

Interne notater

STATISTISK SENTRALBYRÅ

88/6

21. april 1988

LIVSLØPSANALYSE ANVENDT PÅ REKRUTERING TIL OG AVGANG FRA UFØREPENSJONSORDNINGEN.

AV

ELLEN J. AMUNDSEN OG ANDERS RYGH SWENSEN

Innhold.

1. Bakgrunn og formål.....	1
2. Et eksempel.....	2
3. Noen sentrale begreper i livsløpsanalyse.....	10
4. Mer om livsløpsanalyse.....	17
5. Uførepensjonsordningen i et livsløpsperspektiv.....	20
6. Faktorer som påvirker tilgang og opphør fra uførepensjonsordningen.....	25
7. Datakildenes betydning for valg av metode.....	29
8. Anbefalinger og mulig videreføring av prosjektet.....	32
Referanser.....	34
Appendiks A. Noen tekniske utledninger.....	35
Appendiks B. Tabeller over nye uføretilfeller i 1983 og estimert eksponering på grunnlag av Helseundersøkelsen 1985.	37
Appendiks C. Program for kjøring av LOGLIN.....	40

1. Bakgrunn og formål.

I dette notatet skal vi diskutere hvordan ideer fra livsløpsanalyse kan anvendes på data om bruk av trygdeordninger. Vi skal gi en kort oversikt over de viktigste ideene i forløpsanalyse, og belyse hvordan begreper og metoder fra dette feltet kan benyttes til å analysere data om rekruttering til og avgang fra uførepensjonsordningen. Grunnen til det valget er både at dette er et viktig felt der det er ønskelig med større innsikt i faktorer som påvirker antallet uførepensjonister, og at vi ønsker å diskutere metodene i lys av konkrete problemstillinger for å få større innsikt i metodenes styrke og begrensninger.

Notatet er organisert på følgende måte. Kapittel 2 er et eksempel på hvordan en analyse kan gjennomføres, og en diskusjon av begrensningen i de data som står til rådighet og hvilke konsekvenser dette har for tolkingen av resultatene. Eksempelet tar opp variasjoner i tilgangen av nye uførepensjonister for ulike sykdomstyper. Siden disse sykdommene ikke er like utbredt i befolkningen, kan det gi et feilaktig bilde å konsentrere seg om antallet nye uføretrygdede. At det er mange uføretrygdede som har en bestemt diagnose, behøver ikke å bety at det blant de som lider av sykdommen er en høy andel som blir uføretrygd. Det kan ganske enkelt skyldes at sykdommen er svært utbredt. En form for standardisering er derfor nødvendig. Kapittel 3 gir en kortfattet beskrivelse av de viktigste begrepene i livsløpsanalyse. Dette utdypes i kapittel 4 som inneholder mer stoff om hvordan forklaringsvariable kan trekkes inn i modellene ved å benytte regresjonslignende teknikker. Spesiell vekt er lagt på forutsetningene som må være til stede for å kunne bruke standard program for analyse av kontingenstabeller. Kapittel 5 gir en kort oversikt over de formelle sidene ved uførepensjonsordningen. Dessuten påpeker vi noen tidspunkter som er viktige for å beskrive en uføretrygdets livsløp og viser hvordan disse passer inn i begrepsapparatet fra kapittel 3. I kapittel 6 ser vi på en del faktorer som er av betydning for variasjonen i antall nye uførepensjonister og for avgang fra ordningen. Kapittel 7 inneholder en oversikt over registerdata i Rikstrygdeverket og Statistisk Sentralbyrå om de forholdene som ble berørt i de to foregående

kapitlene. Vi påpeker spesielt de begrensningene dataorganiseringen medfører. Notatet er en dokumentasjon av arbeidet i et mindre forprosjekt. I kapittel 8 kommer vi med noen synspunkter på hvordan arbeidet kan videreføres.

2. Et eksempel.

Vi ønsker å se næyere på overgangen fra syk til ufør med utgangspunkt i dataene i SSB over nye uføre i 1983. Det er interessant å studere hvordan tilgangen varierer med kjønn, sykdomstype, alder, sivilstand og sysselsettingsstatus. Dette vil bli gjort med en regresjonslignende teknikk innen de ulike kombinasjonene av kjønn og sykdomstype (4 grupper). Dette gir ialt 8 grupper, og i hver av disse undersøker vi hvordan tilgangen varierer med alder, sivilstand og sysselsettingsstatus.

Slik dataene foreligger kan vi ikke uten videre benytte noen standard metode, men må foreta noen tillempninger. Vi skal i dette kapitlet gi en mer uformell begrunnelse for hvordan dette kan gjøres. I de følgende kapitler vil det bli gjort nærmere rede for hvordan en slik framgangsmåte kan begrunnes med ideer og begreper fra teorien for livsløpsanalyse.

Utgangspunktet er overgangsratene innen hver kombinasjon av verdier for de fem variablene. Disse ratene er forholdet mellom antallet overganger i et bestemt tidsrom ("occurrence"), og den totale observasjonstid i populasjonen i samme tidsrom ("exposure"). I vårt tilfelle vil det si forholdet mellom antall nye uførepenjonister i løpet av et år, og den totale tid det året for personer uten uføretrygd i en bestemt kategori. Disse ratene uttrykker overgangstilbøyelighetene, og tar hensyn til hvor mange som kunne ha skiftet status i tillegg til antallet som faktisk gjør det.

La oss nå se på hvordan ratene beregnes når man har data av den typen vi betrakter i dette eksemplet. "Occurrence" estimatene er relativt ukompliserte og kan baseres på tallene over nye uførepenjonister i 1983. "Exposure" er langt mer problematisk. Vi ønsker innsikt i hvordan tilgangen er for ulike sykdomstyper. Opplysninger om

sykeligheten i befolkningen må hentes fra utvalgsundersøkelser. Dette reiser en rekke problemer:

- Måles "occurrence" og "exposure" for samme populasjon?
- Er definisjonen av bakgrunnskjenntegn den samme i de to datasettene?
- Hva er innflytelsen av å blåse opp tall fra en utvalgsundersøkelse?

Vi skal ikke her diskutere disse forholdene i detalj, men i stedet gå ut fra at de ikke alvorlig influerer på resultatene. Konklusjonene en kommer fram til må derfor vurderes i forhold til hva som er rimelig å forvente. Slående resultater bør en derfor ha et kildekritisk forhold til. Likeledes bør en vokte seg for overfortolkninger.

Vi vil understreke at det ikke er bruken av en formell metodikk som gjør analysen problematisk. Det er begrensningene i data, og hvordan disse skal tolkes som er vanskeligheten. Å stole på mer uformelle vurderinger er ikke noen løsning i så henseende. Sammenligner man gale tabeller, blir konklusjonene gale uansett hvordan sammenligningen utføres. Det kan endog tenkes at en formalisering bidrar til at en tenker nøyere gjennom hele problemstillingen. Dessuten kan det være enklere å etterprøve konklusjonene. Brukt med måte er det verste som kan skje at man med formelle metoder ikke oppdager annet en det som er innlysende med et enklere maskineri.

De kjennetegnene vi har konsentrert oss om i tillegg til overgangsratene er, som nevnt, kjønn, diagnosetype (nervøse lidelser, hjerte- karsykdommer, skjelett-muskelsykdommer og en restgruppe), alder (30-39, 40-49, 50-54, 55-59 og 60-66 år), sivilstand (ikke-gift og gift) samt sysselsetting (yrkesaktiv og ikke yrkesaktiv). Tallene over nye uføre er som sagt hentet fra SSB's data for slike overganger i 1983. Tallene for "exposure" er beregnet på følgende måte. Antall varig syke i Helseundersøkelsen (1985) som framkommer ved å kombinere de ulike kategoriene for alder og bakgrunnsvariable, er justert for å ta hensyn til at uførepensjonister utgjør en del av gruppa. På grunn av definisjonsforskjeller og tilfeldige variasjoner blir en del tall negative på denne måten. De er satt lik null. Femveistabellen er deretter glattet ved å tilpasse en

log-lineær modell uten femte og fjerde ordens samspill. Den glattede tabellen er deretter blåst opp med en faktor beregnet på grunnlag av hvor stor andel helseundersøkelsen utgjør av hele befolkningen. Tabellen med justerte og oppblåste tall samt tabellen over nye uførepensjonister er gjengitt i Appendiks B.

Tallene for eksponering som framkommer på denne måten, er selvsagt svært grove. En mer rimelig måte å estimere eksponering ved bruk av tverrsnittsdata er å ta utgangspunkt i bestanden ved starten og slutten av perioden og anta at alle overgangene skjer midt i perioden. Siden vi bare har opplysninger om bestanden på ett tidspunkt, er det ikke mulig å ta skikkelig hensyn til antall overganger når eksponeringstiden skal estimeres.

En annen svakhet med framgangsmåten ovenfor er at det ikke er mulig å si noe om hvorvidt overgangsratene avhenger av hvor lenge en person har hatt en bestemt kombinasjon av kjennetegn. Det er ingen grunn til å anta at det ikke skulle være en slik avhengighet, men observasjonene har en slik form i dette tilfellet at man ikke kan undersøke om en sammenheng av denne typen eksisterer.

Etter å ha beregnet tallene for "occurrence" og "exposure" innfører vi en parametrisering av overgangsratene, som gjør det mulig å estimere hovedeffekter av alder, sivilstand og sysselsettingsstatus og samspill mellom disse faktorene.

Estimeringen foregår nå ved bruk av et program for estimering i kontingenstabeller slik det er beskrevet i kapittel 3. Et kjøreoppsett med kommentarer for programmet LOGLIN er inkludert i appendiks C. Vi gjennomfører beregningene for alle kombinasjonene av kjønn og diagnose. Parametrene er de som inngår i den vanlige variansanalyse-parametriseringen. Logaritmen til overgangsratene skrives:

$$\log \lambda_{ijk} = u_0 + u_1(i) + u_2(j) + u_3(k) + u_{01}(ij) + u_{02}(ik) + u_{12}(jk) + u_{012}(ijk)$$

der $u_0(1) + \dots + u_0(5) = 0, \dots, u_{012}(ij1) + u_{012}(ij2) = 0, i=1, \dots, 4, j=1, 2$. Tolkningen av parametrene $u_0, u_0(1), \dots, u_{012}(411)$ er den samme som i variansanalyse, og finnes i mange lærebøker om lineære modeller.

Vi gjentar hovedpunktene for noen av parametrene:

u representerer den gjennomsnittlige overgangsraten i alle de

5x2x2 gruppene.

$u_0(i)$ kan skrives som $\frac{1}{4} \sum_{j,k} \log \lambda_{ijk} - \frac{1}{20} \sum_{i,j,k} \log \lambda_{ijk}$, altså gjennomsnittlige rate i alderskategori i minus den gjennomsnittlige virkningen, og kalles ofte hovedeffekt av alder på nivå i .

