

Sammenligning av dødelighetsrater for
mennesker født før, under og etter den
2. verdenskrig i de nordiske land

av

Vanja Kristinedatter Steinholt

MASTEROPPGAVE

for graden

Master i Modellering og dataanalyse

med studieretning

Finans, forsikring og risiko



Mai 2009

*Det matematisk- naturvitenskapelige fakultet
Universitetet i Oslo*

Forord

Med denne masteroppgaven avslutter jeg mitt femårige masterstudium i modellering og dataanalyse ved Universitetet i Oslo. Oppgaven har blitt skrevet i tidsrommet januar 2009 til mai 2009. Det har vært spennende, utfordrende og lærerikt å jobbe med denne oppgaven, og det har fått meg til virkelig å innse at å velge en studieretning innen matematikk og statistikk var det riktige.

Takk til min veileder Sven Ove Samuelsen som alltid satt av tid til innspill og råd. Hans engasjement innen temaet var smittende og engasjerende, og jeg tror det har vært med på å gjøre dette til en bedre oppgave.

I tillegg vil jeg takke alle medstudenter for hyggelig selskap og hjelp i løpet av studieårene. En spesiell takk til Qingsheng Dong for hennes hjelp både faglig og i forbindelse med tips, triks og råd innen R og Latex.

En takk også til mine foreldre Kristine og Jan for korrekturlesing.

Sist men ikke minst vil jeg takke min kjære samboer Erik for oppmuntring og støtte, ikke bare under arbeidet med masteroppgaven, men gjennom hele studiet.

Oslo, Mai 2009,
Vanja Steinholt

Innhold

1	Introduksjon	7
2	Dødelighet og overlevelse	8
2.1	Dødelighetsrater	8
2.1.1	Hvordan beregne dødelighetsrater	8
2.1.2	Dødelighetsrater i Norge, Danmark, Sverige og Finland	12
2.2	Overlevelsesfunksjon	22
2.2.1	Beregning av overlevelsesfunksjoner	22
2.2.2	Overlevelsessannsynlighet i de nordiske landene	22
2.3	Forventet levealder	25
2.3.1	Forventet gjenstående levetid for en x-åring	25
2.3.2	Forventet gjenstående levetid basert på dataene	26
3	Poisson regresjon	31
3.1	Teori	31
3.2	Poisson regresjon benyttet på datasettet	34
4	Generaliserte additive modeller	45
4.1	Hva er en generalisert additiv modell?	45
4.2	Generaliserte additive modeller benyttet på datasettet	46
5	Konklusjoner	51
6	Vedlegg	54
6.1	Resultater	54
6.1.1	Figurer	54
6.1.2	Tabeller	61
6.2	R-kode	65
6.2.1	Beregning og analysing av dødelighetsrater	65
6.2.2	Beregning av overlevelsessannsynligheter	68
6.2.3	Beregning av forventet gjenstående levetid	68
6.2.4	Modellering av Poisson regresjon og relative risikoer	69
6.2.5	Devians	70
6.2.6	Generaliserte additive modeller	71

1 Introduksjon

Denne oppgaven undersøker hypotesen om at dødelighet i voksen alder avhenger av ernæringsforholdene individene levde under i barnealder. Dersom ernæringsforholdene var svært dårlige ventes høy dødelighet, men det har også vært spekulert i om moderat vanskelige forhold faktisk kan føre til en reduksjon i dødeligheten. Dette har blant annet blitt indikert gjennom dyrestudier[10].

Som utgangspunkt for hypotesen skal dødeligheten i de nordiske land undersøkes, da mer spesifikt dødeligheten i Norge, Sverige, Danmark og Finland, i årene før, under og etter 2. verdenskrig. Under krigen var forholdene antageligvis ganske forskjellige i de fire landene. Sverige som var et ikke-okkupert land, og Danmark med sitt omfattende landbruk, hadde nok en vesentlig bedre ernærings situasjon enn Norge. I tillegg skal forholdene i Finland ses nærmere på, et land hvor det var krigstilstand under det meste av krigen, og som dermed kan ha hatt ytterligere dårlige ernæringsforhold. For å sammenligne dødeligheten i landene og for å beskrive hvordan ernæringsforholdene i barnealder påvirker dødeligheten i voksen alder, er dataene inndelt i fødselskohorter for hvert land, og deretter sammenlignes disse. Fødselskohortene er delt inn i individer som er født i årene 1930-1934, 1935-1939, 1940-1945 og 1946-1950.

Dataene som er brukt er dødelighetstall og befolkningstall etter alder, fødselsår og kjønn som er hentet inn fra Human Mortality Database for de ulike landene i perioden 1930-2006. Som følge av at dataene for alle landene kunne innhentes fra samme sted, var de også registrert på tilsvarende måte, noe som var positivt for sammenligningen landene i mellom. Ettersom utgangspunktet er individenes fødselsår, innebærer dette at siste mulige observerte dødsfall er 76 år, for personer født i 1930. Det vil kun ses på dødeligheten for individer født i perioden 1930-1950, inndelt i de fire ovennevnte kohortene for sammenligning. I tillegg til sammenligning mellom kohortene vil også landene sammenlignes.

Metodene som er brukt er beregning og sammenligning av dødelighetsrater, overlevelsesfunksjoner og forventet levetid i de ulike landene. Til disse sammenligningene er det også benyttet Poisson regresjonsteknikker, og generaliserte additive modeller (GAM) er også tatt i bruk. Innenfor Poisson regresjon har det også vært svært aktuelt å se på relative risikoer. Nærmere bestemt omhandler kapittel 2 ulike innfallsvinkler for analysering av dødelighetsrater. Dette innebærer å studere dødelighetsratene ved hjelp av tabeller og figurer, å se på overlevelses sannsynligheter samt beregning av forventede gjenstående levetider. For de forventede levealderene er maksalder satt til 60 år på grunn av at dataene kun er fullstendige fram til 60 år dersom man skal se på alle kohortene. Videre omhandler kapittel 3 Poisson regresjon der det spesielt ses på relative risikoer, både med

og uten interaksjonsledd. De relative risikoene er oppgitt i tabeller i kapitlet. Kapittel 4 tar for seg generaliserte additive modeller (GAM), der det blir sett på trendutviklingen i dødelighet som følge av alder og fødselsår. Sluttkapitlet omtaler resultatene som er funnet og setter dem i sammenheng med hypotesen om at moderat vanskelig ernæringsforhold i barnealder kan føre til en reduksjon i dødeligheten i voksen alder.

Opgaven er lagt opp slik at hver hoveddel, slik som analysering av dødelighetsrater, Poisson regresjon og bruk av generaliserte additive modeller er egne kapitler. Hvert kapittel er deretter inndelt i seksjoner. Gjennomgående består hvert kapittel først av en seksjon med teori om emnet for kapitlet, og deretter kommer en seksjon der teorien er benyttet på datasettet og hva slags resultater det ga. De viktigste og mest interessante og relevante figurene og tabellene er inkludert i selve teksten. Andre supplerende plott og tabeller finnes i vedlegg, men det viktigste fra disse blir også nevnt og diskutert i hovedteksten. Rett etter vedlegget bakerst i teksten finnes også en oversikt med sidehenvisninger til de ulike figurene og tabellene.

2 Dødelighet og overlevelse

Dødelighetsratene beregnes ut fra dødelighets- og befolkningsdataene for hvert land og i hver kohort. Disse brukes deretter for å finne overlevelsesfunksjoner, og videre forventet gjenstående levetid. Etersom ikke alle individene som ble født etter år 1930 har dødd ut enda, vil dette være ufullstendige data. Dataene for alle kohortene i alle landene er komplette opp til alder 60 år.

2.1 Dødelighetsrater

2.1.1 Hvordan beregne dødelighetsrater

Dødelighetsrater kan fortolkes som hazarder, det vil si sannsynligheten for å dø i et lite intervall om t (dividert med Δ) gitt at individet var i live ved tidspunkt t . Δ er her da lengden, eller tiden, fra et tidspunkt vi ser på til det neste. I denne oppgaven vil dette da typisk være ett år. Hazarden $h(t)$ kan dermed skrives som

$$h(t) \approx \frac{1}{\Delta} P(t \leq T < t + \Delta | T \geq t)$$

der T er levetiden til individet. Det er også grunn til å anta at hazardene er stykkevis konstante, det vil si at $h(t)$ er konstant over hvert subintervall.

Man kan videre definere dødelighetsratene ved tidspunkt t_j som ρ_j , antall døde betegnes d_j , antall sensurerte m_j og befolkningsantall r_{j-1} ved tid t_{j-1} . Det vil da si at $r_{j-1} - d_j - m_j$ er antall overlevende som ikke ble sensurert. λ_j er raten for de sensurerte. Samtlige av variablene er betinget på at individet er i live ved starten av intervallet. De følgende resultatene er hentet fra Cox & Oakes[2]. For hvert intervall $[t_{j-1}, t_j)$ må man se på bidraget til den simultane likelihooden $L(\rho, \lambda)$ fra

- antall individer $r_{j-1} - d_j - m_j$ som overlever og ikke blir sensurert
- antall individer d_j som dør
- antall individer m_j som blir sensurert

som har betingede sannsynligheter

- $\exp[-b(\rho + \lambda)]$
- $\int_0^b \rho e^{-\rho v} e^{-\lambda v} dv = \frac{\rho}{\rho + \lambda} \{1 - \exp[-b(\rho + \lambda)]\}$
- $\int_0^b \lambda e^{-\lambda v} e^{-\rho v} dv = \frac{\lambda}{\rho + \lambda} \{1 - \exp[-b(\rho + \lambda)]\}$

der b er lengden av intervallet.

Bidraget til log-likelihooden for intervallet $[t_{j-1}, t_j)$ blir

$$l_j(\rho_j, \lambda_j) = -(r-d-m)b(\rho+\lambda) + d \log\left(\frac{\rho}{\rho+\lambda}\right) + m \log\left(\frac{\lambda}{\rho+\lambda}\right) + (d+m) \log\{1 - \exp[-b(\rho + \lambda)]\}$$

der subskriptene er utelatt på høyre side av likningen for enkelhets skyld. Vi skal nå se på to ulike måter å finne et estimat for dødelighetsratene $\hat{\rho}_j$. Den ene metoden går ut på at vi har dataene for hvert individ, og den andre bygger på at vi har grupperte data(som her). La oss først se på metoden for å finne $\hat{\rho}_j$ for grupperte data.

For å finne maksimum likelihood estimatene $\hat{\rho}_j$ og $\hat{\lambda}_j$ må man løse

$$\frac{\delta l}{\delta \rho} = -(r - d - m)b + \frac{d}{\rho} - \frac{d + m}{\rho + \lambda} + \frac{(d + m)b \exp[-b(\rho + \lambda)]}{1 - \exp[-b(\rho + \lambda)]} = 0$$

$$\frac{\delta l}{\delta \lambda} = -(r - d - m)b + \frac{m}{\lambda} - \frac{d + m}{\rho + \lambda} + \frac{(d + m)b \exp[-b(\rho + \lambda)]}{1 - \exp[-b(\rho + \lambda)]} = 0$$

Subtrahering av de to likningene gir

$$\begin{aligned} & -(r - d - m)b + \frac{d}{\rho} - \frac{d + m}{\rho + \lambda} + \frac{(d + m)b \exp[-b(\rho + \lambda)]}{1 - \exp[-b(\rho + \lambda)]} \\ & + (r - d - m)b - \frac{m}{\lambda} + \frac{d + m}{\rho + \lambda} - \frac{(d + m)b \exp[-b(\rho + \lambda)]}{1 - \exp[-b(\rho + \lambda)]} \\ & = \frac{d}{\rho} - \frac{m}{\lambda} = 0 \\ & \Rightarrow \frac{d}{\rho} = \frac{m}{\lambda} \\ & = \frac{d + m}{\rho + \lambda} \end{aligned}$$

Substituering gir deretter estimatorene

$$\begin{aligned} \hat{\rho} &= -\frac{d}{b(d + m)} \log\left(\frac{r - d - m}{r}\right) \\ \hat{\lambda} &= -\frac{m}{b(d + m)} \log\left(\frac{r - d - m}{r}\right) \end{aligned}$$

Ved hjelp av rekkeutvikling får man til slutt estimatoren

$$\hat{\rho}_j = \frac{d_j}{b_j[r_{j-1} - \frac{1}{2}(d_j + m_j)]} \quad (2.1)$$

som brukes for å estimere hazardraten $\hat{\rho}_j$ i intervallet $[t_{j-1}, t_j)$.

Dersom man har individuelle data er maksimum likelihood estimatoren for ρ_j

$$\hat{\rho}_j = \frac{d_j}{u_j} \quad (2.2)$$

der u_j er total observeringstid. Ved å sammenligne likning (2.1) og (2.2) viser det seg at estimatoren i (2.1) kan ses på som ekvivalent med estimatoren i (2.2) dersom man kan anta at dødsfall og sensureringer forekommer uniformt gjennom intervallet, og dermed at nevneren i (2.1) tilsvarer den totale observasjonstiden $u_j[2]$.

Det kan også være av interesse å se på variansen til de estimerte dødelighetsratene. Dersom man ser bort fra sensurering kan $\hat{\rho}_j$ tilnærmes ved

$$\begin{aligned}\hat{\rho}_j &\approx \frac{d_j}{b_j r_{j-1}} \\ &= \frac{1}{b_j} \hat{p}_j\end{aligned}$$

der \hat{p}_j er sannsynligheten for å dø, $\hat{p}_j = \frac{d_j}{r_{j-1}}$. Hvis man nå videre tenker seg at

$$d_j | r_{j-1} \sim \text{bin}(r_{j-1}, p_j)$$

så er

$$\begin{aligned}\text{Var}(d_j | r_{j-1}) &= r_{j-1} \cdot p_j \cdot (1 - p_j) \\ &\approx r_{j-1} \cdot p_j\end{aligned}$$

når p_j er liten. Dette resulterer i at variansen til $\hat{\rho}_j$ kan estimeres som

$$\begin{aligned}\hat{\text{Var}}(\hat{\rho}_j) &\approx \frac{r_{j-1} \hat{p}_j}{u_j^2} \\ &= \frac{r_{j-1} \frac{d_j}{r_{j-1}}}{u_j^2} \\ &= \frac{d_j}{u_j^2}\end{aligned}$$

Ettersom antall individer som dør som regel er svært få i forhold til den totale befolkningen i samme alder, vil denne høyst sannsynlig være liten. Dette vil vi senere også se at den faktisk er.

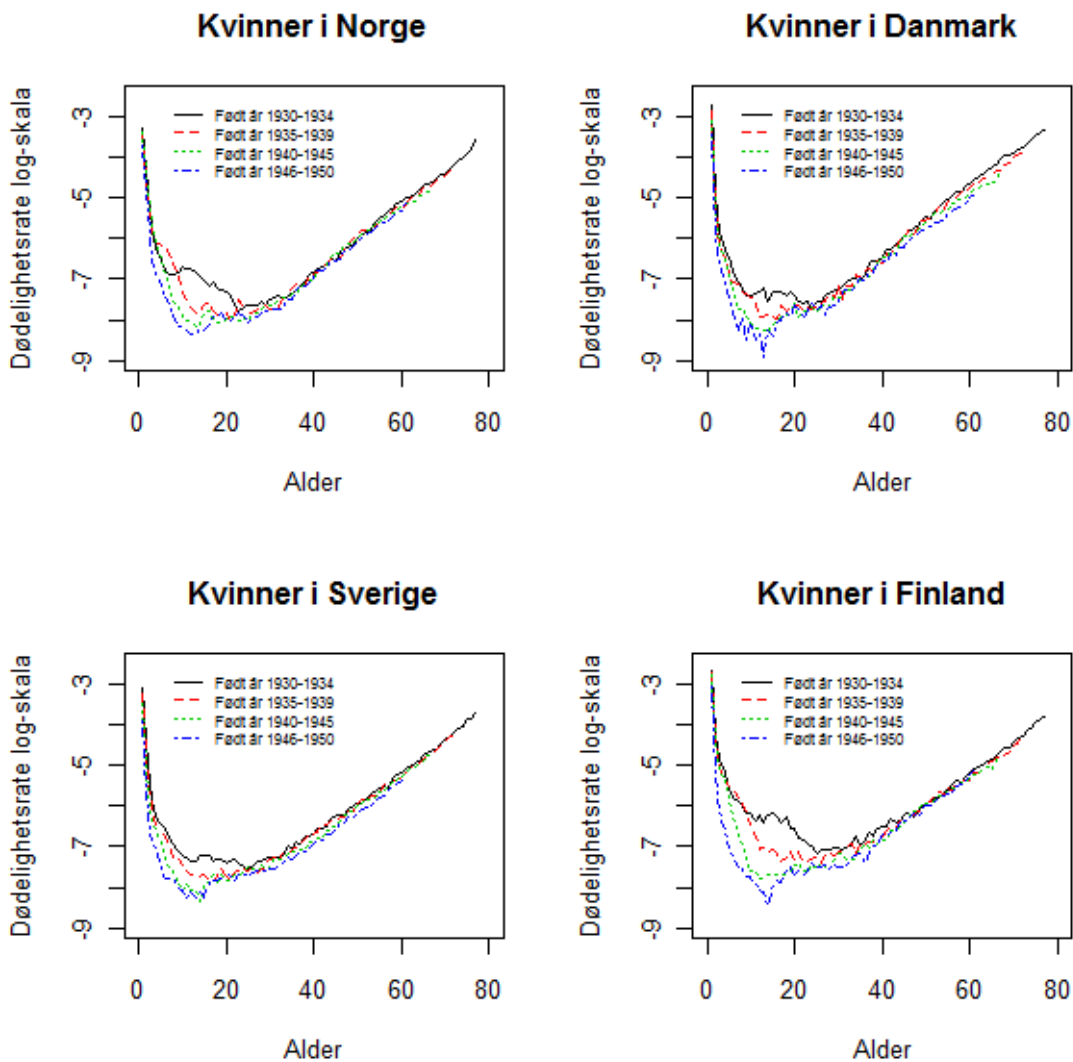
2.1.2 Dødelighetsrater i Norge, Danmark, Sverige og Finland

La oss nå se på dødelighetsratene for de ulike alderene, kohortene og landene. Det er enklere å se forskjellene i dødelighetsratene dersom de er plottet på log-skala, og dette er dermed gjort i figurene 2.1, 2.2, 2.3 og 2.4 for både kvinner og menn. Plottene 2.1 og 2.2 viser utviklingen av dødelighetsratene i hvert enkelt land gjennom de ulike tidsperiodene. Disse figurene er laget for å se om det er den samme utviklingen i de fire ulike landene, eller om det er noen store forskjeller. Plottene 2.3 og 2.4 viser dette derimot på en annen måte, der individene i hvert land er plottet mot hverandre innenfor hver kohort. I disse figurene av dødelighetsratene hadde det også vært av interesse å ha med konfidensintervaller, men på grunn av at det er såpass mange grafer som inngår i hvert plott, ville det nærmest ha blitt umulig å lese av konfidensintervallene. Ikke minst ville det blitt svært uoversiktlig. Dette er årsaken til at konfidensintervallene er utelatt her, men i kapittel 3 som omhandler Poisson regresjon og relative risikoer ses det også på konfidensintervaller. Dermed inngår det på et vis der i stedet.

I vedlegget finnes tabeller for dødelighetsratene. Tabellene 6.1 og 6.2 viser dødelighetsratene for hvert tiår i alderen 0-70 år for kvinner og menn i Norge, Danmark, Sverige og Finland. For de to seneste kohortene har vi ikke fulle data fram til 70 år, og dermed går tabellene 6.3 og 6.4 kun opp til 60 år.

Figur 2.1 og 2.2 viser tydelig at dødeligheten generelt sett har blitt redusert for hver kohort i samtlige av landene, både for kvinner og menn. Altså har det vært en positiv utvikling i dødeligheten. Figur 2.1 viser at det er finske kvinner som har hatt størst endring i dødeligheten i alderen 5-45 år. Det samme er tilfelle for finske menn, som vist i figur 2.2 men da over en kortere periode, fram til omtrent 30-årsalderen. Av kvinnene er danske kvinner den gruppen som har hatt størst endring i dødeligheten i voksen alder. De resterende landene ser ut til ikke å ha hatt noen store endringer i dødeligheten i godt voksen alder, men litt har det naturlig nok vært. Blant disse ser det ut til at svenske kvinner har hatt den største reduksjonen.

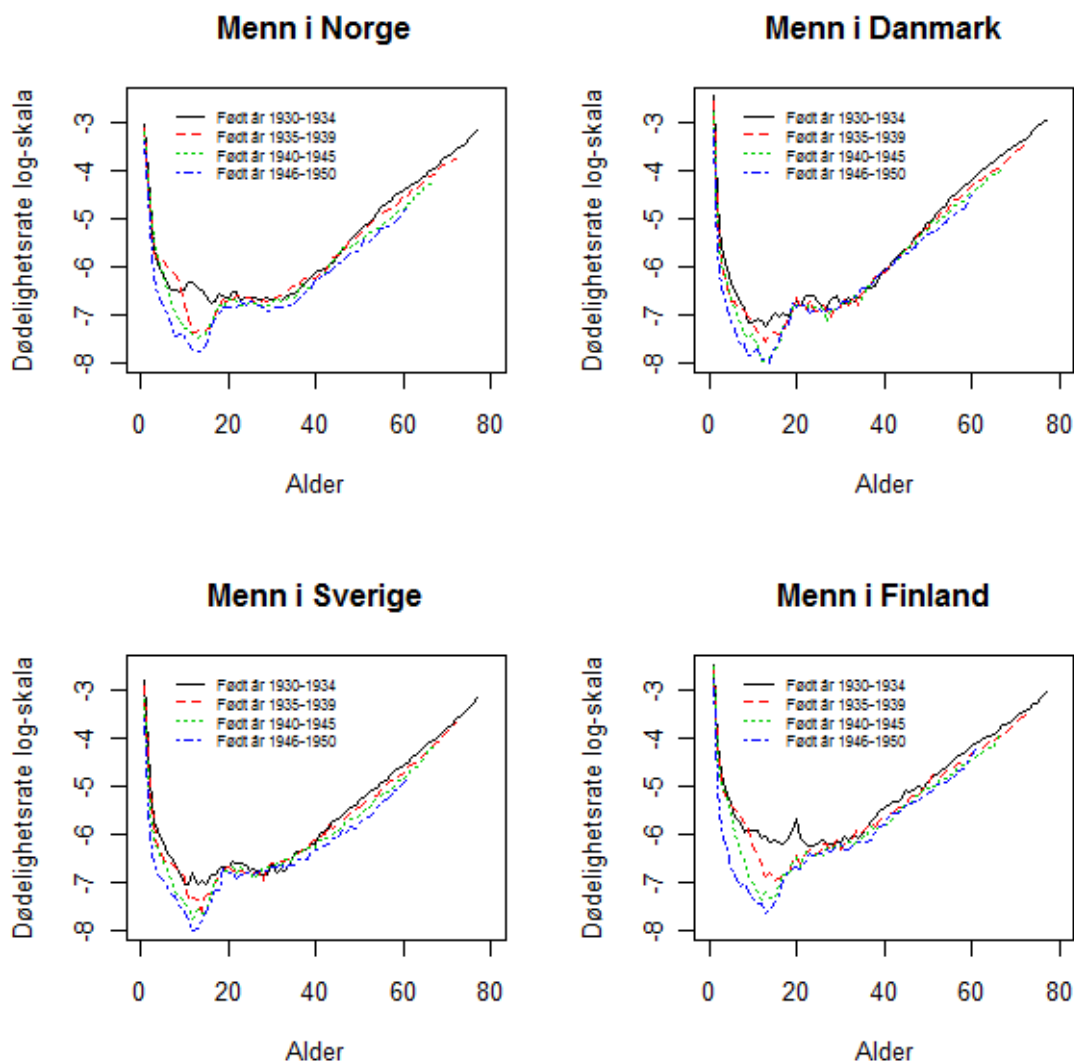
For menn er det derimot annerledes; der har dødeligheten i høy alder sunket mer betydelig for hver kohort i alle fire landene. Dette kan skyldes revolusjonen i utviklingen av medisiner i forbindelse med hjerte- og karsykdommer som kom på 1970-tallet. Det er i størst grad menn som blir utsatt for disse sykdomstypene, og det er kjent at medisinene som nå finnes mot dette har resultert i en kraftig reduksjon av dødsfall blant menn som følge av hjerte- og karsykdommer. For individene som det ses på her, vil de som er født i 1930 da være 40 år i 1970, og de andre være enda yngre. Dette vil si at de har vært i den aktuelle aldersgruppen som er utsatt for hjerte- og karsykdommer (oftest menn over 20 år, men særlig



Figur 2.1: Dødelighetsrater per land for kvinner

godt voksne menn) etter at medisinene mot dette kom, slik at dette absolutt kan ha ført til en reduksjon i dødeligheten blant disse. For både menn og kvinner ser det ut til at utviklingen er relativt lik i alle landene, og ut i fra figurene virker det som om det har vært en nedgang i dødeligheten i alle alderene også.

