Arbeidsnotater

STATISTISK SENTRALBYRÅ

IO 68/15 Oslo, 23. juli 1968

Modell for befolkningsprojeksjoner for norske regioner

av Eivind Gilje[x)]

Arbeidsdokument til Nordisk demografisk symposium i Mattby, Finland, 14. - 16. august 1968.

INNHOLD

x) Jeg vil takke Jan M. Hoem for råd og veiledning og Fridjof Wiese for at han har lest manuskriptkorrektur.

I. Innledning

Statistisk Sentralbyrå har i en tid arbeidet med opplegget til en større personmodell som er tenkt brukt til ulike analyseformål. Tanken er at den skal være en slags parallell til økonomiske modeller. Som et første skritt på veien har vi utarbeidet en prognosemodell som er brukt til å produsere regionale befolkningsprojeksjoner. Det er denne jeg skal presentere her.

Vi har bygd opp en modell som gir faktorprognoser,og den er ikke vesensforskjellig fra klassiske prognosemodeller av denne art. Det viktigste nye ligger i at vi har brukt mer spesifiserte data, og i at vi kan gi langt mer detaljerte resultater enn noen annen modell jeg har sett. Vi har foreløpig avstått fra videre raffinementer. Hovedformålet har i første omgang vært å lage en operabel modell som kan gi oss noenlunde brukbare resultater før vi begynner med statistiske og tekniske forbedringer.

Selv om moderne EDB-teknikk har gjort det mulig å gå svært detaljert til verks, er metodene på mange punkter fremdeles lite tilfredsstillende. Dette vil vi prøve å hjelpe på i den nystartede "Gruppe for personmodeller." Her utvikles for tiden en del demografiske modeller av mindre omfang, og analyser utføres på det datamaterialet som finnes i Byrået. Med tiden håper vi å kunne nytte resultatene av dette arbeid til å forbedre projeksjonsmodellen. Dette vil jeg kommentere nærmere i siste kapittel av denne fremstillingen.

I kapittel II skal jeg gi en beskrivelse av den modellen vi har laget, og her har jeg lagt vekt på å gi en riktig sannsynlighetsteoretisk begrunnelse for formlene i modellen.

Kapittel III omhandler estimeringen av strukturkoeffisientene.

I nest siste kapittel skal vi se litt på datagrunnlaget, og de restriksjoner dette har lagt på modellen.

II. Modellbeskrivelse

1. Konvensjoner og definisjoner

Jeg vil bruke følgende symboler som topp- og fotskrifter:

Venstre toppskrift:

Kjønn: M for menn, F for kvinner.

Høyre toppskrift:

Kommunenr.: k hvor $k = 1,2,\ldots,K$.

Venstre fotskrift:

Varighet: t.

Høyre fotskrift:

Alder: x hvor $x = 0,1,\ldots,\omega$. (ω er høyeste levealder.)

Videre vil n i en parentes etter symbolene stå for datoen 1/1 år n ved tilstandsbegreper og for året n ved strømningsbegreper.

For å belyse symbolbruken vil jeg gi følgende eksempler:

(i) ${}^{F}L_x^k(n)$ betyr faktiske antall x-årige kvinner i kommune k pr. 1/1 år n.

(ii) ${}^{F}D_x^k(n)$ betegner faktisk antall dødsfall i kommune k i året n blant kvinner som var x år gamle pr. 1/1 år n.

(iii) ${}^{F}_{t}q_x^k(n)$ er sannsynligheten for at en kvinne som pr. 1/1 år n bor i kommune k og er x år gammel, skal dø i kommunen innen hun blir x+t år gammel.

Andre tilstandsbegreper defineres analogt med ${}^{F}L_x^k(n)$, og andre strømningsbegreper defineres analogt med ${}^{F}D_x^k(n)$.

Videre vil vi innføre den konvensjonen at hvis X er stokastisk variabel, skal $\tilde{X}$ betegne en prediktor for X. (Amundsen, 1962, bind 2, side 198, og bind 3, side 252.) Således blir f.eks. ${}^{F}\tilde{L}_x^k(n)$ en prediktor for ${}^{F}L_x^k(n)$.