Parameterene $u_{01}(ij)$ kan skrives $\frac{1}{2} \sum_k \{ \log \lambda_{ijk} - \frac{1}{2} \sum_j \log \lambda_{ijk} - \frac{1}{5} \sum_i \log \lambda_{ijk} - \frac{1}{10} \sum_{i,j} \log \lambda_{ijk} \}$. Tallet i parentes kan oppfattes som samspillet mellom alder og sivilstand i sysselsettingskategori k . $u_{01}(ij)$ blir derfor det gjennomsnittlige samspill mellom alder og sivilstand. Her tolker vi "samspill mellom alder og sivilstand i sysselsettingskategori k " som virkning av sivilstand spesifikk for alderskategori i i sysselsettingskategori k , d.v.s

$\log \lambda_{ijk} - \frac{1}{2} \sum_j \log \lambda_{ijk}$, minus den gjennomsnittlige virkning av sivilstand i sysselsettingskategori k , d.v.s

$\frac{1}{5} \sum_i \log \lambda_{ijk} - \frac{1}{10} \sum_{i,j} \log \lambda_{ijk}$.

Endelig kan parametrene $u_{012}(ijk)$ oppfattes som avviket mellom samspill for alder og sivilstand i sysselsettingskategori k og det gjennomsnittlige samspill mellom alder og sivilstand.

Resultatene av estimeringen er gjengitt i tabell 7.1 og tabell 7.2.

Tabell 2.1 Estimerte koeffisienter, diagnosetype ,menn.

diagnose	hoved- gj.snitt	hoved- eff. alder	hoved- eff. sivst.	hoved- eff. syss	samp. alder sivst.	samp. alder syss.	samp. sivst. syss.	2.ord. samp. sivst.
	u	u	u	u	u	u	u	u
		0	1	2	01	02	12	012
			30-39	ikke- gift	yrk.akt.			
					ikke			
			60-66	gift	yrk.akt.			
nerv.		-1.25			-0.07	-0.72		0.04
		-0.79			0.08	-0.71		-0.06
lid.	-1.36	-0.48	1.34	0.01	-0.39	0.33	0.17	0.00
		0.77	-1.34	-0.01	0.42	0.20		0.09
		1.76			-0.04	0.89		-0.08
hjerter-		-1.75			0.82	0.00		0.00
kar	-4.93	-0.27			0.01	-0.78		-0.06
		0.26	0.59	1.05	-0.20	0.34	-0.01	0.28
		0.68	-0.59	-1.05	-0.43	-0.14		-0.01
		1.08			-0.20	0.57		-0.21
skjel.-		-1.80			-0.02	-0.50		0.06
musk.	-4.50	-0.49			0.33	-0.34		-0.22
		0.16	0.43	0.71	-0.13	-0.14	-0.05	0.00
		0.93	-0.43	-0.71	-0.01	0.33		0.09
		1.20			-0.16	0.64		0.07
andre		-1.14			0.31	-0.61		0.11
		-0.37			0.07	-0.74		-0.01
	-3.23	-0.16	0.64	-0.42	-0.22	0.46	-0.18	0.24
		0.45	-0.64	0.42	0.12	0.37		-0.26
		1.22			-0.28	0.52		-0.08

Tabell 2.2 Estimerte koeffisienter, diagnosetype ,kvinner.

diagnose	hoved- gj.snitt	hoved- eff. alder	hoved- eff. sivst.	hoved- eff. syss	samsp. alder sivst.	samsp. alder syss.	samsp. sivst. syss.	2.ord. samsp.
	u	u	u	u	u	u	u	u
		0	1	2	01	02	12	012
		30-39	ikke- gift	yrk.akt.				
		.		ikke				
		60-66	gift	yrk.akt.				
nerv.		-1.40			-0.25	-0.41		0.05
lid.	-2.61	-0.16	0.01	0.87	0.17	0.18	-0.63	-0.16
		0.71	-0.87	-0.17	0.50	0.05	0.33	-0.04
		0.84			0.17	0.66		0.15
		-0.79			0.14	-0.36		0.04
hjerter-		-0.56			0.05	-0.38		-0.20
kar	-4.50	-0.07	0.21	0.14	-0.13	0.19	-0.12	0.35
		0.54	-0.21	-0.14	-0.15	0.05		-0.03
		0.88			0.09	0.50		-0.16
skjel.		-1.60			-0.08	-0.26		0.05
musk.	-3.72	-0.32	0.24	0.01	0.19	0.23	-0.11	-0.08
		0.62	-0.01	-0.19	-0.01	0.17	0.02	0.17
		1.06			0.17	0.53		0.03
		-1.34			0.29	-0.30		-0.08
andre	-3.32	-0.25	0.07	0.63	-0.32	0.36	-0.44	-0.14
		0.42	-0.63	0.32	-0.40	0.16	-0.24	0.33
		1.09			-0.18	0.19		0.05
					-0.07	0.40		-0.15

Som en kan vente, er ratene stigende med alder for alle diagnosegruppene og begge kjønn. Videre er det gjennomgående høyere rater blant ugifte eller før gifte i alle grupper. Ratene blant yrkesaktive er spesielt høye for menn med hjerte-kar sykdommer og med skjelett- muskelsykdommer. Ratene for personer med nervøse lidelser er markert høyere enn for andre grupper, spesielt blant menn. Alt dette er i tråd med det en vet fra tidligere undersøkelser. Det kan imidlertid være en fordel med kvantitative estimater hvis en skal sammenligne resultatene fra flere tilsvarende undersøkelser på ulike tidspunkter.

Blant de mer subtile trekk, som derfor bør vurderes med en viss skepsis, er at det er påfallende hvor like estimatene er for hjerte-kar og skjelett-muskelsykdommer, både for kvinner og menn. For menn gjelder det at det er negative samspill mellom alder og sivilstand i de høyere aldersgruppene, noe som demper inntrykket en kan få ved bare å betrakte hovedeffektene. Koeffisientene er imidlertid ikke store nok til å veie opp hovedeffektene i noen kategori. For kvinner er det derimot positive samspillsledd mellom alder og sysselsetting i de øvre aldersgruppene, noe som betyr at ratene for overgang til uføretrygd er større enn hovedeffektene alene skulle tilsi.

Det er all grunn til å være varsom med bruk av sannsynlighetsutsagn slik vi har gått fram. Men de vanlige mål for tilpasning kan brukes som deskriptive mål for hvor godt de tilpassede modeller føyer seg til de observerte tall. Vi har derfor også tilpasset modeller der samspillsledd settes lik 0. Hovedkonklusjonen er at estimatene for parametrene endrer seg lite. I tabell 2.3 har vi angitt de modellene som etter vanlige kriterier for tilpasning av log-lineære modeller gir akseptabel tilpasning. Modellene karakteriseres ved de høyeste ordens samspill som er inkludert. Siden vi begrenser oss til hierarkiske modeller er dermed alle lavere ordens samspill med. For eksempel betyr [012] en mettett modell, og [01] [2] en modell med hovedeffekt av faktor 2, sysselsetting, samt hovedeffekt og første ordens samspill mellom faktor 0 og 1, alder og sivilstand. Dessuten er størrelsen på sannsynlighetskvoten angitt sammen med antall frie parametre.

Tabell 2.3. Tilpassede modeller etter vanlige kriterier for analyse av kontingenstabeller.

Kjønn	Diagnose	Modell	Sannsynlighetskvt.	Fri.gr.
Menn	nerv. lid.	[01] [02]	11.1	5
	hjerter- kar	[01] [02]	6.3	5
	skjel.-musk.	[01] [02]	2.3	5
	andre	[01],[02],[12]	8.6	4
Kvinner	nerv. lid.	[01],[02],[12]	17.2	1
	hjerter- kar	[01],[02],[12]	16.4	1
	skjel.- musk.	[01],[02],[12]	38.0	1
	andre	[01],[02],[12]	52.3	1

Som man ser er det enklere modeller som gir akseptabel tilpasning for menn enn for kvinner. For kvinner bør en trolig inkludere alle samspillsleddene hvis en skal legge noen vekt på sannsynlighetsutsagnene. Det er verdt å legge merke til at modellen [01] [02] svarer til at samspillsledd mellom sivilstand og sysselsetting settes lik 0, altså til en modell uten samspill mellom sysselsetting og sivilstand innen hver alderskategori. Innen hver alderskategori er det derfor slik at ratene er høyere blant yrkesaktive enn ikke yrkesaktive, høyere for ugifte eller før gifte enn for gifte, og at ratene er høyere i gruppen som har begge kjennetegn enn for gruppene som bare har ett.

3. Noen sentrale begreper i livsløpsanalyse.

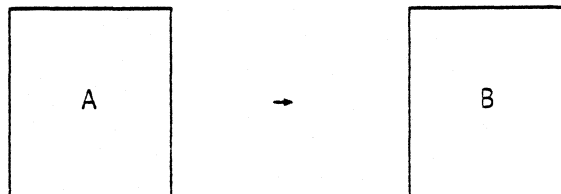
Livsløpsanalyse eller forløpsanalyse tar utgangspunkt i data for hvor lenge individer befinner seg i ulike tilstander. Kjennetegnet på data av denne typen er at variablene av primær interesse, de avhengige variablene, ikke kan anta negative verdier og at såkalt høyre sensurering kan forekomme. Det vil si at observasjonsperioden er for kort til at en kan si nøyaktig når individet skifter tilstand. Man vet bare at det skjer etter et kjent tidspunkt.

Det er av stor interesse å finne ut hvordan fordelingen til responsvariablene avhenger av uavhengige variable. Som et eksempel kan en tenke på sammenligning av ulike medisinske behandlingsopplegg. Responsen er i dette tilfellet tiden fra behandlingen starter til pasienten dør. De uavhengige variablene er i første rekke de ulike behandlingsoppleggene, men det kan også tenkes at variable som kjønn og alder bør tas hensyn til. Sensurering forekommer hvis observasjonen avsluttes før alle personene er døde. Da vet man bare at de har levd lenger enn observasjonsperioden.

Vi skal i dette kapitlet presentere noen sentrale begrep i livsløpsanalyse. Dette vil bli gjort i den enkleste situasjonen og er sentrert rundt beskrivelse av responsvariabelen. Hvordan denne kan relateres til de uavhengige variablene vil bli diskutert i et senere avsnitt. Vi nevner bare at den enkleste måten å foreta slike sammenligninger på, er å dele inn materialet i grupper svarende til ulike uavhengige variable, beskrive responsvariabelen innen hvert strata og foreta sammenligningen på dette grunnlaget.

3.1. Overlevelsesfunksjon, rater og sensurering.

Vi betrakter situasjonen med to tilstander A og B. Hvert individ vil før eller siden bevege seg fra A til B, og en observerer tiden til



overgangen. Hvert enkelt forløp er uavhengig av de andre. For en populasjon av størrelse n vil en derfor ha ikke negative observasjoner, såkalte levetider, X_1, \dots, X_n som kan antas å være uavhengige og identisk fordelte med fordelingsfunksjon F og tetthet f .

Dette kan virke som en nokså uskyldig antagelse, men verdien av livsløpsanalyse avhenger i stor grad av at den er oppfylt. I særdeleshet er det to forhold som er viktige. Det må finnes et naturlig nullpunkt å måle levetiden fra. Dette er særlig viktig hvis det er mulighet for seleksjon. Det vil si at sannsynligheten for at individer blir med i observasjonsmaterialet, avhenger av det forhold en ønsker å undersøke. I så fall må det foretas en korreksjon, som ofte i praksis betyr å innføre et nytt nullpunkt som levetiden måles fra. Videre må en med rimelighet kunne anta at populasjonen er homogen. Hvis en i virkeligheten observerer to populasjoner med svært ulike overgangsmønstre, og ikke er i stand til å fastslå hvilken populasjon individene stammer fra, er det rom for mange feilslutninger. Vi skal komme tilbake til dette problemet senere, og antar inntil videre at forutsetningen om uavhengige identiske fordelte observasjoner er berettiget.

