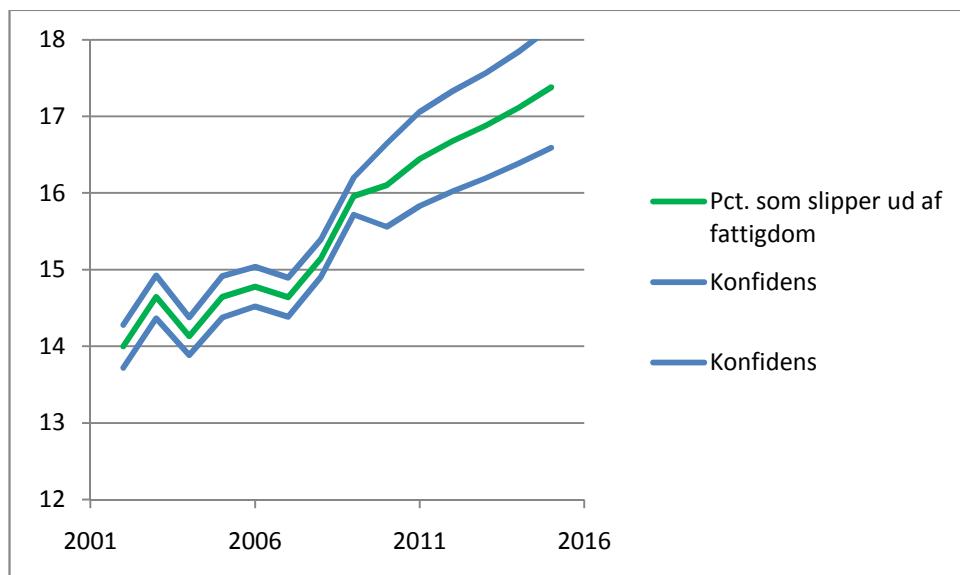
```

Call:
lm(formula = SIM ~ ÅR + TY, data = mydata)

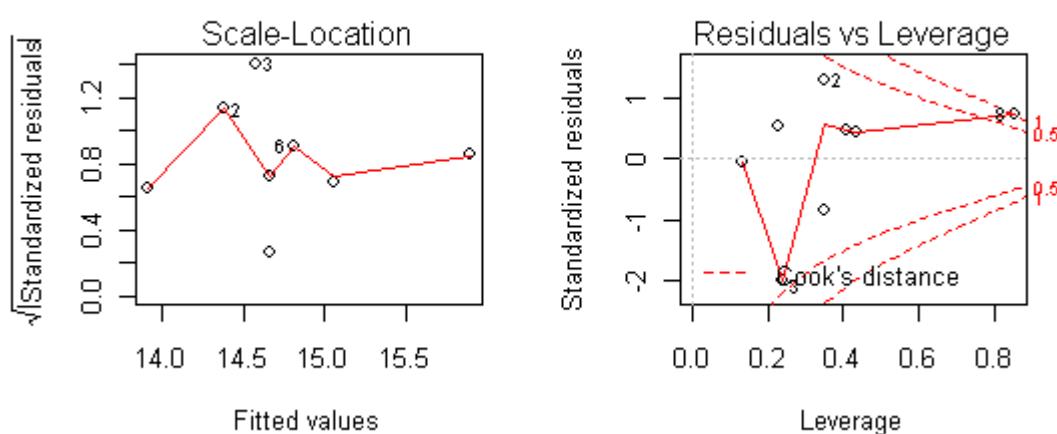
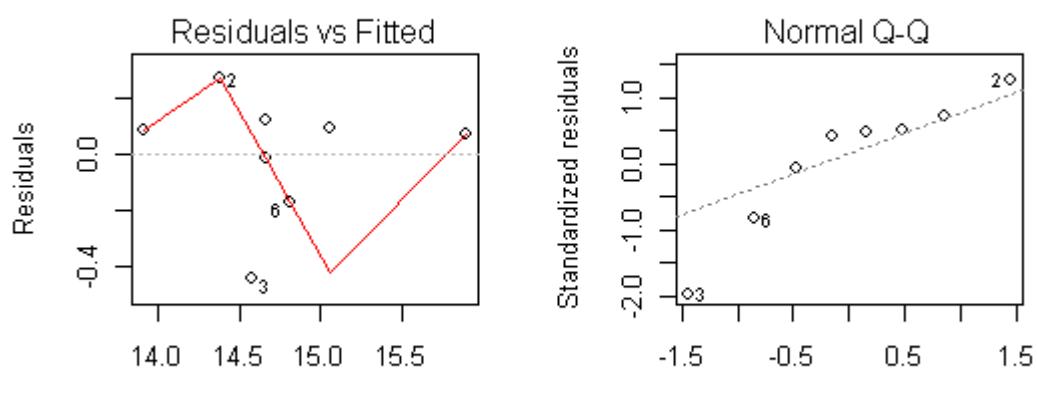
Residuals:
    1     2     3     4     5     6     7     8 
 0.08272 0.26568 -0.44249 -0.01612 0.11872 -0.17246 0.09253 0.07141 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -520.31003   91.80207 -5.668 0.00238 ** 
ÅR          0.26441   0.04528  5.840 0.00208 ** 
TY          0.29249   0.12450  2.349 0.06562 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2573 on 5 degrees of freedom
Multiple R-squared:  0.873,    Adjusted R-squared:  0.8222 
F-statistic: 17.19 on 2 and 5 DF,  p-value: 0.005746
  
```



	5 %	95 %
(Intercept)	-705.29565047	-335.3244103
ÅR	0.17317057	0.3556474
TY	0.04161561	0.5433606



Herefter fremskrives pct. delen af mennesker som slipper ud af fattigdom i forhold til år. Nedenstående viser de statistisk resultater for dette.