Et symbol som ${}^{F}_{t}\hat{q}_x^k(n)$ betegner en estimator for ${}^{F}_{t}q_x^k(n)$.

Mange av de relasjoner vi kommer fram til er analoge for menn og kvinner. Derfor innfører vi den typografiske forenkling at symboler uten kjønnsangivelse representerer ett enkelt, uspesifisert kjønn.

2. Prognosemodellen

I dette avsnittet skal jeg gå igjennom alle relasjonene som er brukt i modellen og begrunnelsen for disse. I avsnitt II.3 vil jeg foreta en oppsummering av modellrelasjonene.

Først skal vi sette opp en helt enkel regnskapslikning som vi vil arbeide videre på. La L betegne den faktiske folkemengden, D antall døde, U antall utflyttere og I antall innflyttere. Vi har da for $0 \leq x \leq \omega$:

$$(2.1) \quad L_{x+1}^k(n+1) = L_x^k(n) - D_x^k(n) - U_x^k(n) + I_x^k(n).$$

Når n refererer seg til fremtidige år vil imidlertid størrelsene på høyre side i (2.1) være ukjente. I en prognosemodell vil man da erstatte dem med predikerte verdier. I overensstemmelse med (2.1) kan vi så stille opp følgende <u>første relasjon i prognosemodellen</u>:

$$(2.2) \quad \tilde{L}_{x+1}^k(n+1) = \tilde{L}_x^k(n) - \tilde{D}_x^k(n) - \tilde{U}_x^k(n) + \tilde{I}_x^k(n)$$

for n = 0,1,... ; k = 1,2,...,K og $0 \leq x \leq \omega$. Her betegner n = 0 et utgangstidspunkt med kjent bestand, slik at $\tilde{L}_x^k(0) = L_x^k(0)$ for $0 \leq x \leq \omega$ og k = 1,2,...,K.

Det er nærliggende å gå ut fra estimatorer for forventningsverdiene til $D_x^k(n)$, $U_x^k(n)$ og $I_x^k(n)$ for å finne prediktorer for disse. Prediktorer for $L_x^k(n)$ beregnes så rekursivt av (2.2).

I det følgende skal vi se på disse estimatorene. La oss da først betrakte $D_x^k(0)$ og $U_x^k(0)$. Disse stokastiske variable er begge binomisk fordelt, og vi innfører $q_x^k(0)$ og $U_x^k(0)$ som henholdsvis ettårige influerte døds- og utflyttingssannsynligheter. (Sverdrup, 1961, side 23.)

De som dør i en kommune kan deles i to kategorier. Den ene omfatter de som bodde i kommunen ved årets begynnelse og så har dødd. I den andre er de som har flyttet inn i kommunen etter 1/1 og dødd på et senere tidspunkt i året. Dødsfallene blant innflytterne i løpet av året har vi antatt har neglisjerbar betydning og derfor blir:

$$(2.3)\quad E\, D_x^k(0) = q_x^k(0)\ .\ L_x^k(0).$$

I analogi med den tilnærmelse som er gjort for de døde, antar vi også at flere flyttinger av samme person i løpet av ett år har neglisjerbar betydning, og dette fører til:

$$(2.4)\quad E\, U_x^k(0) = u_x^k(0)\ .\ L_x^k(0).$$

Denne siste antakelse vil forøvrig bli kommentert nærmere til slutt i dette avsnittet.

Så skal vi finne et uttrykk for innflytterne. Først har vi:

$$(2.5)\quad E\, U_x(0) = \sum_{k=1}^{K} E\, U_x^k(0).$$

La så $i_x^k(0)$ være sannsynligheten for at en utflytter i alder x 1/1 år 0 skal flytte inn i kommune k i løpet av samme år. Da blir:

$$(2.6)\quad E\,\{I_x^k(0)\,|\,U_x(0)\} = i_x^k(0)\ .\ U_x(0)$$

og dermed:

$$(2.7)\quad E\, I_x^k(0) = i_x^k(0)\ .\ E\, U_x(0).$$

Bak denne flytteprosedyre ligger en teori som kan formuleres på følgende måte: En persons tendens til å flytte ut fra den kommune han bor i er bare avhengig av egenskaper ved denne kommunen. Dette fører til (2.4). Hvor personen bestemmer seg til å flytte hen, avhenger derimot ikke av utflyttingskommunen i det hele tatt, men av egenskaper ved de mulige innflyttingskommunene. Gitt at en x-årig person bestemmer seg for å flytte, vil den betingede sannsynligheten for at han skal flytte til kommune k kunne skrives som en funksjon i_x^k som er uavhengig av hvor han bodde ved årets begynnelse. Dette gir en formel som (2.7). (Hoem, 1968, side 14.)

Som prediktorer for $D_x^k(0)$, $U_x^k(0)$ og $I_x^k(0)$ vil vi bruke estimatorer for forventningsverdiene og får så likningene (2.8), (2.9), (2.10) og (2.11) for n = 0. Disse kombineres med (2.2) og dermed kan vi la likningssettene gjelde for alle n.

(2.8) $\tilde{D}_x^k(n) = \hat{q}_x^k \cdot \tilde{L}_x^k(n).$

(2.9) $\tilde{U}_x^k(n) = \hat{u}_x^k \cdot \tilde{L}_x^k(n).$

(2.10) $\tilde{U}_x(n) = \sum_{j=1}^{K} \tilde{U}_x^j(n).$

(2.11) $\tilde{I}_x^k(n) = \hat{i}_x^k \cdot \tilde{U}_x(n).$

der $k = 1,2,\ldots,K$; $0 \leq x \leq \omega$ og $n = 0,1,\ldots$.

Vi har ikke datert estimatorene før strukturkoeffisientene $q_x^k(n)$, $U_x^k(n)$ og $i_x^k(n)$. Grunnen er at vi i modellen lar dødeligheten og flyttetendensen være den samme som i utgangspret gjennom hele prognoseperioden. Vi vil nedenfor gjøre samme antakelse for forventet antall fødte.

Av (2.2),(2.8), (2.9), (2.10) og (2.11) kan vi nå gi prognoser for alle aldersgrupper unntatt de som fødes i prognoseåret.

Så skal vi se litt på disse nyfødte. La $f_x^k(n)$ være det forventede antall levende barn som fødes i året n av en kvinne som pr. 1/1 år n er x år gammel og bor i kommune k (Hoem, 1967, side 57), og la $B_x^k(n)$ være det antall barn disse ${}^F L_x^k(n)$ kvinnene faktisk får i dét året. Da er

$$E\,\{E(B_x^k(n) \mid {}^F L_x^k(n)\} = E\{f_x^k(n) \cdot {}^F L_x^k(n)\}$$

som gir

$$E\,B_x^k(n) = f_x^k(n) \cdot E\,{}^F L_x^k(n)$$

og dermed blir forventet totalt antall fødte i bestanden i år 0

(2.12) $E\,B^k(0) = \sum_{x=15}^{44} f_x^k(0) \cdot {}^F L_x^k(0)$

hvor vi har summert over det vi har regnet som kvinnenes fødedyktige aldre. I tråd med tidligere utledninger finner vi følgende prediktorer:

(2.13) $\tilde{B}^k(n) = \sum_{x=15}^{44} \hat{f}_x^b \cdot {}^F\tilde{L}_x^k(n)$

for $k = 1,2,\ldots,K$; $0 \leq x \leq \omega$ og $n = 0,1,\ldots$. Her må vi være oppmerksomme på at også (2.12) og (2.13) er tilnærmelsesformler. Analogt med tidligere har vi neglisjert fødsler av kvinner som flytter inn i kommune k og føder i samme kommune og samme år.