Estimatet for fordelingsfunksjonen er den vanlige empiriske fordelingsfunksjonen:

$$F(t) = \frac{\{i : X_i \leq t\}}{n},$$

det vil si den relative andelen som har varighet kortere enn t . Den såkalte overlevelsesfunksjonen $S(t) = 1 - F(t)$, estimeres derfor som $S(t) = 1 - F(t) = \{i: X_i > t\}/n$.

En alternativ representasjon av fordelingen til X_1, \dots, X_n er ved den såkalte tidsavhengige intensitet eller overgangsrate, som kan tolkes som sannsynligheten for at et individ skifter tilstand umiddelbart etter tidspunkt x gitt at det har vært i tilstand A inntil tidspunkt x :

$$\lambda(x) = f(x)/(1 - F(x)). \quad (3.1)$$

Omvendt kan F uttrykkes ved λ :

$$F(x) = 1 - \exp\left(-\int_0^x \lambda(t) dt\right) \quad (3.2)$$

Ofte kan intensiteter eller rater som er en annen betegnelse, gi en naturligere beskrivelse av variabelen en er interessert i enn fordelingsfunksjonen. At relasjonene (3.1) og (3.2) gjelder, betyr at man lett kan velge den formen som er mest hensiktsmessig for formålet. Et par eksempler kan være illustrerende.

Eksempel 1. Eksponensiell fordeling med parameter α . Her er overlevelsesfunksjonen gitt ved $\exp(-\alpha x)$, som svarer til konstant overgangsintensitet $\lambda(x) = \alpha$. Sannsynligheten for å skifte tilstand er uavhengig av hvor lenge individet har vært i tilstand A.

Eksempel 2. Weibull fordeling med parametre α og β . Her er overlevelsesfunksjonen og intensiteten gitt ved

$$S(x) = \exp\left(-(\alpha x)^\beta\right) \quad \text{og} \quad \lambda(x) = \beta\alpha(\alpha x)^{\beta-1}.$$

Hvis $\beta > 1$, er intensiteten voksende, noe som betyr at sannsynligheten for å skifte tilstand øker med tiden. For $\beta < 1$ vil det omvendte forhold gjøre seg gjeldende.

Vi nevnte ovenfor at den populasjonen en observerer kan bestå av delpopulasjoner med svært ulike overgangsmønstre. Hvis det ikke er mulig å fastslå hvilken gruppe et individ kommer fra, kan overgangsratene anta mange ulike former avhengig av sammensetningen av hele populasjonen. Vi skal se på et eksempel:

Eksempel 3. Anta populasjonen består av to grupper som har overgangsrater α_1 og α_2 og at den relative størrelsen av gruppe 1 er π i utgangspunktet. Ved tidspunkt x vil den relative størrelsen av den første gruppa være den relative andel som er igjen i den første gruppa dividert på den relative andel i hele populasjonen som fremdeles er i tilstand A:

$$p(x) = \pi \exp(-\alpha_1 x) / (\pi \exp(-\alpha_1 x) + (1-\pi) \exp(-\alpha_2 x)).$$

Intensiteten blir derfor:

$$\lambda(x) = \alpha_1 p(x) + (1-p(x))\alpha_2,$$

som er en avtagende funksjon i x , og nærmer seg den minste av størrelsene α_1 og α_2 når x vokser. Forklaringen på dette er at på grunn av den høyere overgangsraten i den ene gruppa vil individene i tilstand A etter en tid hovedsakelig være av den typen som har lavest overgangsrate.

Vi skal kort se på hvordan F og λ kan estimeres hvis det ikke foreligger fullstendige levetidsforløp for hele materialet. Vi antar at for en del av materialet vet man bare at levetiden er lengre enn et kjent størrelse, såkalt høyresensurering. I medisinsk sammenheng vil avslutning av et behandlingsopplegg være et eksempel på slik sensurering. For personer som fremdeles er under observasjon når behandlingen avsluttes, vet en bare at levetiden er lenger enn observasjonsperioden. I demografi vil flytting til og fra området en studerer kunne resultere i sensurerte observasjoner, siden en ofte bare vil ha opplysninger om personer i den tiden de er i det aktuelle området. Hvis overlevelsesfunksjonen bare estimeres på grunnlag av den delen av materialet det foreligger fullstendige opplysninger for, får man et skjevt estimat. Lange levetider har større sjanse for sensurering. Å ignorere dem vil derfor bety at overlevelsesfunksjonen underestimeres.

Observasjonsmaterialet er nå $(Y_1, \delta_1), \dots, (Y_n, \delta_n)$ der

$$Y_i = \begin{cases} X_i & \text{hvis hele levetiden observeres,} \\ T_i & \text{som er tiden for sensurering, hvis levetiden} \\ & \text{sensureres.} \end{cases}$$

$$\delta_i = \begin{cases} 1 & \text{hvis hele levetiden observeres,} \\ 0 & \text{hvis levetiden sensureres.} \end{cases}$$

Hvis τ_1, \dots, τ_n er de distinkte tidspunktene for overganger, definerer vi følgende størrelser for $j=1, \dots, n$:

d_j : antallet som skifter tilstand på tidspunkt τ_j ,

r_j : antallet som fremdeles er i live eller ikke er sensurert på tidspunkt τ_j .

Under rimelige antagelser er følgende størrelse et estimat for overlevelsesfunksjonen på tidspunkt t :

$$S(t) = \prod_{\{i: \tau_i \leq t\}} (1 - d_i / r_i).$$

Dette er den såkalte produkt grense (product-limit) estimatoren, også kalt Kaplan-Meier estimatoren. Legg merke til at uttrykket for $S(t)$ kan tolkes som et produkt av betingede sannsynligheter. På tidspunkt τ_j estimeres sannsynligheten for å overleve som $1 - d_j / r_j$, altså som forholdet mellom antallet som overlevde og antallet som fremdeles observeres

Ved å benytte relasjonen (3.2) kan en nå estimere den kumulative intensiteten $\int_0^t \lambda(s) ds$ ved $-\log S(t)$. En mer vanlig estimator som gir omtrent samme resultat i de fleste tilfeller er:

$$\sum_{\{j: \tau_j \leq t\}} d_j / r_j.$$

Estimat for intensiteten $\lambda(t)$ kan en nå få ved å benytte teknikker som ligner de som brukes i tetthetsestimering.

3.2 Stykkevis konstante intensiteter.

Ofte vil det være rimelig å inndelegge tidsaksen i intervaller og anta at intensiteten er konstant over intervallene. Ved å velge passende intervallgrenser gir det tilstrekkelig fleksibilitet for mange formål. La $[t_1, t_2)$, $[t_2, t_3)$, \dots , $[t_k, \infty)$ være intervallene der intensiteten antas å være konstant lik $\lambda_1, \dots, \lambda_k$. Vi skal se på to situasjoner av denne typen. Først antar vi at det er de samme individene som observeres hele tiden, såkalt kohortanalyse. Derneft antar vi at det er ulike individer som er grunnlag for observasjonene i hvert intervall.

For å knytte dette til problemstillingen i dette notatet vil det i det første tilfellet bety at en følger en bestemt aldersgruppe og registrerer andelen som uføretrygdes etterhvert som tiden går. I det

andre tilfellet vil en betrakte hva som skjer i løpet av en bestemt periode, for eksempel et år, og se hvor mange som blir uføretrygdet i de ulike aldersgruppene. Dette er den situasjonen vi betraktet i eksemplet i kapittel 2.

Estimatorene er i begge tilfeller basert på det såkalte sannsynlighetsmaksimeringsprinsippet. For et sett observasjoner X med tetthet $f(x, \theta)$ er sannsynlighetsmaksimeringsestimatoren den verdien av θ som gjør $f(X, \theta)$ størst mulig. I den situasjonen vi betrakter, der sensurering er mulig, kan en ikke uten videre skrive opp tettheten til observasjonene siden fordelingen til sensureringstidene er ukjente. Det er imidlertid mulig med en modifikasjon, som vi kan kalle sannsynlighetskvoten. Et viktig poeng er at denne modifikasjonen har samme form i de to tilfellene vi nevnte ovenfor. Det betyr at sannsynlighetsmaksimeringsestimatorene også får samme form, selv om det egentlig dreier seg om helt ulike sannsynlighetsmodeller.

3.2.1 Faktisk kohort.

For de n individene i populasjonen observeres (X_i, δ_i) , $i=1, \dots, n$ der X_i er tiden til overgang eller sensurering, og δ_i er 1 hvis overgang observeres, 0 hvis sensurering observeres. I appendikset gis det en begrunnelse for at sannsynlighetskvoten kan skrives

$$\prod_{j=1}^k \lambda_j^{d_j} \exp(-\lambda_j T_j) \quad (3.3)$$

der $d_j = \sum_{i=1}^n \delta_{ji}$ er det totale antall observerte overganger i intervallet $[t_j, t_{j+1})$, og $T_j = \sum_{i=1}^n X_{ji}$ er den totale tiden individene er observert i intervallet.

Sannsynlighetsmaksimeringsestimatoren for λ_j finner en nå ved å maksimere (3.3). Resultatet er $\lambda_j = d_j / T_j$, $j=1, \dots, k$, med andre ord, den vanlige "occurrence/exposure" estimatoren.

Uttrykket (3.3) er proporsjonalt med sannsynlighetskvoten til k uavhengige Poissonfordelte variable med forventning $\lambda_1 T_1, \dots, \lambda_k T_k$. Det kan utnyttes til beregning av estimatorene. Har man et program for å finne sannsynlighetsmaksimeringsestimatoren for uavhengige

Poissonfordelte variable, kan det benyttes til å finne sannsynlighetsmaksimeringsestimatoren i den forenklete livsløpsanalysemodellen vi har betraktet i dette avsnittet. Så lenge disse estimatorene har en enkel form som forholdet mellom to observerte størrelser er ikke dette noen særlig besparing. Hvis vi derimot antar at det er en sammenheng mellom ratene, kan estimatorene bli svært kompliserte funksjoner av observasjonene og da er den analogien vi har påpekt særdeles nyttig. Vi kommer tilbake til dette i neste kapittel.

3.2.2 Syntetisk kohort.

I forrige avsnitt var det de samme individene som ble observert over intervallene $[t_1, t_2), \dots, [t_k, \infty)$. Anta nå det er ulike grupper som observeres. Til intervallet $[t_j, t_{j+1})$ vil det nå svare de n_j observasjonene (X_{ji}, δ_{ji}) , $i=1, \dots, n_j$. X_{ji} er tiden individ i observeres, det vil si tiden inntil overgang eller sensurering, og δ_{ji} er som før 1 eller 0 avhengig om en overgang observeres eller sensurering forekommer.

I appendiks A blir det vist at sannsynlighetskvoten også i dette tilfellet kan skrives på formen (3.3), med $d_j = \sum_{i=1}^{n_j} \delta_{ij}$ og $T_j = \sum_{i=1}^{n_j} X_{ij}$, der d_j og T_j er henholdsvis antallet observerte overganger og total observert eksponering i intervallet $[t_j, t_{j+1})$. Estimatoren for λ_j blir derfor også nå "occurrence-exposure" estimatoren.