Call:

```
lm(formula = PCT ~ ÅR, data = mydata)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.2879	-0.1566	0.1591	0.2923	1.0284

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	979.0130	241.7318	4.050	0.00673 **
ÅR	-0.4759	0.1205	-3.948	0.00755 **

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

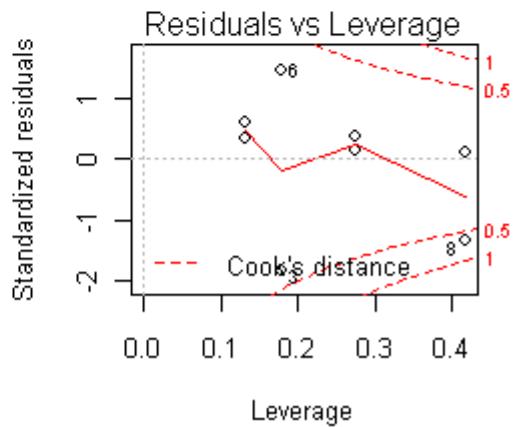
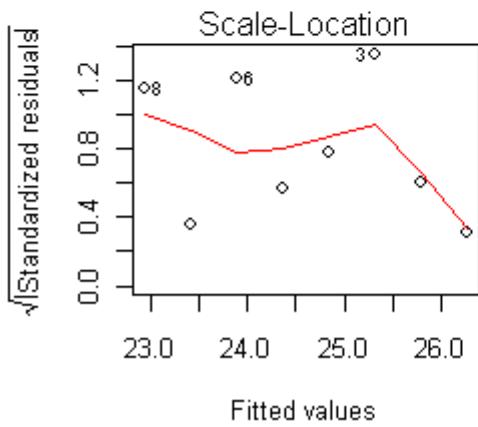
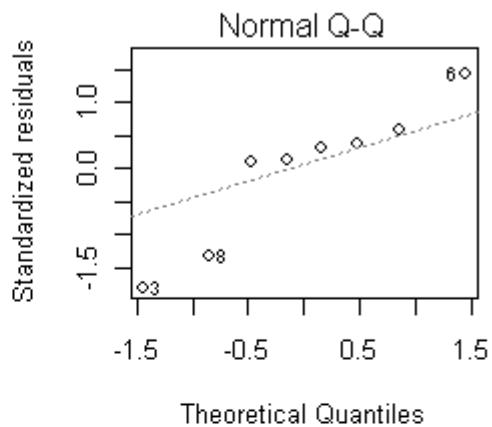
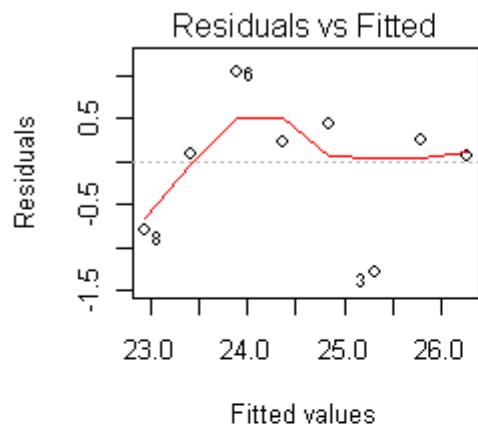
Residual standard error: 0.7812 on 6 degrees of freedom

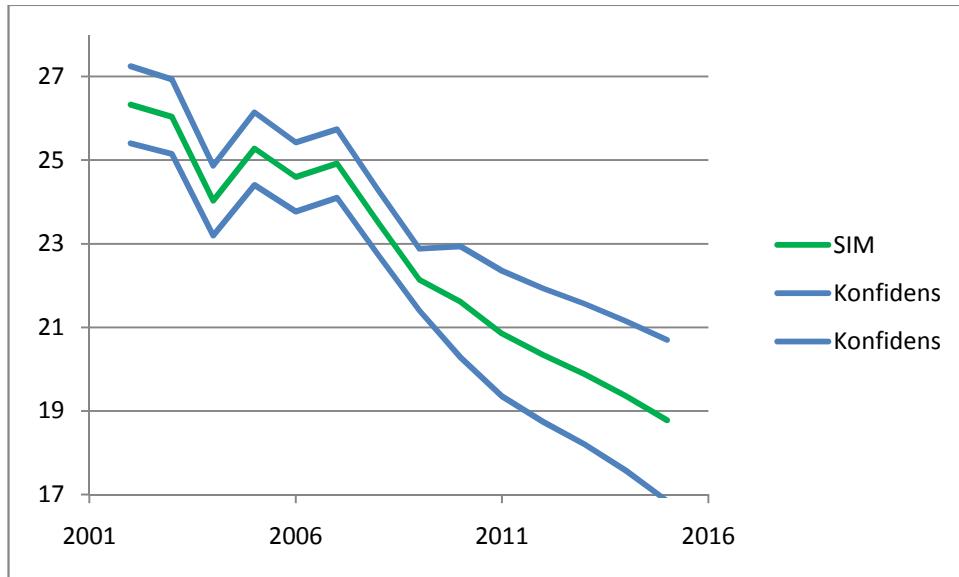
Multiple R-squared: 0.7221, Adjusted R-squared: 0.6758

F-statistic: 15.59 on 1 and 6 DF, p-value: 0.007553

5 %        95 %

(Intercept)	509.284573	1448.7415260
ÅR	-0.710116	-0.2416760





Til sidst fremskrives pct.delen af mennesker som slipper ud af fattigdom i forhold til år og procentdel af bruttonationalprodukt brugt på indkomstoverførsler. Nedenstående viser de statistisk resultater for dette.