La ${}^{F}c$ betegne andelen av piker i et fødselskull. Når vi regner med at fødslene"i gjennomsnitt finner sted midt i året", kan vi stille opp formelen:

$$(2.14) \quad {}^{F\sim k}L_0(n+1) = \{1-({}^{F\hat{}k}_{\frac{1}{2}}q_0 + {}^{F\hat{}k}_{\frac{1}{2}}U_0)\} \cdot {}^{F}c \cdot \tilde{B}^{k}(n) + {}^{F\hat{}k}i_0 \cdot {}^{F\sim}U_0(n);\ k=1,2,\ldots K;\ n=0,1,\ldots\ .$$

Her er ${}^{F\sim k}L_0(n+1)$ prediktor for antallet nyfødte piker som lever og bor i kommune k ved slutten av det kalenderår de er født i. Tilsvarende prediktorformel gjelder for gutter. I det siste leddet her skulle vi hatt med en korreksjonsfaktor for dødsfall og utflytting etter innflytting. Da den gjennomsnittlige tid de fødte er under risiko for å dø eller flytte ut igjen etter innflytting er liten, regner vi imidlertid med at korreksjonsfaktoren vil ha uvesentlig betydning.

Flere steder i modellen har vi gjort liknende forutsetninger om at flere "begivenheter" ikke kan inntreffe for en og samme person i ett og samme år. ("Begivenheter" må her tolkes som noe vi tar hensyn til i modellen.) Vi skal kort oppsummere og kommentere disse.

Vi regner ikke med at en person både kan flytte og dø i ett kalenderår. Et enkelt regneeksempel gir et bilde av hvor mye dette kan bety. Tallene er relevante for situasjonen i Norge i dag. Av 200 000 flyttere regner vi med at de aller fleste har alder mellom 20 og 40 år. Den ett-årige dødssannsynligheten er maksimalt ca. 0,002 for disse aldersgrupper. Regner vi med at flyttingene "i gjennomsnitt" skjer midt i året, vil altså maksimalt ca. 200 av de 200 000 flyttere dø i resten av året.

I modellen tar vi heller ikke hensyn til at personer kan flytte flere ganger i løpet av et år. Nå er vi egentlig bare interessert i hvor en person befinner seg i slutten av hvert kalenderår. Vi kan derfor omgå dette problem ved å se bort fra alle "mellomlandinger" under estimeringen av flyttesannsynligheten. Som et spesialtilfelle av dette vil vi ikke registrere noen flytting for en person som til slutt havner i samme kommune som han først flyttet ut fra.

En kvinne som har født, kan enten dø eller flytte etter fødselen, men i modellen tar vi ikke hensyn til at hun kan føde etter flytting. Etter innflytting vil den gjenstående tid kvinnene er under risiko for å føde"i gjennomsnitt" være så vidt lang som et halvt år. Derfor kan vi ikke regne med at denne utelatelse har neglisjerbar betydning. Ved estimeringen av forventet antall fødte har det imidlertid ikke vært mulig å følge den <u>enkelte</u> kvinne fra årets begynnelse til årets slutt. Alle de fødsler som blir registrert i en kommune er derfor tilskrevet de kvinnene som var tilstede i begynnelsen av året. Dette gjelder også fødsler av kvinner som har flyttet inn i løpet av året. Denne feilestimering virker i motsatt retning av den ovennevnte utelatelse og skulle derfor kompensere helt eller delvis for denne.

De fleste av de her omtalte forenklinger er gjort av programmeringstekniske årsaker. Problemene kunne kanskje teoretisk vært løst på en tilfredsstillende måte, men kjøretiden på datamaskinen ville i så tilfelle blitt en god del lengre. Faren ligger ikke i de feilene som gjøres for det enkelte prognoseår, men i at feilene akkumulerer seg opp. Derfor kan de nok få betydning hvis vi lager prognoser for mange år framover.