Aritmetikken blir derfor den samme i de to tilfellene, enten estimatene er basert på kohort eller tversnittsdata. Tolkningen av parameterestimatene blir derimot forskjellig. I det første tilfellet har det umiddelbar mening å konstruere en levetidsfordeling basert på overgangsratene. Det vil være et estimat for levetidsfordelingen i den observerte kohorten.

Gjør man noe tilsvarende for overgangsratene basert på tversnittsdata er tolkningen langt mer problematisk, siden det her vil dreie seg om en hypotetisk populasjon.

4. Mer om livsløpsanalyse.

Vi skal i dette kapitlet se nærmere på noen mer tekniske sider ved livsløpsanalyse. I første omgang skal vi beskrive hvordan bakgrunnsvariables innflytelse på responsvariabelen kan kartlegges.

4.1. Innføring av forklaringsvariable.

Ofte vil en være interessert i å fastslå hvordan levetider og overgangsrater avhenger av bakenforliggende variable. For uføretrygdede vil det være av interesse å fastslå hvordan tilgangen avhenger av faktorer som alder, kjønn og forhold på arbeidsmarkedet. I tillegg vil det være av interesse å sammenligne de ulike diagnosegruppene og se om noen av disse skiller seg ut i forhold til de andre slik vi gjorde i eksemplet i kapittel 2. Her skal vi gi en nærmere begrunnelse for framgangsmåten som ble benyttet.

Som nevnt tidligere er den enkleste måten å dele inn materialet etter den variabelen en ønsker å undersøke virkningen av, beskrive responsvariabelen for hver gruppe og deretter foreta en mer uformell sammenligning. For mange formål vil det være tilstrekkelig. Det kan imidlertid være situasjoner der en ønsker en mer detaljert vurdering, og da er det mest hensiktsmessig å formulere en mer omfattende modell. På samme måte som i vanlig lineær regresjon tenker en seg at den relevante bakgrunnsinformasjon om en observasjon kan uttrykkes ved en q -dimensjonal vektor. Siden tidsavhengighet er vesentlig i levetidsanalyse, tillates den å variere med tiden. La $\underline{z}(t) = (z_1(t), \dots, z_q(t))$ betegne denne uavhengige variabelen. Hver komponent representerer en forklaringsfaktor.

I den enkleste situasjonen ønsker en å sammenligne to grupper som karakteriseres ved et kjennetegn som er det samme for hvert individ i hele observasjonsperioden. Da lar en $\underline{z}(t)$ være konstant. For individer i den ene gruppen kan $\underline{z}(t) = 1$, for de som hører til den andre $\underline{z}(t) = 0$. For faktorer som varierer over tid, som f. eks. inntekt, lar en $\underline{z}(t)$ uttrykke denne variasjonen.

Vi skal se på en vanlig måte å estimere innflytelsen fra forklaringsvariable, og en forenkling som tar utgangspunkt i at intensiteten antas å være konstant i bestemte intervaller.

4.1.1 Proposjonal overgangsrater.

Her antar en at et individ med forklaringsvektor $\underline{z}(t)$ har en intensitet som kan skrives

$$h_0(t)\psi(\underline{z}(t),\underline{\beta}) \quad (4.1)$$

der ψ er en kjent funksjon som vi foreløpig antar er uspesifisert. Derimot er h_0 en genuint ukjent størrelse, som kan oppfattes som en slags parameter. Betegnelsen proposjonal henspeiler på den multiplikative formen. En fortolkning av rater av denne typen er at h_0 representerer en felles grunnleggende overgangsrater for alle individer i populasjonen. Denne grunnraten moduleres gjennom funksjonen $\psi(\underline{z}(t),\underline{\beta})$ avhengig av kombinasjonen av forklaringsvariable til et bestemt individ uttrykt ved $\underline{z}(t)$, og av hvor stor innflytelse forklaringsvariablene har uttrykt ved parametrene $\underline{\beta} = (\beta_1, \dots, \beta_q)$. En vanlig spesifisering av $\psi(\underline{z}(t),\underline{\beta})$ er

$$\psi(\underline{z}(t),\underline{\beta}) = \exp(\beta_1 z_1(t) + \dots + \beta_q z_q(t)) \quad (4.2)$$

Intensiteter av formen (4.2) betegnes vanligvis som Cox-modeller. Parametrene β_1, \dots, β_q kan estimeres ved en spesiell metode som går under navnet "partial likelihood". Vi skal ikke gå inn på nærmere detaljer her, men heller henviser til lærebøker i livsløpsanalyse, f. eks. Cox og Oakes (1984).

4.1.2. Stykkevis konstante intensiteter og kategoriske kovariablene.

På samme måte som ved estimering av overlevelsesfunksjonen som ble diskutert i kapittel 3, kan en oppnå en vesentlig forenkling ved å anta at intensiteten $h_0(t)$ i (4.1) er konstant over intervallene $[t_j, t_{j+1})$, $j = 1, \dots, k$ og at samme forutsetning holder for kovariablene. La nå (i_0, \dots, i_q) indeksere de mulige kombinasjonene av tidspunkt og verdi av kovariablene. Den første komponenten $i_0 = 1, \dots, k$ betegner tiden. La antallet overganger i intervallet $[t_{i_0}, t_{i_0+1})$ blant individer med kovariat kombinasjon $i_1 \dots i_q$ være $d_{i_0 i_1 \dots i_q}$, og la $T_{i_0 i_1 \dots i_q}$ betegne den totale eksponering for denne individgruppen i samme tidsrom. Hvis $\lambda_{i_0 i_1 \dots i_q}$ er overgangsraten for individgruppen i angjeldende tidsrom, er sannsynlighetskvoten som vist i kapittel 3

$$i_0, i_1, \dots, i_q \prod_{i_0, i_1, \dots, i_q} \lambda_{i_0 i_1 \dots i_q}^{d_{i_0 i_1 \dots i_q}} \exp(-\lambda_{i_0 i_1 \dots i_q} T_{i_0 i_1 \dots i_q}). \quad (4.3)$$

Som funksjon av $\underline{\lambda}$ er (4.3) proporsjonal med

$$i_0, i_1, \dots, i_q \prod_{i_0, i_1, \dots, i_q} m_{i_0 i_1 \dots i_q}^{d_{i_0 i_1 \dots i_q}} \exp(-m_{i_0 i_1 \dots i_q}), \quad (4.4)$$

der

$$m_{i_0 i_1 \dots i_q} = \lambda_{i_0 i_1 \dots i_q} T_{i_0 i_1 \dots i_q}. \quad (4.5)$$

Verdien av $\underline{\lambda}$ som maksimerer uttrykkene (4.3) og (4.4) er derfor den samme. Ved å utnytte at (4.4) er sannsynlighetskvoten til uavhengige Poisson fordelte variable med forventning $m_{i_0 i_1 \dots i_q}$ kan en nå gjøre bruk av estimeringsprogrammer for denne typen modeller.

En reparametrisering som åpner for interessante tolkningsmuligheter, er den vanlige variansanalyseparametriseringen av $\log \lambda_{i_0 i_1 \dots i_q}$.

Vi skriver da

$$\log \lambda_{i_0 i_1 \dots i_q} = u_0 + u_0(i_0) + \dots + u_q(i_q) + u_{01}(i_0 i_1) + \dots + u_{0\dots q}(i_0 \dots i_q)$$

med følgende restriksjoner for å unngå overparametrisering:

$$\sum_{i_0} u_0(i_0) = \dots = \sum_{i_0} u_{01}(i_0 i_1) = \sum_{i_0} u_{01}(i_0 i_1) = \dots = 0.$$

En del viktige spesialtilfeller kan nå identifiseres:

Hvis tidsavhengigheten, som er indeksert ved i_0 , bare forekommer som hovedeffekt, svarer det til Cox modellen fra forrige avsnitt. Verdiene til h_0 på intervallene $[t_1, t_2), \dots, [t_k, \infty)$ er nå gitt ved $\exp(u_0 + u_0(1)), \dots, \exp(u_0 + u_0(k))$.

Hvis alle høyere ordens samspill kan antas å være neglisjerbare, betyr det at alle kovariater har en virkning som er den samme uavhengig av verdiene av de andre kovariablene. Virkningen på overgangsraten vil altså være den samme ved tilstedeværelse av en bestemt faktor for alle mulige kombinasjoner av de andre kovariatene.

Estimering av modellen kan foretas ved bruk av programmer for log-lineære modeller for kategoriske data, med ett forbehold som vi

skal komme tilbake til. Strategien for valg av passende modell vil være den samme som for analyse av kontingenstabeller. Man tilpasser suksessivt enklere modeller ved å sammenligne differensen mellom sannsynlighetskvotene med en χ^2 fordeling.

Forbeholdet som ble antydnet ovenfor, gjelder kravet (4.5). For å sikre at det er oppfylt kan det vises at det er tilstrekkelig å starte de vanlige iterasjonsprosedyrene med verdier som tilfredsstillende (4.5). Dette krever estimeringsprogrammer der startverdien kan bestemmes av brukeren. Programmet LOGLIN kan brukes, likeledes rutinen IPF i PROC MATRIX i SAS. Derimot tillater ikke PROC CATMOD i SAS å lese inn startverdier. For en nærmere begrunnelse av hvordan programmene for log-lineære modeller for kontingenstabeller kan brukes i livsløpsanalyse kan man konsultere en artikkel av Laird og Oliver (1981)

5. Uførepensjonsordningen i et livsløpsperspektiv

Uførepensjonen er etterhvert blitt en viktig inntektskilde for en stor gruppe personer i Norge og en stor utgiftspost for de offentlige budsjetter. Det er derfor viktig å klarlegge rekruttering til og den tid personer blir i ordningen. Vi skal først beskrive de formelle rammer som uførepensjonsordningen er underlagt.

5.1 Formelle sider ved uførepensjonsordningen

Folketrygdens uførepensjondel ble innført i 1967, med endel mindre justeringer senere. Idag kan personer fra 16 til 66 år få uførepensjon. Til grunn for pensjonen skal det være en varig sykdom eller skade som fører til at personer taper mer enn 50 prosent av sin ervervsevne. Det skal altså foreligge en medisinsk diagnose og samtidig et tap av evnen til å forsørge seg selv ved inntektsgivende arbeid. Dessuten skal det være prøvd hensiktsmessig utføring.

En yrkesaktiv som blir syk pleier først å gå på sykepenger i ett år. Deretter må det søkes om utføringspenger eller annen aktiv utføring. Dersom sykdommen blir varig og det er umulig å fortsette i arbeid eller finne annet arbeid, kan behandlende lege søke om uførepensjon. Søknaden blir sendt det lokale trygdekontor og tas senere opp i fylkesnemder til første avgjørelse. Her avgjøres de fleste saker, men noen videresendes Rikstrygdeverket og noen ender opp i trygderetten.

Først etter et positivt vedtak sendes det melding til Rikstrygdeverket om registrering av uførepensjonen og det er først etter en slik registrering at personen telles med i statistikken over uførepensjonister.

De fleste uførepensjonister går ut av uførepensjonsordningen når de blir 67 år og går over på alderspensjon. Dette omfatter ca. tre firedeler av de som går ut. Resten går ut pga. død. Det er bare 1-3 prosent av uførepensjonistene som går ut av andre årsaker.