```

Call:
lm(formula = PCT ~ ÅR + TY, data = mydata)

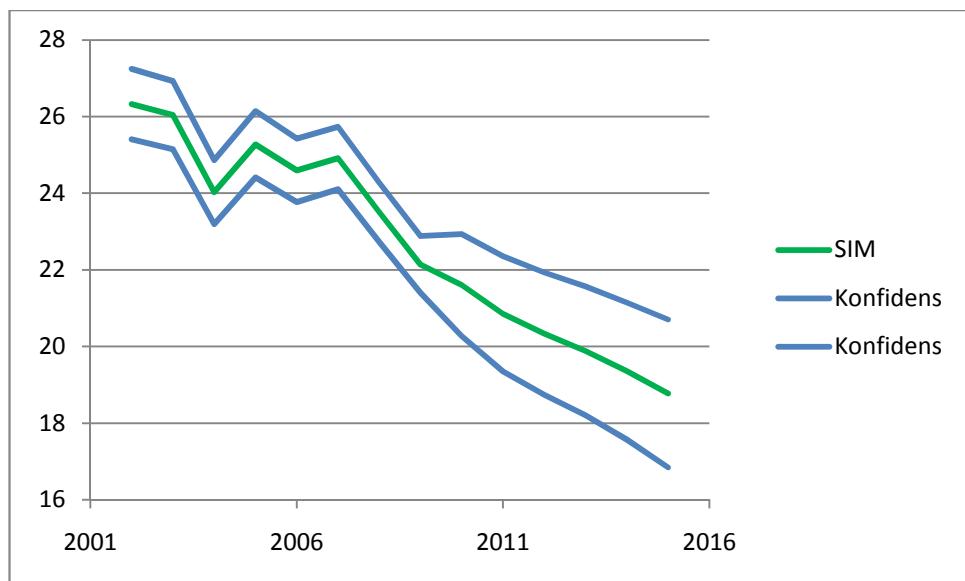
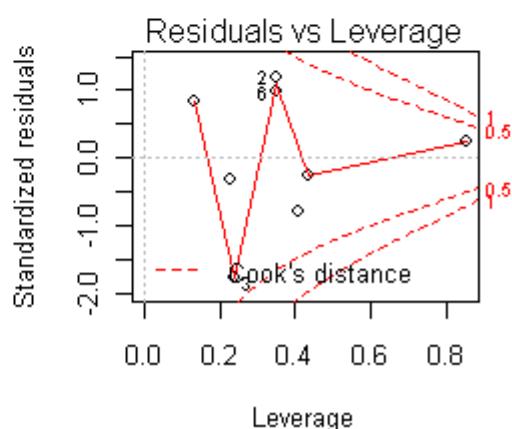
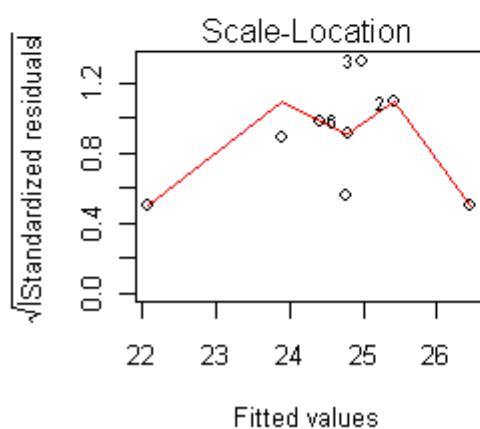
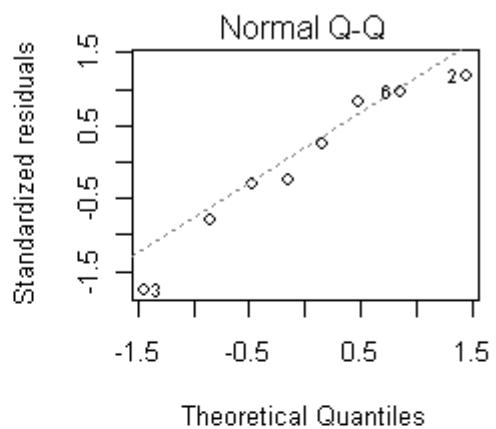
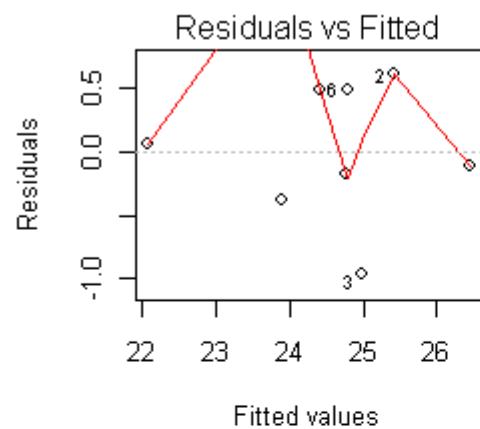
Residuals:
    1     2     3     4     5     6     7     8 
-0.11834  0.60587 -0.96757  0.48912 -0.17448  0.49353 -0.38909  0.06096 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1208.9343   224.8541   5.377  0.00300 ***
ÅR          -0.5854    0.1109  -5.279  0.00325 ***
TY          -0.6264    0.3049  -2.054  0.09515 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6302 on 5 degrees of freedom
Multiple R-squared:  0.8493,    Adjusted R-squared:  0.789 
F-statistic: 14.09 on 2 and 5 DF,  p-value: 0.008821

```

	5 %	95 %
(Intercept)	755.8423604	1662.02629034
ÅR	-0.8089096	-0.36196245
TY	-1.2408401	-0.01189792



## Deskriptiv statistik

### ESS Data

Data anvendt er fra runderne 1, 2, 3 og 4 der svarer til henholdsvis årene 2002, 2004, 2006 og 2008.