3. En oppsummering av relasjonene i modellen

Prognosemodellen inneholder altså følgende relasjoner:

(2.2) $\tilde{L}^{k}_{x+1}(n+1) = \tilde{L}^{k}_{x}(n) - \tilde{D}^{k}_{x}(n) - \tilde{U}^{k}_{x}(n) + \tilde{I}^{k}_{x}(n).$

(2.8) $\tilde{D}^{k}_{x}(n) = \hat{q}^{k}_{x} \cdot \tilde{L}^{k}_{x}(n).$

(2.9) $\tilde{U}^{k}_{x}(n) = \hat{U}^{k}_{x} \cdot \tilde{L}^{k}_{x}(n).$

(2.10) $\tilde{U}_{x}(n) = \sum_{j=1}^{K} \tilde{U}^{j}_{x}(n).$

(2.11) $\tilde{I}^{k}_{x}(n) = \hat{i}^{k}_{x} \cdot \tilde{U}_{x}(n).$

(2.13) $\tilde{B}^{k}(n) = \sum_{x=15}^{44} \hat{f}^{k}_{x} \cdot {}^{F}\tilde{L}^{k}_{x}(n).$

(2.14[1]) $\tilde{L}^{k}_{0}(n+1) = \{1 - ({}_{\frac{1}{2}}\hat{q}^{k}_{0} + {}_{\frac{1}{2}}\hat{U}^{k}_{0})\} \cdot c \cdot \tilde{B}^{k}(n) + \hat{i}^{k}_{0} \cdot \tilde{U}_{0}(n)$

for $k = 1,2,\dots,K;\ 0 \leq x \leq \omega$ og $n = 0,1,\dots$.

III. Estimering av døds- og flyttesannsynligheter og forventet antall fødte

Vi skal først se på døds- og utflyttingssannsynlighetene som estimeres analogt.

La den gjenstående levetid for en person i alder x være T_x. Vi innfører nå fordelingsfunksjonen F_x ved

(3.1) $F_x(t) = P(T_x = t) = {}_{t}q_x,$

og dødsintensitetsfunksjonen (Hoem, 1967, side 97):

$$(3.2)\quad \mu_x(t) = \frac{\frac{d}{dt} F_x(t)}{1 - F_x(t)} .$$

Så antar vi at $\mu_x(t)$ er uavhengig av t innenfor et tidsintervall av ett år. Da er åpenbart:

(3.4) $F_x(t) = 1 - e^{-\mu_x \cdot t}$

for $0 \leq t < 1$ som er den eksponentielle fordelingsfunksjon.

La oss så betrakte en person i tiden $[t,1]$, dvs. i en periode av lengde 1-t (t er fremdeles mindre enn 1). Antall ganger personen dør i perioden er ${}_tM$, og antall ganger han flytter er ${}_tN$. Vi har $({}_tM,{}_tN) \in \{(0,0), (0,1), (1,0)\}$. Videre er σ utflyttingsintensiteten og definert analogt med μ. Vi kan foreløpig anta at μ og σ er konstanter uavhengig av kjønn, alder, kommune og år. Da er

$$ {}_{1-t}q_{x+t} = E\,{}_tM = P({}_tM=1) = \int_0^{1-t} e^{-(\mu+\sigma)\tau}\,\mu\,d\tau = \frac{\mu}{\mu+\sigma}\left(1-e^{-(\mu+\sigma)(1-t)}\right), $$

og spesielt for t=0

$$ (3.5) \qquad q_x = \frac{\mu}{\mu+\sigma}\left(1-e^{-(\mu+\sigma)}\right). $$

Etter Sverdrup (1961) lønner det seg å gå ut fra estimatorer for μ og σ når vi skal estimere q og u. La oss si at vi har data for hvert av årene m til (m+t-1). Estimatorer for gjennomsnittlige døds- og flytteintensiteter i perioden finnes da av:

$$ (3.6) \qquad \hat{\mu}_x^k(m,t) = \frac{\sum_{j=1}^{t} D_x^k(m+j-1)}{\sum_{j=1}^{t} M_x^k(m+j-1)} $$

og

$$ (3.7) \qquad \hat{\sigma}_x^k(m,t) = \frac{\sum_{j=1}^{t} U_x^k(m+j-1)}{\sum_{j=1}^{t} M_x^k(m+j-1)} $$

$M_x^k(m+j-1)$ er den observerte, aggregerte levetid i kommune k og kalenderår (m+j-1) av personer i alder x ved begynnelsen av året. Etter Sverdrup (1961) kan denne størrelsen approksimeres ved

$$ (3.8) \qquad M_x^k(m+j-1) \approx L_x^k(m+j-1) - \tfrac{1}{2}\left(D_x^k(m+j-1) + U_x^k(m+j-1) - I_x^k(m+j-1)\right). $$