Spesielt bør den høye spedbarnsdødeligheten nevnes. Alle landene har høy barne-dødelighet, men spesielt stor er den i Danmark og Finland. Dette er tilfelle for alle kohortene og gjelder både kvinner og menn. Norge er landet med lavest spedbarnsdødelighet i perioden 1930-1939, men blir forbigått av Sverige i resten



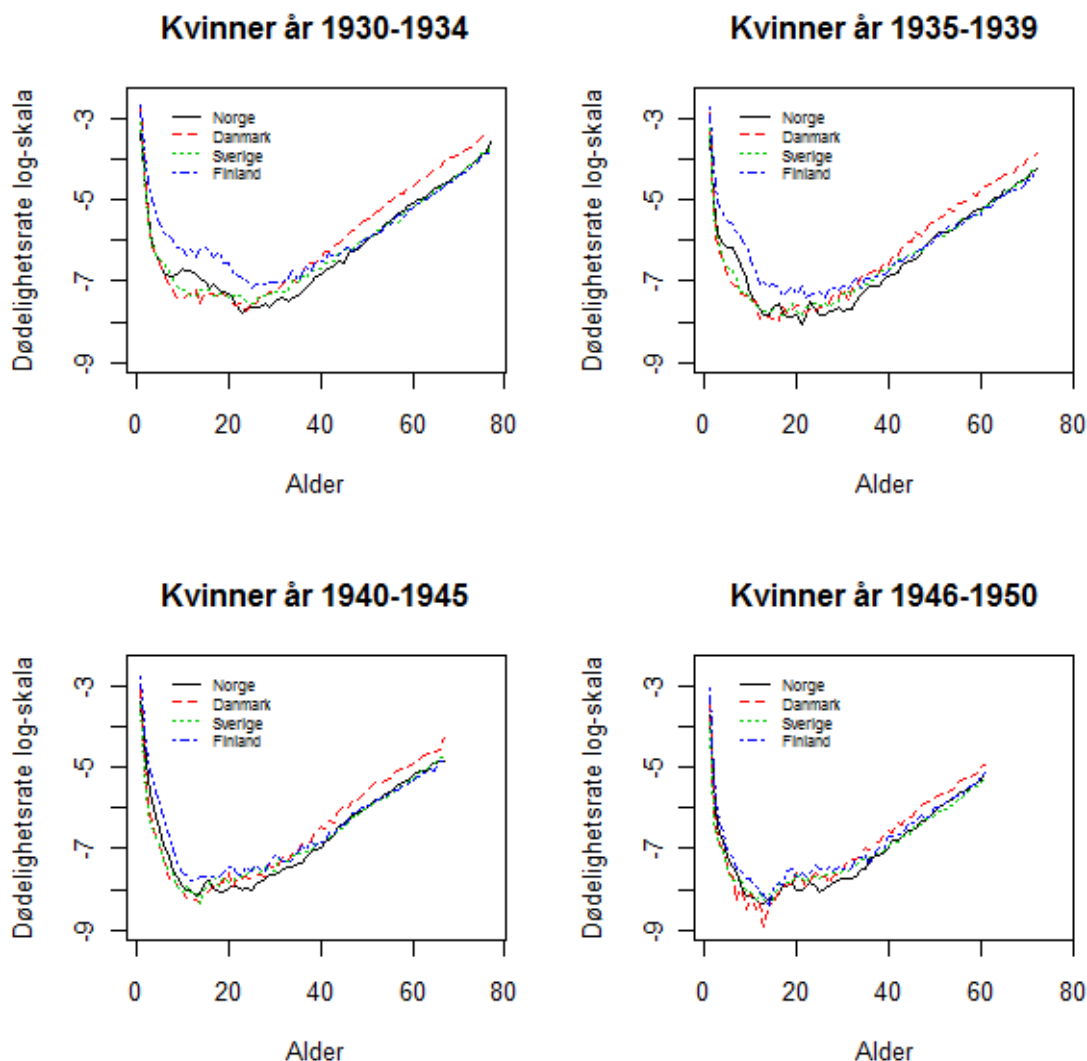
Figur 2.2: Dødelighetsrater per land for menn

av perioden. De konkrete tallene er oppgitt i tabellene 6.1, 6.2, 6.3 og 6.4, som er å finne i vedlegget. For jenter i årene 1930-1934 kommer Norge best ut med en dødelighet på 3.66%, mens danske og svenske jenter har en spedbarnsdødelighet på henholdsvis 6.62% og 6.9%. For guttene i samme periode ligger danskene høyest med en spedbarnsdødelighet på 8.76% og deretter kommer finnene med en dødelighet på 8.39%. Norske nyfødte gutter har en dødelighet på 4.69% i denne perioden, og svenskene ligger også forholdsvis høyt med 6.03%. For individer født i løpet av de neste fem årene reduseres spedbarnsdødeligheten litt i alle landene.

Det kan være interessant å se på spedbarnsdødeligheten i krigsårene 1940-1945. For norske jenter og gutter har dødeligheten ikke gått spesielt mye ned, og er henholdsvis 0.0334 og 0.0409. For svenske jentebarn har dødeligheten blitt redusert til 0.0304, og blant guttene til 0.04. Finland har fortsatt den høyeste dødeligheten med 6.37% hos jentene og 7.89% hos guttene. Danskene ligger også fortsatt høyt med en spedbarnsdødelighet på 0.0457 blant jenter og 0.0609 hos gutter. Disse tallene viser at Sverige, som ikke hadde krigstilstander, kom best ut i disse årene. Dette virker fornuftig ettersom svenskene trolig hadde de beste oppvekstvilkårene.

Etter krigen reduseres spedbarnsdødeligheten mer betydelig i alle land. Norske og svenske jentebarns dødelighet har blitt redusert til å befinne seg litt over to prosent i årene 1946-1950. Norske gutter ligger på 3.31%, mens svenske gutter har en dødelighet på 2.68%. For danske jenter er dødeligheten kommet ned mot 3%, og guttene nærmer seg 4%. Finnene har fortsatt høyest dødelighet blant spedbarn, med 4.78% for jenter og 6.03% for gutter. Totalt sett har det vært størst nedgang i spedbarnsdødeligheten hos danskene fra perioden 1930-1934 til 1946-1950, deretter følger svenskene og så finnene. Norge er det landet som har hatt minst endring, men det var også det landet som startet ut med den laveste dødeligheten.

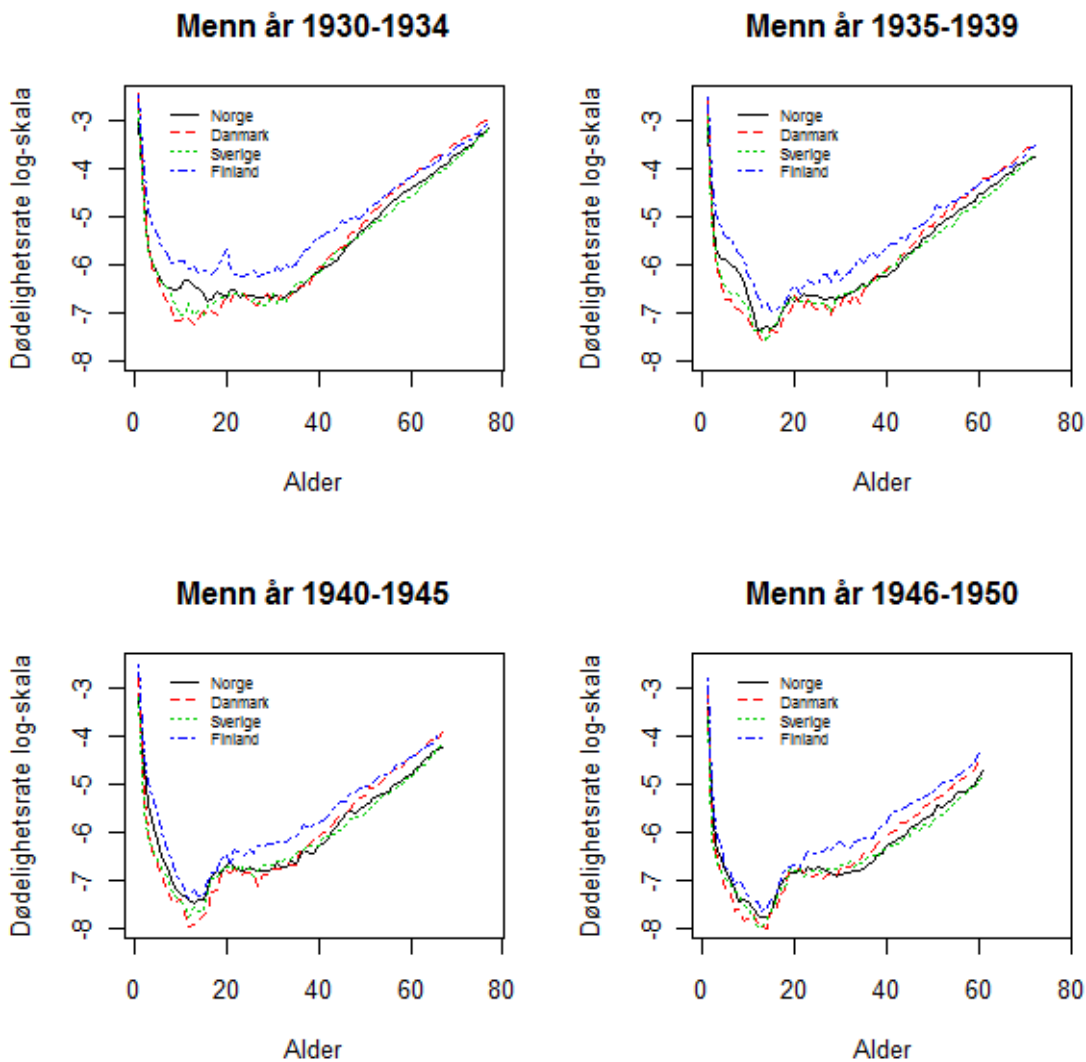
Vi har nå sett mye på forskjellen landene i mellom når det gjelder spedbarnsdødelighet. Vel så interessant er det å se på ulikhetene også i andre aldre. Dette vises godt i figur 2.3 og 2.4, da henholdsvis for kvinner og menn. La oss se på kvinnene først. Man kan tydelig se av plottene at Finland skiller seg ut med høy dødelighet i lavere aldre, fram til cirka 30-40-årsalderen med adskillig høyere dødelighet enn i de andre landene. Deretter er de mer på nivå med norske og svenske kvinner, mens danske kvinner helt klart toppe dødelighetsstatistikken i aldre over 40 år. Til tross for dette har de danske kvinnene den laveste dødeligheten fram til 20-årsalderen, tett etterfulgt av Sverige og etterhvert også Norge. Det er verdt å merke seg at norske kvinner har den laveste dødeligheten i alderen 20 til 40 år, i samtlige av kohortene, og faktisk opp til omtrent 50 år i årene 1930-1939. Dødeligheten er stort sett på sitt laveste i 10-30-årsalderen, men etter dette øker den mer eller mindre kontinuerlig. Plottene viser også at forskjellene landene i mellom minker for hver kohort, og i etterkrigsårene er det ikke store ulikheter. For kohort fire har vi kun data opp til 60 år, men det vil trolig være trenden også utover ettersom det ikke er så store forskjeller landene i mellom i dag når det gjelder slike ting som blant annet utvikling, økonomi og livsstil. For å se på krigsårene spesielt er det fram til 20-årsalderen ikke store forskjeller mellom norske, svenske og danske kvinner, mens finske kvinner ligger litt høyere. Fra 20-årsalderen og opp til 40-årene har norske kvinner den laveste dødeligheten. Etter dette er dødeligheten i Norge, Sverige og Finland mer eller mindre lik, mens danskene ligger høyere med eksempelvis dødelighetsrate på 0.0042 for 50-åringer mot rater på 0.0026-0.0028 for de tre andre landene. For 60-åringer har



Figur 2.3: Dødelighetsrater kvinner for hver kohort

dødeligheten blant danske kvinner da økt til 0.0081, mens dødeligheten blant kvinner i de tre andre landene ligger på 0.0053-0.0058. Nøyaktige tall for utvalgte aldre er gitt i tabell 6.1, 6.2, 6.3 og 6.4.

Videre er det mennene som står for tur. Figur 2.4 viser at dødeligheten blant menn generelt sett har ligget høyere enn hos kvinner. Det er et kjent fenomen at kvinner generelt lever lengre enn menn, og dette resultatet er dermed ikke overraskende. Figuren viser at til forskjell fra for kvinner gjelder det for menn, at finske menn topper dødelighetsstatistikken gjennom alle periodene fram til



Figur 2.4: Dødelighetsrater menn for hver kohort

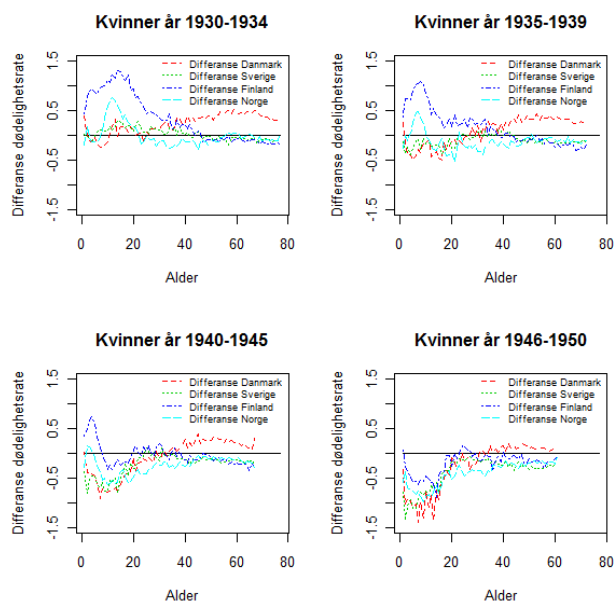
omtrent fylte 60 år. Etter dette ser det ut til at også de blir forbigått av danskene. Av plottet for menn født før 1935 i figur 2.4 er det en topp for dødelighetsraten i 20-årsalderen for finske menn som skiller seg ut fra de andre landene i denne perioden. Dette vil si at de døde i årene 1950-1955, og det er uvisst hva dette kan skyldes. Plottene viser også at dødelighetsratene blant svenske og danske menn utvikler seg forholdsvis likt, men starter å avvike ved cirka 50 år. Også norske menn følger i stor grad den samme utviklingen som svenskene og finnene her, men ligger litt høyere fram til 20-årsalderen. En studie som er gjort på dødsfall som følge av ulykker blant barn for årene 1971-1989 [9], viser at svenske barn

helt klart har lavere dødelighet enn barn i de andre tre landene. Danske barn har også lavere ulykkesdødelighet enn norske barn. Disse resultatene omfatter barn i alderen 0-14 år. Selv om denne studien er gjort for litt senere år enn dataene som er brukt i denne oppgaven, kan det likevel godt tenkes at det samme var tilfelle noen år tidligere. Altså kan dette være en årsak til at dødeligheten blant unge i Norge, det gjelder da særlig gutter, er høyere enn i Danmark og Sverige.

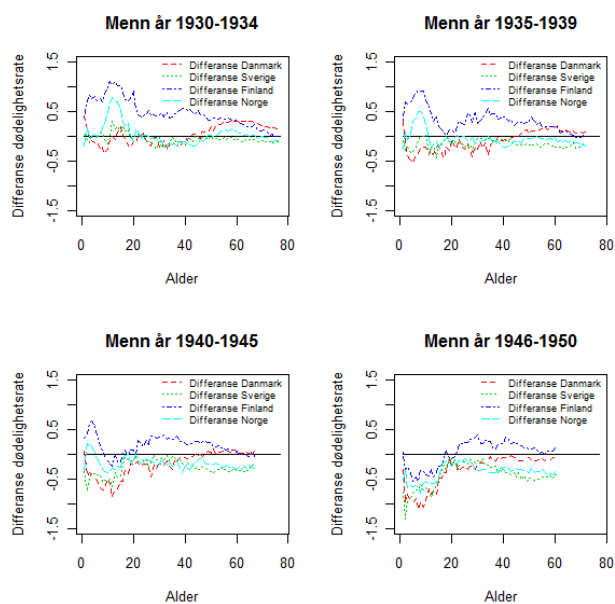
Figur 2.4 viser at forskjellene i dødelighetsratene hos menn landene i mellom i 0-20-årsalderen ser ut til å minke for hver kohort. I den midtre aldersgruppen virker det som om det ikke er så store endringer i forholdet mellom landene ettersom årene går. For høye aldre derimot ser det ut til at det blir større sprik for hver kohort, i motsetning til hva som var tilfelle hos kvinnene. Om man ser spesielt på krigsårene så følger Norge og Sverige hverandre forholdsvis godt, og ligger ganske lavt for alle aldre. I ung alder er det danskene som har den laveste dødeligheten, for så å øke og passere både svenskene og nordmennene i 30-årsalderen, og videre finnene i 60-årene. Finnene har den høyeste dødelighetsstatistikken fram til 60-årsalderen, og det var jo også de som trolig hadde de vanskeligste ernæringsforholdene under krigen. Likevel er det ikke nødvendigvis dette som er årsaken til at de har den høyeste dødeligheten i årene 1940-1945 fordi plottene for de andre periodene også viser at finnene generelt har hatt den høyeste dødeligheten av de fire nordiske landene.

I forbindelse med dødelighetsratene er det også interessant å ta i betraktning differansen i dødelighetsratene i forhold til totalsnittet. Totalsnittet vil her da si gjennomsnittet for alle landene for alle kohortene. Det er naturligvis to totalsnitt, ett for kvinner og ett for menn. Det er laget to ulike typer plot for disse. Figur 6.1 og 6.2 som er å finne i vedlegget, viser den faktiske differansen mellom alle landenes dødelighetsrater og totalsnittet, da henholdsvis for kvinner og menn. På grunn av svært store utslag som følge av høy barnedødelighet er de første ti leveårene utelukket her for bedre å synliggjøre resultatene. Plottene starter dermed fra alder 10 år. Den andre typen plot er vist i figurene 2.5 og 2.6. Disse viser differansen mellom logaritmen av dødelighetsratene og logaritmen av de totale gjennomsnittlige dødelighetsratene. I disse plottene er også alder 0-10 år inkludert. Et annet argument for å se på differansen i logaritmene, er at logaritmer også inngår i beregninger i de senere kapitlene. Figurene 2.5 og 2.6 er derfor tatt med i selve teksten, mens figurene 6.1 og 6.2 er inkludert i vedlegget.

La oss først se på plottene for kvinner. Både figur 6.1 og figur 2.5 viser de samme resultatene som vi har observert tidligere, bortsett fra at de store avvikene kommer tydeligere frem. Plottene viser at Norge, Sverige og Finland er ganske like i høye aldre og ligger under totalsnittet fra cirka 40-årsalderen. Danmark skiller seg tydelig ut ved å ligge et godt stykke over totalsnittet fra 40-årsalderen i de tre første kohortene, men nærmer seg totalsnittet i årene etter krigen. Man ser



Figur 2.5: Differansen mellom logaritmen av dødelighetsrater for kvinner og logaritmen av totalt gjennomsnittlige dødelighetsrater for kvinner



Figur 2.6: Differansen mellom logaritmen av dødelighetsrater for menn og logaritmen av totalt gjennomsnittlige dødelighetsrater for menn

tydelig Danmarks nedgang i dødelighet i høye aldre, og dette kan også leses av tabellene 6.1, 6.2, 6.3 og 6.4. Det er verdt å merke seg at Finland går fra å ha den høyeste dødeligheten i ung alder i forhold til totalsnittet, til å være det landet med den laveste dødelighetsdifferansen for godt voksne og eldre kvinner i både kohort 1, 2 og 3. I årene etter krigen er det derimot svenske kvinner som ligger lengst under totalsnittet i denne aldersgruppen. I kohort fire har avviket fra totalsnittet blitt lite for samtlige av landene.

Figur 6.2 viser at det er større variasjoner landene i mellom blant menn enn hos kvinner. Man ser også at svenskene for det meste ligger under totalsnittet i alle aldre og kohorter, og også nordmennene holder seg forholdsvis lavt. Finnene skiller seg ut ved å ligge godt over totalsnittet fram til rundt 60-årsalderen, spesielt i årene før krigen. I de senere årene nærmer de i større grad seg de andre landene og totalsnittet. I 60-årene blir finnene forbigått av danskene for individer født i alle årene bortsett fra etterkrigsårene, men også disse nærmer seg totalsnittet mer og mer for hver kohort. I figur 2.6 vises det spesielt godt at barnedødeligheten i alle landene har blitt kraftig redusert. Spesielt i Norge og Finland går den fra å ligge et godt stykke over snittet til å havne godt under. Plottene viser også at danskene går fra å ha den laveste dødeligheten i forhold til totalsnittet i ung alder til å ha den høyeste fra rundt 60 år.

Variansen til de estimerte dødelighetsratene har også blitt beregnet, som nevnt i slutten av seksjon 2.2.1, der også beregningsmetoden ble beskrevet. Det var antatt at variansen ville bli liten, noe som også viste seg å være tilfelle. Det er heller ingen store forskjeller mellom landene. For norske kvinner født i årene 1930-1934 er den minste verdien for variansen $2.04 \cdot 10^{-8}$ og den største $1.80 \cdot 10^{-6}$. For norske menn født i samme periode er de tilsvarende tallene $5.20 \cdot 10^{-8}$ og $3.39 \cdot 10^{-6}$. Videre er eksempelvis den minste verdien til variansen til de estimerte dødelighetsratene for kvinner født i årene etter krigen $7.53 \cdot 10^{-9}$ og den største verdien $7.87 \cdot 10^{-7}$. De tilsvarende tallene for norske menn er $1.28 \cdot 10^{-8}$ og $1.00 \cdot 10^{-6}$. I vedlegget finnes det en tabell over variansene for norske kvinner og menn, tabell 6.5. Etersom variansene for de andre landene ikke avviker noe særlig fra de norske er disse ikke tatt med. Figur 6.3 i vedlegget viser hvordan variansene til de estimerte dødelighetsratene endrer seg med alder for norske kvinner og menn. Etersom det er flest mennesker som dør enten i ung alder eller som eldre mennesker, er variansen størst der.

La oss nå forsøke å sette disse resultatene i sammenheng med hypotesen om forskjeller etter ernæringsforhold i de ulike landene. Hovedsakelig vil det ses på dødeligheten for kvinner og menn født i krigsårene 1940-1945, men naturligvis også tas i betraktning utviklingen i dødeligheten i landene både før og etter krigen. Dette for å undersøke om dødeligheten kun følger trenden eller om det er noe annet som har hatt betydning.

La oss først se på Sverige, som var et ikke-okkupert land under krigen og som trolig hadde en rimelig bra ernærings situasjon. Sverige har lave dødelighetsrater i perioden, men så kan man også se at Sverige generelt har lave dødelighetsrater i alle periodene som er sett på her.

Videre har danskene lave dødelighetsrater i tidlige aldre, men høye dødelighetsrater i de øvre aldersgruppene. Dette er også noe som ser ut til å være en trend, siden det samme er tilfelle for både menn og kvinner også i årene før krigen. Likevel kan det muligens tyde på at det omfattende landbruket i Danmark kan ha hatt en positiv effekt på ernærings situasjonen slik at det var lettere for barn og unge å få i seg nok næring til å holde seg i live. Den høye dødelighetsstatistikken i godt voksen alder kan trolig skyldes det danske kostholdet, som består av mye fet mat og alkohol, hvilket som kjent kan føre til en økning i dødeligheten. I tillegg til dette kan også røyking være en sterk medvirkende årsak. Nyere studier der dødeligheten i Danmark har blitt sammenlignet med den lavere dødeligheten i Sverige[6] konkluderer med at mange av de for tidlige dødsfallene i Danmark kan relateres til usunn kost, stort alkoholforbruk og røyking. De baserer seg på data for årene 1997-2001 ettersom dette er den seneste tilgjengelige perioden med opplysninger om dødsårsaker for de to landene. Likevel kan disse resultatene trolig også strekkes lenger tilbake i tid.

Når det gjelder Finland så har de en generelt høy dødelighet både for personer født før krigen og under krigen, da særlig for menn. For kvinner har mortaliteten vært høyest i lave aldre og deretter i større grad nærmet seg de andre landenes dødelighet. Finland hadde det vanskelig under krigen og det kan forklare at dødeligheten var høyest blant barn og unge der i forhold til de andre landene. Samtidig kan man se at dette har vært tilfelle for finnene også i tidligere år, så det er ikke sikkert at den høye dødeligheten i årene 1940-1945 skyldes krigstilstandene.