#### Behandling af data for ESS 2008:

For at vurderer hvorvidt en husstand lever under 70% medianen eller ej tages udgangspunkt i dokumentation for husstandens indkomst, der er kodet som variablen der tager nedenstående værdier:

**ESS R4 Household net income 2008**

ESS R4	Strata	incom min corrected to 2008	incom max corrected to 2008	incom min corrected to 2008 rounded	incom max corrected to 2008 rounded
J	1	-233 889 185	109 966	0	109 999
R	2	109 966	147 886	110 000	147 999
C	3	147 887	186 306	148 000	185 999
M	4	186 307	226 272	186 000	225 999
F	5	226 273	277 349	226 000	276 999
S	6	277 350	337 747	277 000	337 999
K	7	337 748	393 429	338 000	392 999
P	8	393 430	450 200	393 000	449 999
D	9	450 202	533 573	450 000	533 999
H	10	533 574	426 161 831	534 000	eller derover
refuse	77				
dont now	88				

Grænse for at blive klassificeret som værende under 70% af medianen til året 2008.

Maks. Disponibele Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	1
R	2	0
C	1	4
C	2	2
C	3	0
M	1	5
M	2	3
M	3	1
M	4	0

Grænse for at blive taget med i betragtning til den logistiske regression (mht. 70%-medianen):

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
R	1	0
C	1	1
C	2	0
M	1	4
M	2	2
M	3	0
F	1	5
F	2	3
F	3	1
F	4	0

Grænse for at blive klassificeret som værende under 60% af medianen til året 2008.

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	1
R	2	0
C	1	2
C	2	2
C	3	0
M	1	5
M	2	2
M	3	1
M	4	0

Grænse for at blive taget med i betragtning til den logistiske regression (mht. 60%-medianen):

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
R	1	0
C	1	1
C	2	0
M	1	2
M	2	2
M	3	0
F	1	5
F	2	2
F	3	1
F	4	0

Behandling af data for ESS 2006, ESS 2004 og ESS 2002:

For at vurderer hvorvidt en husstand lever under 70% medianen eller ej tages udgangspunkt i dokumentation for husstandens indkomst, der er kodet som variablen der tager nedenstående værdier:

Din husstands samlede nettoindkomst			
	Ca pr. uge	Ca. pr. måned/år	Ca. pr. år
J	Mindre end 13.300 kr.	Mindre end 13.300 kr.	Mindre end 13.300 kr.
R	Mellem 13.300 og 26.600 kr.	Mellem 13.300 og 26.600 kr.	Mellem 13.300 og 26.600 kr.
C	Mellem 26.600 og 44.400 kr.	Mellem 26.600 og 44.400 kr.	Mellem 26.600 og 44.400 kr.
M	Mellem 44.400 og 88.800 kr.	Mellem 44.400 og 88.800 kr.	Mellem 44.400 og 88.800 kr.
F	Mellem 88.800 og 133.200 kr.	Mellem 88.800 og 133.200 kr.	Mellem 88.800 og 133.200 kr.
S	Mellem 133.200 og 177.600 kr.	Mellem 133.200 og 177.600 kr.	Mellem 133.200 og 177.600 kr.
K	Mellem 177.600 og 222.000 kr.	Mellem 177.600 og 222.000 kr.	Mellem 177.600 og 222.000 kr.
P	Mellem 222.000 og 266.400 kr.	Mellem 222.000 og 266.400 kr.	Mellem 222.000 og 266.400 kr.
D	Mellem 266.400 og 444.000 kr.	Mellem 266.400 og 444.000 kr.	Mellem 266.400 og 444.000 kr.
H	Mellem 444.000 og 666.000 kr.	Mellem 444.000 og 666.000 kr.	Mellem 444.000 og 666.000 kr.
U	Mellem 666.000 og 888.000 kr.	Mellem 666.000 og 888.000 kr.	Mellem 666.000 og 888.000 kr.
N	Over 888.000 kr.	Over 888.000 kr.	Over 888.000 kr.