5.2 Uførepensjon og livsløp

Tilstanden "ufør" vil være en mellomstasjon mellom tilstanden "syk" (uten uførepensjon) og tilstanden "død" eller "alderspensjonist". Det er flere tidspunkter som markerer viktige sider ved slike overganger: sykdomstidspunktet, tidspunktet for varig tap av ervervsevnen (uføretidspunkt), søkertidspunktet, virkningstidspunktet, registreringstidspunktet og opphørstidspunktet for uførepensjonen. Det er problemstillingen en skal studere som er avgjørende for hvilke tidspunkter som skal benyttes.

Ved studier av varighet som frisk kan tiden i tilstanden "frisk" måles ved alder og overgangen til tilstanden "syk" fastsettes ved et sykdomstidspunkt. Dette sykdomstidspunktet er igjen et nullpunkt i tilstanden "syk" ved videre studier av overgang til varig sykdom og uførepensjon. Ved stadige skiftninger mellom tilstanden frisk og tilstanden syk kan det være et problem å fastsette både et generelt sykdomstidspunkt og et startpunkt i tilstanden "varig syk". Det er trolig bare i kontrollerte forsøk med en løpende registrering at det er mulig å få gode registreringer av slike overganger.

Tidspunktet for varig tap av mer enn 50 prosent av ervervsevnen kalles i folketrygdloven for uføretidspunktet. Legen og trygdeetaten fastsetter et anslag på dette tidspunktet ved søknad om uførepensjon. Fra dette tidspunkt har en i prinsippet rett til uførepensjon. For yrkesaktive vil dette være starten på sykemeldingsperioden(e) som leder fram til uførepensjonering. For ikkeyrkesaktive vil det være legens vurdering av når sykdommen ble varig og arbeidsevnen nedsettes med 50 prosent eller mer som er avgjørende for fast settelse av uføretidspunkt. Det er i dag vanlig å gå på sykepenger et helt år før en eventuelt går over på attføringspenger. Dersom attføringstiltak

mislykkes eller ikke prøves vil den syke gå over på uførepensjon hvis en søknad innvilges. Uføretidspunktet er derfor ikke et mål for når uførepensjonen begynner å løpe.

Det kan naturligvis være uenighet om når en sykdom kan anses for å være så alvorlig at ervervsevnen er nedsatt med minst halvparten. Tidspunktet har betydning for størrelsen av tilleggspensjonen idet inntekten i perioden umiddelbart før uføretidspunktet bestemmer hvor stor en eventuell tilleggspensjon skal være. Det er derfor viktig for søkeren av uførepensjon at uføretidspunktet fastsettes så nær opp til noen gode inntektsår som mulig. For personer som ikke er yrkesaktive vil det være evnen til å utføre husarbeid som skal være avgjørende for fastsettelse av uføretidspunktet. Dette gir naturligvis rom for en større grad av skjønn. Uføretidspunktet, slik det er registrert i Rikstrygdeverkets registre, er derfor et anslag for når en blir så varig syk at det ikke er mulig å arbeide mer enn 50 prosent.

Overgang til uførepensjon kan måles etter flere ulike kriterier. Virkningstidspunktet eller starttidspunktet for når uførepensjonen begynner å løpe i kroner og øre, er tidligst 3 måneder før søknadstidspunkt. Det er behandlende lege som skal søke om uførepensjon for pasienten. I helt spesielle tilfelle kan pensjon etterbetales inntil 3 år før søkertidspunkt. Forholdet mellom tidsavstandene for uføretidspunkt, søkertidspunkt og virkningstidspunkt er i seg selv et interessant tema, men det sier lite om sykdomshistorien. Det har imidlertid betydning for studier av "trygdekarrierer" og budsjetter for trygdeetaten.

Først etter at en søknad om pensjon er innvilget, blir den registrert i RTVs registre. Statistikk over nye uførepensjonister blir dannet på grunnlag av registreringstidspunkt, ikke uføretidspunkt eller virkningstidspunkt. Årsaken til det er delvis at en må vente flere år etter registreringsår før en kan lage god statistikk med de to siste tidspunkter som basis. Det kommer jo stadig til nye registrerte uførepensjoner med uføretidspunkt flere år bakover. Dessuten har RTV alltid tidligere laget statistikk med registreringstidspunkt som basis.

Ved avgang fra uførepensjonsordningen på grunn av død kan det være dødsmeldinger som kommer sent inn og registerne vil derfor til enhver tid inneholde endel døde. Dette vil stort sett gjelde ved alle

tidspunkter og er derfor ingen tungtveiende innvending mot å bruke registerdata.

Ved fylte 67 år konverteres uførepensjon automatisk til alderspensjon. Annen avgang kan være at personer blir friske igjen og begynner å arbeide eller at de finner arbeid med sin sykdom. Noen svært få har gått ut og inn av ordningen flere ganger. Det er imidlertid vanlig at uføregraden blir endret, spesielt at den økes til full pensjon for personer som først har fått en gradert pensjon.

5.3 Sensurering og seleksjoner

Ved overgang fra en tilstand til den neste vil det være ulike typer av sensurering eller seleksjoner av observasjonene. Vi begrenser oss til å se på tilgang til og avgang fra uførepensjonen. En sensurering vil i hovedtrekk bety at vi observerer en person bare i deler av det livsløp som studeres, enten det er startfasen som mangler eller avslutningen. Metodene i livsløpsanalysen trekker inn slike faktorer i estimeringen. En seleksjon av livsløp betyr imidlertid at det er enkelte livsløp som ikke observeres i det hele tatt slik at levetiden feilestimeres av denne grunn.

Ved studier av overgangen fra tilstanden "varig syk" til "uførepensjonist", vil endel personer dø før uførepensjonering. Eller rettere sagt at en ikke vet om de ville blitt uførepensjonert. Dette kaller vi høyresensurering fordi tidsaksen gjerne tegnes opp fra venstre mot høyre. Venstresensurering opptrer når vi ikke har starten av tiden som varig syk.

Ved studier basert på registerdata av tiden fra uføretidspunkt til registreringstidspunkt vil vi miste mange med lang tid mellom de to tidspunktene pga. at tidspunktene registreres samtidig i registrene. Uføretidspunkt kjennes bare når registreringstidspunkt er kjent. Lange varigheter før uførepensjonering selekteres derfor bort i en tidsbegrenset studie. En annen følge av at uføretidspunkt registreres samtidig med uførepensjonen er at personer som dør etter at de er erklært ervervsmessig uføre, men som ikke har fått søkt om uførepensjon, faller ut. Korte levetider som ervervsufør selekteres derfor bort.

Ved studier av tid i uførepensjonsordningen før død må det fastsettes

hva som er den relevante start for uførepensjonen. Deretter måler en tiden fram til avgang ved død. Et forsøksopplegg vil omfatte data både for avsluttede og ikke avsluttede livsløp. For de løpende uførepensjoner ved avslutning av forsøket er sensurtidene målt i forhold til starttidspunktet.

5.4 Mulighet for å estimere varighetsavhengighet

Det er ikke alle forsøksopplegg som gir anledning til å måle levetider i tilstandene "varig syk" og "ufør" og overganger til hhv. "ufør" og "død". Den vanligste måten å presentere data på er som tverrsnitt, f. eks. antall uføre pr. 31.12. Figuren nedenfor skisserer 4 ulike typer dataorganiseringer hvor bare en er ideell. De tre andre forsøksopplegg vil på ulike måter legge begrensninger på hvilke typer konklusjoner som det er mulig å trekke.

	Varighetsavhengig	Ikke varighetsavhengig
Kohortundersøkelse	Ideell	
Tverrsnittsundersøkelse		X

En kohortundersøkelse hvor vi kjenner varighet i tilstanden varig syk eller som uførepensjonist er den ideelle situasjon. Ved tverrsnittsundersøkelser er det noe mer problematisk. Hvis vi kjenner varigheten i tilstanden, kan vi trekke inn dette i estimeringen for hver aldersgruppe. En vil da ha data om antall overganger for ulike aldersgrupper og estimerer overgangsrater for disse gruppene. Det en egentlig bør gjøre er å estimere ratene innen grupper definert både ved hvor lenge sykdommen har vart og individets alder idet sykdommen ble varig. Da er det mulig å estimere fordelinger for hvor lenge det går fra individet blir varig syk til det uførepensjoneres og forstå hvordan denne fordelingen avhenger av individets alder.

Når vi ikke kjenner varigheten i tilstanden "syk" eller "ufør", og i tillegg har tverrsnittsdata kan en fortsatt beregne overgangsrater for ulike aldersgrupper. Disse kan så legges "etter hverandre" til en syntetisk populasjon, se avsnitt 3.2.2.slik at vi får en levetidsfordeling. Men vi mister da muligheten for å trekke inn betydning av individets alder ved hhv. sykdomsstart og uførestart.

De forhold som påvirker overgang til uførepensjon endrer seg lite i løpet av korte tidsperioder, f. eks. ett år. Virkninger av endringer i behandlingstilbud for ulike varige sykdommer, endringer i samlivsmønster og i forhold på arbeidsmarkedet slår ikke ut i kraftige endringer i løpet av ett år. En lovendring eller vedtektsendring kan imidlertid få umiddelbar innflytelse på rekruttering eller avgang.

En studie som følger varig syke i 1975 vil imidlertid kunne ha andre nivåer på de stykkevis konstante intensitetene enn en studie som følger varig syke i 1985. Dette betyr jo bare at "overlevelsesfunksjonen" som varig syk før overgang til uførepensjon endrer seg.

Forutsetningene om at levetiden for hvert individ har samme fordelingsfunksjon er trolig ikke holdbar og vi må enten dele inn i grupper eller bruke mer avanserte metoder for å få gode estimater for overlevelsesfunksjoner i de ulike tilstander, se kapittel 4. I neste kapittel skal vi se på viktige faktorer som bør trekke inn ved undersøkelser av rekruttering til og avgang fra uførepensjonsordningen.

6. Faktorer som påvirker tilgang til og opphør fra uførepensjonsordningen.

I kapittel 3 ble det lagt vekt på at den populasjon man betrakter må være rimelig homogen. Det betyr at den ikke består av grupper som har svært ulike overlevelsesfunksjoner. Da "tappes" populasjonen først av de som har korte overlevelsestider og deretter de som har lange overlevelsestider. Hvis man i en senere undersøkelse har en annen sammensetning av grupper, kan det være denne endring som gjenspeiles i en endret levetidsfunksjon og ikke en reell endring i de enkelte grupperes levetidsfunksjon.

6.1 Tilgang til uførepensjonsordningen

Tilgang til ordningen reguleres av mange ulike faktorer som f. eks. sykkeligheten i befolkningen og helsevesenets "reparerende" egenskaper, arbeidsmarkedets fleksibilitet overfor å ta vare på og ta inn igjen

syke arbeidstakere og holdninger og ressurser hos den enkelte, hos behandlende lege, i fylkesnemda og i trygderetten. Det vil for de aller fleste søkere av uførepensjon i teorien være et valg om en vil gå over på uførepensjon eller ei, men i praksis blir valget ofte avgjort av manglende mulighet for rehabilitering til arbeid.

En persons alder har betydning for sykелighet og type sykелighet. Eldre mennesker pådrar seg lettere sykdommer og det tar lengre tid å bli frisk igjen. Personer over 50 år vil dessuten få større problemer med å finne annet arbeid av lettere type hvis en blir delvis ufør. Sjansen for å gå over på uførepensjon varierte i 1983 fra 0.1 prosent for personer i aldersgruppen 20-29 år, til nesten 6 prosent i aldersgruppen 65-66 år. Sjansen er beregnet som antall nye registrerte uførepensjoner i 1983 dividert med antall personer uten uførepensjon. Varig sykdom er ikke trukket inn.