For Norge er dødeligheten for de som ble født i 1940-1945 lav, og da særlig for kvinner. Dødeligheten i Norge ligger på samme nivå som svenskene, og i alderen 20-45 år (30-45 år for menn) er dødeligheten faktisk lavere hos nordmennene enn hos svenskene. Dette kan muligens være en liten indikasjon på at det ligger noe i hypotesen om at moderat vanskelige forhold faktisk kan føre til en reduksjon i dødeligheten.

2.2 Overlevelsesfunksjon

2.2.1 Beregning av overlevelsesfunksjoner

Å se på overlevelsesfunksjonene er på et vis ikke noe annet enn en annen måte å se på dødelighetsratene på. Overlevelsesfunksjonene finnes ved å ta utgangspunkt i dødelighetsratene. Ettersom disse var gitt som hazarder $h(t)$

$$h(t) \approx \frac{1}{\Delta} P(t \leq T < t + \Delta | T \geq t)$$

kan overlevelsesfunksjonen da skrives som

$$\begin{aligned} S(t) &= \exp\left(-\int_0^t h(s) ds\right) \\ &= \exp(-H(t)) \end{aligned}$$

der

$$H(t) = \int_0^t h(s) ds$$

er den kumulative hazarden.

Ettersom vi har stykkevis konstante hazarder kan $H(t)$ i stedet skrives på formen

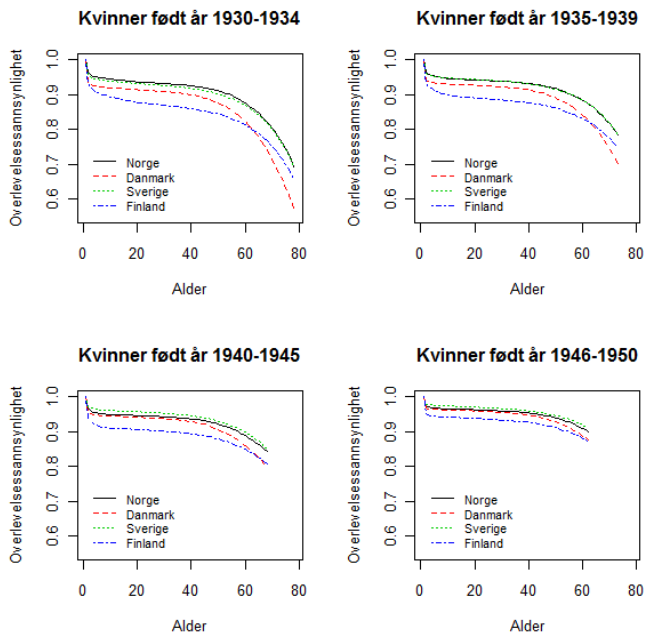
$$H(t) = \sum_{u \leq t} h_u$$

og overlevelsesfunksjonen blir dermed

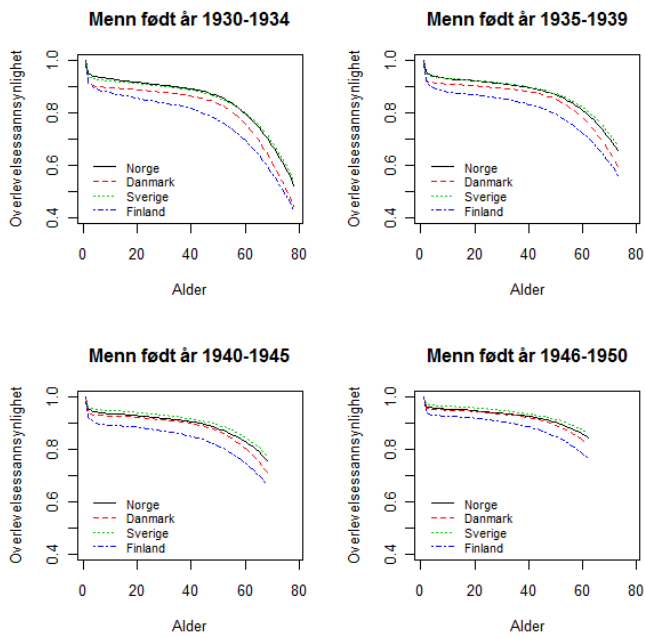
$$S(t) = e^{-\sum_{u \leq t} h_u} \quad (2.3)$$

2.2.2 Overlevelsessannsynlighet i de nordiske landene

Overlevelsessannsynligheten for kvinner og menn i hver alder for hver av periodene er vist i figur 2.7 og figur 2.8. Man ser av figurene at det er Norge som har høyest overlevelsessannsynlighet for både menn og kvinner som er født i årene 1930-1934, mens Sverige kommer best ut fra og med år 1940. I tidsperioden 1935-1939 ligger Norge og Sverige ganske så likt på toppen, men svenske menn



Figur 2.7: Overlevelsessannsynlighet kvinner



Figur 2.8: Overlevelsessannsynlighet menn

har litt bedre overlevelsessannsynlighet enn de norske i den eldre aldersgruppen. Finske menn har den dårligste overlevelsessannsynligheten for alle aldre i alle periodene. Likevel ser man at de i årene etter andre verdenskrig i større grad kommer opp på nivå med de andre landene. Det samme er tilfelle for de finske kvinnene. Danske menn har den nest dårligste sannsynligheten for å overleve gjennom alle årene 1930-1950. Likevel ligger de nærmere svenske og norske menn enn det de finske gjør. Mer eller mindre det samme er tilfelle for de danske kvinnene. Den eneste forskjellen er at danske kvinner har lavest overlevelsessannsynlighet etter fylte 60 år, og har dermed passert finnene. Forskjellene mellom disse tre landene med størst overlevelsessannsynlighet reduseres tydelig i årene under og etter krigen, mens finnene fortsatt befinner seg litt under. Dersom man sammenligner årene før 1940 med årene etter ser man at det, i tillegg til hos finnene, har vært en klar forbedring i overlevelsessannsynligheten hos danskene, både for kvinner og menn. Det viser seg også at spredningen i overlevelsessannsynligheten er større for menn enn for kvinner etter 2. verdenskrig. Dette skyldes hovedsakelig at avstanden mellom finnenes overlevelsessannsynlighet og de andre landenes overlevelsessannsynlighet er større hos mennene enn hos kvinnene.

For å sette dette i sammenheng med hypotesen om at moderat vanskelige forhold kan føre til en reduksjon i dødeligheten, kan man se av figurene at for individer som ble født i krigsårene 1940-1945 er overlevelsessannsynligheten blant danske menn og kvinner mindre enn for de norske. Dette kan være et understøttende argument for at hypotesen kan stemme, ettersom Danmark trolig hadde bedre ernæringsforhold enn Norge under krigen på grunn av sitt landbruk. På den annen side går det imot hypotesen at landet med høyest overlevelsessannsynlighet er Sverige, som holdt seg nøytral under andre verdenskrig og dermed trolig hadde gode ernæringsforhold. For årene før 1940 har nordmenn tilnærmet lik sannsynlighet som svenskene for å overleve, og for kvinnene ligger faktisk norske kvinner litt høyere. Siden dette da endrer seg for personer født i krigsårene, strider dette imot hypotesen om at moderate vanskelige ernæringsforhold fører til en reduksjon i dødeligheten. Finland som hadde det ekstra vanskelig under krigen, har forholdsvis lavere overlevelsessannsynlighet enn de andre landene for personer født i denne perioden. Likevel må man se på det som en helhet ved også å se på tidsepokene før 1940, der finnene også hadde en overlevelsessannsynlighet som var adskillig lavere enn for innbyggerne i de andre tre landene. Dermed er det ikke sikkert at de dårlige forholdene i Finland under krigen førte til en reduksjon i overlevelsessannsynligheten.

2.3 Forventet levealder

2.3.1 Forventet gjestående levetid for en x-åring

Vi starter med å definere de ettårige døds-og overlevelsessannsynlighetene til en x-åring som q_x og p_x henholdsvis. Disse har følgende relasjon

$$p_x + q_x = 1$$

Overlevelsessannsynligheten p_x er definert som

$$p_x = e^{-\int_0^1 \mu_{x+s} ds} \quad (2.4)$$

der μ_x er dødsintensiteten til en x-åring [4]. Ut i fra dette kan vi definere sannsynligheten for at en x-åring overlever minst t år som

$${}_t p_x = e^{-\int_0^t \mu_{x+s} ds} \quad (2.5)$$

Fra boken av Gerber[4] har vi at den framtidige levetiden T er en tilfeldig variabel med en fordelingsfunksjon

$$G(t) = Pr(T \leq t), \quad t \geq 0$$

Dette innebærer at

$${}_t q_x = G(t)$$

og dermed

$${}_t p_x = 1 - G(t)$$

Dersom man nå antar at G er kontinuerlig og har en sannsynlighetstetthet $g(t) = G'(t)$, kan man si at

$$g(t)dt = Pr(t < T < t + dt)$$

er sannsynligheten for at døden vil inntreffe i det lille tidsintervallet fra t til $t + dt$. Forventet gjestående levetid for en x -åring kan dermed uttrykkes ved

$$E_x = \int_0^{\infty} t g(t) dt \quad (2.6)$$

Det kan vises at dette er det samme som

$$E_x = \int_0^\infty [1 - G(t)] dt = \int_0^\infty {}_t p_x dt \quad (2.7)$$

${}_t p_x$ er en betinget sannsynlighet, og uttrykt ved hjelp av overlevelsesfunksjonen blir denne da

$$\begin{aligned} {}_t p_x &= P(T > x + t | T > x) \\ &= \frac{S(x + t)}{S(x)} \end{aligned}$$

Resultatet blir dermed at forventet gjenstående levetid kan uttrykkes som følger

$$E_x = \int_0^\infty \frac{S(x + t)}{S(x)} dt \quad (2.8)$$

2.3.2 Forventet gjenstående levetid basert på dataene

Ettersom vi har registrerte dødsfall i alle kohortene kun opp til 60 år, må vi for å finne forventet gjenstående levetid da anta at maksalderen er 60 år, istedenfor 105, 111 eller 120 år som den ofte blir satt til i dag. Å si at maksalderen er 60 år vil her si at man kan fylle 60 år, men ikke 61 år, nærmere bestemt at individene dør i intervallet $[0, 61)$. Det er naturligvis urealistisk å anta en maksalder på 60 år, men mangelen på data gjør det nødvendig å beregne forventet gjenstående levealder på denne måten. Dermed blir forventet gjenstående levetid gitt i antall år for en x -åring

$$E_x = \int_0^{60} {}_t p_x dt = \int_0^{60} \frac{S(x + t)}{S(x)} dt$$

Alder kvinner	Norge	Danmark	Sverige	Finland
0	58.58	58.42	58.67	56.61
10	49.76	49.43	49.73	49.12
20	40.15	39.72	40.02	39.88
30	30.32	29.92	30.24	30.19
40	20.51	20.21	20.49	20.48
50	10.78	10.65	10.79	10.79
60	0.993	0.990	0.994	0.994
Alder menn	Norge	Danmark	Sverige	Finland
0	57.08	57.39	57.51	54.43
10	48.57	48.65	48.76	47.16
20	39.18	39.05	39.24	38.15
30	29.62	29.46	29.65	28.83
40	20.02	19.86	20.06	19.56
50	10.55	10.47	10.61	10.43
60	0.987	0.984	0.989	0.984

Tabell 2.1: Forventet gjenstående levetid for personer født i 1930-1934, når mak-salder er antatt å være 60 år

Alder kvinner	Norge	Danmark	Sverige	Finland
0	58.79	58.82	59.07	57.25
10	49.99	49.62	49.89	49.67
20	40.17	39.80	40.08	40.01
30	30.33	29.98	30.27	30.25
40	20.52	20.26	20.51	20.51
50	10.79	10.68	10.80	10.80
60	0.994	0.990	0.995	0.994
Alder menn	Norge	Danmark	Sverige	Finland
0	57.30	57.87	57.92	55.26
10	48.86	48.84	48.97	47.92
20	39.24	39.17	39.34	38.42
30	29.67	29.53	29.73	29.03
40	20.08	19.93	20.15	19.72
50	10.61	10.53	10.66	10.49
60	0.989	0.985	0.990	0.986

Tabell 2.2: Forventet gjenstående levetid for personer født i 1935-1939, når mak-salder er antatt å være 60 år

Alder kvinner	Norge	Danmark	Sverige	Finland
0	58.98	59.06	59.41	58.15
10	50.05	49.73	49.97	49.88
20	40.20	39.88	40.14	40.09
30	30.34	30.05	30.32	30.30
40	20.52	20.31	20.53	20.52
50	10.80	10.71	10.81	10.80
60	0.994	0.992	0.994	0.995
Alder menn	Norge	Danmark	Sverige	Finland
0	57.64	58.14	58.34	56.26
10	49.05	48.95	49.10	48.21
20	39.43	39.23	39.45	38.62
30	29.82	29.59	29.85	29.21
40	20.18	19.99	20.26	19.82
50	10.67	10.57	10.70	10.54
60	0.991	0.988	0.991	0.988

Tabell 2.3: Forventet gjenstående levetid for personer født i 1940-1945, når mak-salder er antatt å være 60 år

Alder kvinner	Norge	Danmark	Sverige	Finland
0	59.52	59.46	59.71	59.23
10	50.12	49.85	50.07	49.96
20	40.26	39.99	40.23	40.14
30	30.39	30.15	30.40	30.32
40	20.56	20.40	20.59	20.53
50	10.81	10.75	10.83	10.81
60	0.994	0.993	0.995	0.994
Alder menn	Norge	Danmark	Sverige	Finland
0	58.47	58.52	58.81	57.46
10	49.26	49.04	49.31	48.36
20	39.58	39.33	39.62	38.72
30	29.95	29.68	30.00	29.30
40	20.29	20.10	20.37	19.87
50	10.71	10.63	10.74	10.56
60	0.991	0.989	0.992	0.986

Tabell 2.4: Forventet gjenstående levetid for personer født i 1946-1950, når mak-salder er antatt å være 60 år

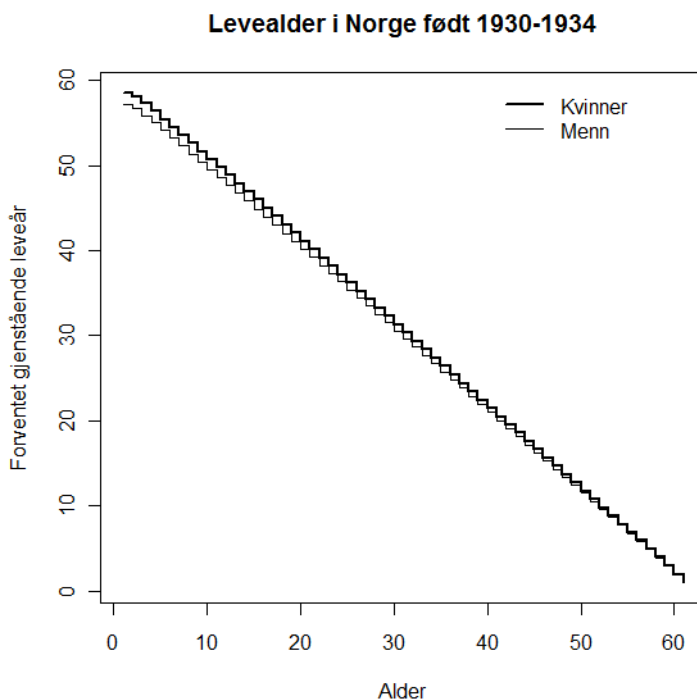
Til tross for at en maksalder på 60 år ikke er realistisk, forteller de beregnede forventede gjenstående levealderene noe om forskjellene landene i mellom, og kan dermed sies å være en annen måte å se på det samme som dødelighetsratene og overlevelsesfunksjonene viste. De forventede gjenstående leveårene for ulike aldre er gitt i tabellene 2.1, 2.2, 2.3 og 2.4, for henholdsvis personer født i 1930-1934, 1935-1939, 1940-1945 og 1946-1950. Ettersom maksalder 60 år ikke representerer den faktiske maksalderen for individene vil dette resultere i at forventet antall gjenstående leveår vil bli en god del høyere enn om den faktiske maksalderen hadde vært kjent. Grunnen til dette er at overlevelsessannsynlighetene for de her antatte siste leveår er mye høyere enn det de ellers ville ha vært for de faktiske siste leveår, ettersom en person på 50-60 år har en forholdsvis høy overlevelses-sannsynlighet.

Tabellene viser at det ikke er store forskjeller landene i mellom, bortsett fra Finland som klart har lavere forventet gjenstående levetid enn de andre landene. Det kommer også tydelig fram av tabellene at menn i samtlige av landene har kortere forventet gjenstående levetid enn det kvinnene har. Dersom man sammenligner tabellene for de ulike kohortene ser man at den forventede levealderen har økt for hver kohort. Økningen er naturlig nok ikke stor ettersom vi her ser på et såpass kort tidsintervall, men man ser likevel klart at det har vært en økning.

For nyfødte i årene 1930-1934 var den forventede levealderen over 58 år for jenter og over 57 år for gutter i Norge, Danmark og Sverige, mens finske jenter kunne forvente å bare leve til de var 56.61 år og guttene til 54.43 år. I den neste kohorten har svenske nyfødte jenter klart å oppnå en forventet levealder på 59.07 år, mens norske og danske ligger nesten likt på omtrent 58.8 år. Finnene henger fortsatt etter, men har klart å klatre seg oppover til 57.25 år for jentene og 55.26 år for guttene. Av de tre andre landene er det de norske guttene som har lavest forventet levetid på 57.30 år, mens svenskene og danskene i større grad har nærmet seg 58 år. I årene under krigen er det svenskene som har den høyeste forventede levealderen med 59.41 år for nyfødte jenter og 58.34 for gutter. Deretter følger danskene, nordmennene og til slutt finnene hos både menn og kvinner. I årene etter krigen har de finske kvinnene kommet seg opp på nivå med de øvrige landene med en forventet levalder på 59.23 år. Svenske jentebarn har fortsatt den lengste forventede levetiden med 59.71 år. Norge og Danmark har en rimelig lik forventet levealder disse årene, og det gjelder både for gutter og jenter. Finske guttebarn ligger fortsatt dårligst an med en forventet levetid på 57.46 år, mens nyfødte svenske gutter har et forventet liv på 58.81 år. Man ser også av tabellene at de forventede levetidene landene i mellom spriker helt klart mest i de tidlige leveårene, for deretter å jevne seg mer ut.

Et mulig problem ved kalkulering av forventet levetid for nyfødte er at den er svært sensitiv for høy dødelighet i de første leveårene. I tillegg er det slik at barn

som dør i løpet av første leveår ofte dør allerede de første dagene eller ukene etter fødselen, og vi har dermed egentlig ikke en uniform fordeling. Disse problemene vil imidlertid bli ekskludert når man ser på forventet gjenstående levetid for høyere aldre, for eksempel for barn som er blitt fem eller ti år gamle, ettersom disse da ikke lenger baserer seg på dødeligheten i de tidligere leveårene. Det er heller ikke uvanlig å se en klar økning i forventet gjenstående levetid ved nådd pubertetsalder.



Figur 2.9: Forventet gjenstående levetid for norske kvinner og menn født i 1930-1934

Tabellene viser at etterhvert som man blir eldre så øker også den forventede totale levealderen. Dette skyldes naturlig nok at dersom man har overlevd en viss alder så er det større sannsynlighet for å overleve til en høyere alder. Når man kommer opp i forventet resterende levetid for 60-åring er det mer eller mindre ingen forskjell hverken mellom menn og kvinner eller landene i mellom, og det var nødvendig å ta med tre desimaler for å vise at disse ikke var helt identiske. For kvinner i alle landene og i alle kohortene ligger forventet gjenstående levetid ved 60 år på over 0.99 år. For menn ligger det noe lavere, de holder seg over 0.98, mens svenske og norske menn klatrer opp til 0.99 for individer født i henholdsvis 1935-1939 og 1940-1945. En såpass høy forventet gjenstående levetid ved fylte 60 år når maksalder er satt slik at man ikke kan fylle 61, skyldes at dette ikke er den faktiske

maksalderen og dermed baserer seg på for høye overlevelsessannsynligheter.

Figur 2.9 viser at funksjonen for forventet gjenstående levetid, slik de er beregnet her, er en trappefunksjon (også kalt step-funksjon), og viser den forventede resterende levetiden i hver alder for norske kvinner og menn født i perioden 1930-1934. Funksjonene for de andre landene og kohortene er også trappefunksjoner, og ser mer eller mindre tilsvarende ut. Ettersom en trappefunksjon ikke gir helt nøyaktige tall, ved at den ikke får med seg alle punktene mellom hvert steg, kunne det vært bedre å ha brukt en interpolert levetidsfunksjon i stedet. En interpolert levetidsfunksjon ville beregnet nye datapunkter mellom to eller flere eksisterende datapunkter, og dette kunne muligens ha gitt en bedre og mer nøyaktig tilpassning. Likevel ville den trolig ikke gitt noen store avvik fra metoden som er brukt her.

3 Poisson regresjon

Begrepet regresjonsanalyse vil si bruk av teknikker for modellering og analyse av sammenhenger mellom en avhengig variabel og en eller flere uavhengige variable. Poisson regresjon er en type regresjon benyttet på Poisson-fordelte data, og brukes ofte i forbindelse med analyse av telldata.

3.1 Teori

Definisjon 1. *de Jong & Heller, Rice. En Poisson-fordelt tilfeldig variabel Y har punktsannsynlighet $f(y) = P(Y = y)$, gitt som*

$$P(Y = y) = \frac{e^{-\mu} \mu^y}{y!}$$

der μ er forventningsverdien til Y , $E(Y) = \mu$.

Ved Poisson regresjon er forventningen μ forklart ved hjelp av forklaringsvariable x via en passende link-funksjon $g(\mu)$.

Definisjon 2. *de Jong & Heller. Poisson regresjonsmodellen er*

$$y \sim P(\mu), \quad g(\mu) = \beta' X$$

der X er en vektor av forklaringsvariable og β er en vektor av tilhørende koeffisienter.

Ved modellering av Poisson regresjon benyttes generaliserte lineære modeller (GLM). En GLM består av tre elementer

1. En fordelingsfunksjon f , fra den eksponensielle familien.
2. En lineær prediktor $\eta = X\beta$
3. En link-funksjon g slik at $E(Y) = \mu = g^{-1}(\eta)$

Forventningen til den Poisson-fordelte tilfeldige variabelen kan skrives som

$$E(Y) = \mu = e^{\alpha + \beta_1 x_1 + \dots + \beta_p x_p}$$

Ettersom logaritmen av denne funksjonen gir en lineær kombinasjon av prediktorer sies denne modellen å ha en log-link funksjon. Det vil si at den funksjonen som forbinder forventningen til den lineæriserte prediktoren, er log-funksjonen.

En vanlig situasjon for Poisson regresjon er med en respons som er dødsfall i en gruppe j , stratum i . Vi kan definere denne responsen som y_{ij} , og dødelighetsraten er gitt ved

$$r_{ij} = \frac{y_{ij}}{p_{ij}}$$

der p_{ij} er akkumulert personår i gruppe j , stratum i . Dette tilsvarer jo da dødelighetsratene som er funnet for hver alder i , $i = 0, \dots, 110$, og hver kohort j , $j = 1, 2, 3, 4$. Der kohort 1 er individer født i 1930-1934, kohort 2 er individer født i perioden 1935-1939, kohort 3 er personer født i 1940-1945 og kohort 4 er personer født etter andre verdenskrig, som her vil si årene 1946-1950.

Ved modellering av tellevariable, slik som antall dødsfall i en gruppe, kreves det korrigerer for antall individer under risiko, n . Man kan si at n tilsvarer akkumulert antall personår, p_{ij} , som beskrevet over. Vi har at den såkalte "occurrence/

exposure raten” er gitt ved

$$\frac{\mu}{n}$$

Når g da er log-link funksjonen har man at

$$\ln\left(\frac{\mu}{n}\right) = \beta'x \quad \Rightarrow \quad \ln(\mu) = \ln(n) + \beta'x$$

Her kalles $\ln(n)$ for ”offset”. Offsets brukes for å korrigere for forskjeller i tidsperioder eller gruppestørrelser. Et offset kan sies å være en annen x -variabel i regresjonsmodellen, med en β -parameter som er lik 1. Dersom man nå definerer forventet antall døde som μ_d kan denne skrives som

$$\begin{aligned}\mu_d &= ne^{\beta'x} \\ &= e^{\ln(n)+\beta'x}\end{aligned}$$

Et viktig hjelpemiddel for å analysere resultatene av Poisson regresjon og hypotese-testing er devians. Deviansen er generelt definert ved

$$\Delta = 2(\tilde{l} - l)$$

der \tilde{l} er log-likelihooden for en mettet modell. En modell kalles mettet dersom den har like mange parametere som observasjoner, det vil si den modellen som gir best tilpasning. Hvis l forventes å ligge nærme (men ikke være større enn) \tilde{l} så gir modellen en god tilpasning. En stor verdi for deviansen indikerer at modellen passer dårlig. Det er også slik at minimering av Δ er ekvivalent med maksimering av likelihooden. Spesielt er deviansen for Poisson fordelingen

$$\Delta = 2 \sum_{i=1}^n \left\{ y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i) \right\}$$

Det er også vanlig å se på to ulike typer devianser; nulldevians og residual devians.