Grænse for at blive klassificeret som værende under 70% af medianen:

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	0
C	1	0
M	1	0
F	1	1
F	2	0
S	1	4
S	2	2
S	3	0
K	1	5
K	2	3
K	3	1
K	4	0

Grænse for at blive taget med i betragtning til den logistiske regression (mht. 70%-medianen):

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	0
C	1	0
M	1	0
F	1	0
S	1	1
S	2	0
K	1	4
K	2	2
K	3	0
P	1	5
P	2	3
P	3	1
P	4	0

Grænse for at blive klassificeret som værende under 60% af medianen:

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	0
C	1	0
M	1	0
F	1	1
F	2	0
S	1	2
S	2	1
S	3	0
K	1	4
K	2	2
K	3	1
K	4	0

Deskriptiv statistik af sammenlagte ESS Data efter behandling:

Husstandstype	Under 70%-Medianen				
	2002	2004	2006	2008	Alle år
En voksen under 65	32	28	43	31	134
En voksen over 65	36	27	27	20	110
En voksen med et eller flere børn	10	9	7	6	32
To voksne under 65	19	20	27	29	95
To voksne, hvoraf mindst den ene er over 65	47	41	30	57	175
To voksne med et barn	0	3	5	10	18
To voksne med to børn	3	7	3	6	19
To voksne med tre eller flere børn	3	8	4	8	23
Tre eller flere voksne	6	9	15	21	51
Tre eller flere voksne med børn	5	3	8	9	25
Alle husstande	161	155	169	197	682

Husstandstype	Anvendt til logistisk regression, mht. 70% medianen				
	2008	2006	2004	2002	Alle år
En voksen under 65	68	54	83	73	278
En voksen over 65	66	78	65	50	259
En voksen med et eller flere børn	17	17	13	13	60
To voksne under 65	33	30	48	48	159
To voksne, hvoraf mindst den ene er over 65	82	81	79	107	349
To voksne med et barn	3	7	6	14	30
To voksne med to børn	9	10	7	14	40
To voksne med tre eller flere børn	5	10	8	9	32
Tre eller flere voksne	13	14	28	37	92
Tre eller flere voksne med børn	11	10	20	20	61
Alle husstande	307	311	357	385	1360

	Anvendt til logistisk regression (mht. 60% medianen)				
Husstandstype	2008	2006	2004	2002	Alle år
En voksen under 65	32	54	83	73	169
En voksen over 65	36	78	65	50	179
En voksen med et eller flere børn	10	17	13	13	40
To voksne under 65	21	30	48	48	99
To voksne, hvoraf mindst den ene er over 65	41	81	79	107	201
To voksne med et barn	3	7	6	14	16
To voksne med to børn	9	10	7	14	26
To voksne med tre eller flere børn	5	10	8	9	23
Tre eller flere voksne	13	14	28	37	55
Tre eller flere voksne med børn	10	10	20	20	40
Alle husstande	180	311	357	385	848

	Under 60%-Medianen				
Husstandstype	2008	2006	2004	2002	Alle år
En voksen under 65	32	28	43	31	134
En voksen over 65	36	27	27	20	110
En voksen med et eller flere børn	14	14	9	10	47
To voksne under 65	19	20	27	29	95
To voksne, hvoraf mindst den ene er over 65	47	41	30	57	175
To voksne med et barn	0	7	6	14	27
To voksne med to børn	5	8	4	8	25
To voksne med tre eller flere børn	3	8	4	8	23
Tre eller flere voksne	6	9	15	21	51
Tre eller flere voksne med børn	5	3	8	9	25
Alle husstande	167	165	173	207	712

Husstandstype	Samples i alt				
	2002	2004	2006	2008	Alle år
En voksen under 65	160	176	190	168	694
En voksen over 65	124	121	114	102	461
En voksen med et eller flere børn	49	31	27	19	126
To voksne under 65	389	298	318	309	1314
To voksne, hvoraf mindst den ene er over 65	267	291	264	303	1125
To voksne med et barn	135	109	100	102	446
To voksne med to børn	169	155	111	90	525
To voksne med tre eller flere børn	67	56	44	35	202
Tre eller flere voksne	135	173	192	255	755
Tre eller flere voksne med børn	115	95	127	123	460
Alle husstande	1610	1505	1487	1506	6108

Grænse for at blive taget med i betragtning til den logistiske regression (mht. 60%-medianen):

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	0
C	1	0
M	1	0
F	1	0
S	1	1
S	2	0
K	1	2
K	2	1
K	3	0
P	1	4
P	2	2
P	3	1
P	4	0

**ESS Uddannelsesniveauer**

Side 52, POPULATION STATISTICS, ESS4-2008 ed. 3.0 (ESS4AppendixA1\_e03.0.pdf)

**Highest obtained education**

Approx.  
equivalence

with EDULVL Education

2	10 GRUNDSKOLE
3	20 ALMENGYMNASIAL UDDANNELSER
3	25 ERHVERVSGYMNASIAL UDDANNELSER
4	35 ERHVERVSUDDANNELSER
5	40 KORTE VIDEREGRÅENDE UDDANNELSER
6	50 MELLEMLANGE VIDEREGRÅENDE UDDANNELSER
6	60 BACHELOR
7	65 LANGE VIDEREGRÅENDE UDDANNELSER
7	70 FORSKERUDDANNELSER
	90 Don't know

## SSM - Udeladte undermodeller

Nedenstående er analyser som er udeladt fra den originale rapport, da deres viden ikke er essentielt for ovenstående metoder og konklusioner. I stedet er de medtaget som appendiks for at underbygge forståelsen af hvordan de sociale processer for dødelighed og separation/parforhold fungere for de lavt uddannede og økonomisk dårligst stillede danskere fra et statistisk synspunkt.

