




ARTIKLER

38



**PERSONMODELLER,
PERSONREGNSKAPSSYSTEMER
OG PERSONDATAARKIVER**

Av Svein Nordbotten

**POPULATION MODELS,
POPULATION ACCOUNTING SYSTEMS
AND INDIVIDUAL DATA BANKS**

OSLO 1970

ARTIKLER FRA STATISTISK SENTRALBYRÅ NR. 38

**PERSONMODELLER,
PERSONREGNSKAPSSYSTEMER
OG PERSONDATAARKIVER**

Av Svein Nordbotten

**POPULATION MODELS,
POPULATION ACCOUNTING SYSTEMS
AND INDIVIDUAL DATA BANKS**

OSLO 1970

Forord

Den stadig økende materielle velstand har ført til at politikere og samfunnsplanleggere i stigende grad retter søkelyset mot de menneskelige, sosiale og ikke-økonomiske forhold. Det skaper et behov for tallmessig å kunne beskrive og forklare disse forhold innen rammen av et logisk integrert begreps- og grupperingssystem.

Blant statistikere har det vært drøftet om sosio-demografiske regnskaps-systemer etter mønster av de økonomiske nasjonalregnskapssystemer ville være en hensiktsmessig måte å foreta den nødvendige logiske integrering og tallmessige sammenkøpling på. I land som Norge med individuelle data-arkiver under utvikling, kan integreringen og sammenkøplingen alternativt foretas på individualdata.

Artikkelen drøfter noen tenkelige typer av personmodeller for forklaring av menneskelige forhold og hvilke krav til data modellbyggerne vil ha. Den viser hvordan et personregnskap kan utvikles og hvordan individuell integrering av data kan utføres innen et arkivstatistisk system. Personregnskapstabeller og persondataarkiver vurderes deretter ut fra de muligheter de gir for å gi samfunnsforskerne de data de har behov for i sin modellbygging.

Statistisk Sentralbyrå, Oslo, 25. mai 1970.

Petter Jakob Bjerve

Preface

With increasing material living conditions, politicians and planning authorities have focused their attention on human, social and non-economic aspects of the society. This has created a need for statistical description and explanation within the framework of a logically integrated system of concepts and classifications.

Statisticians have discussed the development of a socio-demographic accounting system as a suitable tool for realizing such an integration and linkage of statistics. In countries, such as Norway, with individual data banks being developed, an alternative may be to perform the integration and linkage at the level of individual data.

This article discusses several types of population models and the needs for data which the construction of such models will generate. It indicates how a population accounting system may be constructed and how individual data integration can be performed within an individual data bank system. Population accounting tables and individual data files are evaluated as to their respective potentials for satisfying the needs of the model builder.

Central Bureau of Statistics, Oslo, 25 May 1970.

Petter Jakob Bjerve

Innhold

	Side
1. Innledning	7
2. Personmodeller	8
2.1 Modeller	8
2.2 Makro-modeller	10
2.3 Aggregerte mikro-modeller	13
3. Statistiske regnskapssystemer	18
3.1 Sosio-demografiske tabeller	18
3.2 Personregnskapstabeller	19
4. Persondataarkiver	23
4.1 Arkivstatistisk system	23
4.2 Integreerte individualdata	24
5. Sammenfatning	25
Sammendrag på engelsk	26

Contents

	Page
1. Introduction	7
2. Population models	8
2.1 Models	8
2.2 Macro models	10
2.3 Aggregated micro models	13
3. Statistical accounting systems	18
3.1 Socio-demographic tables	18
3.2 Population accounting tables	19
4. Individual data banks	23
4.1 Statistical data banks	23
4.2 Integrated individual data	24
5. Conclusions	25
English summary	26

1. Innledning

I de senere år har politikere, samfunnsforskere og -planleggere i stigende grad vært opptatt med de ikke-økonomiske forholds betydning for den menneskelige velferd. Betydningen av trivsel, god helse, gode boligforhold, variert utdanningstilbud, adgang til varierende fritidssysler og naturomgivelser etc. understrekes sterkt som en motvekt til kravet om maksimal økonomisk vekst. En måte å beskrive de sosiale forhold kvantitativt på, er statistisk å belyse hvordan befolkningen på gitte tidspunkter fordeler seg på forskjellige klasser, for eksempel hvordan den fordeler seg på helse-tilstandsklasser.