- **Nulldevians** er devians med modellen $\mu_i = \mu$ eller $g(\mu_i) = \beta_0$, det vil si en modell bestående kun av konstantledd.
- **Residual devians** er devians i den aktuelle modellen $g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$, det vil si deviansen innsatt maksimum likelihood estimatoren $\hat{\beta}$.

3.2 Poisson regresjon benyttet på datasettet

For dataene kan man sette opp følgende regresjonsmodell

$$\log(\mu_{ijk}) = \eta_0 + \alpha_i + \beta_j + \gamma_k$$

der α er kovariaten alder, β er kovariaten kohort og γ er kovariaten land. Regresjonsfunksjonene α_i , β_j og γ_k beskriver hvordan kovariatene påvirker hazard raten ved tid t . η_0 kan for eksempel svare til landet Norge, kohort 1930-1934 og alder 1 år. Dette vil da si at ettersom regresjonsmodellen består av α -parameteren som står for alder, β for kohort og γ for land og disse i tillegg inngår i η_0 , så er modellen egentlig overparametrisert.

Det er interessant å se på relative risikoer(RR), som er definert som forholdet mellom to sannsynligheter[8]. Først har vi at intensiteten er gitt ved

$$\mu_{ijk} = e^{\eta_0 + \alpha_i + \beta_j + \gamma_k}$$

og den estimerte intensiteten er da

$$\hat{\mu}_{ijk} = e^{\hat{\eta}_0 + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k}$$

der $\hat{\eta}_0$, $\hat{\alpha}_i$, $\hat{\beta}_j$ og $\hat{\gamma}_k$ er de estimerte verdiene for η_0 , α_i , β_j og γ_k .

Dermed blir for eksempel relativ risiko for kohort

$$\begin{aligned} RR_j &= \frac{\mu_{ijk}}{\mu_{i0k}} \\ &= \frac{e^{\eta_0 + \alpha_i + \beta_j + \gamma_k}}{e^{\eta_0 + \alpha_i + 0 + \gamma_k}} \\ &= e^{\beta_j} \end{aligned}$$

og estimert relativ risiko for kohort blir da

$$\hat{RR}_j = e^{\hat{\beta}_j}$$

På tilsvarende måte blir estimert relativ risiko for alder

$$\hat{RR}_i = e^{\hat{\alpha}_i}$$

og estimert relativ risiko for land

$$\hat{RR}_k = e^{\hat{\gamma}_k}$$

Det er også av interesse å se om det er noen interaksjon mellom land og hvilken tidsperiode personene er født i. Dette gjøres ved å utvide modellen til

$$\log(\mu_{ijk}) = \eta_0 + \alpha_i + \beta_j + \gamma_k + \beta_j : \gamma_k$$

der leddet $\beta_j : \gamma_k$ er interaksjonen mellom kohort j og land k . Den estimerte intensiteten blir da betegnet som

$$\hat{\mu}_{ijk} = e^{\hat{\eta}_0 + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\beta}_j : \hat{\gamma}_k}$$

og de relative risikoene finnes på tilsvarende måte som for modellen uten interaksjonsledd.

Ettersom dødeligheten generelt er svært høy i første leveår, tas dette ikke med i beregningen. Det tas da altså kun i betraktning aldere større enn null for å få mer fornuftige resultater.

Regresjonsmodellen kan ses i sammenheng med modellen for proporsjonale hazarder, på grunn av at den har stykkevis konstante hazarder. Man snakker ofte om proporsjonale hazarder i forbindelse med Cox regresjon. Cox regresjon benyttes som regel på individdata, mens for grupperte data slik som vi har her er det bedre å bruke Poisson regresjon.

La oss utdype litt mer hva en proporsjonal hazard modell er, der de følgende resultatene er hentet fra boken av Aalen et al.[1] Dersom man antar at en vektor av kovariater $\mathbf{x}_i(t)$ for individ i er relatert til hazard raten $\alpha(t | \mathbf{x}_i)$ til individet ved relasjonen

$$\alpha(t | \mathbf{x}_i) = \alpha_0(t) r(\beta, \mathbf{x}_i(t))$$

der $r(\beta, \mathbf{x}_i(t))$ er relativ risikofunksjon, $\beta = (\beta_1, \dots, \beta_p)^T$ er en vektor av regresjonskoeffisienter og $\alpha_0(t)$ er en baseline hazard rate. Dersom man ser på to individer, indeksert som 1 og 2, henholdsvis med vektor av kovariater $\mathbf{x}_1(t)$ og $\mathbf{x}_2(t)$ er forholdet mellom deres hazard rater

$$\frac{\alpha(t | \mathbf{x}_2)}{\alpha(t | \mathbf{x}_1)} = \frac{r(\beta, \mathbf{x}_2(t))}{r(\beta, \mathbf{x}_1(t))} \quad (3.1)$$

Dersom alle kovariatene er faste er dette forholdet konstant over tid, og modellen (3.1) kalles en proporsjonal hazard modell. Dersom man antar at alle komponentene til $\mathbf{x}_1(t)$ og $\mathbf{x}_2(t)$ er like, bortsett fra den j -te komponenten, vil (3.1) for Cox modellen, det vil si når $r(\beta, \mathbf{x}) = e^{\beta' \mathbf{x}}$, bli

$$\frac{\alpha(t | \mathbf{x}_2)}{\alpha(t | \mathbf{x}_1)} = \exp \{ \beta^T (\mathbf{x}_2(t) - \mathbf{x}_1(t)) \} = e^{\beta_j}$$

Dermed finner man effekten ved én enhets økning i den j -te kovariatene, når alle andre kovariater holdes faste, ved å multiplisere hazard raten med e^{β_j} .

For å se på devianser og relative risikoer, finner man disse ved å ta utgangspunkt i Poisson regresjon. Koeffisientene man får ved Poisson regresjon brukes videre for å finne de relative risikoene. Residualdeviansene får man ut direkte i R ved å bruke *anova*-kommandoen.

Aldersgruppe	Faktorer	Frihetsgrader	Residual devians	Endring devians
1-25 år	NULL		93228	
	Alder	24	29668	63560
	Kohort	3	19885	9783
	Land	3	9336	10548
	Land:Kohort	9	8698	638
26-50 år	NULL		33005	
	Alder	24	5298	27707
	Kohort	3	4708	589
	Land	3	2710	1999
	Land:Kohort	9	2662	47
51-76 år	NULL		100087	
	Alder	25	15461	84626
	Kohort	3	14186	1275
	Land	3	2344	11842
	Land:Kohort	9	2048	296

Tabell 3.1: Residual devians for kvinner

La oss se på tabellene for deviansene, tabell 3.1 og 3.2. Disse er inndelt i tre ulike aldersgrupper; en for unge, en for middelaldrende, og en for de litt eldre. Dette er fordi det er grunn til å anta at dødeligheten vil være forskjellig disse tre aldersgruppene i mellom. Tabellene viser hvilken betydning faktorene alder, kohort, land og interaksjonen mellom land og kohort har. Tabellen brukes for å undersøke om alle faktorene er signifikante, og for å se om noen av dem eventuelt ikke har så stor innflytelse på modellen og dermed ikke trenger å være med slik at vi får en enklere modell. Det er to ulike måter man kan se om faktorene er signifikante på. Den første er å se på p-verdien.

Aldersgruppe	Faktorer	Frihetsgrader	Residual devians	Endring devians
1-25 år	NULL		88084	
	Alder	24	30782	57302
	Kohort	3	22037	8745
	Land	3	10961	11076
	Land:Kohort	9	10314	647
26-50 år	NULL		56772	
	Alder	24	14065	42706
	Kohort	3	12720	1346
	Land	3	3787	8933
	Land:Kohort	9	3598	189
51-76 år	NULL		154411	
	Alder	25	20800	133611
	Kohort	3	14922	5878
	Land	3	3731	11192
	Land:Kohort	9	3187	544

Tabell 3.2: Residual devians for menn

For menn er p-verdiene knøttsmå og tilnærmet lik null for samtlige av faktorene. Dette viser at de alle er signifikante. For kvinner er også alle p-verdiene så og si lik null, bortsett fra for interaksjonen mellom land og kohort i alderen 26-50 år der p-verdien er $3.206 \cdot 10^{-7}$. Dette indikerer at det er større grunnlag for å ta leddet med interaksjon mellom kohort og land i aldersgruppen 26-50 år for kvinner ut av modellen. Likevel, selv om denne ikke er like lav som de andre er den likevel svært lav, og er dermed også helt klart signifikant. Dermed blir konklusjonen at også for kvinnene er alle faktorene signifikante for samtlige av aldersgruppene.

Den andre måten er å se på endringen i deviansen i forhold til forventningen i kjikvadratfordelingen. For en variabel X som er kjikvadratfordelt med k frihetsgrader, $X \sim \chi_k^2$, så er forventningen til denne gitt ved

$$E(X) = k$$

Dette vil si at dersom endringen i deviansen er større enn forventningen i kjikvadratfordelingen, som da vil si antall frihetsgrader, så er faktoren signifikant. Det kommer fram av tabellene 3.1 og 3.2 at dette er tilfelle for alle faktorene i alle aldersgruppene for både kvinner og menn.

Videre kan vi se på de relative risikoene. Som nevnt tidligere finnes de relative risikoene ved å bruke koeffisientene man får fra Poisson regresjonen. I tillegg må man legge inn hva offsetet skal være. Her blir det da befolkningsstørrelsen for henholdsvis kvinner og menn. I tillegg til de relative risikoene er det også ønskelig å få tak i konfidensintervallene til disse. Uavhengig av hvor mange faktorer man ønsker å se på, dersom man nå definerer en samlebetegnelse for alle koeffisientene, *koeff*, og sier at standardfeilen til disse er σ , så kan 95% konfidensintervallene til

den relative risikoen skrives som

$$RR_{ki} = e^{koeff \pm 1.96\sigma}$$

For å finne den relative risikoen til interaksjonsleddene er det ikke fullt så rett fram som for de enkeltstående faktorene. Det er to måter å finne den relative risikoen til interaksjonen mellom land og kohort på. Den ene er å multiplisere den relative risikoen for land, den relative risikoen for kohort og den relative risikoen for land og kohort med hverandre, som følger

$$RR_{interaksjon\ land:kohort} = RR_{land} \cdot RR_{kohort} \cdot RR_{land:kohort}$$

Den andre varianten får man ved å definere en ny variabel "LandKohort" som er

$$LandKohort = Kohort + 4 \cdot (Land - 1)$$

Denne variabelen får da 16 verdier som definerer alle de ulike kombinasjonene av land og kohort. Deretter brukes "LandKohort" i Poisson regresjonen i stedet for de tre faktorene land, kohort og interaksjonen mellom land og kohort. Videre finnes de relative risikoene på samme måte som tidligere. Fordelen med denne metoden er at man får de relative risikoene med konfidensintervall direkte ut i R, og derfor er det denne som er brukt i denne oppgaven.

De relative risikoene uten å ta i betraktning interaksjon mellom land og kohort er gitt med 95% konfidensintervall i tabellene 3.3, 3.4, 3.5 og 3.6. De to første viser den relative risikoen for land, mens de to siste viser den relative risikoen for kohortene. Tabellene 3.7 og 3.8 viser de relative risikoene for interaksjonen mellom landene og kohortene, også disse med 95% konfidensintervall, for henholdsvis kvinner og menn. Alle tabellene er inndelt i tre aldersgrupper, på samme måte og av samme grunn som for devians-tabellene.

La oss først studere nærmere de relative risikoene for land og kohort uten å ta i betraktning interaksjon. Tabell 3.3 viser kvinnes relative risiko landene i mellom, og tallene er også justert for alder og kohort. For kvinner i alderen 1-25 år ser man at både svenske og danske kvinner har over 20% lavere dødelighet relativt til det norske kvinner har. Likevel er det i denne aldersgruppen helt klart finnene som kommer dårligst ut, for de har en relativ risiko på hele 1.79 i forhold til Norge. Det er også her verdt å merke seg at finnene har et bredere

Aldersgruppe	Land	Relativ risiko	RR nedre	RR øvre
1-25 år	Norge	1	-	-
	Danmark	0.79	0.77	0.81
	Sverige	0.79	0.78	0.81
	Finland	1.79	1.75	1.83
26-50 år	Norge	1	-	-
	Danmark	1.48	1.45	1.51
	Sverige	1.09	1.07	1.11
	Finland	1.17	1.15	1.20
51-76 år	Norge	1	-	-
	Danmark	1.51	1.49	1.53
	Sverige	0.96	0.95	0.97
	Finland	0.93	0.92	0.94

Tabell 3.3: Relativ risiko for land (kvinner) med 95% konfidensintervall, justert for alder og kohort

Aldersgruppe	Land	Relativ risiko	RR nedre	RR øvre
1-25 år	Norge	1	-	-
	Danmark	0.74	0.72	0.75
	Sverige	0.77	0.76	0.79
	Finland	1.50	1.48	1.53
26-50 år	Norge	1	-	-
	Danmark	1.15	1.13	1.17
	Sverige	0.98	0.97	0.99
	Finland	1.63	1.61	1.65
51-76 år	Norge	1	-	-
	Danmark	1.28	1.27	1.29
	Sverige	0.92	0.91	0.92
	Finland	1.26	1.25	1.28

Tabell 3.4: Relativ risiko for land (menn) med 95% konfidensintervall, justert for alder og kohort

konfidensintervall enn de andre landene. I alderen 26-50 år ser vi at alle de tre andre nordiske landene har relativt høyere dødelighet enn Norge. Likevel har svenske kvinner tross alt ikke mye høyere dødelighet relativt til den blant norske kvinner, nærmere bestemt en relativ risiko på 1.09. Danske kvinner derimot er de som har den høyeste relative risikoen i forhold til de norske på 1.48. For kvinner i aldersgruppen 51-76 år, har både Sverige og Finland dødeligheter som er lavere relativt til dødeligheten blant norske kvinner. Faktisk ligger finnene lavest, med

en dødelighet som er 7% lavere enn for norske kvinner. Danskene har derimot en dødelighet i denne aldersgruppen på 1.51 relativt til de norske.

Videre kan vi se på menns relative risiko landene i mellom. Tabell 3.4 viser at svenske menn har relativ risiko til Norge som er lavere enn 1 i samtlige av aldersgruppene. Dette er noe annet enn det som kunne ses hos kvinnene. I aldersgruppen 1-25 år er det danskene som har lavest dødelighet i forhold til norske menn, med en relativ risiko på 0.74. Finnene har derimot en relativ risiko på 1.50, som tilsier at de har betraktelig høyere dødelighet relativt til de norske i samme aldersgruppe. I alderen 26-50 år har finnenes relative risiko økt ytterligere, til 1.63. I tillegg har også danske menn fått en dødelighet som er 15% høyere enn for de norske. Det er verdt å merke seg at den relative risikoen for svenskene blir mindre fra aldersgruppen 1-25 år til 26-50 år, der den går fra 0.77 til 0.98. Dette vil si at svenskene og nordmennene har fått mer lik dødelighet i 26-50-årsalderen. I aldersgruppen 51-76 år har Danmark og Sverige ganske like relative risikoer, på henholdsvis 1.28 og 1.26. Dette vil si at Danmarks relative risiko til Norge har økt i forhold til de yngre aldersgruppene, mens finnenes relative risiko til norske menn har blitt redusert. Svenskene har fortsatt den laveste dødeligheten, og ligger 8% under Norge.

Tabell 3.5 viser kvinnes relative risiko for kohortene. Verdiene her er også justert for alder og land. I alderen 1-25 år ser vi at dødeligheten har blitt redusert med hele 63% fra de som er født i årene 1930-1934 til de som er født i 1946-1950! Vi ser også at dødeligheten har blitt redusert mye for hver kohort i denne aldersgruppen. Den store reduksjonen i dødeligheten her skyldes trolig at det har vært en reduksjon i barnedødeligheten. Til tross for at spedbarn er utelatt her vet man at små barn også har høyere dødelighet enn unge og voksne. Nye medisiner og økt kunnskap har resultert i at barnedødeligheten har blitt redusert med årene. I aldersgruppen 26-50 år har det også vært en nedgang i dødeligheten. Reduksjonen i dødeligheten fra kvinner født i 1930-1934 til kvinner født i 1946-1950 er på nesten 20%. Likevel er dette den aldersgruppen som har hatt den minste dødelighetsreduksjonen. For kvinner i alderen 51-76 år har dødeligheten totalt blitt redusert med 22% fra kohort 1 til kohort 4. Dette er ganske mye tatt i betraktning at dette kun er over cirka 20 år. Når det er snakk om reduksjon i dødelighet er det ofte denne aldersgruppen man er mest interessert i, rett og slett fordi vi er interesserte i om vi lever lenger, og i så tilfelle; hvor lenge.

Mennenes relative risiko for kohortene er vist i tabell 3.6. Denne tabellen viser at også menn har hatt en stor reduksjon i dødeligheten i alderen 1-25 år. Deres reduksjon er på 52% fra menn født i 1930-1934 til menn som er født i 1946-1950. Selv om denne reduksjonen ikke er like stor som det som er tilfelle for kvinnene, skyldes den trolig det samme. I aldersgruppen 26-50 år har dødeligheten blitt redusert omtrent like mye som for kvinnene i tilsvarende aldersgruppe. Derimot

Aldersgruppe	Kohort	Relativ risiko	RR nedre	RR øvre
1-25 år	Født år 1930-1934	1	-	-
	Født år 1935-1939	0.79	0.78	0.81
	Født år 1940-1945	0.59	0.58	0.60
	Født år 1946-1950	0.37	0.37	0.38
26-50 år	Født år 1930-1934	1	-	-
	Født år 1935-1939	0.95	0.93	0.97
	Født år 1940-1945	0.89	0.88	0.91
	Født år 1946-1950	0.81	0.80	0.83
51-76 år	Født år 1930-1934	1	-	-
	Født år 1935-1939	0.91	0.90	0.92
	Født år 1940-1945	0.86	0.85	0.86
	Født år 1946-1950	0.78	0.77	0.79

Tabell 3.5: Relativ risiko for kohort (kvinner) med 95% konfidensintervall, justert for alder og land

Aldersgruppe	Kohort	Relativ risiko	RR nedre	RR øvre
1-25 år	Født år 1930-1934	1	-	-
	Født år 1935-1939	0.83	0.82	0.85
	Født år 1940-1945	0.67	0.66	0.68
	Født år 1946-1950	0.48	0.47	0.49
26-50 år	Født år 1930-1934	1	-	-
	Født år 1935-1939	0.94	0.93	0.96
	Født år 1940-1945	0.87	0.86	0.89
	Født år 1946-1950	0.80	0.79	0.81
51-76 år	Født år 1930-1934	1	-	-
	Født år 1935-1939	0.87	0.86	0.88
	Født år 1940-1945	0.77	0.76	0.77
	Født år 1946-1950	0.69	0.68	0.69

Tabell 3.6: Relativ risiko for kohort (menn) med 95% konfidensintervall, justert for alder og land

har det i alderen 51-76 år skjedd en reduksjon fra årene 1930-1934 til 1946-1950 på hele 31%. Denne reduksjonen er en god del større enn det den var for kvinner. Det er fra før kjent at menns dødelighet i høye aldre har hatt en større reduksjon enn det som har vært tilfelle for kvinner de siste årene. Hovedårsaken til dette er mest sannsynlig reduksjonen av dødsfall som følge av hjerte-og karsykdommer, grunnet utviklingen i medisin og kunnskap i forhold til denne typen sykdommer.

Aldersgruppe	Kohort	Norge	Danmark	Sverige	Finland
1-25 år	Født år 1930-1934	1	0.84 (0.81, 0.88)	0.90 (0.86, 0.93)	2.13 (2.05, 2.22)
	Født år 1935-1939	0.86 (0.82, 0.90)	0.65 (0.62, 0.68)	0.67 (0.64, 0.70)	1.72 (1.66, 1.79)
	Født år 1940-1945	0.73 (0.70, 0.77)	0.54 (0.52, 0.57)	0.50 (0.48, 0.53)	1.18 (1.13, 1.23)
	Født år 1946-1950	0.47 (0.45, 0.50)	0.39 (0.37, 0.41)	0.39 (0.37, 0.40)	0.60 (0.58, 0.63)
26-50 år	Født år 1930-1934	1	1.60 (1.53, 1.67)	1.19 (1.14, 1.25)	1.29 (1.24, 1.35)
	Født år 1935-1939	1.00 (0.95, 1.06)	1.52 (1.46, 1.59)	1.14 (1.09, 1.19)	1.18 (1.13, 1.24)
	Født år 1940-1945	0.99 (0.95, 1.04)	1.43 (1.37, 1.49)	1.05 (1.01, 1.09)	1.10 (1.05, 1.15)
	Født år 1946-1950	0.91 (0.87, 0.95)	1.29 (1.24, 1.35)	0.94 (0.90, 0.98)	1.05 (0.01, 1.10)
51-76 år	Født år 1930-1934	1	1.59 (1.56, 1.62)	0.97 (0.95, 0.98)	0.93 (0.91, 0.95)
	Født år 1935-1939	0.92 (0.90, 0.95)	1.41 (1.38, 1.44)	0.90 (0.88, 0.91)	0.85 (0.83, 0.87)
	Født år 1940-1945	0.89 (0.86, 0.91)	1.25 (1.23, 1.28)	0.86 (0.84, 0.88)	0.82 (0.80, 0.84)
	Født år 1946-1950	0.83 (0.80, 0.86)	1.09 (1.06, 1.12)	0.76 (0.74, 0.78)	0.83 (0.80, 0.85)

Tabell 3.7: Relativ risiko interaksjon med 95% konfidensintervall for kvinner, aldersjustert

La oss videre se på tabellene for de relative risikoene som inkluderer interaksjon mellom land og kohort. Dette er vist i tabellene 3.7 og 3.8. Alle verdiene i tabellene er målt i forhold til verdiene for Norge og individer født i 1930-1934. Tabellen for kvinner viser at i aldersgruppen 1-25 år har dødeligheten i Norge blitt redusert med hele 53% fra kvinner født i årene 1930-1934 til kvinner som er født i årene 1946-1950. I den samme aldersgruppen har Sverige og Danmark en dødelighet som er lavere enn den blant norske kvinner i alle kohortene. Finland utmerker seg spesielt i årene 1930-1934 med en relativ risiko på hele 2.13. For finnene faller ikke den relative risikoen under én før i årene 1946-1950, men da faller den også godt under. Dette innebærer at dødeligheten blant finske kvinner i krigsårene var høyere enn det den var for norske kvinner i årene 1930-1934.

I aldersgruppen 26-50 år har dødeligheten for norske kvinner blitt redusert med nesten 10% fra kohort en til kohort fire. Man ser også at den største endringen har skjedd fra krigsårene til etterkrigsårene. Det ses også at helt fram til og med krigsårene er dødeligheten høyere i Danmark, Sverige og Finland enn det den er i

Aldersgruppe	Kohort	Norge	Danmark	Sverige	Finland
1-25 år	Født år 1930-1934	1	0.80 (0.77, 0.83)	0.84 (0.81, 0.87)	1.80 (1.75, 1.86)
	Født år 1935-1939	0.91 (0.87, 0.94)	0.64 (0.62, 0.67)	0.69 (0.67, 0.72)	1.48 (1.43, 1.53)
	Født år 1940-1945	0.82 (0.79, 0.85)	0.55 (0.53, 0.57)	0.57 (0.55, 0.59)	1.11 (1.07, 1.14)
	Født år 1946-1950	0.56 (0.54, 0.59)	0.44 (0.42, 0.46)	0.46 (0.44, 0.47)	0.68 (0.66, 0.70)
26-50 år	Født år 1930-1934	1	1.09 (1.06, 1.13)	1.01 (0.99, 1.05)	1.72 (1.67, 1.77)
	Født år 1935-1939	0.99 (0.96, 1.03)	1.07 (1.03, 1.10)	0.96 (0.94, 0.99)	1.55 (1.50, 1.59)
	Født år 1940-1945	0.90 (0.87, 0.93)	1.05 (1.01, 1.08)	0.88 (0.85, 0.90)	1.40 (1.36, 1.45)
	Født år 1946-1950	0.80 (0.77, 0.83)	0.99 (0.96, 1.02)	0.77 (0.75, 0.79)	1.34 (1.31, 1.38)
51-76 år	Født år 1930-1934	1	1.27 (1.25, 1.29)	0.91 (0.90, 0.93)	1.20 (1.18, 1.21)
	Født år 1935-1939	0.87 (0.86, 0.89)	1.08 (1.07, 1.10)	0.79 (0.78, 0.80)	1.07 (1.05, 1.08)
	Født år 1940-1945	0.73 (0.71, 0.74)	0.97 (0.96, 0.99)	0.68 (0.67, 0.69)	0.98 (0.96, 1.00)
	Født år 1946-1950	0.64 (0.62, 0.65)	0.83 (0.81, 0.85)	0.58 (0.56, 0.59)	0.98 (0.96, 1.00)

Tabell 3.8: Relativ risiko interaksjon med 95% konfidensintervall for menn, aldersjustert

Norge i årene 1930-1934. Spesielt utmerker danskene seg med en relativ risiko på 1.60 i 1930-1934, og fortsatt så mye som 1.29 i etterkrigsårene. I årene 1946-1950 har svenskene kommet seg mer ned på nivå med de norske kvinnene, og deres relative risiko har minket til 0.94.