## Pardannelse og separation

### Motivation

Man kan spørge sig selv hvorfor pardannelse og separation er relevant for ulighed. Sociologiske teorier fortæller os at ægteskab og længere parforhold er et resultat af værdier hos de deltagende parter og deres bevidste handlingerne for at gå på kompromis og skabe en fremtid sammen. Sociologiske teorier fortæller os også ligeledes at status, indkomst, uddannelse og en lang række andre egenskaber kan være alt afgørende for denne proces. Smits et al og Hamplová bekræfter grundigt at der er stærk uddannelses homogami i Danmark (det vil sige at langt de fleste mennesker indgår i parforhold, hvor deres partner har samme uddannelsesniveau). Smits et al kommer blandt andet frem til at uddannelsesniveauet er langt mere afgørende end indkomsten for partner valg. Som konsekvens af dette vil individer med lavt uddannelsesniveau finde sammen med andre individer som også kun har et lavt uddannelsesniveau. Det forstærker naturligvis uligheden i samfundet. Forstærkelsen gælder både i forhold til antallet af børn parforholdet får, men også i forhold til indkomst og muligheden for at finde job i tilfælde af, at en part skulle blive arbejdsløs. Det er derfor afgørende at undersøge denne variabel så udstrakt vi kan.

### Faktorer og model

Baseret på ESS data fra 2002 udvikler Hamplová en solid statistisk model til at bestemme frekvenserne for parforhold der afspejler begge parters uddannelsesniveau. Vi vil benytte disse frekvenser til at udregne sandsynligheden for hvilket parforhold et individ ender i. Nedenstående viser frekvenstablerne for de Skandinaviske lande:

Table 7: Estimated cell frequencies from Model 17

Man	Woman							
	Marriage				Cohabitation			
	primary	lower secondary	upper secondary	Higher	primary	lower secondary	upper secondary	higher
Scandinavia								
Primary	46.00	31.61	16.29	11.11	24.00	24.76	22.24	14.99
Lower secondary	22.30	69.10	63.98	40.63	25.33	73.90	60.86	43.90
Upper secondary	11.75	65.43	177.24	97.58	19.72	52.74	203.76	96.78
Higher	5.95	30.87	72.50	237.68	8.95	25.60	65.13	237.32

Vi antager at de sociale processer mellem mænd og kvinder, når det kommer til pardannelse og separation, er ens for alle de skandinaviske lande, og tillader derfor os selv at bruge modellens resultater. Det svarer i grove træk til at postulere at parforholds-”mentaliteten” er ens i alle de Skandinaviske lande. Det er klart at den aldrig helt vil være ens, men overordnet mener vi at dette er en mild antagelse. Man skal ikke forveksle denne antagelse om ”ens mentalitet” med f.eks. løndiskrimination eller valg af uddannelses... Det er klart at både løndiskrimination og valg af uddannelse på baggrund af køn er forskellige i Skandinavien. Men vores antagelse holder ikke kønnene op mod hinanden, men går i stedet på det sammenspil der er mellem mand og kvinde.

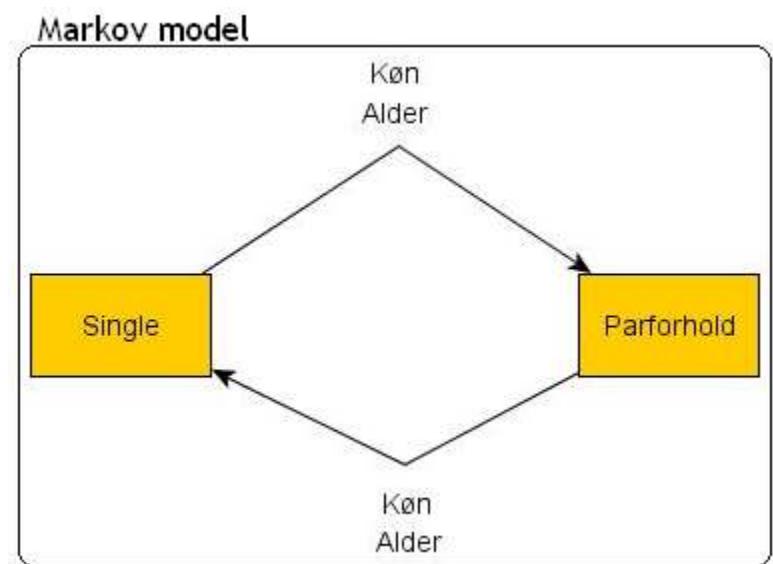
Vi antager endvidere at der ikke kan dannes parforhold mellem samme køn. Denne antagelse er meget mild, men nødvendig, da tilstrækkelig data ikke forefindes for andre parforhold.