Menn og kvinner har ulik sykелighet, ulikt forhold til arbeidsmarkedet og ulike holdninger til det å bli forsørget av eventuell ektefelle. Tradisjonelt har kvinner hatt en lavere sjanse for å gå over på uførepensjon enn menn. Men dette er i ferd med å snu seg. I 1983 var sjansen for å gå over på uførepensjon 0.98 prosent for menn og 0.92 prosent for kvinner. Kvinner har større sjanse for å gå over på uførepensjon enn menn i aldersgrupper under 60 år.

Det er klart at ulike forhold på arbeidsmarkedet vil ha innflytelse på om syke mennesker kan bli i jobbene sine eller eventuelt få andre jobber. Det er gjort endel undersøkelser rundt disse forhold, men det er fortsatt mange uavklarte sider ved problemet. Såkalte retrettstillinger ved høy alder er mindre vanlig enn før, nå "sluses" eldre arbeidstakere ut på førtidspensjon. For personer som kan sykemeldes er sykepengene og uførepensjon et rimeligere alternativ for bedriftene enn en ren bedriftspensjon. Hvis slike forhold skal trekkes inn i en analyse, er det flere måter å forholde seg på. Utifra kunnskap om arbeidsmarkedet kan en fastsette tidspunkter (år) hvor det var vanskelig å få eller holde på arbeid og trekke dette inn i forhold til når personen ble varig syk. Opplysninger om yrkesaktivitet før uførepensjon kan utnyttes på ulike måter og i en intervjuundersøkelse kan en få konkrete opplysninger om betydningen av problemer i forhold til å holde på arbeidet sitt eller få nytt arbeid når varig sykdom har oppstått.

Helsevesenets rolle er også delvis klarlagt. Det viser seg at jo lenger tid det tar å få rehabilitert en syk person, jo større er sjansen for å gå over på uførepensjon. Dårlig behandlingstilbud, køer for operasjoner etc. vil påvirke personens muligheter for å gå tilbake i arbeid. Ved en intervjuundersøkelse kan det kartlegges hvordan rehabiliteringsperioden har forløpt for de enkelte personene.

Type sykdom og behandlingsmuligheter for de ulike sykdommer påvirker også sjansen for å gå over på uførepensjon. Personer med nervøse lidelser har f. eks. mye større sjanse for å gå over på uførepensjon enn personer med hjertekarsykdommer eller skjelett-muskelsykdommer. Hvis vi deler populasjonen inn i grupper etter type sykdom, vil vi ha fanget opp mye av den forskjell i levetid som eksisterer innen hele populasjonen av varig syke.

Personlige ressurser og støtte fra en primærgruppe påvirker sjansen for å bli syk og også sjansen for å gå over på uførepensjon. Det er f. eks. påvist at gifte har en lavere sjanse for å gå over på uførepensjon enn ikke gifte, for begge kjønn, type sykdom og i ulike aldersgrupper. Et annet og foreløpig åpent spørsmål er om endring i ekteskapelig status fører til kortere eller lengre tid som varig syk før uførepensjonering.

Trygdeetatens organisering og det dårlige tilbudet om attføring blir også ansett for å være medvirkende til at så mange ender opp med uførepensjon. Hvis det er mulig å kartlegge variasjoner i dette over år, kan dette trekkes inn som en forklaringsvariabel. Ellers er det mulig ved intervju å kartlegge enkelt personers attføringstilbud og be om en vurdering av om dette har hatt inn flytelse på tiden som varig syk.

Forhold mellom inntekt før og etter pensjonering blir trukket inn som et diskusjonstema i debatten om trygd er et "attraktivt" valg eller ei. Dersom en persons økonomiske forpliktelser ikke kan oppfylles med uførepensjonens størrelse, vil det for mange være viktig å fortsette i arbeidslivet. En persons dekningsgrad, dvs. forhold mellom uførepensjon og inntekt før sykdom, kan derfor trekkes inn som forklaring på eventuell overgang til uførepensjon. En mer realistisk dekningsgrad er vel uførepensjonens størrelse sett i forhold til den teoretiske inntekt etter sykdom. Den siste størrelse er imidlertid svært vanskelig å måle.

Det er store regionale forskjeller i rekrutteringen til uførepensjon, men det er ikke avklart om dette skyldes forskjeller i strukturen av andre kjennemerker nevnt over eller om det er en egeneffekt av region eller bostedsstrøk.

6.2 Avgang fra uførepensjonsordningen

Det kan umiddelbart se litt enklere ut å få innsikt i hvilke prosesser som påvirker avgang fra uførepensjonsordningen. Tre firedeler av opphørte uførepensjoner omfatter personer som fyller 67 år og dermed går over på alderspensjon. Den siste firedelen er personer som dør, minus en til tre prosent hvis pensjon opphører av andre årsaker.

Avgang fra uførepensjonsordningen vil imidlertid også kunne variere med sammensetningen av tilgangen til ordningen. Nye grupper med en annen dødelighet kommer inn og påvirker dødsintensitetene. Det er derfor ikke umiddelbart enkelt å si hva som er årsak til endringer i levetiden. Det er derfor nødvendig å se på tilleggsfaktorer ved studier av dødelighet for uførepensjonister.

SSB analyserer dødelighet i befolkningen, og vi vet at alder, kjønn, type sykdom og type yrke eller sosial status har stor betydning. Livsstil er blitt et begrep når en skal forklare sykkelighet og død og en tenker da på røykevaner, spisevaner, sovevaner og stress på jobb og hjemme. Slike opplysninger kan skaffes ved intervju.

Uførepensjonister hadde i 1977 en høyere dødelighet enn befolkningen. For aldersgruppen 17-66 år var dødeligheten for uførepensjonister 8 ganger høyere enn dødeligheten for befolkningen i samme aldersgruppe. Forskjellen var noe større for kvinner enn for menn. Og forskjellen mellom dødeligheten for uførepensjonister og befolkningen var størst blant de yngste. Disse beregningene er et estimat på overgangssintensiteten i 1977 hvor en har delt inn i grupper etter kjønn og alder. Det er ikke trukket inn tid i ordningen.

7. Datakildenes betydning for valg av metode

Vi har startet med å beskrive den ideelle situasjon hvor vi for hvert individ har en fortløpende registrering for den aktuelle populasjon av de faktorer som kan ha en betydning for levetidsfordelingene. Muligheten for å bruke tverrsnittsdata til å estimere levetider er også nevnt. Vi skal her se på de mulige datakilder som foreligger og se hvilke begrensninger som dette setter på valg av metode.

7.1 Datakilder i Rikstrygdeverket

Rikstrygdeverket har to hoveddatakilder for vårt formål. Den første er den rekonstruerte folketrygdbasen (RF-basen). Her ligger det pensjonsopplysninger om alders-, uføre-, og etterlattepensjonister. Dessuten ligger det opplysninger om pensjongivende inntekt for hele befolkningen fra 1967 og til siste år med slike opplysninger fra ligningsetaten.

Den andre hoveddatakilden er dynamisk statistikkregister (DSR). Dette er også en database og den inneholder noe færre opplysninger enn RF-basen for hver pensjonist. I tillegg har den opplysninger om grunn- og hjelpestønad som RFbasen ikke har. De to basene blir matet samtidig med opplysninger, men det er noen færre som går inn i DSR enn i RF-basen. For de uføre ligger det her opplysninger om uføretidspunkt, pensjonens størrelse (rettighetsbeløp), sykdomsdiagnoser, grad av uførhet etc.

De to basene ble etablert 1. november 1983. Tidligere registre ble konvertert inn i det nye systemet. Opplysninger om uføretidspunkt for uførepensjonister fulgte med, men det er dessverre oppdaget en feil for 5-20 prosent av uførepensjonistene. Denne feilen må rettes opp før det blir aktuelt å ta ut opplysninger til forløpsanalyser med uføretidspunkt som starttidspunkt.

Fra og med november 1983 er det mulighet for å lage en fullstendig uførehistorikk. Det ligger opplysninger om uføretidspunkt, registreringstidspunkt av pensjonen og eventuelt opphørstidspunkt. Ingen personer blir tatt ut selv om pensjonen opphører. Rikstrygdeverket har imidlertid få ressurser til å kontrollere opplysningene og foreløpig (for statistikkåret 1986) stoler de bare på tilgangsupplysningene og ikke avgangsupplysningene. Gruppen av

opphørte uførepensjoner for 1986 etableres fra DSR ved å plukke ut personer med uførepensjon som fylte 67 år i 1986 (og dermed gikk over på alderspensjon) og uførepensjonister som døde i 1986.

Fra DSR blir det hver måned kjørt ut en fil med pensjonister og personer med grunn- eller hjelpestønad, kalt datagrunnlag 1 (GR1). Denne filen danner grunnlag for Rikstrygdeverkets månedsstatistikk, kvartalsstatistikk og årsstatistikk. Et annet datagrunnlag, GR2, dannes også hver måned, men brukes bare hvert kvartal til statistikk over ektepar hvor begge har pensjon eller hvor det ytes forsørgingstillegg for ektefelle. Et tredje datagrunnlag, GR3, dannes ved utgangen av hvert år til Rikstrygdeverkets årsstatistikk. Denne filen inneholder bare personer som er registrert i DSR pr. 31.12, mens de to andre datagrunnlagene også inneholder tilgangs- og avgangsupplysninger.

Statistisk Sentralbyrå har betalt for et program som henter ut uføregrad, pensjonsgivende inntekt, pensjonspoeng og beregnet uførepoeng for enkelt personer for alle år fra 1967 til siste inntektsår. Dette programmet kan brukes til å hente ut opplysninger om den del av befolkningen som har hatt inntektsgivende arbeid en eller annen gang siden 1967 eller som har vært eller er uførepensjonert. Det er imidlertid begrenset hvilke tilleggsopplysninger som er tilgjengelige.

7.2 Datatkilder i Statistisk Sentralbyrå

Statistisk Sentralbyrå har laget statistikk over uførepensjonister for årene 1977, 1980 og 1983. Det arbeides nå med datauttak for statistikkåret 1986. I SSB foreligger GR1 for 31.12 1985 og 1986. Rikstrygdeverket stoler ikke på registreringen av tilgang på uførepensjoner på denne fil for 1985, mens de velger å gjøre det for 1986. Avgangsupplysningene stoler de ikke på for noen av årene.

SSB har hatt sin egen måte å finne omfanget av tilgangen på for alle statistikkårene, og denne metoden vil det være mulig å bruke også for 1986. For 1986 mistes imidlertid pensjoner som kom til og opphørte i løpet av året.

Til uførestatistikken kobles RTVs data med opplysninger fra SSBs personregister(BSR), ligningsregisteret og for 1983 fra bostøtte-

registeret. Dette gir oss opplysninger om bl.a. ekteskapelig status, og inntekt i statistikkåret. Fra den koblede folketellingsfil med opplysninger fra 1960-70-80 er det mulig å hente elementer av en yrkeshistorie.