På samme måte som en tidligere savnet og derfor bygde ut indikatorer som gav oversiktlig orientering om endringene i de økonomiske forhold, drøftes nå behovet for forskjellige sosiale indikatorer som kan belyse utviklingen på de ikke-økonomiske områder. Med indikator forstår en da et summarisk uttrykk (f. eks. en indeks) for situasjonen på et område. Mens beskrivelse av for eksempel befolkningens fordeling etter helsetilstand krever flere tall, og således gir en relativt detaljert beskrivelse som krever tilsvarende innsats fra brukeren for å fortolke, sammenfattes utviklingen i en helseindikator ved et enkelt tall.

Målet må imidlertid ikke bare være å kunne følge utviklingen, men å forstå de viktigste mekanismer som bestemmer den, slik at nødvendige tiltak kan settes i verk for å motvirke en uønsket utvikling. Dette krever at data og statistikk legges til rette for samfunnsforskerne. Med gode erfaringer fra utbyggingen av det økonomiske nasjonalregnskap, foreslås det utviklet sosio-demografiske regnskapssystemer som en ramme for utvikling og presentasjon av statistikk for dette formål.

Denne framstillingen tar sikte på å peke på noen enkle typer av modeller til forklaring av sosio-demografisk utvikling, drøfte i hvilken grad et sosio-demografisk regnskapssystem vil kunne tilfredsstillende kravene til data fra slike modeller og klargjøre deres sammenheng med det arkivstatistiske systemet.

2. Personmodeller

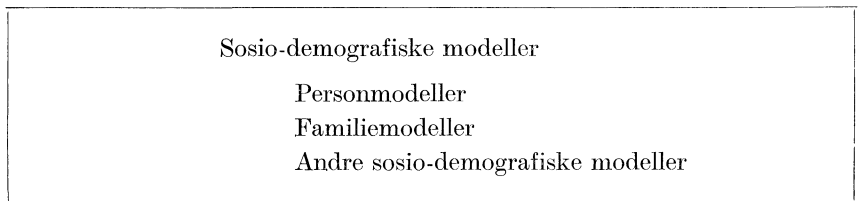
2.1 Modeller

Modeller av den type som drøftes her, er en forenklet, idealisert og symbolsk «forklaring» av størrelsen på visse variable, som kan observeres i samfunnet.

Vi nytter begrepsapparatet fra den økonomiske modellbygging og sondrer mellom de *endogene* variable som modellen forklarer størrelsen på, og de *eksogene* variable, som får sin størrelse fastlagt utenfor modellen, men som antas å påvirke de endogene variable. Alle de modeller vi drøfter her er *dy-namiske*, dvs. en eller flere av de endogene variable inngår også i modellen med allerede forklart eller gitt verdi. Disse variable er betegnet *predetermi- nerte* variable.

Modellene er sammensatt av matematiske relasjoner mellom de forskjellige variable. Det kan være to typer av relasjoner i en modell, definisjonsrela- sjoner og strukturrelasjoner. Modellens definisjonsrelasjoner må også gjelde i det statistiske system modellbyggingen henter sine data fra.

En gruppe av modeller til forklaring av ikke-økonomiske forhold i sam- funnet kan vi som i Figur a betegne *sosio-demografiske modeller*. Disse modellene søker å forklare hvordan personer, familier og andre grupper av personer endrer seg under bestemte forhold.



Figur a: Sosio-demografiske modeller

Alle de modeller vi skal behandle, har til formål å gi forklaring på vari- able hvis størrelse måles i «tallet på personer». Modellene er derfor kalt *personmodeller*. Det meste av det som sies om disse modellene gjelder også for andre typer av sosio-demografiske modeller.

Vi drøfter to typer av personmodeller som bygger på vesensforskjellige forutsetninger. I *makro-modellene* tillegges en *klasse* av personer en identitet og atferd som gir seg uttrykk i at klassen vokser, forblir uendret eller avtar i størrelse uavhengig av de enkelte personers atferd. I *mikro-modellene* for- utsetter vi derimot at det er den enkelte *person* som har atferd. Ved hjelp av mikro-modellene kan vi ved aggregering også forklare hvordan forskjellige klassers størrelser endrer seg som følge av enkeltpersonenes atferd.

Det stilles forskjellige krav med hensyn til tilgjengelige data for modellene. Datakravene er vanligvis størst når modellenes struktur skal tallfestes eller estimeres. Når modellene er numerisk bestemt, krever bruken av modellene for analytiske formål langt færre og mindre detaljerte data.

Befolkningen, N , forutsettes klassifiserbar etter ett eller flere *klassifiseringssystemer*, f. eks. etter helsetilstand, arbeidsaktivitet, utdanningsaktivitet, etc. I denne framstillingen forutsetter vi at klassifiseringen i prinsippet kan betraktes ved utgangen av hver periode. Det tidspunkt som modellen søker å forklare, betegnes som utgangen av periode t . For å forenkle framstillingen skal vi her begrense oss til klassifiseringer etter helsetilstand og til modeller som forklarer hvordan fordelingene etter helsetilstand utvikles. Det som sies, kan imidlertid uten vanskeligheter også omsettes til arbeidsaktivitet, utdanningsaktivitet, etc.