Vi antager endvidere at uddannelses homogamien ikke ændrer sig over tid. Smits et al bekræfter at uddannelses homogamien ændrer sig relativt lidt over både 70'er og 80'er, og faktisk at den eneste ændring de kan observere er forstærkelse af den. Visse denne tendens forsætter er vores antagelse til fordel for de laveste sociale klasser, da parforhold mellem individer i lave indkomstgrupper og højere indkomstgrupper skaber mindre fattigdom.

#### Pardannelse og separation analyse

Givet vores tidligere antagelser og diskussioner kan vi udregne sandsynligheder for pardannelse og separation som en stationær kontinuert markov proces. Diagrammet til højre illustrerer modellen og de faktorer som indgår. Se appendiks for ligninger til denne model. Bemærk: da alt data anvendt har været store samples fra survey data af høj kvalitet, må vi antage at det ikke kun er repræsentativt men også er præcist for hele den danske befolkning. Derudover har vi givet vores tidligere antagelser at samtlige resultater må være fuldstændige præcise. Således kan vi ikke udregne konfidens-intervaller eller på anden måde måle usikkerheden på disse data. For reelt at vurdere usikkerheden på disse data, skal man enten ind og kigge på usikkerheden af vores eller kigge på indsamlingen af data.

De statistiske resultater er givet i nedenstående tabeller:



Mænd	Er i ægteskab efter 5-årig periode, givet at vedkommende ikke var i ægteskab før.	Er ikke i ægteskab efter 5-årig periode, givet at vedkommende var det før.
20-24	0,023888993	0,045325814
25-29	0,105012409	0,043392166
30-34	0,134475389	0,042675214
35-39	0,072822155	0,045051213
40-44	0,03423445	0,031792718
45-49	0,019405333	0,023882661
50-54	0,013081387	0,014228183
55-59	0,006473214	0,01427589
60+	0,001549702	0,014311365

Kvinder	Er i ægteskab efter 5-årig periode, givet at vedkommende ikke var i ægteskab før.	Er ikke i ægteskab efter 5-årig periode, givet at vekommende var det før.
20-24	0,053113288	0,044635757
25-29	0,143537734	0,042453009
30-34	0,118050996	0,043075901
35-39	0,052734292	0,04553941
40-44	0,024763919	0,031949178
45-49	0,01466451	0,023940632
50-54	0,009097246	0,014256959
55-59	0,003593564	0,014296646
60+	0,000601679	0,01431819

Kilder:

More or Less Educational Homogamy? A Test of Different Versions of Modernization Theory Using Cross-Temporal Evidence for 60 Countries  
 Author(s): Jeroen Smits, Wout Ultee, Jan Lammers  
 Source:

Educational Homogamy in 65 Countries: An Explanation of Differences in Openness Using Country-Level Explanatory Variables  
 Author(s): Jeroen Smits, Wout Ultee, Jan Lammers  
 Source:

Educational Homogamy in Marriage and Cohabitation

in Selected European Countries **Dana Hamplová**

## Dødsfald

### Motivation

Man kan spørge sig selv hvorfor vi overhovedet gider bøvle med dødsfald. Svaret er, at dødeligheden er væsentlig højere for lave indkomstgrupper. I diskussion om social mobilitet er det derfor vigtigt at medtage dødelighed. Derudover arbejder vi med husstande i vores ESS data, hvorfaf flere medlemmer som ikke er i arbejde sagtens være i en ældre aldersgruppe og således have stor dødelighed. Samtidig måler vi også fattigdommen i forhold til hvor mange individer der er i denne husstand, det er derfor vigtigt at sådanne dødsfald i husstande bliver taget i betragtning.

### Faktorer og model

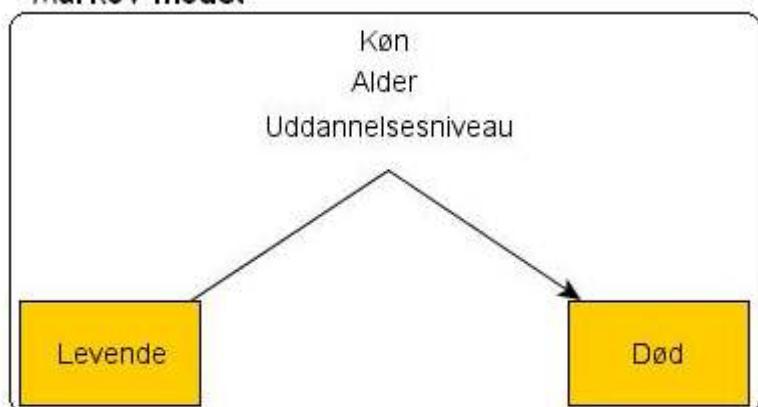
Vi har dog kun begrænset data til rådighed vedrørende dødsfald som omfatter hele den danske befolkning. Derfor kan vi ikke kontrollere sandsynligheden for dødsfald i forhold til indkomsten. I stedet kontrollere vi sandsynlighederne i forhold til uddannelsesniveau. Derudover kontrollerer vi også i forhold til alder og køn, som det er almen anerkendt spiller en stor rolle for dødeligheden.