7.3 Dataenes innflytelse på valg av metode

En analyse av rekruttering til uførepensjon bør ta utgangspunkt i en populasjon av varig syke mennesker hvor vi kjenner startpunktene i tilstanden varig syk. Deretter følges personer over tid og en registerer om personer dør eller går over på uførepensjon. Et slikt forsøksopplegg er ikke mulig utifra de registre som er tilgjengelig. Det må derfor lages en ny langvarig dataregistrering for at forløpsanalysemetodene kan brukes direkte.

Tverrsnittsdataene i SSB kan utnyttes som beskrevet i avsnitt 2. Situasjonen vi betrakter svarer til den minst attraktive rute fra figuren på side 24 , og i tillegg er ikke "occurence" (nye uførepensjonister) og "exposure" (tiden varig syke mennesker er eksponert) fra samme populasjon. Nye uførepensjonister er hentet fra RTVs registre for 1983 og tiden varig syke mennesker er eksponert for overgang i ett år er beregnet med grunnlag i Helseundersøkelsen 1985. Det kan tenkes at det resultat vi får ikke avviker så mye fra det vi ville fått ved et bedre forsøksopplegg, men vi har ingen mulighet for å kontrollere dette.

Avgangsstudier på grunnlag av SSBs tverrsnittsdata har et noe bedre grunnlag, ihverfall hvis en ønsker å bruke uføretidspunkt som start. Her blir imidlertid størrelsen av utvalget trolig for lite til å få pålitelige resultater bare basert på ett års avganger. Vi kan heller ikke få vurdert om det "vindu" vi ser gjennom gir oss et godt bilde på den løpende prosess.

Registrene i RTV kan gi et noe bedre grunnlag for å studere hvilke faktorer som har innflytelse på overgangsratene til uførepensjon. Etter en fem års tid med løpende tilgangs og avgangsregistrering vil vi ha et større "vindu" å kikke inn gjennom. Men først må det avklares om en kan stole på de registreringer som er foretatt.

8. Anbefalinger og mulig videreføring av prosjektet.

Vi har i dette notatet gått gjennom noen sentrale begreper i overlevelsesanalyse, sett på noen enkle modeller og diskutert hvordan disse kan estimeres ved å bruke data fra Rikstrygdeverket og SSB. Et hovedpoeng er at overlevelsesanalyse i sin ideelle form forutsetter data om individuelle livsløp. Mangler dette, for eksempel ved at data bare foreligger på aggregert og i særdeleshet på tverrsnittsform, kan ideer og begreper fra overlevelsesanalyse fortsatt med fordel brukes. Det blir imidlertid mer problematisk å tolke resultatene. Med alle sine begrensninger vil slike analyser likevel utvilsomt være av verdi. Et eksempel på hva en kan oppnå er gitt i kapittel 2 ovenfor. Slike analyser er meget enkle å gjennomføre og vi vil anbefale at noe lignende gjøres i tilsvarende undersøkelser ellers i SSB. Hvorvidt man skal presentere resultatene i form av estimater er en annen sak. Poenget er at en kan bli oppmerksom på trekk ved observasjonsmaterialet, som det kan være vanskelig å få øye på ved bare å sammenligne krysstabeller.

Data om uførepensjonister i SSB er stort sett på tverrsnittsform. Dette begrenser muligheten for analyser basert på metoder fra overlevelsesanalyse. På den andre siden er det klart at Rikstrygdeverket, i alle fall på litt sikt, vil ha store mengder data som med fordel kan egne seg for analyse med slike metoder. Det bør derfor vurderes i hvilken utstrekning slike data bør hentes ut av RTV's databaser til bruk i statistikkproduksjon og analyseformål. En annen mulighet er å lagre tverrsnittsdata, men på en slik form at individuelle forløp kan rekonstrueres. Det er selvfølgelig et problem i denne sammenheng at hvilke kjennetegn som bør tas med, og formen bør data lagres på, avhenger av hvilke forhold en ønsker å belyse.

Dette forprosjektet kan videreføres på flere ulike måter. En studie

som kan gjennomføres uten innsamling av nye data fra Rikstrygdeverket eller andre, er å se på tilgangen i 1986 på samme måte som i avsnitt 2. Utover de koblinger som allikevel vil bli gjort til uførestatistikken for 1986, vil det kreve en kobling med folketellingen 1980 for å skaffe opplysninger om yrkesaktivitet. En kan da også utvide og trekke inn andre kjennemerker.

Avgangen av uførepensjonister i 1986, som hentes inn til uførestatistikken for samme år, kan analyseres med de samme metoder som er brukt for tilgangen. Her er det også mulighet for å trekke inn et startpunkt og dermed tid som uførepensjonist.

En større analyse basert på tilgang og avgang i en lengre periode, f. eks. 1.1.84 - 31.12.88, fra Rikstrygdeverkets registre, vil gi et bedre grunnlag for å estimere overganger. Hvis det blir alvor av en budsjettmodell for uførepensjonister, er det rimelig å koordinere en innsats på dette feltet med arbeidet i Gruppen for arbeidsmarked, skatt og offentlig økonomi.

En annen mulighet som foreligger, er å ta utgangspunkt i Helseundersøkelsen 1985 og følge opp de intervjuede personer med koblinger med Rikstrygdeverkets registre. Da vil vi vite hvem som er uførepensjonister, og hvem som blir det f.eks i en 5-årsperiode fram til 1990. En slik studie gir mye bakgrunnsinformasjon, også om varigheten som syk. Fordelen er at alle dataene foreligger i SSB, ulempen at vi må vente til begynnelsen av 1991 for å få ut resultater.

Det er altså flere mulige måter å følge opp forprosjektet på. De to første forslagene krever lite resurser i sin enkleste form, og bør gjennomføres for å vedlikeholde kompetansen samtidig som resultatene er interessante. Det tredje forslaget krever mer innsats, men vil også kaste mer av seg. Det siste forslaget er interessant, men resultatene vil ikke foreligge på lang tid.

Referanser.

Cox, D. R. og Oakes, D. (1984). Analysis of Survival Data. Chapman and Hall, London.

Laird, N. og Oliver, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. J. Amer. Statist. Assoc. vol. 76 231-240.

Appendiks A. Noen tekniske utledninger.

A.1 Utledning av sannsynlighetskvoten for faktisk kohort.

Vi betrakter situasjonen fra avsnitt 2.2.1.

For å knytte observasjonen (X_i, δ_i) til intervallene definerer vi størrelsene (X_{1i}, \dots, X_{ki}) og $(\delta_{1i}, \dots, \delta_{ki})$. X_{ji} er tiden individ i tilbringer i intervallet $[t_j, t_{j+1})$ før overgang, slik at $X_i = \sum_{j=1}^k X_{ji}$. δ_{ji} er lik 1 hvis det observeres en overgang i intervallet $[t_j, t_{j+1})$, 0 ellers. For sensurerte observasjoner er derfor δ_{ji} lik 0 for alle $j=1, \dots, k$. Også $\delta_i = \sum_{j=1}^k \delta_{ji}$.

Vi skal nå utlede sannsynlighetskvoten til observasjonene. Hvis det foregår en overgang i intervallet $[t_j, t_{j+1})$ er $X_{j+1i} = \dots = X_{ki} = 0$, og $\delta_{ji} = 1$. Bidraget til sannsynlighetskvoten er derfor $\lambda_j \exp(-\sum_{j=1}^k \lambda_j X_{ji})$,

som kan skrives $\prod_{j=1}^k \lambda_j^{\delta_{ji}} \exp(-\lambda_j X_{ji})$. Hvis det er sensurering i intervallet $[t_j, t_{j+1}]$, er også $X_{j+1i} = \dots = X_{ki} = 0$, men $\delta_{ji} = 0$. Bidraget til sannsynlighetskvoten er i dette tilfellet $\exp(-\sum_{j=1}^k \lambda_j X_{ji})$, som også

kan skrives $\prod_{j=1}^k \lambda_j^{\delta_{ji}} \exp(-\lambda_j X_{ji})$.

Sannsynlighetskvoten for alle observasjonene er derfor

$$\prod_{i=1}^n \prod_{j=1}^k \lambda_j^{\delta_{ji}} \exp(-\lambda_j X_{ji}) = \prod_{j=1}^k \lambda_j^{d_j} \exp(-\lambda_j T_j)$$

der $d_j = \sum_{i=1}^n \delta_{ji}$ er det totale antall observerte overganger i intervallet $[t_j, t_{j+1})$, og $T_j = \sum_{i=1}^n X_{ji}$ er den totale tiden individene er observert i intervallet.

A.2 Utledning av sannsynlighetskvoten for en virtuell kohort.

La situasjonen være som i avsnitt 2.2.2. I dette tilfellet blir det individuelle bidraget til sannsynlighetskvoten

$$\lambda_j^{\delta_{ji}} \exp(-\lambda_j X_{ji}), \quad i=1, \dots, n_j, \quad j=1, \dots, k.$$

For alle observasjonene blir derfor sannsynlighetskvoten

$$\prod_{i=1}^n \prod_{j=1}^k \lambda_j^{\delta_{ji}} \exp(-\lambda_j X_{ji}) = \prod_{i=1}^n \prod_{j=1}^k \lambda_j^{\delta_{ji}} \exp(-\lambda_j X_{ji}) = \prod_{j=1}^k \lambda_j^{d_j} \exp(-\lambda_j T_j)$$

der $d_j = \sum_{i=1}^n \delta_{ji}$ og $T_j = \sum_{i=1}^n X_{ji}$. Dette gir igjen forholdet mellom antall observerte overganger og den totalt observerte tid for eksponering som estimator for overgangsraten λ_j .

Appendiks B. Tabeller over nye uføretilfeller i 1983 og estimert eksponering på grunnlag av helseundersøkelsen 1985.

Tabell B.1. Antall nye uføretilfeller i 1983.

Alder	Ekteskapelig status	Yrkesaktivitet	Nervøse lidelser	Hjerte- og kar-sykdommer	Skjelett- og muskel-sykdommer	Andre sykdommer	
30-39	UF-FG	YAKT	217	14	25	130	
30-39	UF-FG	IYAKT	51	0	3	20	
30-39	GIFT	YAKT	57	5	45	93	
30-39	GIFT	IYAKT	5	0	1	5	
40-49	UF-FG	YAKT	200	37	78	125	
40-49	UF-FG	IYAKT	30	3	5	15	
40-49	GIFT	YAKT	114	65	166	175	
40-49	GIFT	IYAKT	8	2	2	10	
50-54	UF-FG	YAKT	113	65	79	123	
50-54	UF-FG	IYAKT	19	2	9	11	
50-54	GIFT	YAKT	112	175	237	224	
50-54	GIFT	IYAKT	3	2	2	7	
55-59	UF-FG	YAKT	158	105	166	176	
55-59	UF-FG	IYAKT	18	5	5	14	
55-59	GIFT	YAKT	191	453	482	419	
55-59	GIFT	IYAKT	9	5	3	3	
60-66	UF-FG	YAKT	208	283	348	332	
60-66	UF-FG	IYAKT	12	14	13	20	
60-66	GIFT	YAKT	548	1414	1519	1278	
60-66	GIFT	IYAKT	12	13	19	25	
<hr/>							
30-39	UF-FG	YAKT	110	5	38	73	
30-39	UF-FG	IYAKT	55	4	15	50	
30-39	GIFT	YAKT	96	12	107	128	
30-39	GIFT	IYAKT	41	9	50	66	
40-49	UF-FG	YAKT	146	12	141	122	
40-49	UF-FG	IYAKT	93	5	29	47	
40-49	GIFT	YAKT	164	45	349	186	
40-49	GIFT	IYAKT	64	15	99	80	
50-54	UF-FG	YAKT	106	27	142	79	
50-54	UF-FG	IYAKT	45	5	35	24	
50-54	GIFT	YAKT	159	65	369	170	
50-54	GIFT	IYAKT	40	19	95	101	
55-59	UF-FG	YAKT	130	50	221	128	
55-59	UF-FG	IYAKT	65	21	52	43	
55-59	GIFT	YAKT	201	137	550	257	
55-59	GIFT	IYAKT	55	41	143	111	
60-66	UF-FG	YAKT	231	198	500	360	
60-66	UF-FG	IYAKT	62	100	175	131	
60-66	GIFT	YAKT	231	299	923	521	
60-66	GIFT	IYAKT	72	136	233	159	

M
e
n
n

K
v
i
n
n
e
r

UF-FG: Ugift/Før gift
YAKT: Yrkesaktiv
IYAKT: Ikke yrkesaktiv

Tabell B.2. Estimert eksponering på grunnlag av Helseundersøkelse (1985), Folkemengden etter alder og ekteskapelig status pr 31.12.83 (NOS B473), Folketellingen 1980, hefte II (NOS B299) og Trygdestatistikk for uføre 1983 (NOS B646).