I alderen 51-76 år er den totale nedgangen i dødelighet for norske kvinner 17%. Finnene har den samme relative risikoen som de norske kvinnene i årene 1946-1950. Disse hadde også en relativ risiko på 0.93 i årene 1930-1934, noe som da vil si at de ikke har hatt like stor nedgang som de norske. Det er også verdt å merke seg for finnenes vedkommende at den relative risikoen ikke har endret seg i noen særlig grad i årene etter 1935. Dette vil dermed si at dødeligheten blant finske kvinner ser ut til å ha stabilisert seg, og ikke reduseres noe særlig mer. Svenske kvinner er de som kommer best ut av det i etterkrigsårene med en relativ risiko på 0.76, men de lå også bedre an enn de norske kvinnene allerede i årene 1930-1934. Dårligst er det for de danske kvinnene i denne aldersgruppen, som selv i etterkrigsårene fortsatt har en dødelighet som er høyere enn det den var i Norge

i 1930-1934.

For menn i alderen 1-25 år har dødeligheten for de norske blitt redusert med 44% totalt. Dette er ganske mye, men likevel ikke like mye som reduksjonen var for de norske kvinnene. Også for menn i denne aldersgruppen har danskene og svenskene relativt lavere dødelighet enn de norske i årene 1930-1934. Finnene ligger derimot godt over de norske, akkurat som for kvinnene, med en relativ risiko på 1.80 i samme periode. I årene 1946-1950 er det danske og svenske menn som har de laveste relative risikoene, på henholdsvis 0.44 og 0.46. Da må det også tas i betraktning at disse startet ut lavere enn de norske. Finske menns dødelighet reduseres betydelig fra krigsårene og til årene etter krigen, der de ender opp med en relativ risiko på 0.68.

I aldersgruppen 26-50 år har dødeligheten for norske menn blitt redusert med 20% fra individer født i 1930-1934 til individer født i 1946-1950. Dette er en større nedgang enn det vi så hos kvinnene. I denne aldersgruppen har finnene ganske mye høyere dødelighet i alle tidsperiodene relativt til dødeligheten for norske menn født i årene 1930-1934, og den relative risikoen i årene 1946-1950 er på så mye som 1.34, noe som er en god del høyere enn det som var tilfelle for de finske kvinnene. Danske menn har også en dødelighet som er høyere enn for de norske, og det er ikke før i etterkrigsårene at deres dødelighet kommer på nivå med dødeligheten i Norge i årene 1930-1934. Svenskene ligger forholdsvis likt med de norske mennene i denne aldersgruppen, men de har en dødelighet som reduseres litt hurtigere enn den for de norske, og de ender i årene 1946-1950 med en relativ risiko på 0.77.

I alderen 51-76 år har det forekommet en reduksjon i dødeligheten på 36% for norske menn fra perioden 1930-1934 til årene 1946-1950. Svenske menn har den laveste dødeligheten i årene 1930-1934, som er nesten 10% lavere enn dødeligheten for norske menn i samme periode. I etterkrigsårene er det også svenskene som har den laveste relative risikoen, da på 0.58. I de første fem årene på 1930-tallet har danskene og finnene en dødelighet som er henholdsvis 27% og 20% høyere enn for norske menn. I de to neste kohortene har danskene og finnene mer eller mindre like relative risikoer, mens det endrer seg for årene etter andre verdenskrig. Da er det Finland som kommer dårligst ut med en relativ risiko på 0.98, mens danskene har redusert sin til 0.83. Generelt sett ser det dermed ut til at finske menn faktisk kommer dårligst ut i alle aldersgruppene.

Videre kan vi nå forsøke å koble disse resultatene til hypotesen om at moderat vanskelige ernæringsforhold i barnealder kan føre til en reduksjon i dødeligheten i voksen alder. Sverige, som var forventet å ha en lav dødelighet i krigsårene, viser seg også å ha det. Både for den yngste og eldste aldersgruppen har svenskene en lavere dødelighet relativt til den for nordmennene. Hos svenske menn har

dødeligheten også vært lavere enn dødeligheten i Norge i aldersgruppen 26-50 år, mens for kvinnene har den vært rimelig lik, og faktisk noe høyere. Dette ser man er noe som også er tilfelle i årene før og etter krigen. Danskene, som trolig ikke hadde det like ille som nordmennene under krigen på grunn av at de har et omfattende jordbruk, har lav dødelighet relativt til de norske i den yngste aldersgruppen, men fra 26-årsalderen kommer de dårligere ut. Norge har den laveste dødeligheten i den midtre aldersgruppen, mens det er Sverige, og også Finland hos kvinnene, som har lavest dødelighet i alderen 51-76 år. Det faktum at nordmenn hadde lavere dødelighet enn danskene i voksen alder skyldes trolig ikke en reduksjon i dødeligheten blant norske menn og kvinner som følge av vanskelige forhold under krigen, men muligens kan heller danskenes usunne kosthold og hyppige forbruk av tobakk[6] ha vært en sterk medvirkende årsak. En bakgrunn for denne tolkningen er at man ser det samme utfallet også i årene før og etter andre verdenskrig. Finnene, som hadde det hardt i krigsårene, har også en generelt høy dødelighet relativt til de andre landene, da særlig blant menn hvor de kommer dårligst ut i bortimot samtlige aldersgrupper og kohorter. Hos kvinnene kommer de derimot bedre ut enn danskene helt fra aldersgruppen over 25 år.

4 Generaliserte additive modeller

4.1 Hva er en generalisert additiv modell?

En generalisert additive modell(GAM) er en statistisk modell som setter sammen egenskapene til en generalisert lineær modell med egenskapene til en additiv modell. Fordelingen til responsen til en generalisert additiv modell skal være fra den eksponensielle familien, som her tilfredsstilles av Poisson fordelingen. Forventningen til responsen har tilknytning til forklaringsvariablene gjennom en linkfunksjon

$$g(\mu) = \beta_0 + s_1(x_1) + \dots + s_p(x_p)$$

der s_j -ene er glatte funksjoner av forklaringsvariablene x_j , $j = 1, \dots, p$.

En glattingsmetode som ofte er brukt for spredningsplot er *splines*. Hensikten med en slik glattingsmetode er å avsløre om det er en ikke-lineær tendens. En spline er en funksjon

$$s(t) : [a, b] \rightarrow \mathbb{R}$$

som består av stykkevise polynomer

$$P_i(t) : [t_i, t_{i+1}) \rightarrow \mathbf{R}$$

der $a = t_0 < t_1 < \dots < t_{k-2} < t_{k-1} = b$.

Dette gir da splinefunksjonen

$$s(t) = \begin{cases} P_0(t) & , t_0 \leq t < t_1 \\ P_1(t) & , t_1 \leq t < t_2 \\ \dots & , \dots \\ P_{k-2}(t) & , t_{k-2} \leq t < t_{k-1} \end{cases}$$

En glattingsoperator brukes for å oppsummere trenden til en responsvariabel som en funksjon av en eller flere forklaringsvariable.

4.2 Generaliserte additive modeller benyttet på datasettet

Ettersom kovariaten alder kan innta mange ulike varianter, er det fornuftig å redusere antall frihetsgrader som benyttes for denne i modelleringen av GAM for å få en mer glattet funksjon. Dermed oppstår det et spørsmål om hvor mange frihetsgrader det er fornuftig å bruke. De følgende resultatene er hentet fra Hastie et al.[5] Ved valg av glattingsparameter trenger vi ikke å prøve å minimere forventet kvadratfeil for hver x_i , men istedet se på gjennomsnittlig forventet kvadratfeil

$$MSE(\lambda) = \frac{1}{n} \sum_{i=1}^n E \left\{ \hat{f}_\lambda(x_i) - f(x_i) \right\}^2$$

der λ er glattingsparameter, og \hat{f}_λ er den tilpassede funksjonen.

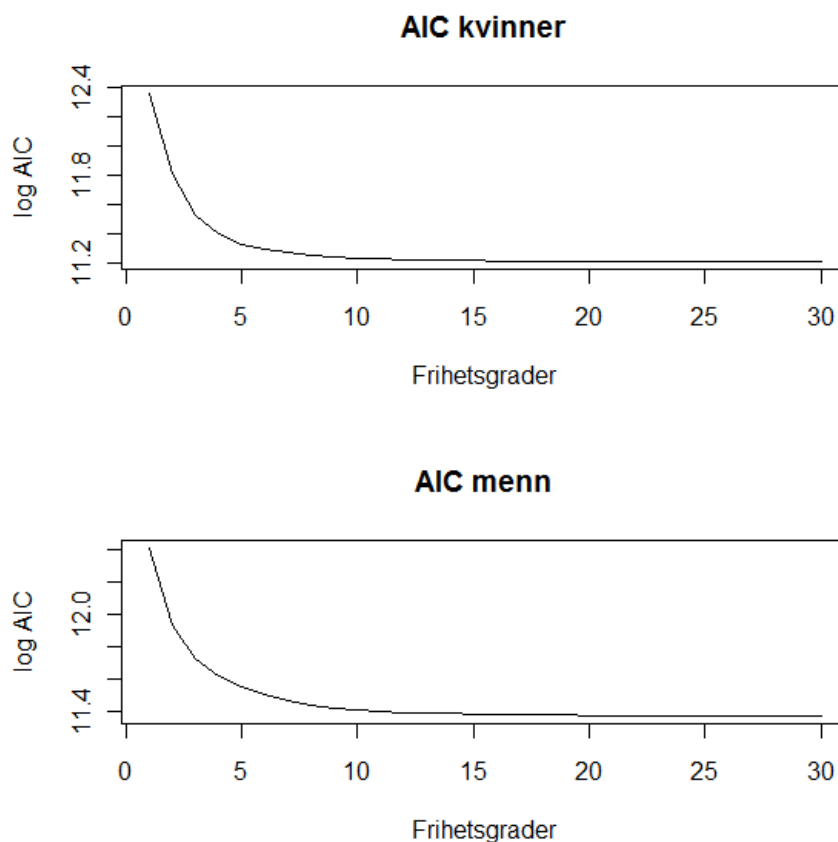
For å avgjøre hvilken glattingsparameter som bør benyttes kan man bruke kryss-validering. Kryss-validering fungerer ved å utelate punkter (x_i, y_i) , ett av gangen, og estimere glattingen ved x_i basert på de gjenstående $n - 1$ punktene. Man får da kryss-valideringskvadratsummen

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left\{ y_i - \hat{f}_\lambda^{-i}(x_i) \right\}^2$$

der $\hat{f}_\lambda^{-i}(x_i)$ indikerer tilpasningen ved x_i , beregnet ved å utelate datapunkt nummer i . $CV(\lambda)$ blir da beregnet for flere verdier av glattingsparameteren λ over

et passende spenn og deretter velges den minimerende λ . Dersom vi nå sier at $\hat{\mu}_i^{-i}$ er den tilpassede verdien i det i -te punktet, oppnådd ved å utelate det i -te punktet i utvalget, så er den kryss-validerte deviansen gitt ved

$$CV = \frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_i^{-i})$$



Figur 4.1: Utviklingen i AIC som følge av endring i frihetsgrader for alder

I tilfellet med én glatting finnes det ulike tilnærminger som er enklere å implementere. Et alternativ er å bruke AIC (Akaike Information Criterion), som kan defineres som

$$AIC = \frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_i) + \frac{2\phi}{n} df \quad (4.1)$$

der df er antall frihetsgrader og ϕ er spredningsparameteren. I tilfellet for Poisson

fordelingen er $\phi = 1$. Dette gir da at for våre data blir

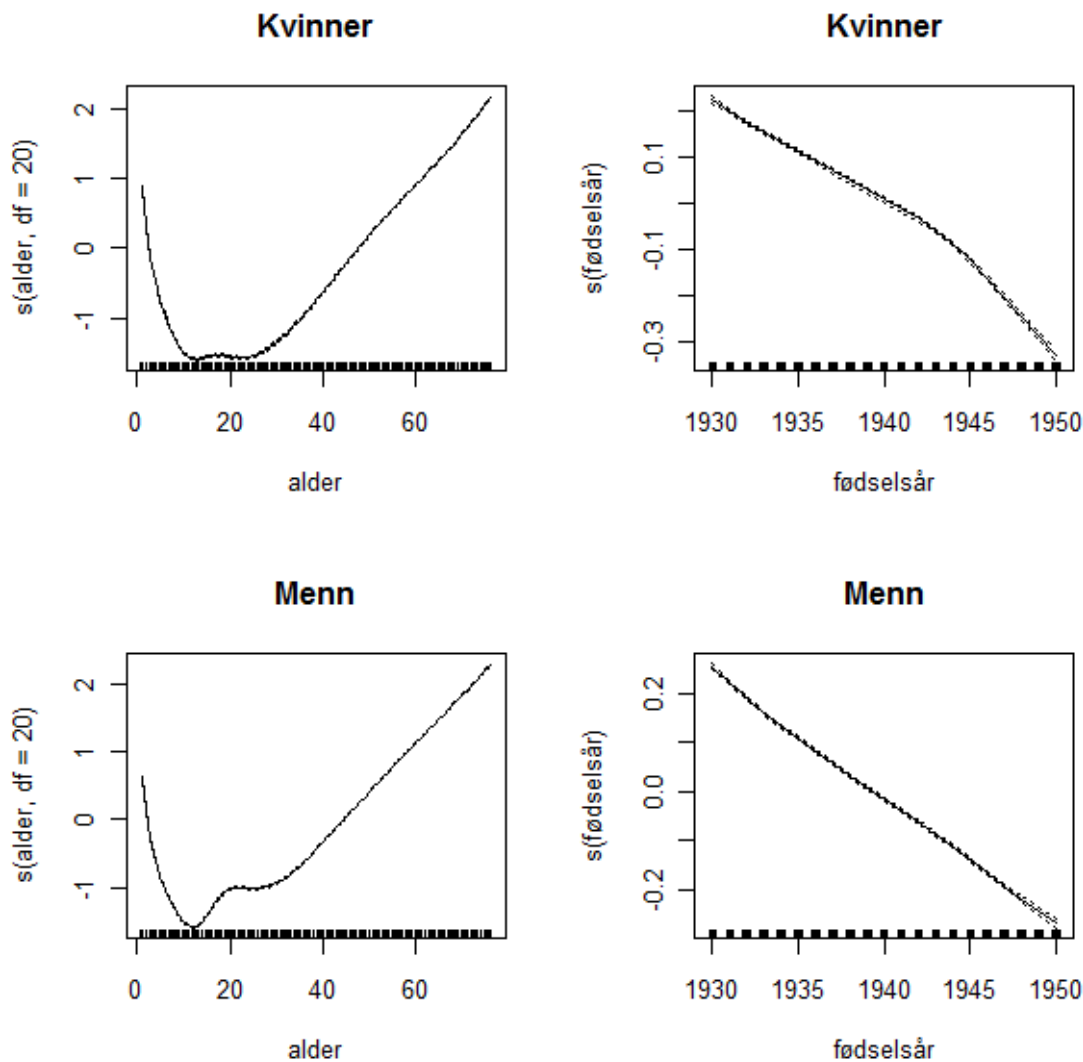
$$AIC = \frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_i) + \frac{2}{n} df \quad (4.2)$$

Deretter velger man det antallet frihetsgrader som minimerer denne. For valg av frihetsgrader til modelleringen av GAM vil det holde å se på hvor mange frihetsgrader som er nødvendig for at AIC skal stabilisere seg, det vil si at den ikke endrer seg noe særlig om frihetsgradene justeres litt opp eller ned. Ved modellering av GAM i programvaren R er det mulig å få ut AIC-verdien direkte. Dette er da gjort på datasettet for ulike frihetsgrader, og det viser seg at AIC stabiliserer seg mellom 15 og 20 frihetsgrader som vist i figur 4.1. Ut i fra dette har valget falt på å bruke antall frihetsgrader lik 20.

Ettersom alle landene har en forholdsvis lik utvikling i dødeligheten både i forhold til alder og kohort, er det rimelig å se på alle landene under ett. Dette er da vist i figur 4.2. Figuren viser at kvinnene har en trend i dødeligheten der dødeligheten er avtagende fra fødsel og fram til omtrent 10-15-årsalderen. Deretter holder den seg lav fram til rundt 25-årsalderen, med en liten midlertidig økning i dødeligheten i alderen 15-25 år, for så kun å øke etter dette. Det at det er to bunnpunkter skyldes at danske og svenske kvinner når sitt bunnpunkt i dødeligheten ved 10-15-årsalderen, mens i Norge og Finland har kvinnene lavest dødelighet omtrent ved alder 25 år. En annen medvirkende årsak til denne toppen er trolig ulykkesdødelighet, som ofte forekommer blant individer i denne alderen.

For menn derimot utvikler det seg noe annerledes. De har en nedgang i dødeligheten fra fødselen og fram til 10-15-årsalderen, der de når en bunn. Deretter øker dødeligheten ganske kraftig igjen. Det er verdt å merke seg at de har en "dump" i 20- til 30-35-årsalderen der økningen i dødeligheten stopper litt opp, for så å fortsette igjen etter dette. Det at mennene har en økning i dødeligheten tidligere enn kvinnene, og at de har denne dumpen, skyldes trolig ulykkesdødelighet, slik som for eksempel ved bilkjøring. Gutter er ofte mer utagerende enn jenter i denne alderen, og dør dermed dessverre også oftere i ulykker.

Vi ser også at det har vært en bratt nedgang i dødeligheten både for menn og kvinner i løpet av årene 1930 til 1950, selv om menns dødelighet har blitt redusert noe raskere. Samtidig ser det ut til at reduksjonen i dødelighet blant kvinner var større for kvinner født i årene etter andre verdenskrig. Ved å lese av plottene ser man at det har vært en total reduksjon på omtrent 0.4-0.5. Dette tilsier en relativ risiko på mellom $e^{-0.45} = 0.64$ og $e^{-0.5} = 0.61$, altså en forbedring i den totale dødeligheten på rundt 35-40% for individer født i år 1930 til individer født i 1950.



Figur 4.2: Generaliserte additive modeller for alder og fødselsår for alle landene sett under ett

Selv om den totale endringen i dødeligheten for alle landene sett under ett gir et godt bilde av den generelle utviklingen i dødeligheten, kan det også være av interesse å se på noen forskjeller blant landene. GAM-plottene for de ulike landene er å finne i vedlegget, figur 6.4, 6.5, 6.6 og 6.7. Dødelighetstrenden for alder blant menn ser ut til å være rimelig lik for alle landene, bortsett fra finnene som skiller seg litt ut. Hos dem er nemlig ikke "dumpen" i 20-30-årene like markant som hos de andre, og de har i større grad en log-lineær trend i dødeligheten fra 15-årsalderen. Det er også verdt å merke seg at når dødeligheten blant menn er på

et bunnivå i ti-årsalderen er det danske menn som har den laveste dødeligheten. For kvinnene er det litt større forskjeller i dødelighetstrenden for alder. Danske og svenske kvinner har en rimelig lik trendutvikling, og det samme har norske og finske kvinner. De er like i den forstand at danskene og svenskene har lavest dødelighet i alder omtrent ti år, mens norske og finske kvinner når bunnen først i 20-25-årsalderen. Ved nådd bunnivå er det norske og danske kvinner som har den laveste dødeligheten.

Plottene som viser utviklingen i dødeligheten per år for individer født i årene fra 1930 til 1950 viser at norske kvinner har hatt den slakeste nedgangen i dødelighet blant kvinnene, men i etterkrigsårene har den blitt redusert kraftigere enn tidligere. Også de svenske kvinnene har hatt en større reduksjon i årene etter krigen enn i de tidligere årene. Finske kvinner har hatt en kraftig reduksjon i dødeligheten gjennom alle årene, og har totalt hatt, grunnet en relativ risiko på mellom $e^{-0.6} = 0.55$ og $e^{-0.7} = 0.50$, en reduksjon på så mye som 45 – 50%. Mer eller mindre det samme er tilfelle for finske menn, men deres reduksjon er nærmere 45%. Spesielt for disse er også at de ser ut til å ha hatt den kraftigste reduksjonen hos individer født i årene 1930-1934. Norske menn har hatt sin kraftigste nedgang i dødelighet i årene som fulgte etter andre verdenskrig. De har også hatt en større nedgang enn de norske kvinnene. For Danmark og Finland kommer det også tydelig fram av plottene at kvinnene kommer ut med lavere dødelighet enn menn i år 1950. I tillegg ser man at generelt sett har menn hatt en hyppigere nedgang i dødeligheten enn det kvinner har, noe som også kommer fram av figur 4.2.

Det er verdt å merke seg at GAM-plottene også inneholder konfidensintervaller, selv om de ikke er så enkle å se. De smale konfidensintervallene tilsier at kurven er en god tilpasning. Det viser seg også at for de siste fødselsårene som er tatt med er konfidensintervallet litt videre, noe som indikerer større usikkerhet her. Dette er tilfelle både for figur 4.2 og for figurene av GAM-plottene for hvert enkelt av landene. I tillegg ses det av figurene for hvert av landene, figur 6.4, 6.5, 6.6 og 6.7, at de har større konfidensintervall enn figur 4.2 som viser alle landene sett under ett. Etersom GAM-plottet i figur 4.2 baserer seg på mer data vil det gi sikrere resultater, og dermed er konfidensintervallene smalere der.

5 Konklusjoner

Målet var å undersøke hypotesen om at dødelighet i voksen alder avhenger av ernæringsforholdene individene levde under i barnealder. Bakgrunnen for hypotesen er at dersom ernæringsforholdene var svært dårlige ventes det høy dødelighet, men det har også vært spekulert i om moderat vanskelige forhold faktisk kan føre til en reduksjon i dødeligheten. Ved å ta utgangspunkt i dødelighetsratene i årene før, under og etter 2. verdenskrig i de fire nordiske landene Norge, Danmark, Sverige og Finland har resultatene indikert at det ikke er direkte grunnlag for å tro på hypotesen. Alle landene ser ut til å ha hatt en rimelig lik utvikling i dødeligheten gjennom de fire tidsperiodene vi har sett på her, men i løpet av årene har forskjellene landene i mellom minket. For individer som var født i krigsårene 1940-1945 var dødeligheten i landene som følger:

- Lav dødelighet i Sverige og Norge.
- Lav dødelighet i Danmark i barnealder og høy dødelighet i voksen alder.
- Omvendt i Finland, med høy dødelighet fram til 40-årsalderen for kvinner og 60-årsalderen for menn, og deretter lav dødelighet. Blant finnene var det ventet høy dødelighet ettersom de trolig hadde de hardeste ernæringsforholdene under krigen, og det kan derfor sies å være noe overraskende at det ikke er tilfelle i høye aldere, da spesielt for kvinner.

La oss sette disse resultatene i sammenheng med hypotesen

- Den lave dødeligheten i Norge, som kan sies å være et land som hadde moderat vanskelige forhold under krigen sammenlignet med de andre landene, kan indikere at det kan ligge noe i hypotesen.
- Fra rundt 20-årsalderen hos kvinner og omkring 40-årsalderen hos menn er dødeligheten i Norge lavere enn hos danskene. Ettersom danskene trolig hadde bedre ernæringsforhold under krigen, kan dette muligens sies å støtte oppunder hypotesen.
- Samtidig har Sverige, som holdt seg nøytral under krigen og dermed trolig hadde de beste ernæringsforholdene av de fire landene, stort sett lavere dødelighet enn Norge. Dette kan dermed på den annen side tilsa at hypotesen om at moderat vanskelige ernæringsforhold kan føre til større reduksjon i dødeligheten enn det gode ernæringsforhold gjør, er gal.