Vi laver den antagelse at dødsfald er uafhængig af andre faktorer, samt at vi kan modellere den nøjagtigt med en stationær kontinuert markov proces. Det er klart at menneskers biologi er arbejder i kontinuert tid, og derfor vil kontinuitet være den bedste approksimation. Derudover har vi også splittet sandsynlighederne ind i aldersgrupper, hvorfor kravet om stationæritet (time-homogeneus) kun er afgrænset til disse aldersgrupper, hvilket er en meget mild antagelse. Derudover har vi igen forsimplet situationen og skabt en sort-hvid verden: enten lever individet eller også er individet dødt. Der findes ingen mellem stadier, som f.eks. permanent hurtigt-dødelig sygdom. Samtlige antagelser er milde og skaber en langt mere præcis model end vi reelt har behov for. Hvis der skulle være en bias ville den påvirke den sociale mobilitet positivt for de laveste indkomstgrupper, således at vi har overvurderet den sociale mobilitet. Dette bekræftes også af en række forskere som finder at selvmords risikoen er op til 5 gange højere for lave indkomstgrupper (Qin et al, Ferda et al).

### Dødsfald analyse

Givet vores tidlige antagelser og diskussioner kan vi udregne sandsynligheder for dødsfald som en stationær kontinuert markov proces. Diagrammet til venstre illustrerer modellen og de faktorer som indgår. Se appendiks for ligninger til denne model. Bemærk: da alt data anvendt har været register data af høj kvalitet, må vi antage at det ikke kun er repræsentativt men også er præcist for hele den danske befolkning. Derudover har vi givet vores tidlige antagelser at samtlige resultater må være

### Markov model



fuldstændige præcise. Således kan vi ikke udregne konfidens-intervaller eller på anden måde måle usikkerheden på disse data. For reelt at vurdere usikkerheden på disse data, skal man enten ind og kigge på usikkerheden af vores antagelser (som er meget milde) eller kigge på indsamlingen af data.

De statistiske resultater er givet i nedenstående tabeller:

ISCED97	Alder	Mænd	Kvinder
Pre-primary, primary and lower secondary education - levels 0-2 (ISCED 1997)	0-4	0	0
	5-9	0	0
	10-14	0,013414634	0,003996803
	15-19	0,000476716	0,000193907
	20-24	0,001192418	0,000329985
	25-29	0,001633082	0,000603415
	30-34	0,002147658	0,001165583
	35-39	0,002588935	0,001609036
	40-44	0,003540153	0,002339122
	45-49	0,004951225	0,003238638
	50-54	0,008172071	0,005352824
	55-59	0,011345102	0,00743473
	60-64	0,016582317	0,010264722
	65-69	0,024899296	0,015878353
	70-74	0,037680927	0,024575718
	75-79	0,056936648	0,040292579
	80-84	0,09626255	0,06482581
	85-89	0,171856465	0,119511636
	90-94	0,170731707	0,166153846
	95-99	0,2	0,393939394
	100-104	1	0
	105-109	0	0
	110+	0	0

ISCED97	Alder	Mænd	Kvinder
Upper secondary and post-secondary non-tertiary education - levels 3-4 (ISCED 1997)	0-4	0	0
	5-9	0	0
	10-14	0	0
	15-19	0	0
	20-24	0,000454534	6,83979E-05
	25-29	0,000327613	0,000186876
	30-34	0,000622497	0,000305802

35-39	0,000752598	0,000628439
40-44	0,001611517	0,001074013
45-49	0,003433466	0,00189122
50-54	0,005034235	0,003045167
55-59	0,007998253	0,004773237
60-64	0,013517382	0,008006522
65-69	0,022312088	0,012677947
70-74	0,035176166	0,021320446
75-79	0,064624006	0,039923476
80-84	0,107641939	0,062517986
85-89	0,307998013	0,208518519
90-94	0,142857143	0,342105263
95-99	1	0,2
100-104	0	1
105-109	0	0
110+	0	0

De sidste 6 rækker er markeret med rød da data er fejlfyldt her.  
Endvidere over 85 år antages at have samme dødelighed som individer mellem 80 og 84 år.

ISCED97	Alder	Mænd	Kvinder
Tertiary education - levels 5-6 (ISCED 1997)	0-4	0	0
	5-9	0	0
	10-14	0	0
	15-19	0	0
	20-24	0	7,5375E-05
	25-29	0,000183722	9,75499E-05
	30-34	0,00034493	0,000271904
	35-39	0,000587643	0,000519136
	40-44	0,000799753	0,000648138
	45-49	0,001642562	0,001201923
	50-54	0,003382286	0,002098636
	55-59	0,004566114	0,004140916
	60-64	0,011087533	0,007045107
	65-69	0,016849947	0,011020168
	70-74	0,028263607	0,017662469
	75-79	0,049421238	0,03305169
	80-84	0,098411602	0,063300402
	85-89	1,058394161	1,38
	90-94	0,645833333	3,25
	95-99	1,5	0

100-104	2	0
105-109	0	0
110+	0	0

Kilder:

Suicide Risk in Relation to Socioeconomic, Demographic,  
Psychiatric, and Familial Factors: A National Register–  
Based Study of All Suicides in Denmark, 1981–1997

Af Ping Qin, M.D., Ph.D.

Esben Agerbo, M.Sc.

Preben Bo Mortensen,

Dr.Med.Sc.

Determinants of Suicides in Denmark:

Evidence from Time Series Data

Af HALICIÖGLU, Ferda and Andres, Antonio

Yeditepe University

20