Av (3.5) følger:

$$ (3.9) \qquad \hat{q}_x^k(m,t) = \frac{\hat{\mu}_x^k(m,t)}{\hat{\mu}_x^k(m,t) + \sigma_x^k(m,t)}\left(1-\exp(-\hat{\mu}_x^k(m,t) + \hat{\sigma}_x^k(m,t)\right) $$

og

$$ (3.10) \qquad \hat{u}_x^k(m,t) = \frac{\hat{\sigma}_x^k(m,t)}{\hat{\mu}_x^k(m,t) + \hat{\sigma}_x^k(m,t)}\left(1-\exp(-\hat{\mu}_x^k(m,t) - \hat{\sigma}_x^k(m,t)\right). $$

$\hat{q}_x^k(m,t)$ og $\hat{u}_x^k(m,t)$ i (3.9) og (3.10) er estimatorer for gjennomsnittlige døds- og utflyttingssannsynligheter i årene m til (m+t-1). For detaljer angående estimatorenes sannsynlighetsteoretiske egenskaper henvises til annen litteratur (f.eks. Sverdrup (1961)).

Som estimator for $i_x^k(m,t)$ har vi brukt:

$$(3.11) \quad i_x^k(m,t) = \frac{\sum_{j=1}^{t} I_x^k(m+j-1)}{\sum_{j=1}^{t} U_x(m+j-1)} .$$

At denne er forventningsrett sees av:

$$E\{\hat{i}_x^k(m,t) \mid \sum_{j=1}^{t} U_x(m+j-1)\} = \frac{1}{\sum_{j=1}^{t} U_x(m+j-1)} E\{\sum_{j=1}^{t} I_x^k(m+j-1) \mid \sum_{j=1}^{t} U_x(m+j-1)\} = i_x^k(m,t)$$

og dermed er $E\,\hat{i}_x^k(m,t) = i_x^k(m,t)$. Likningen gjelder bare for $\sum_{j=1}^{t} U_x(m+j-1)>0$, men for $\sum_{j=1}^{t} U_x(m+j-1) = 0$ blir selvsagt $\hat{i}_x^k(m,t) = 0$ for alle k.

$f_x^k(m,t)$ estimeres ved:

$$(3.12) \quad \hat{f}_x^k(m,t) = \frac{\sum_{j=1}^{t} B_x^k(m+j-1)}{\sum_{j=1}^{t} {}^{F}L_x^k(m+j-1)}$$

og på samme måte som ovenfor finner vi at $\hat{f}_x^k(m,t)$ er forventningsrett.

IV. Litt om dataene

Under utarbeidelsen av modellen har vi selvsagt måttet ta hensyn til den datamasse vi har til disposisjon. Dataene skulle oppfylle to krav: De skulle være detaljerte og forholdsvis lette å bearbeide. På grunn av disse krav har vi f.eks. vært nødt til foreløpig å se bort fra en oppdeling etter ekteskapelig status. Likeledes har vi ikke sett oss i stand til å ta hensyn til emigrasjon og immigrasjon. Både oppdeling i sivilstandsgrupper og emigrasjon/immigrasjon håper vi å få med i nye utgaver av modellen.

Som vi skal se har vi likevel mer data til disposisjon enn de fleste andre land.

Initialbestanden er den vi hadde pr. 1/1 1966. Denne har vi fått fra det nyopprettede personregister for Norge. Hovedoppdelingen er på kommuner. Innen hver kommune har vi antall personer innen hver ett-årig aldersgruppe fordelt på kjønn.

Samme detaljoppdeling finnes for alt det datamaterial som er brukt.

Til grunn for beregning av estimater for døds-, flyttesannsynligheter og forventet antall fødte ligger statistikk for det ene år 1966. (t i formlene (3.6) - (3.12)er altså lik 1). Vi kunne her ha tenkt oss å bruke tall for flere år, men på grunn av kommunesammenslåinger har tall for de enkelte år fram til 1966 ikke vært direkte sammenlignbare. Derfor har det ikke vært mulig å finne en tilfredsstillende måte å løse dette problemet på.