- I tillegg er dødelighetsutviklingen som følge av aldring rimelig lik i hvert av landene både i årene før, under og etter krigen. Det viser seg også at forholdet mellom dødelighetene landene i mellom i årene 1940-1945 også har vært tilfelle i årene før og etter andre verdenskrig, selv om differansen har minnet med årene.

Dermed kan man trolig konkludere med at Norges lave dødelighet ikke var spesielt for individer født i, eller like før, krigsårene, men en generelt sett lav dødelighetstrend som man kan se for personer født både før og etter krigen. Den lave dødeligheten for svenskene er også en trend som man kan se gjennom alle tidsperiodene. Det samme gjelder for finnene med den høye dødeligheten fram til henholdsvis 40 og 60 år for kvinner og menn etterfulgt av en nedgang i dødeligheten i forhold til de andre landene. Den voldsomme endringen i danskernes dødelighet, som endrer seg fra å ligge lavest i barnealder til å bli høyest fra omtrent alder 40 år for kvinner og 60 år for menn, er i stor grad et resultat av danskernes usunne kosthold i form av mye fet mat og alkohol, i tillegg til røyking[6]. Dette gjelder hovedsakelig personer i voksen alder og forventes å kunne føre til en økning i dødeligheten. Dermed kan det være en rimelig konklusjon å hevde at som følge av en forholdsvis lik utvikling i dødeligheten i landene i hver av periodene, er det i større grad en trend som har dominert dødeligheten i landene framfor store forskjeller i ernæringsforhold og -situasjoner.

Det var også av interesse å se på forskjellene og endringene i dødeligheten i ulike aldersgrupper, og ikke bare generelt. Aldersgruppene ble inndelt i en ung gruppe for individer fra 1-25 år, en middelaldrende gruppe fra 26-50 år og en eldre gruppe på 51-76 år. Resultatene viser at

- Det har vært størst nedgang i dødeligheten i den yngste aldersgruppen, hovedsakelig på grunn av den kraftige reduksjonen i barnedødelighet.
- Nedgangen har vært minst i den midtre aldersgruppen.
- Både Sverige og Danmark har hatt lavere dødelighet enn Norge i aldersgruppen 1-25 år, mens Finland har hatt adskillig høyere.
- I den midtre aldersgruppen er det danskene som kommer dårligst ut blant kvinnene, mens blant menn er det finnene. I denne aldersgruppen er det de norske innbyggerne som har den laveste dødeligheten, tett etterfulgt av svenskene hos mennene.
- I den eldste aldersgruppen er det svenske og deretter norske menn som har den laveste dødeligheten, mens danskene og finnene ligger et godt stykke

over. For kvinnene i denne aldersgruppen kommer også de norske og svenske godt ut, men her har også finske kvinner mer eller mindre like lav dødelighet.

- Disse resultatene viser at svenskene har lav dødelighet i alle aldersgruppene for både kvinner og menn.

Danskene har høyere dødelighet enn nordmennene i de to eldre aldersgruppene, og dette kunne tenkes å være et understøttende argument for hypotesen ettersom danskene trolig hadde bedre ernæringstilstander med sitt omfattende jordbruk enn det nordmennene hadde. Likevel er det nok fornuftig å si at argumentet med at dette har vært tilfelle også både før og etter krigen veier tyngre, og at det derfor ikke er grunnlag for å ha tiltro til hypotesen. I tillegg strider det faktum at svenskene hovedsakelig har lavere dødelighet enn nordmennene imot hypotesen.

En del av oppgaven var også å undersøke hvor mye større datasettet burde være for å påvise relevante sammenhenger dersom disse ikke skulle komme fram i det benyttede datasettet. Metodene og beregningene som er utført bygger på mye data, og selv om konklusjonen ut ifra disse tilsier at det ikke er noen klare sammenhenger mellom ernæringsforhold i barnealder og dødelighet, er det ikke min oppfatning at en utvidelse av datasettet ville endret konklusjonen. Det hadde vært større grunnlag for å si at datasettet ikke var stort nok dersom det hadde vært større tvil rundt konklusjonen, men jeg syns det kommer rimelig tydelig fram at dødeligheten i de ulike landene i mye større grad avhenger av trend og den generelle dødelighetsutviklingen i de ulike landene enn av ernæringsforholdene for individene.

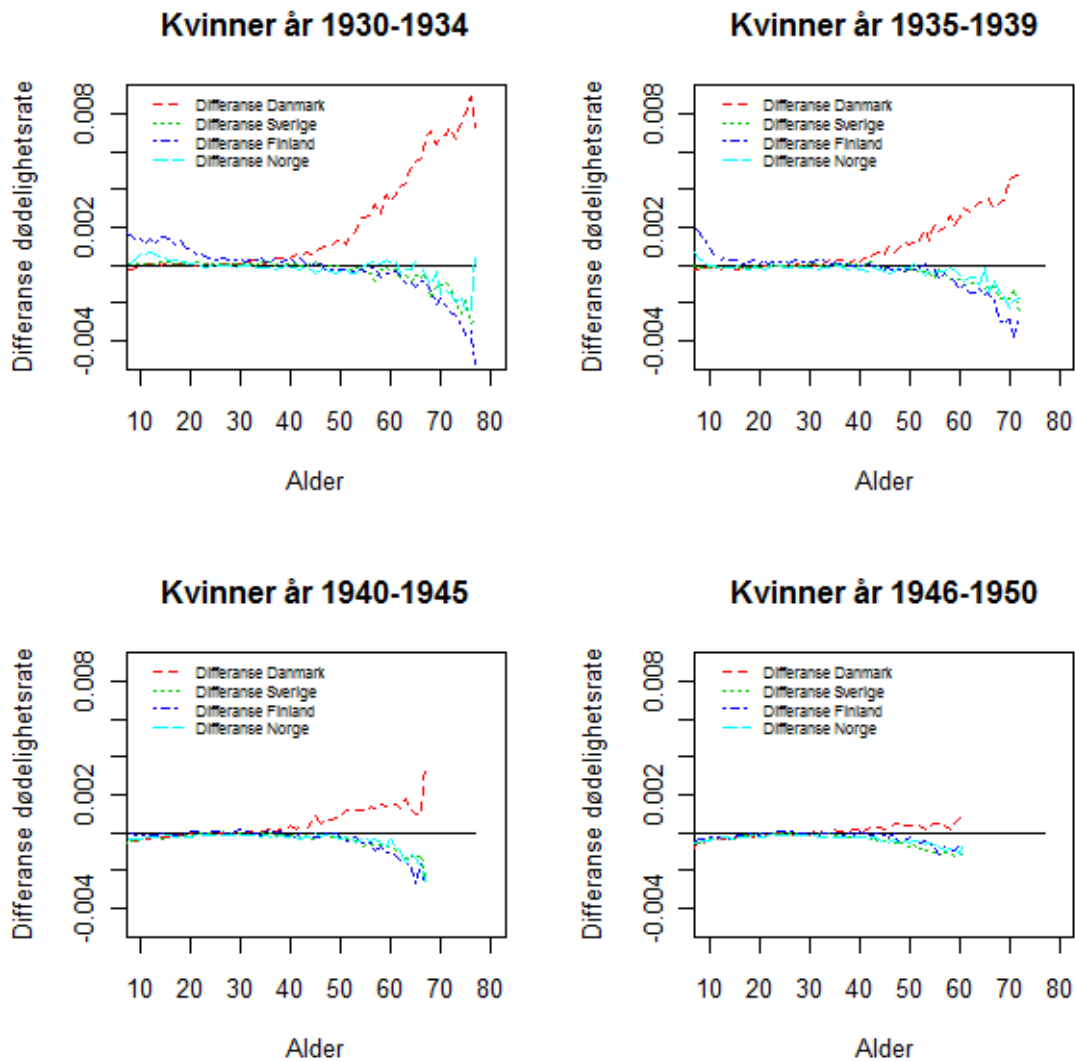
Samtidig kan det muligens argumenteres for at til tross for at det finnes forskjeller landene i mellom er de likevel litt for like, både i forhold til klima, økonomi, politikk og livsstil, slik at dette kan være årsaken til at hypotesen ikke understøttes. Dette er det vanskelig å si noe om her, ettersom det ikke er sett på land med større ulikheter.

I tillegg kan endringer i befolkningen som følge av innvandring og emigrering forårsake problemer slik at datasettet ikke blir like optimalt som man skulle ønske. Til tross for at innvandring og utflytting ikke var like utbredt da som det er i dag, vet vi at det også forekom i disse årene, både utenfor og innen de nordiske landene. Et annet mulig problem kan være i forbindelse med tidspunktet dataene ble registrert. For eksempel er det usikkert om befolkningsantallet er blitt registrert ved årets begynnelse eller slutt, og dette vil også ha innvirkning på resultatene.

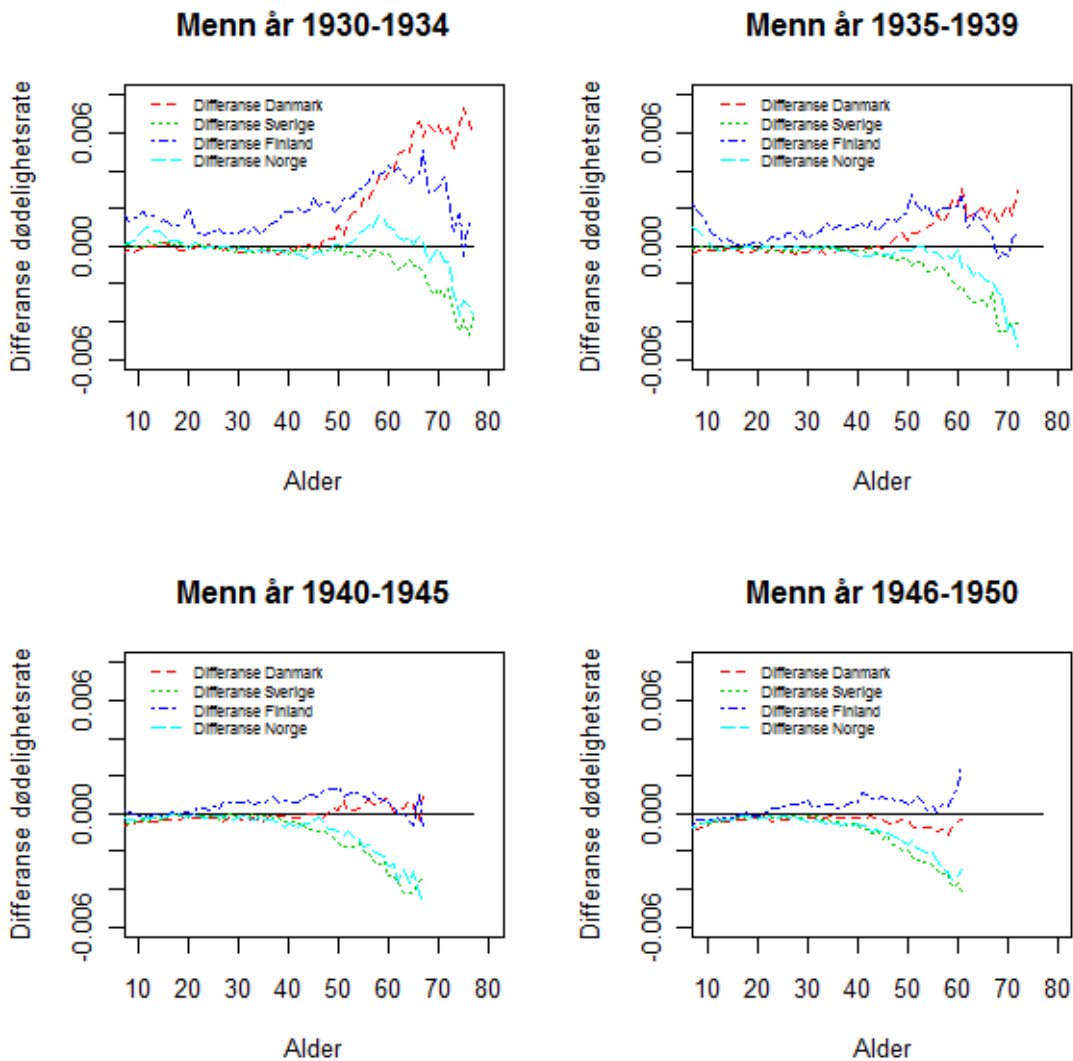
6 Vedlegg

6.1 Resultater

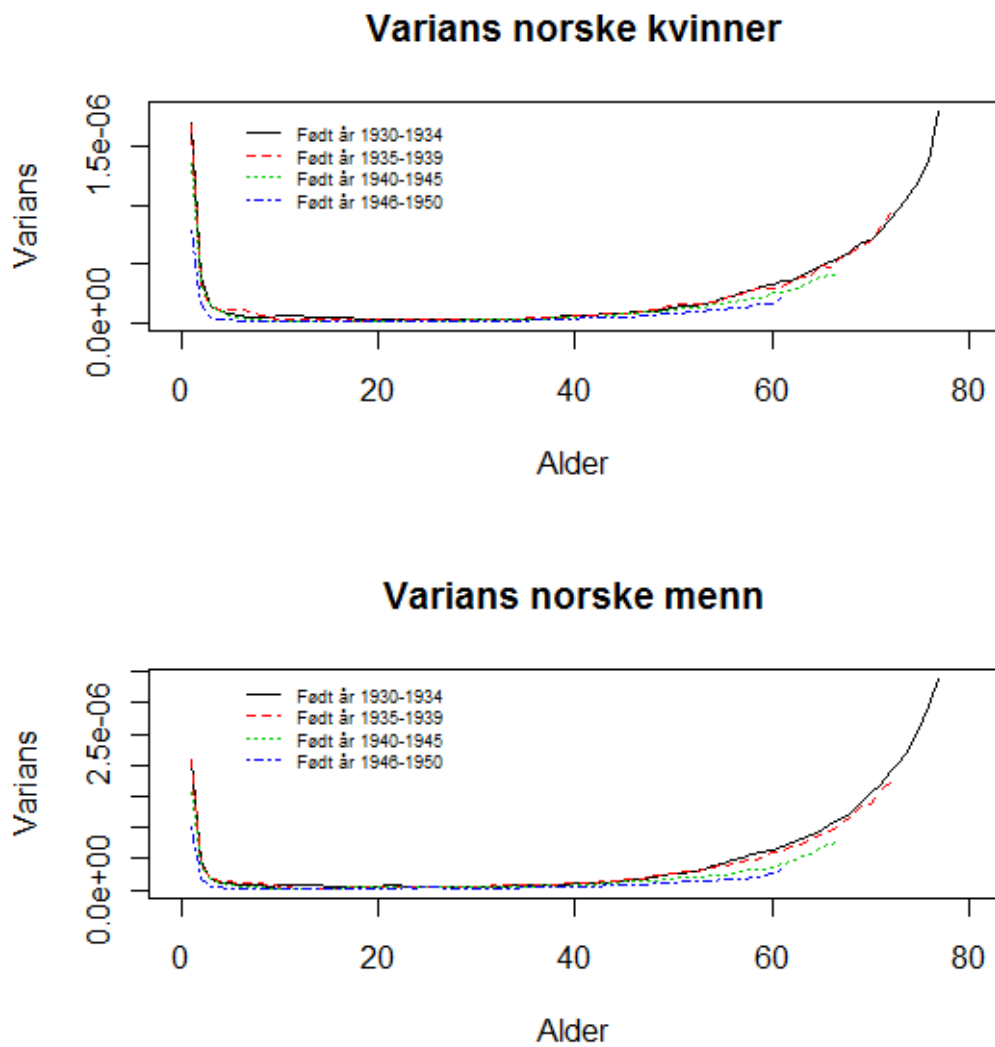
6.1.1 Figurer



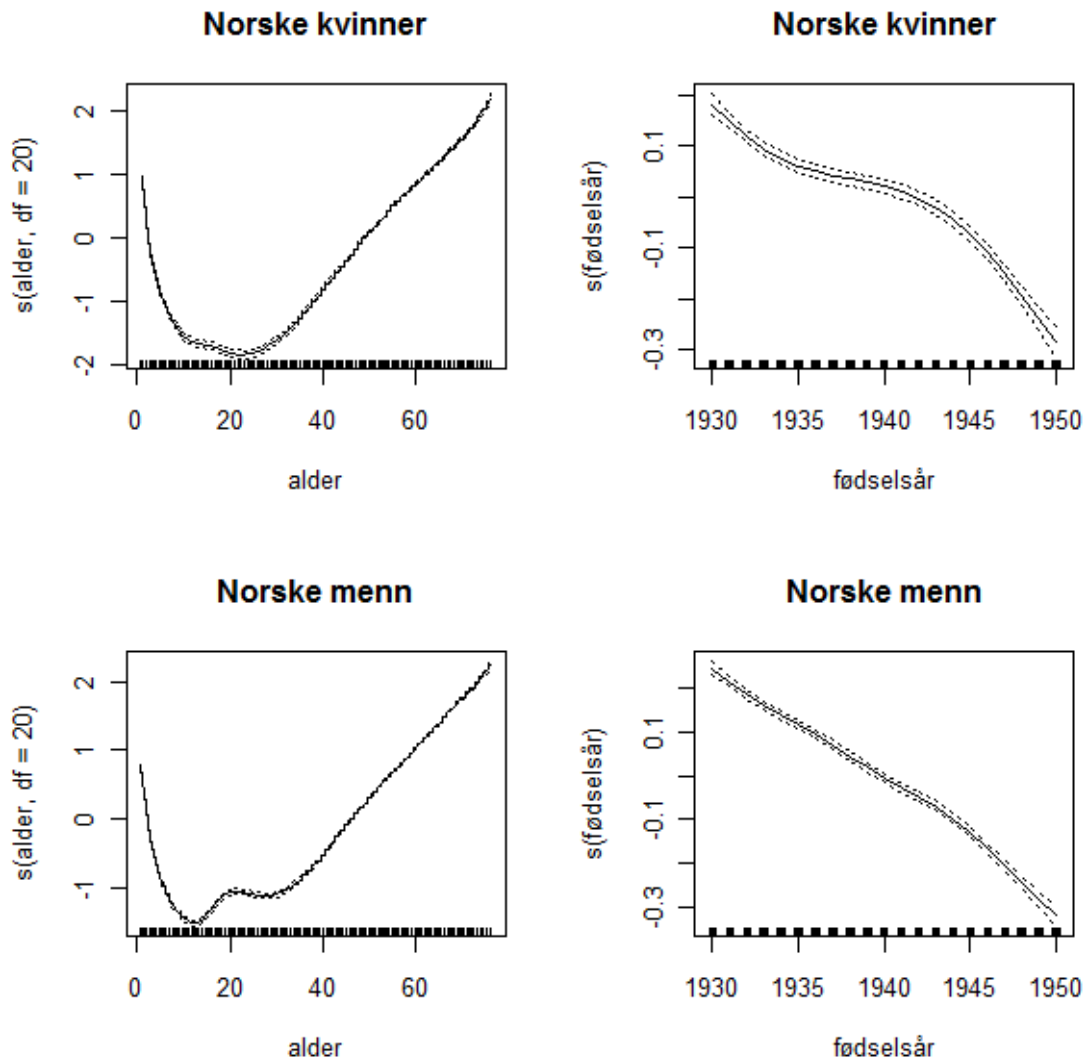
Figur 6.1: Differansen mellom dødelighetsrater for kvinner og totalt gjennomsnittlige dødelighetsrater for kvinner



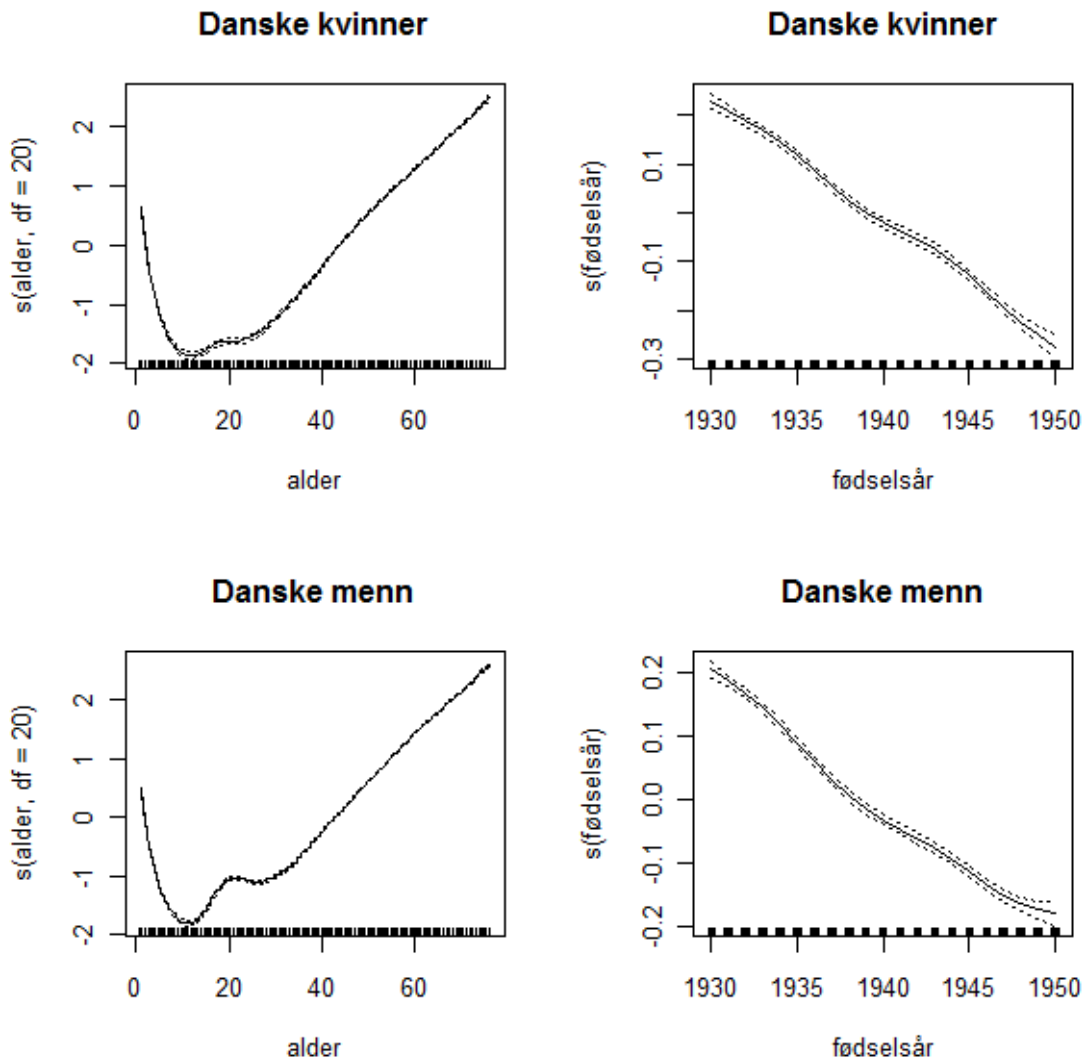
Figur 6.2: Differansen mellom dødelighetsrater for menn og totalt gjennomsnittlige dødelighetsrater for menn