Alder	Ekteskapelig status	Yrkesaktivitet	Nervøse lidelser	Hjerte- og kar-sykdommer	Skjelett- og muskel-sykdommer	Andre sykdommer	
30-39	UF-FG	YAKT	5.247	1.960	20.875	26.196	
30-39	UF-FG	IYAKT	-0.286	0.929	3.860	1.960	
30-39	GIFT	YAKT	16.736	14.957	76.750	135.373	
30-39	GIFT	IYAKT	0.566	0.957	1.875	-0.495	
40-49	UF-FG	YAKT	0.713	10.347	15.552	10.569	
40-49	UF-FG	IYAKT	0.105	-0.135	0.715	1.412	
40-49	GIFT	YAKT	25.113	35.507	69.001	93.585	
40-49	GIFT	IYAKT	0.571	3.539	3.062	-0.634	
50-54	UF-FG	YAKT	-0.964	1.790	5.662	8.941	
50-54	UF-FG	IYAKT	0.540	0.603	2.058	-1.332	
50-54	GIFT	YAKT	4.286	24.069	35.799	23.400	M
50-54	GIFT	IYAKT	-0.057	3.412	0.175	1.920	e
55-59	UF-FG	YAKT	0.034	6.007	2.400	0.014	n
55-59	UF-FG	IYAKT	-4.901	1.687	2.162	-1.058	n
55-59	GIFT	YAKT	5.661	37.462	25.641	25.476	
55-59	GIFT	IYAKT	-1.373	3.540	0.117	-0.350	
60-65	UF-FG	YAKT	-1.325	7.335	1.074	8.512	
60-65	UF-FG	IYAKT	-1.346	3.570	2.592	0.602	
60-65	GIFT	YAKT	1.610	39.472	39.602	22.458	
60-65	GIFT	IYAKT	-3.485	15.458	0.375	-2.153	
30-39	UF-FG	YAKT	11.307	0.952	22.703	22.292	
30-39	UF-FG	IYAKT	1.760	2.890	0.593	0.739	
30-39	GIFT	YAKT	20.545	11.909	60.350	118.963	
30-39	GIFT	IYAKT	10.792	3.804	24.937	31.903	
40-49	UF-FG	YAKT	4.563	2.899	15.125	16.770	
40-49	UF-FG	IYAKT	1.352	2.319	2.158	-1.039	
40-49	GIFT	YAKT	24.910	26.780	53.175	64.559	
40-49	GIFT	IYAKT	0.021	7.417	23.604	13.656	
50-54	UF-FG	YAKT	0.777	4.725	15.911	3.044	
50-54	UF-FG	IYAKT	0.026	2.625	4.459	-1.757	
50-54	GIFT	YAKT	7.973	14.673	31.300	20.715	K
50-54	GIFT	IYAKT	0.520	0.516	4.104	7.332	v
55-59	UF-FG	YAKT	0.375	7.574	6.673	6.092	n
55-59	UF-FG	IYAKT	-2.721	1.710	0.190	-0.904	n
55-59	GIFT	YAKT	7.060	13.320	25.710	17.560	e
55-59	GIFT	IYAKT	7.590	8.675	9.413	10.243	r
60-65	UF-FG	YAKT	-0.356	9.619	13.065	4.471	
60-65	UF-FG	IYAKT	1.055	10.517	11.218	0.057	
60-65	GIFT	YAKT	4.013	15.581	16.710	15.554	
60-65	GIFT	IYAKT	0.974	30.405	22.035	14.044	

UF-FG: Ugift/Før gift
YAKT: Yrkesaktiv
IYAKT: Ikke yrkesaktiv

Tabell B.3. Glattede og oppblåste tall for eksponering.

Alder	Ekteskapsstatus	Yrkesaktivitet	Nervøse lidelser	Hjerte- og kar-sykdommer	Skjelett- og muskel-sykdommer	Andre sykdommer		
30-39	UF-FG	YAKT	1347	989	7283	11352		
30-39	UF-FG	IYAKT	120	256	1345	202		
30-39	GIFT	YAKT	7060	5523	29941	50083		
30-39	GIFT	IYAKT	93	390	1005	449		
40-49	UF-FG	YAKT	745	3001	4801	3753		
40-49	UF-FG	IYAKT	35	366	609	70		
40-49	GIFT	YAKT	9240	15195	27247	34354		
40-49	GIFT	IYAKT	127	926	1155	234		
50-54	UF-FG	YAKT	162	899	2645	2178		
50-54	UF-FG	IYAKT	78	754	869	236		
50-54	GIFT	YAKT	1551	8834	13070	10430	M	
50-54	GIFT	IYAKT	58	974	311	308	e	
55-59	UF-FG	YAKT	31	2584	1315	2154	n	
55-59	UF-FG	IYAKT	9	725	345	64		
55-59	GIFT	YAKT	2127	14505	9504	9872		
55-59	GIFT	IYAKT	90	1055	440	153		
60-66	UF-FG	YAKT	14	2229	1825	2027		
60-66	UF-FG	IYAKT	6	1819	1056	88		
60-66	GIFT	YAKT	608	15434	14306	9580		
60-66	GIFT	IYAKT	67	5714	2572	389		
<hr/>			<hr/>					
30-39	UF-FG	YAKT	4020	944	5412	7871		
30-39	UF-FG	IYAKT	1524	411	3275	806		
30-39	GIFT	YAKT	11231	3901	23529	46102		
30-39	GIFT	IYAKT	3284	2231	9253	12914		
40-49	UF-FG	YAKT	1543	2556	6136	5049		
40-49	UF-FG	IYAKT	252	419	1333	197		
40-49	GIFT	YAKT	9634	8734	22264	25890		
40-49	GIFT	IYAKT	2286	3563	8254	5218		
50-54	UF-FG	YAKT	779	1383	5669	1722		
50-54	UF-FG	IYAKT	645	771	1515	452		
50-54	GIFT	YAKT	2499	6155	12353	7177	K	
50-54	GIFT	IYAKT	739	2231	1514	2549	v	
55-59	UF-FG	YAKT	143	2376	3253	2043	i	
55-59	UF-FG	IYAKT	167	1082	1752	355	n	
55-59	GIFT	YAKT	2993	5435	9050	7216	n	
55-59	GIFT	IYAKT	2641	3193	4306	3352	e	
55-59	GIFT	YAKT	216	3884	4075	2599	r	
60-66	UF-FG	YAKT	285	3943	3737	509		
60-66	UF-FG	IYAKT	1623	4010	6753	5311		
60-66	GIFT	YAKT	2832	13732	9575	4373		
60-66	GIFT	IYAKT						

UF-FG: Ugift/Før gift
 YAKT: Yrkesaktiv
 IYAKT: Ikke yrkesaktiv

Appendiks C. Program for kjøring av programmet LOGLIN.

C.1. Kommandoer for estimeringsprogrammet som beskrevet i kapittel 7.

```

OPTIONS ECHO INLINE 60 DUTLINE 120
1  READ ARRAY OCC (2 2 5 ) UNIT 2 FORMAT
   ( F10.0)
2  READ ARRAY EXP (2 2 5 ) UNIT 3 FORMAT
   ( F10.0)
3  READ LABELS OCC UNIT 1
   REWIND UNIT 1
   READ LABELS EXP UNIT 1
4  FIT OCC (1 2 3 ) START EXP EST PRINT MAXITS 10
5  FIT OCC (1 2/1 3/2 3) &
   START EXP EST PRINT MAXITS 10
6  FIT OCC (1 2/1 3 ) &
   START EXP EST PRINT MAXITS 10
   FIT OCC (1 2/ 2 3) &
   START EXP EST PRINT MAXITS 10
   FIT OCC ( 1 3/2 3) &
   START EXP EST PRINT MAXITS 10
7  FIT OCC ( 1/2/ 3 ) &
   START EXP EST PRINT MAXITS 10
   PRINT ARRAY OCC EXP
   QUIT

```

Kommentarer:

- 1-2. Leser inn henholdsvis matrise for "occurrence" og "exposure" som ligger på UNIT 2 og UNIT 3 med format F10.0, d.v.s. observasjonene ligger i de ti første feltene ordnet slik at første faktor (yrkesaktivitet med to nivåer) varierer raskest, deretter sivilstand og tilslutt faktor 3 (alder med fem nivåer).
3. Leser inn variabelnavn for bruk i utskrift.
4. Tilpasser en mettet modell.
5. Tilpasser en modell der 2. ordens samspill er satt lik null.
6. Tilpasser en modell der 2. ordens samspill og 1. ordens samspill melleom faktor 2 og 3 settes lik null.
7. Tilpasser en modell med bare hovedeffekter.

C2. JCL kort for kjøring av LOGLIN programmet i C1.

```
    /* JCL-KORT FOR
    /* KJØRING AV PROGRAMMET LOGLIN
    /*
1   //GO          EXEC PGM=LOGLIN,REGION=2048K
    //STEPLIB DD DSN=L0810KR.LOGLIN.LOAD,DISP=SHR
    //          DD DSN=SYS1.VFORTLIB,DISP=SHR
2   //FT01F001 DD DSN=TK51.S4164.AUXLAB,DISP=OLD
3   //FT02F001 DD DSN=TK51.S4164.AUXOCC,DISP=OLD
4   //FT03F001 DD DSN=TK51.S4164.AUXEXP,DISP=OLD
5   //FT05F001 DD DSN=TK51.S4164.AUXCDE,DISP=OLD
    //FT06F001 DD SYSOUT=(*)
    //SYSOUT DD SYSOUT=*
    //SYSIN DD DUMMY
    /*
```

Kommentarer:

1. Starter programmet LOGLIN som ligger på fil L0810KR.LOGLIN.LOAD
2. Fil med variabelnavn svarende til UNIT 1 i LOGLIN koden i C1.
3. Fil med antall hendelser ("occurence") svarende til UNIT 2 i LOGLIN koden.
4. Fil med total eksponering ("exposure") svarende til UNIT 3 i LOGLIN koden.
5. Fil inneholdene LOGLIN koden fra C1.