Meningen er i det løpende prognosearbeid å gjøre bruk av ny statistikk så snart denne foreligger.

V. Om modellens anvendbarhet og planlagte forbedringer

Under utarbeidingen av denne modellen har vi hele tiden vært klar over at vi ville være interessert i forbedringer så snart den første utgaven forelå. Dette har vi tatt hensyn til under system- og programmeringsarbeidet, og resultatet er blitt et, etter vår mening, ganske fleksibelt system. I stedet for store programmer som utfører mange operasjoner på én gang, har vi brukt en teknikk med en oppdeling i blokker hvor forandringer i en blokk får liten eller ingen betydning for resten av systemet.

Vi har ikke bare hatt forbedringer i tankene ved utformingen av systemet. Erfaringsmessig vil det alltid være interesse for alternative projeksjoner på nye sett av data. Nå holder vi f.eks. på å produsere projeksjoner hvor vi regner som om flyttingene mellom kommunene er opphørt. Flytterelasjonene i modellen har da ingen interesse. For å spare arbeid og maskintid kan vi lett ta de blokkene som beregner flytting mellom kommunene ut av programmet, uten at dette får innvirkning på resten av systemet.

Ved de første testkjøringene viste det seg at estimatene for de regionale dødssannsynligheter var ubrukelige for små kommuner. Selv Oslo med et datagrunnlag på ca. 500 000 mennesker viste store svingninger i de eldste aldersklasser. Dette har ført til at vi foreløpig har brukt den dødeligheten som gjaldt for hele landet i årene 1961-1965 (Folkemengdens Bevegelse 1965; Tab. XXII) som estimator for dødssannsynligheten i de enkelte kommuner. Da vi i Norge har liten emigrasjon, vil disse estimatene være estimater for noe som tilnærmet er partielle dødssannsynligheter. Når vi så predikerer dødsfall i en kommune får vi egentlig også med dødsfall blant de som har flyttet ut av kommunen. Vi har ikke tillagt denne inkonsistens noen særlig vekt.

De første forbedringer vi har på programmet bunner nettopp i lite datagrunnlag for estimering av døds-, og flyttesannsynligheter og forventet antall fødte i hver kommune. Selv om vi hadde hatt data for, la oss si 5 år, ville variansen på et gjennomsnitt over årene være svært stor for små kommuner. Vår tanke er nå å slå kommunene sammen i grupper, for på denne måten å skaffe et større datagrunnlag pr. region. Hvordan vi skal danne disse kommunegruppene har vi ennå ikke tatt standpunkt til, men det er satt i gang en undersøkelse angående avhengighet mellom flyttinger og variable av både demografisk og ikke-demografisk natur.

De fremtidige fødselstall er også et stort usikkerhetsmoment i befolkningsprojeksjoner. Det ser ut til at fødselskullene vil stige mer enn det er rimelig å tro i de projeksjonene vi nå har produsert. Vi har visse teorier om årsakene til dette, men også her vil vi foreta undersøkelser før vi griper inn i modellen med forbedringer.

VI. Litteraturhenvisninger

Amundsen, H.T. (1962): "Innføring i teoretisk statistikk Bind 1-3." Memorandum av 28/5 1962 fra Sosialøkon.inst., Univ. i Oslo, og Universitetsforlaget, Oslo.

Hoem, Jan M. (1967): "Grunnbegreper i formell befolkningslære." Memorandum av 7/2 1967 fra Sosialøkon.inst., Univ. i Oslo.

Hoem, Jan M. (1968): "Befolkningsprognosemodellens flyttingsrelasjoner. I." Upublisert arbeidsnotat IO 68/11 fra Statistisk Sentralbyrå, Oslo.

Norges Offisielle Statistikk XII 220 (1967): "Folkemengdens Bevegelse 1965." Statistisk Sentralbyrå.

Sverdrup, Erling (1961): "Statistiske metoder ved dødelighetsundersøkelser." Inst. for matem.fag, Univ. i Oslo.