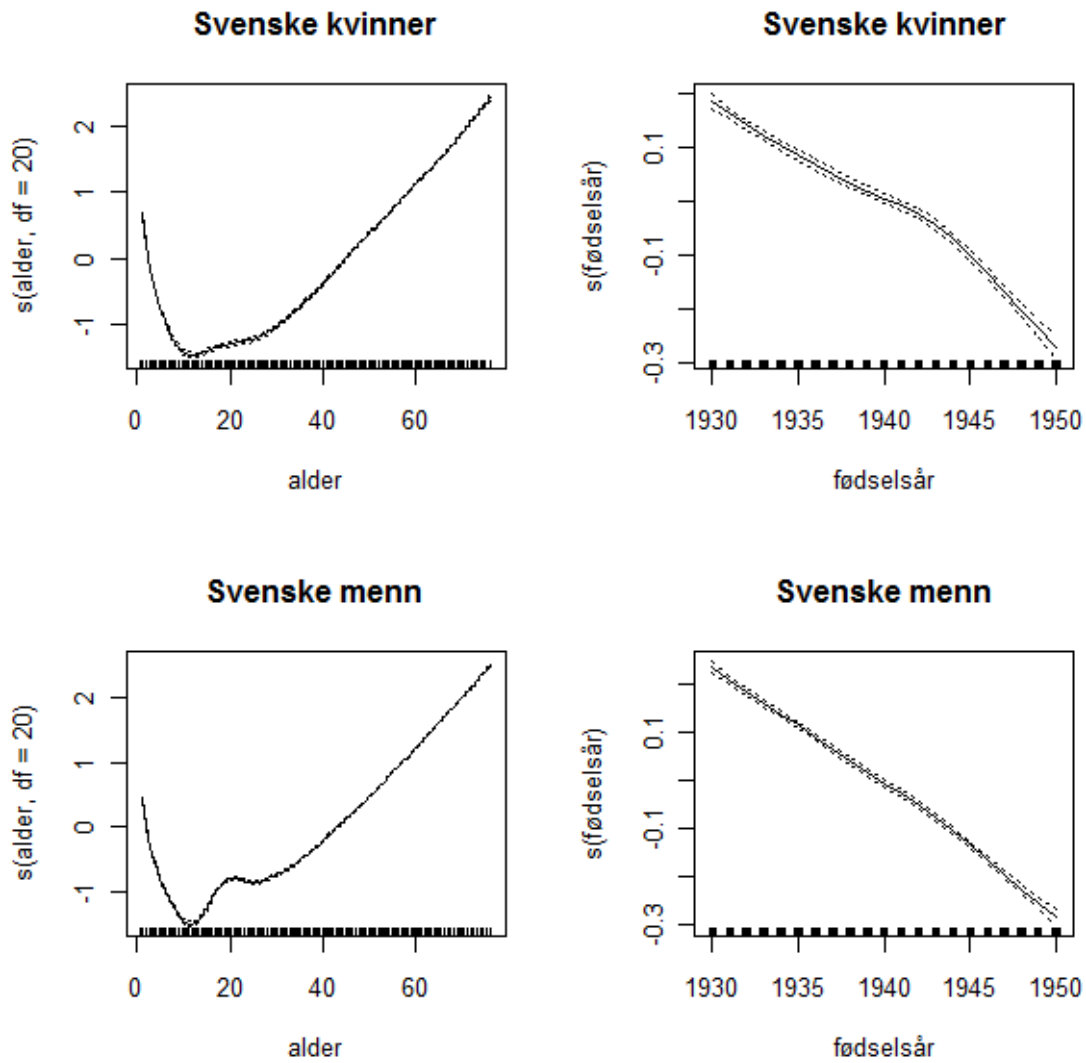
Figur 6.3: Variansen til de estimerte dødelighetsratene for norske kvinner og menn



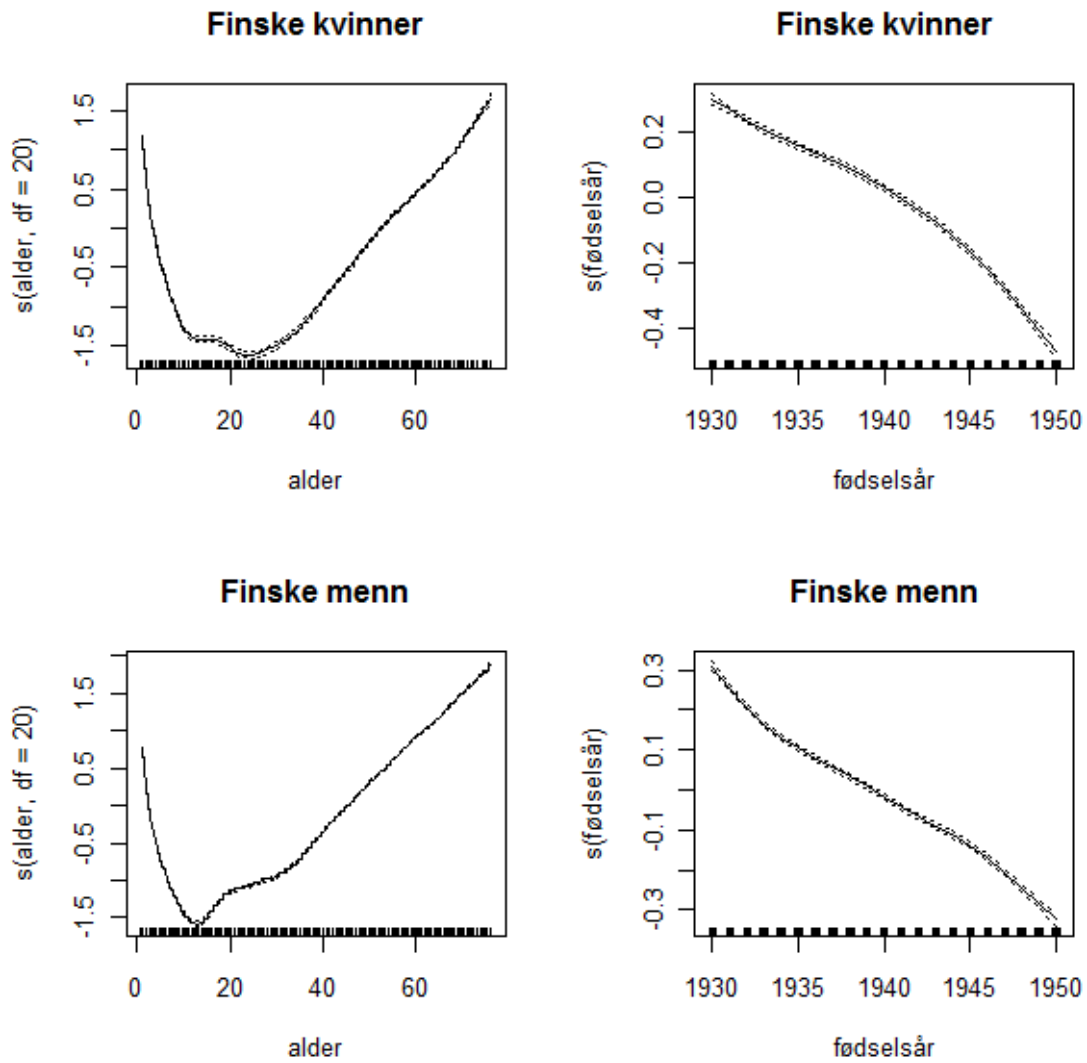
Figur 6.4: Generaliserte additive modeller for alder og fødselsår for Norge



Figur 6.5: Generaliserte additive modeller for alder og fødselsår for Danmark



Figur 6.6: Generaliserte additive modeller for alder og fødselsår for Sverige



Figur 6.7: Generaliserte additive modeller for alder og fødselsår for Finland

6.1.2 Tabeller

Alder kvinner	Norge	Danmark	Sverige	Finland
0	0.0366	0.0662	0.0456	0.0690
10	0.0012	0.0006	0.0007	0.0018
20	0.0006	0.0005	0.0006	0.0011
30	0.0006	0.0008	0.0007	0.0009
40	0.0012	0.0020	0.0014	0.0015
50	0.0028	0.0040	0.0027	0.0027
60	0.0067	0.0103	0.0063	0.0062
70	0.0137	0.0219	0.0141	0.0129
Alder menn	Norge	Danmark	Sverige	Finland
0	0.0469	0.0876	0.0603	0.0839
10	0.0018	0.0008	0.0009	0.0027
20	0.0015	0.0011	0.0014	0.0022
30	0.0013	0.0012	0.0011	0.0023
40	0.0023	0.0024	0.0024	0.0044
50	0.0056	0.0063	0.0051	0.0081
60	0.0127	0.0160	0.0110	0.0159
70	0.0258	0.0325	0.0240	0.0303

Tabell 6.1: Dødelighetsrater født år 1930-1934

Alder kvinner	Norge	Danmark	Sverige	Finland
0	0.0343	0.0583	0.0394	0.0649
10	0.0006	0.0006	0.0005	0.0012
20	0.0003	0.0005	0.0004	0.0008
30	0.0005	0.0006	0.0007	0.0008
40	0.0011	0.0017	0.0013	0.0014
50	0.0031	0.0042	0.0029	0.0027
60	0.0057	0.0096	0.0056	0.0056
70	0.0132	0.0197	0.0138	0.0113
Alder menn	Norge	Danmark	Sverige	Finland
0	0.0446	0.0772	0.0525	0.0801
10	0.0009	0.0006	0.0006	0.0016
20	0.0013	0.0011	0.0012	0.0014
30	0.0013	0.0013	0.0014	0.0020
40	0.0021	0.0025	0.0024	0.0038
50	0.0055	0.0063	0.0046	0.0084
60	0.0107	0.0149	0.0099	0.0145
70	0.0225	0.0282	0.0226	0.0271

Tabell 6.2: Dødelighetsrater født år 1935-1939

Alder kvinner	Norge	Danmark	Sverige	Finland
0	0.0334	0.0457	0.0304	0.0637
10	0.0003	0.0003	0.0004	0.0005
20	0.0004	0.0004	0.0004	0.0006
30	0.0005	0.0006	0.0007	0.0007
40	0.0011	0.0015	0.0011	0.0011
50	0.0028	0.0042	0.0026	0.0026
60	0.0058	0.0081	0.0056	0.0053
Alder menn	Norge	Danmark	Sverige	Finland
0	0.0409	0.0609	0.0400	0.0789
10	0.0006	0.0005	0.0005	0.0007
20	0.0013	0.0011	0.0012	0.0012
30	0.0013	0.0012	0.0014	0.0020
40	0.0020	0.0024	0.0022	0.0032
50	0.0046	0.0064	0.0039	0.0066
60	0.0092	0.0120	0.0086	0.0122

Tabell 6.3: Dødelighetsrater født år 1940-1945

Alder kvinner	Norge	Danmark	Sverige	Finland
0	0.0246	0.0318	0.0205	0.0478
10	0.0003	0.0002	0.0003	0.0004
20	0.0003	0.0004	0.0004	0.0005
30	0.0004	0.0007	0.0005	0.0005
40	0.0011	0.0014	0.0011	0.0012
50	0.0025	0.0034	0.0021	0.0025
60	0.0060	0.0070	0.0054	0.0060
Alder menn	Norge	Danmark	Sverige	Finland
0	0.0331	0.0422	0.0268	0.0603
10	0.0005	0.0004	0.0004	0.0006
20	0.0011	0.0011	0.0012	0.0012
30	0.0010	0.0012	0.0012	0.0018
40	0.0019	0.0025	0.0018	0.0037
50	0.0043	0.0050	0.0034	0.0063
60	0.0089	0.0106	0.0077	0.0145

Tabell 6.4: Dødelighetsrater født år 1946-1950

Alder	Fødselsår				
	Kvinner	1930-1934	1935-1939	1940-1945	1946-1950
0		$1.70 \cdot 10^{-6}$	$1.69 \cdot 10^{-6}$	$1.35 \cdot 10^{-6}$	$7.87 \cdot 10^{-7}$
10		$5.61 \cdot 10^{-8}$	$2.76 \cdot 10^{-8}$	$1.33 \cdot 10^{-8}$	$8.40 \cdot 10^{-9}$
20		$2.94 \cdot 10^{-8}$	$1.56 \cdot 10^{-8}$	$1.48 \cdot 10^{-8}$	$1.07 \cdot 10^{-8}$
30		$2.79 \cdot 10^{-8}$	$2.35 \cdot 10^{-8}$	$2.12 \cdot 10^{-8}$	$1.46 \cdot 10^{-8}$
40		$5.70 \cdot 10^{-8}$	$5.47 \cdot 10^{-8}$	$4.51 \cdot 10^{-8}$	$3.64 \cdot 10^{-8}$
50		$1.38 \cdot 10^{-7}$	$1.60 \cdot 10^{-7}$	$1.21 \cdot 10^{-7}$	$8.37 \cdot 10^{-8}$
60		$3.52 \cdot 10^{-7}$	$3.12 \cdot 10^{-7}$	$2.62 \cdot 10^{-7}$	$2.16 \cdot 10^{-7}$
Menn	1930-1934	1935-1939	1940-1945	1946-1950	
0		$2.07 \cdot 10^{-6}$	$2.09 \cdot 10^{-6}$	$1.56 \cdot 10^{-6}$	$1.00 \cdot 10^{-6}$
10		$7.92 \cdot 10^{-8}$	$4.16 \cdot 10^{-8}$	$2.55 \cdot 10^{-8}$	$1.53 \cdot 10^{-8}$
20		$6.75 \cdot 10^{-8}$	$6.15 \cdot 10^{-8}$	$5.13 \cdot 10^{-8}$	$3.39 \cdot 10^{-8}$
30		$6.20 \cdot 10^{-8}$	$6.50 \cdot 10^{-8}$	$5.27 \cdot 10^{-8}$	$3.18 \cdot 10^{-8}$
40		$1.09 \cdot 10^{-7}$	$1.06 \cdot 10^{-7}$	$8.06 \cdot 10^{-8}$	$6.00 \cdot 10^{-8}$
50		$2.78 \cdot 10^{-7}$	$2.84 \cdot 10^{-7}$	$1.91 \cdot 10^{-7}$	$1.34 \cdot 10^{-7}$
60		$6.91 \cdot 10^{-7}$	$6.06 \cdot 10^{-7}$	$4.14 \cdot 10^{-7}$	$3.14 \cdot 10^{-7}$

Tabell 6.5: Varians til de estimerte dødelighetsratene for norske kvinner og menn opp til alder 60 år

Frihetsgrader	Kvinner	Menn
1	12.36054	12.41601
2	11.81391	11.93789
3	11.53275	11.72274
4	11.39890	11.61802
5	11.33368	11.55248
6	11.29649	11.50330
7	11.27225	11.46547
8	11.25557	11.43758
9	11.24398	11.41812
10	11.23587	11.40490
11	11.23023	11.39613
12	11.22627	11.39021
13	11.22340	11.38601
14	11.22127	11.38289
15	11.21963	11.38040
16	11.21831	11.37832
17	11.21723	11.37652
18	11.21632	11.37493
19	11.21553	11.37353
20	11.21485	11.37230
21	11.21427	11.37120
22	11.21376	11.37026
23	11.21332	11.36943
24	11.21294	11.36873
25	11.21261	11.36812
26	11.21233	11.36760
27	11.21209	11.36715
28	11.21188	11.36676
29	11.21171	11.36643
30	11.21155	11.36614

Tabell 6.6: Utviklingen i logaritmen av AIC som følge av endring i frihetsgrader for kovariaten alder, for alle landene sett under ett

6.2 R-kode

6.2.1 Beregning og analysering av dødelighetsrater

```
#Norske data(tilsvarende for datasettene for de andre landene)
#-----
#Innlesing av data
dod_nor=read.table("C:\\Users\\Vanja\\Documents\\Masteroppgave
\\dod_Norge.txt", header=T)
bef_nor=read.table("C:\\Users\\Vanja\\Documents\\Masteroppgave
\\befolkning_Norge.txt", header=T)

nor_rate_k <- 0*(1:length(dod_nor$Female))
nor_rate_m <- 0*(1:length(dod_nor$Male))

for (i in 1:length(dod_nor$Female)) {
  if (dod_nor$Female[i]==0 && bef_nor$Female[i]==0) {
    nor_rate_k[i] <- 0
  } else {
    nor_rate_k[i] <- dod_nor$Female[i]/bef_nor$Female[i]
    if (nor_rate_k[i]==Inf) {
      nor_rate_k[i]=1
    }
  }
}

for (i in 1:length(dod_nor$Male)) {
  if (dod_nor$Male[i]==0 && bef_nor$Male[i]==0) {
    nor_rate_m[i] <- 0
  } else {
    nor_rate_m[i] <- dod_nor$Male[i]/bef_nor$Male[i]
    if (nor_rate_m[i]==Inf) {
      nor_rate_m[i]=1
    }
  }
}

#Finner ut når individene ble født
nor_fodt <- dod_nor$Year-dod_nor$Age

#Grupperer de som er født før 1935, de som er født 1935-1939, 1940-1945,
#og etter 1945
nor_fodt_gr=cut(nor_fodt, breaks=c(1820,1930,1935,1940,1946,1951,2007)
, right=F, labels=0:5)

teller_Norge <- 0*(1:length(nor_fodt)) + 1

#Samler alt i en matrise
nor_matrise <- cbind(dod_nor, nor_fodt, nor_fodt_gr, nor_rate_k, nor_rate_m
, teller_Norge)

#Henter ut dødelighetsratene
rateKvinner_Norge <- 0*(1:111) %%% t(1:4)
rateMenn_Norge <- 0*(1:111) %%% t(1:4)
ant_Norge <- 0*(1:111) %%% t(1:4)

for (i in 1:4) {
  for (j in 1:111) {
    rateKvinner_Norge[j, i] <- sum(nor_matrise[nor_matrise[,7]==i
& nor_matrise[,2]==(j-1),8])
    rateMenn_Norge[j, i] <- sum(nor_matrise[nor_matrise[,7]==i
& nor_matrise[,2]==(j-1),9])

    ant_Norge[j, i] <- sum(nor_matrise[nor_matrise[,7]==i
& nor_matrise[,2]==(j-1),10])
  }
}

#Dødelighetsrater
nor_dodsrate_k <- rateKvinner_Norge/ant_Norge
nor_dodsrate_m <- rateMenn_Norge/ant_Norge

#Finner varians
test_matrise <- cbind(bef_nor, dod_nor[,3:4], nor_fodt_gr, teller_Norge)

d_kvinner_nor <- 0*(1:111) %%% t(1:4)
d_menn_nor <- 0*(1:111) %%% t(1:4)
u_kvinner_nor <- 0*(1:111) %%% t(1:4)
u_menn_nor <- 0*(1:111) %%% t(1:4)
antall_nor <- 0*(1:111) %%% t(1:4)
```

```

for (i in 1:4) {
  for (j in 1:111) {
    u_kvinner_nor[j, i] <- sum(test_matrise[test_matrise[,8]==i
& test_matrise[,2]==(j-1),3])
    u_menn_nor[j, i] <- sum(test_matrise[test_matrise[,8]==i
& test_matrise[,2]==(j-1),4])

    d_kvinner_nor[j, i] <- sum(test_matrise[test_matrise[,8]==i
& test_matrise[,2]==(j-1),6])
    d_menn_nor[j, i] <- sum(test_matrise[test_matrise[,8]==i
& test_matrise[,2]==(j-1),7])

    antall_nor[j, i] <- sum(test_matrise[test_matrise[,8]==i
& test_matrise[,2]==(j-1),9])
  }
}

uu_kvinner_nor <- u_kvinner_nor/antall_nor
uu_menn_nor <- u_menn_nor/antall_nor
dd_kvinner_nor <- d_kvinner_nor/antall_nor
dd_menn_nor <- d_menn_nor/antall_nor

#Estimert varians
var_kk_nor <- dd_kvinner_nor/(uu_kvinner_nor)^2
var_mm_nor <- dd_menn_nor/(uu_menn_nor)^2

#Sammenligner landene
#-----

#Dødelighetsrater for kvinner i gruppe 1 i alle landene
#Tilsvarende for menn
matplot(cbind(log(nor_dodsrate_k[1:77,1]), log(dan_dodsrate_k[1:77,1]),
log(sve_dodsrate_k[1:77,1]), log(fin_dodsrate_k[1:77,1])), type='l',
ylim=c(-9,-2.5), xlab="Alder", ylab="Dødelighetsrate log-skala",
main="Kvinner år 1930-1934")
legend(5,-2.5, legend=c("Norge", "Danmark", "Sverige", "Finland"), cex=0.6,
col = 1:4, lty=1:4, bty="n")

#Dødelighetsrater for kvinner i gruppe 2 i alle landene
matplot(cbind(log(nor_dodsrate_k[1:77,2]), log(dan_dodsrate_k[1:77,2]),
log(sve_dodsrate_k[1:77,2]), log(fin_dodsrate_k[1:77,2])), type='l',
ylim=c(-9,-2.5), xlab="Alder", ylab="Dødelighetsrate log-skala",
main="Kvinner år 1935-1939")
legend(5,-2.5, legend=c("Norge", "Danmark", "Sverige", "Finland"), cex=0.6,
col = 1:4, lty=1:4, bty="n")

#Dødelighetsrater for kvinner i gruppe 3 i alle landene
matplot(cbind(log(nor_dodsrate_k[1:77,3]), log(dan_dodsrate_k[1:77,3]),
log(sve_dodsrate_k[1:77,3]), log(fin_dodsrate_k[1:77,3])), type='l',
ylim=c(-9,-2.5), xlab="Alder", ylab="Dødelighetsrate log-skala",
main="Kvinner år 1940-1945")
legend(5,-2.5, legend=c("Norge", "Danmark", "Sverige", "Finland"), cex=0.6,
col = 1:4, lty=1:4, bty="n")

#Dødelighetsrater for kvinner i gruppe 4 i alle landene
matplot(cbind(log(nor_dodsrate_k[1:77,4]), log(dan_dodsrate_k[1:77,4]),
log(sve_dodsrate_k[1:77,4]), log(fin_dodsrate_k[1:77,4])), type='l',
ylim=c(-9,-2.5), xlab="Alder", ylab="Dødelighetsrate log-skala",
main="Kvinner år 1946-1950")
legend(5,-2.5, legend=c("Norge", "Danmark", "Sverige", "Finland"), cex=0.6,
col = 1:4, lty=1:4, bty="n")

#Dødelighetsrater for kvinner i Norge (tilsvarende for menn)
matplot(log(dan_dodsrate_k), type='l', xlim=c(0,80), ylim=c(-9,-2.5),
xlab="Alder", ylab="Dødelighetsrate log-skala", main="Kvinner i Norge")
legend(5,-2.5, legend=c("Født år 1930-1934", "Født år 1935-1939",
"Født år 1940-1945", "Født år 1946-1950"), cex=0.6, col = 1:4, lty=1:4,
bty="n")

#Dødelighetsrater for kvinner i Danmark
matplot(log(dan_dodsrate_k), type='l', xlim=c(0,80), ylim=c(-9,-2.5),
xlab="Alder", ylab="Dødelighetsrate log-skala", main="Kvinner i Danmark")
legend(5,-2.5, legend=c("Født år 1930-1934", "Født år 1935-1939",
"Født år 1940-1945", "Født år 1946-1950"), cex=0.6, col = 1:4, lty=1:4,
bty="n")

#Dødelighetsrater for kvinner i Sverige
matplot(log(sve_dodsrate_k), type='l', xlim=c(0,80), ylim=c(-9,-2.5),
xlab="Alder", ylab="Dødelighetsrate log-skala", main="Kvinner i Sverige")
legend(5,-2.5, legend=c("Født år 1930-1934", "Født år 1935-1939",
"Født år 1940-1945", "Født år 1946-1950"), cex=0.6, col = 1:4, lty=1:4,
bty="n")

#Dødelighetsrater for kvinner i Finland

```

```

matplot(log(fin_dodsrate_k), type='l', xlim=c(0,80), ylim=c(-9,-2.5),
        xlab="Alder", ylab="Dødelighetsrate log-skala", main="Kvinner i Finland")
legend(5,-2.5,legend=c("Født år 1930-1934","Født år 1935-1939",
        "Født år 1940-1945","Født år 1946-1950"),cex=0.6,col = 1:4,lty=1:4,
        bty="n")

#Gjennomsnittlige dødelighetsrater
#-----
#Gjennomsnitt for hver periode
gjsnitt_k <- (nor_dodsrate_k + sve_dodsrate_k + dan_dodsrate_k
        + fin_dodsrate_k)/4
gjsnitt_m <- (nor_dodsrate_m + sve_dodsrate_m + dan_dodsrate_m
        + fin_dodsrate_m)/4

#Gjennomsnitt for alle periodene for kvinner(tilsvarende for menn)
totalsnitt_nor_k1 <- (nor_dodsrate_k[1:61,1]+nor_dodsrate_k[1:61,2]
        +nor_dodsrate_k[1:61,3]+nor_dodsrate_k[1:61,4])/4
totalsnitt_dan_k1 <- (dan_dodsrate_k[1:61,1]+dan_dodsrate_k[1:61,2]
        +dan_dodsrate_k[1:61,3]+dan_dodsrate_k[1:61,4])/4
totalsnitt_sve_k1 <- (sve_dodsrate_k[1:61,1]+sve_dodsrate_k[1:61,2]
        +sve_dodsrate_k[1:61,3]+sve_dodsrate_k[1:61,4])/4
totalsnitt_fin_k1 <- (fin_dodsrate_k[1:61,1]+fin_dodsrate_k[1:61,2]
        +fin_dodsrate_k[1:61,3]+fin_dodsrate_k[1:61,4])/4

totalsnitt_nor_k2 <- (nor_dodsrate_k[62:67,1]+nor_dodsrate_k[62:67,2]
        +nor_dodsrate_k[62:67,3])/3
totalsnitt_dan_k2 <- (dan_dodsrate_k[62:67,1]+dan_dodsrate_k[62:67,2]
        +dan_dodsrate_k[62:67,3])/3
totalsnitt_sve_k2 <- (sve_dodsrate_k[62:67,1]+sve_dodsrate_k[62:67,2]
        +sve_dodsrate_k[62:67,3])/3
totalsnitt_fin_k2 <- (fin_dodsrate_k[62:67,1]+fin_dodsrate_k[62:67,2]
        +fin_dodsrate_k[62:67,3])/3

totalsnitt_nor_k3 <- (nor_dodsrate_k[68:72,1]+nor_dodsrate_k[68:72,2])/2
totalsnitt_dan_k3 <- (dan_dodsrate_k[68:72,1]+dan_dodsrate_k[68:72,2])/2
totalsnitt_sve_k3 <- (sve_dodsrate_k[68:72,1]+sve_dodsrate_k[68:72,2])/2
totalsnitt_fin_k3 <- (fin_dodsrate_k[68:72,1]+fin_dodsrate_k[68:72,2])/2

totalsnitt_nor_k4 <- (nor_dodsrate_k[73:77,1])/1
totalsnitt_dan_k4 <- (dan_dodsrate_k[73:77,1])/1
totalsnitt_sve_k4 <- (sve_dodsrate_k[73:77,1])/1
totalsnitt_fin_k4 <- (fin_dodsrate_k[73:77,1])/1

#Lager vektorer for hvert land
totalsnitt_nor_k <- c(totalsnitt_nor_k1, totalsnitt_nor_k2, totalsnitt_nor_k3,
        totalsnitt_nor_k4)
totalsnitt_dan_k <- c(totalsnitt_dan_k1, totalsnitt_dan_k2, totalsnitt_dan_k3,
        totalsnitt_dan_k4)
totalsnitt_sve_k <- c(totalsnitt_sve_k1, totalsnitt_sve_k2, totalsnitt_sve_k3,
        totalsnitt_sve_k4)
totalsnitt_fin_k <- c(totalsnitt_fin_k1, totalsnitt_fin_k2, totalsnitt_fin_k3,
        totalsnitt_fin_k4)

#Finner de totale gjennomsnittlige dødelighetsratene
totalsnitt_k <- (totalsnitt_nor_k + totalsnitt_dan_k + totalsnitt_sve_k
        + totalsnitt_fin_k)/4
totalsnitt_m <- (totalsnitt_nor_m + totalsnitt_dan_m + totalsnitt_sve_m
        + totalsnitt_fin_m)/4

#Dødelighetsrater differanse til totalsnitt(log)

#Kvinner(tilsvarende for menn)
matplot(cbind(0, log(dan_dodsrate_k[1:77,1]) - log(totalsnitt_k),
        log(sve_dodsrate_k[1:77,1]) - log(totalsnitt_k), log(fin_dodsrate_k[1:77,1])
        -log(totalsnitt_k),log(nor_dodsrate_k[1:77,1]) - log(totalsnitt_k)), type='l',
        ylim=c(-1.5,1.5), xlab="Alder", ylab="Differanse dødelighetsrate",
        main="Kvinner år 1930-1934")
legend(38,1.7,legend=c("Differanse Danmark","Differanse Sverige",
        "Differanse Finland","Differanse Norge"),cex=0.6,col = 2:5,lty=2:5, bty="n")

matplot(cbind(0, log(dan_dodsrate_k[1:77,2]) - log(totalsnitt_k),
        log(sve_dodsrate_k[1:77,2]) - log(totalsnitt_k), log(fin_dodsrate_k[1:77,2])
        -log(totalsnitt_k),log(nor_dodsrate_k[1:77,2]) - log(totalsnitt_k)),
        type='l', ylim=c(-1.5,1.5), xlab="Alder", ylab="Differanse dødelighetsrate",
        main="Kvinner år 1935-1939")
legend(38,1.7,legend=c("Differanse Danmark","Differanse Sverige",
        "Differanse Finland","Differanse Norge"),cex=0.6,col = 2:5,lty=2:5, bty="n")

matplot(cbind(0, log(dan_dodsrate_k[1:77,3]) - log(totalsnitt_k),
        log(sve_dodsrate_k[1:77,3]) - log(totalsnitt_k), log(fin_dodsrate_k[1:77,3])
        -log(totalsnitt_k),log(nor_dodsrate_k[1:77,3]) - log(totalsnitt_k)),
        type='l', ylim=c(-1.5,1.5), xlab="Alder", ylab="Differanse dødelighetsrate",
        main="Kvinner år 1940-1945")
legend(38,1.7,legend=c("Differanse Danmark","Differanse Sverige",
        "Differanse Finland","Differanse Norge"),cex=0.6,col = 2:5,lty=2:5, bty="n")

matplot(cbind(0, log(dan_dodsrate_k[1:77,4]) - log(totalsnitt_k),

```

```

log(sve_dodsrate_k[1:77,4]) - log(totalsnitt_k), log(fin_dodsrate_k[1:77,4])
-log(totalsnitt_k), log(nor_dodsrate_k[1:77,4]) - log(totalsnitt_k)),
type='l', ylim=c(-1.5,1.5), xlab="Alder", ylab="Differanse dødelighetsrate",
main="Kvinner år 1946-1950")
legend(38,1.7,legend=c("Differanse Danmark","Differanse Sverige",
"Differanse Finland","Differanse Norge"),cex=0.6,col = 2:5,lty=2:5, bty="n")

#Dødelighetsrater faktisk differanse til totalsnitt

#Kvinner(tilsvarende for menn)
matplot(cbind(0, dan_dodsrate_k[1:77,1]-totalsnitt_k, sve_dodsrate_k[1:77,1]
-totalsnitt_k, fin_dodsrate_k[1:77,1]-totalsnitt_k, nor_dodsrate_k[1:77,1]
-totalsnitt_k), type='l', xlim=c(10,80), ylim=c(-0.005, 0.009),xlab="Alder",
ylab="Differanse dødelighetsrate", main="Kvinner år 1930-1934")
legend(10,0.0095,legend=c("Differanse Danmark","Differanse Sverige",
"Differanse Finland","Differanse Norge"),cex=0.6,col = 2:5,lty=2:5, bty="n")

matplot(cbind(0, dan_dodsrate_k[1:77,2]-totalsnitt_k, sve_dodsrate_k[1:77,2]
-totalsnitt_k, fin_dodsrate_k[1:77,2]-totalsnitt_k, nor_dodsrate_k[1:77,2]
-totalsnitt_k), type='l', xlim=c(10,80), ylim=c(-0.005, 0.009), xlab="Alder",
ylab="Differanse dødelighetsrate", main="Kvinner år 1935-1939")
legend(10,0.0095,legend=c("Differanse Danmark","Differanse Sverige",
"Differanse Finland","Differanse Norge"),cex=0.6,col = 2:5,lty=2:5, bty="n")

matplot(cbind(0, dan_dodsrate_k[1:77,3]-totalsnitt_k, sve_dodsrate_k[1:77,3]
-totalsnitt_k, fin_dodsrate_k[1:77,3]-totalsnitt_k, nor_dodsrate_k[1:77,3]
-totalsnitt_k), type='l', xlim=c(10,80), ylim=c(-0.005, 0.009), xlab="Alder",
ylab="Differanse dødelighetsrate", main="Kvinner år 1940-1945")
legend(10,0.0095,legend=c("Differanse Danmark","Differanse Sverige",
"Differanse Finland","Differanse Norge"),cex=0.6,col = 2:5,lty=2:5, bty="n")

matplot(cbind(0, dan_dodsrate_k[1:77,4]-totalsnitt_k, sve_dodsrate_k[1:77,4]
-totalsnitt_k, fin_dodsrate_k[1:77,4]-totalsnitt_k, nor_dodsrate_k[1:77,4]
-totalsnitt_k), type='l', xlim=c(10,80), ylim=c(-0.005, 0.009), xlab="Alder",
ylab="Differanse dødelighetsrate", main="Kvinner år 1946-1950")
legend(10,0.0095,legend=c("Differanse Danmark","Differanse Sverige",
"Differanse Finland","Differanse Norge"),cex=0.6,col = 2:5,lty=2:5, bty="n")

```

6.2.2 Beregning av overlevelsessansynligheter

```

#Overlevelsesfunksjon

#Matrise
A_nor_k <- 0*(1:111) %>% t(1:4)
sum_nor_k <- 0*(1:111) %>% t(1:4)
A_nor_m <- 0*(1:111) %>% t(1:4)
sum_nor_m <- 0*(1:111) %>% t(1:4)

A_nor_k[1,] <- nor_dodsrate_k[1,]
sum_nor_k[1,] <- nor_dodsrate_k[1,]
A_nor_m[1,] <- nor_dodsrate_m[1,]
sum_nor_m[1,] <- nor_dodsrate_m[1,]

#Finner A(t)
for (i in 2:111) {
  for (j in 1:4) {
    sum_nor_k[i,j] <- sum_nor_k[i-1, j] + nor_dodsrate_k[i, j]
    sum_nor_m[i, j] <- sum_nor_m[i-1, j] + nor_dodsrate_m[i, j]
    A_nor_k[i, j] <- sum_nor_k[i, j]
    A_nor_m[i, j] <- sum_nor_m[i, j]
  }
}

#Overlevelsesfunksjon menn og kvinner
S_nor_k <- exp(-A_nor_k)
S_nor_m <- exp(-A_nor_m)

```

6.2.3 Beregning av forventet gjestående levetid

```

#Input alder, der alder 0 år tilsvarer alder=1
#Input bestand; kohort 1,2,3 eller 4
#Input overlevelse er ønsket overlevelsesfunksjon, f.eks S_nor_k for

```

```

#overlevelsesfunksjonen til norske kvinner
levealder <- function(alder, bestand, overlevelse) {
  tPx <- overlevelse[alder:61,bestand]/overlevelse[alder,bestand]
  forventet_levaalder <- sum(tPx)
  return(forventet_levaalder)
}

#Forventet gjenstående levetid for 60-åringer
levealder60 <- function(overlevelse) {
  tPx <- overlevelse[61,]/overlevelse[60,]
  return(tPx)
}

#Eksempel
round(levealder(1,4,S_nor.k),2)
round(levealder60(S_nor.k),3)

```

6.2.4 Modellering av Poisson regresjon og relative risikoer

```

#Poisson regresjon
#Land: Norge=1, Danmark=2, Sverige=3, Finland=4
land_nor <- 0*(1:8547) + 1
land_dan <- 0*(1:8547) + 2
land_sve <- 0*(1:8547) + 3
land_fin <- 0*(1:8547) + 4

nor_matrise_land <- cbind(nor_matrise, land_nor)
dan_matrise_land <- cbind(dan_matrise, land_dan)
sve_matrise_land <- cbind(sve_matrise, land_sve)
fin_matrise_land <- cbind(fin_matrise, land_fin)

#Etter å ha samlet dataene for alle landene til én stor matrise:
pois_alle_k <- glm(k.dod ~ factor(alder) + factor(kohort)
  + factor(land)+ offset(logbef.k), family=poisson(link=log),
  data=data_alle, subset=(kohort>0 & kohort<5 & alder>0))
pois_alle_m <- glm(m.dod ~ factor(alder) + factor(kohort)
  + factor(land)+ offset(logbef.m), family=poisson(link=log),
  data=data_alle, subset=(kohort>0 & kohort<5 & alder>0))

#Med vektorer i stedet for offset
pois_alle_k_b <- glm(I(k.dod/k.bef) ~ factor(alder) + factor(kohort)
  + factor(land), weights=(k.bef), family=poisson, data=data_alle,
  subset=(k.bef>0 & kohort>0 & kohort<5 & alder>0))
pois_alle_m_b <- glm(I(m.dod/m.bef) ~ factor(alder) + factor(kohort)
  + factor(land), weights=(m.bef), family=poisson, data=data_alle,
  subset=(m.bef>0 & kohort>0 & kohort<5 & alder>0))

#Finner relativ risiko (RR) og 95% konfidensintervall
RRCItab<-function(glmfit){
  sumglm<-summary(glmfit)$coef
  RRL<-exp(sumglm[,1])
  RRL<-exp(sumglm[,1]-1.96*sumglm[,2])
  RRU<-exp(sumglm[,1]+1.96*sumglm[,2])
  cbind(RR,RRL,RRU)
}

#Uten interaksjonsledd
RR_alle_k <- RRCItab(pois_alle_k)
RR_alle_m <- RRCItab(pois_alle_m)

#Sjekker interaksjon
pois_interaksjon_k <- glm(k.dod ~ factor(alder) + factor(kohort)
  + factor(land)+ offset(logbef.k) + factor(land):factor(kohort),
  family=poisson(link=log), data=data_alle,
  subset=(kohort>0 & kohort<5 & alder>0))
pois_interaksjon_m <- glm(m.dod ~ factor(alder) + factor(kohort)
  + factor(land)+ offset(logbef.m) + factor(land):factor(kohort),
  family=poisson(link=log), data=data_alle,
  subset=(kohort>0 & kohort<5 & alder>0))

RR_ki.k <- RRCItab(pois_interaksjon_k)
RR_ki.m <- RRCItab(pois_interaksjon_m)

#Ser på land, kohort og alder hver for seg i tre ulike aldersgrupper
pois_landkohort_k1 <- glm(k.dod ~ factor(alder) + factor(kohort)
  + factor(land) + offset(logbef.k), family=poisson(link=log),
  data=data_alle, subset=(kohort>0 & kohort<5 & alder>0 & alder<26))
pois_landkohort_k2 <- glm(k.dod ~ factor(alder) + factor(kohort)

```

```

+ factor(land) + offset(logbef.k), family=poisson(link=log),
data=data_alle, subset=(kohort>0 & kohort<5 & alder >25 & alder <51))
pois_landkohort_k3 <- glm(k.dod ~ factor(alder) + factor(kohort)
+ factor(land) + offset(logbef.k), family=poisson(link=log),
data=data_alle, subset=(kohort>0 & kohort<5 & alder >50 & alder <77))

pois_landkohort_m1 <- glm(m.dod ~ factor(alder) + factor(kohort)
+ factor(land) + offset(logbef.m), family=poisson(link=log),
data=data_alle, subset=(kohort>0 & kohort<5 & alder >0 & alder <26))
pois_landkohort_m2 <- glm(m.dod ~ factor(alder) + factor(kohort)
+ factor(land) + offset(logbef.m), family=poisson(link=log),
data=data_alle, subset=(kohort>0 & kohort<5 & alder >25 & alder <51))
pois_landkohort_m3 <- glm(m.dod ~ factor(alder) + factor(kohort)
+ factor(land) + offset(logbef.m), family=poisson(link=log),
data=data_alle, subset=(kohort>0 & kohort<5 & alder >50 & alder <77))

RR_ki_landkohort_k1 <- RRCItab(pois_landkohort_k1)
RR_ki_landkohort_k2 <- RRCItab(pois_landkohort_k2)
RR_ki_landkohort_k3 <- RRCItab(pois_landkohort_k3)

RR_ki_landkohort_m1 <- RRCItab(pois_landkohort_m1)
RR_ki_landkohort_m2 <- RRCItab(pois_landkohort_m2)
RR_ki_landkohort_m3 <- RRCItab(pois_landkohort_m3)

#Sette sammen land og kohort til én faktor
#Annen måte å se på interaksjon
LandKohort <- data_alle$kohort + 4*(data_alle$land - 1)

#Inndelt i tre ulike aldersgrupper
pois_alle_kk1 <- glm(k.dod ~ factor(alder) + factor(LandKohort)
+ offset(logbef.k), family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder >0 & alder <26))
pois_alle_kk2 <- glm(k.dod ~ factor(alder) + factor(LandKohort)
+ offset(logbef.k), family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder >25 & alder <51))
pois_alle_kk3 <- glm(k.dod ~ factor(alder) + factor(LandKohort)
+ offset(logbef.k), family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder >50 & alder <77))

pois_alle_mm1 <- glm(m.dod ~ factor(alder) + factor(LandKohort)
+ offset(logbef.m), family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder >0 & alder <26))
pois_alle_mm2 <- glm(m.dod ~ factor(alder) + factor(LandKohort)
+ offset(logbef.m), family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder >25 & alder <51))
pois_alle_mm3 <- glm(m.dod ~ factor(alder) + factor(LandKohort)
+ offset(logbef.m), family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder >50 & alder <77))

RR_alle_kk1 <- RRCItab(pois_alle_kk1)
RR_alle_kk2 <- RRCItab(pois_alle_kk2)
RR_alle_kk3 <- RRCItab(pois_alle_kk3)

RR_alle_mm1 <- RRCItab(pois_alle_mm1)
RR_alle_mm2 <- RRCItab(pois_alle_mm2)
RR_alle_mm3 <- RRCItab(pois_alle_mm3)

```

6.2.5 Devians

```

#Devianstabell inndelt i tre aldersgrupper

#Alder 1–25 år, kvinner
pois_interaksjon_kk1 <- glm(k.dod ~ factor(alder) + factor(kohort)
+ factor(land)+ offset(logbef.k) + factor(land):factor(kohort),
family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder >0 & alder <26))

#Alder 26–50 år, kvinner
pois_interaksjon_kk2 <- glm(k.dod ~ factor(alder) + factor(kohort)
+ factor(land)+ offset(logbef.k) + factor(land):factor(kohort),
family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder >25 & alder <51))

#Alder 51–76 år, kvinner
pois_interaksjon_kk3 <- glm(k.dod ~ factor(alder) + factor(kohort)
+ factor(land)+ offset(logbef.k) + factor(land):factor(kohort),
family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder >50 & alder <77))

```

```

#Alder 1-25 år, menn
pois_interaksjon_mm1 <- glm(m.dod ~ factor(alder) + factor(kohort)
+ factor(land)+ offset(logbef.m) + factor(land):factor(kohort),
family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder>0 & alder<26))

#Alder 26-50 år, menn
pois_interaksjon_mm2 <- glm(m.dod ~ factor(alder) + factor(kohort)
+ factor(land)+ offset(logbef.m) + factor(land):factor(kohort),
family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder>25 & alder<51))

#Alder 51-76 år, menn
pois_interaksjon_mm3 <- glm(m.dod ~ factor(alder) + factor(kohort)
+ factor(land)+ offset(logbef.m) + factor(land):factor(kohort),
family=poisson(link=log), data=data_alle,
subset=(kohort>0 & kohort<5 & alder>50 & alder<77))

#Skriver ut deviansene
anova(pois_interaksjon_kk1, test="Chisq")
anova(pois_interaksjon_kk2, test="Chisq")
anova(pois_interaksjon_kk3, test="Chisq")

anova(pois_interaksjon_mm1, test="Chisq")
anova(pois_interaksjon_mm2, test="Chisq")
anova(pois_interaksjon_mm3, test="Chisq")

```

6.2.6 Generaliserte additive modeller

```

#Generaliserte additive modeller

library(gam)

#Finner ut hvor mange frihetsgrader som bør brukes i glattingen
f <- 30
AIC_kk <- 0*(1:f)
AIC_mm <- 0*(1:f)

for (i in 1:f) {
  gamfit_mm <- gam(m.dod ~ s(alder, df=i) + s(fodselsaar)+offset(logbef.m),
family=poisson, data=data_alle, subset=(kohort>0 & kohort<5 & alder>0))
  gamfit_kk <- gam(k.dod ~ s(alder, df=i) + s(fodselsaar)+offset(logbef.k),
family=poisson, data=data_alle, subset=(kohort>0 & kohort<5 & alder>0))
  AIC_kk[i] <- summary(gamfit_kk)$aic
  AIC_mm[i] <- summary(gamfit_mm)$aic
}

#Plot for utviklingen i AIC som følge av endring av frihetsgrader
#Ved bruk av log ser det ut til at 20 frihetsgrader bør benyttes
fig2 <- par(mfrow=c(2, 1))
  matplot(log(AIC_kk), type='l', xlab="Frihetsgrader", ylab="log AIC", main="AIC kvinner")
  matplot(log(AIC_mm), type='l', xlab="Frihetsgrader", ylab="log AIC", main="AIC menn")
par(fig2)

#Ser ut til å passe bra å bruke 20 frihetsgrader for alder
gamfit_k <- gam(k.dod ~ s(alder, df=20) + s(fodselsaar)+offset(logbef.k),
family=poisson, data=data_alle, subset=(kohort>0 & kohort<5 & alder>0))
gamfit_m <- gam(m.dod ~ s(alder, df=20) + s(fodselsaar)+offset(logbef.m),
family=poisson, data=data_alle, subset=(kohort>0 & kohort<5 & alder>0))

#Menn og kvinner i samme plot
par2par <- par(mfrow=c(2, 2))
  plot(gamfit_k, se=T, main="Kvinner")
  plot(gamfit_m, se=T, main="Menn")
par(par2par)

```

Figurer

2.1	Dødelighetsrater per land for kvinner	13
2.2	Dødelighetsrater per land for menn	14
2.3	Dødelighetsrater kvinner for hver kohort	16
2.4	Dødelighetsrater menn for hver kohort	17
2.5	Differansen mellom logaritmen av dødelighetsrater for kvinner og logaritmen av totalt gjennomsnittlige dødelighetsrater for kvinner	19
2.6	Differansen mellom logaritmen av dødelighetsrater for menn og logaritmen av totalt gjennomsnittlige dødelighetsrater for menn	19
2.7	Overlevelsessannsynlighet kvinner	23
2.8	Overlevelsessannsynlighet menn	23
2.9	Forventet gjenstående levetid for norske kvinner og menn født i 1930-1934	30
4.1	Utviklingen i AIC som følge av endring i frihetsgrader for alder	47
4.2	Generaliserte additive modeller for alder og fødselsår for alle landene sett under ett	49
6.1	Differansen mellom dødelighetsrater for kvinner og totalt gjennomsnittlige dødelighetsrater for kvinner	54
6.2	Differansen mellom dødelighetsrater for menn og totalt gjennomsnittlige dødelighetsrater for menn	55
6.3	Variansen til de estimerte dødelighetsratene for norske kvinner og menn	56
6.4	Generaliserte additive modeller for alder og fødselsår for Norge	57
6.5	Generaliserte additive modeller for alder og fødselsår for Danmark	58
6.6	Generaliserte additive modeller for alder og fødselsår for Sverige	59
6.7	Generaliserte additive modeller for alder og fødselsår for Finland	60

Tabeller

2.1	Forventet gjenstående levetid for personer født i 1930-1934, når maksalder er antatt å være 60 år	27
2.2	Forventet gjenstående levetid for personer født i 1935-1939, når maksalder er antatt å være 60 år	27
2.3	Forventet gjenstående levetid for personer født i 1940-1945, når maksalder er antatt å være 60 år	28
2.4	Forventet gjenstående levetid for personer født i 1946-1950, når maksalder er antatt å være 60 år	28
3.1	Residual devians for kvinner	36
3.2	Residual devians for menn	37
3.3	Relativ risiko for land (kvinner) med 95% konfidensintervall, justert for alder og kohort	39
3.4	Relativ risiko for land (menn) med 95% konfidensintervall, justert for alder og kohort	39
3.5	Relativ risiko for kohort (kvinner) med 95% konfidensintervall, justert for alder og land	41
3.6	Relativ risiko for kohort (menn) med 95% konfidensintervall, justert for alder og land	41
3.7	Relativ risiko interaksjon med 95% konfidensintervall for kvinner, aldersjustert	42
3.8	Relativ risiko interaksjon med 95% konfidensintervall for menn, aldersjustert	43
6.1	Dødelighetsrater født år 1930-1934	61
6.2	Dødelighetsrater født år 1935-1939	61
6.3	Dødelighetsrater født år 1940-1945	62
6.4	Dødelighetsrater født år 1946-1950	62
6.5	Varians til de estimerte dødelighetsratene for norske kvinner og menn opp til alder 60 år	63
6.6	Utviklingen i logaritmen av AIC som følge av endring i frihetsgrader for kovariaten alder, for alle landene sett under ett	64

Referanser

- [1] Odd O. Aalen, Ørnulf Borgan, Håkon K. Gjessing *Survival and Event History Analysis* 2008: Springer
- [2] D. R. Cox, D. Oakes *Analysis of Survival Data* 1984: Cambridge University Press
- [3] Piet De Jong, Gillian Z. Heller *Generalized Linear Models for Insurance Data* 2008: Cambridge University Press
- [4] Hans U. Gerber *Life Insurance Mathematics* 1997: Springer
- [5] T.J. Hastie, R.J. Tibshirani *Generalized Additive Models* 1990: Chapman and Hall
- [6] Knud Juel *Middellevetid og dødelighed i Danmark sammenlignet med i Sverige* 2008: Ugeskrift for Læger
- [7] John A. Rice *Mathematical Statistics and Data Analysis* 1995: Duxbury Press
- [8] Sven Ove Samuelsen *Introduksjon til Generaliserte Lineære Modeller (GLM)* Forelesningsfoiler STK3100 høst 2008
- [9] Sven Ove Samuelsen, Anne Inger Borge, Per Magnus, Leiv S. Bakketeig *Temporal and regional trends in fatal childhood injuries in Norway 1971-1989* 1993: Scand J Soc Med
- [10] Byung Pal Yu, Beong Ou Lim, Michihiro Sugano *Dietary Restriction Down-regulates Free Radical and Lipid Peroxide Production: Plausible Mechanism for Elongation of Life Span* 2002: J Nutr Sci Vitaminol
- [11] www.mortality.org *Human Mortality Database*