La oss anta at enhver levende person, f. eks. person i , på tidspunkt t , har en *helsetilstand* h_i som må tilhøre en av følgende klasser:

H_1 = frisk

H_2 = behov for legebehandling utenfor sykehus

H_3 = behov for behandling i sykehus

H_4 = behov for varig pleie

Vi innfører videre følgende tilleggsklasser for tidligere og ennå ikke eksisterende personer:

H_D = Døde i tiden mellom $t - 1$ og t

H_F = fødte i tiden mellom t og $t + 1$

For enkelthets skyld ser vi bort fra ut- og innflyttere. Disse kunne vært klassifisert i klasser analogt med døde og fødte.

Tallet på personer på tidspunkt t i de 5 forskjellige klasser, de endogene variable, betegnes med $N_{.1}$, $N_{.2}$, $N_{.3}$, $N_{.4}$ og $N_{.D}$, mens de predeterminerte variable på tidspunkt $t - 1$ er $N_{1.}$, $N_{2.}$, $N_{3.}$ og $N_{4.}$. Tallet på fødte i år t oppfattes her som en eksogen variabel og betegnes med $N_{F.}$. Denne variabel kunne vi tenke oss forklart i en modell til forklaring av fruktbarheten som på sin side kan ha helsetilstandsvariabler som eksogene variable.

Dersom vi er opptatt av personenes overgang mellom klassene, betegner vi med N_{hk} tallet på personer som på tidspunkt t tilhører klasse k , mens de på tidspunkt $t - 1$ tilhørte klasse h . Med dette innfører vi 25 nye endogene variable. Helsetilstanden kan vi tenke oss simultant å påvirke og bli påvirket av andre forhold som knytter seg til personene, for eksempel deres yrke. Skiller vi mellom A forskjellige yrker, Y_1, \dots, Y_A , vil vi ved å kryssklassifisere med helsetilstand få $5 \cdot A$ mulige befolkningsklasser eller endo-

gene variable. En klasse vil her f. eks. kunne tenkes å være kontorfunksjonærer med behov for legehjelp.

Innen noen av modellene vil vi ha behov for et sett av eksogene variable $\bar{X}_1, \bar{X}_2, \bar{X}_3, \bar{X}_4$ og \bar{X}_F som representerer verdien av de helsemessige tiltak pr. person som ble foretatt i klassene H_1, H_2, H_3, H_4 og H_F i t . Dette forutsetter at verdien av forskjellige helsefremmende tiltak blir klassifisert i den økonomiske statistikken etter mottakerklasser som svarer til de helsestilstandsklasser som nyttes i personstatistikken. Andre modeller krever at vi kan klassifisere personer etter verdien av mottatte helsetjenester.

Vi kan tenke oss flere alternative helseindikatorer som bygger på befolkningens fordeling etter helsetilstand. Hvilken som velges, vil i stor utstrekning avhenge av hvilken velferdsoppfatning eller -modell som legges til grunn. Eksempelvis kan vi tenke oss at velferdsbetraktningene leder til valg av en indikator hvor tallet på friske personer, $N_{.1}$, settes i forhold til personer i alt, $\Sigma N_{.h}$, pluss de døde, $N_{.D}$, i siste periode:

$$I = N_{.1}/(\Sigma N_{.h} + N_{.D})$$

Denne indikatoren legger vekt på hvor stor del av befolkningen er friske, men sier lite om status hos dem som ikke er friske.

For å kunne trekke statistiske slutninger om de forklaringer modellene gir, forutsetter vi at de representerer spesifikasjoner av sannsynlighetsfordelinger, og at de variable vi kan observere kan betraktes som realisasjoner fra denne. I denne framstillingen vil vi betrakte fordelinger som bare er betinget av den nærmest forutgående situasjon og av de eksogene variable.

Vi er nå istand til å ta fatt på presentasjonen av våre modeller. Modellene må oppfattes som eksempler på modeller det er rimelig å anta det vil bli arbeidd med. Det understrekes at de på ingen måte gir noe fullstendig bilde av alle eller de «beste» personmodeller.

2.2 Makro-modeller

I de følgende to modeller oppfattes de endogene variablene $N_{.h}$, $h = 1, 2, 3, 4$ og D , som stokastiske med en sannsynlighetsfordeling

$$P(N_{.1}, N_{.2}, N_{.3}, N_{.4}, N_{.D})/N_{.1}, N_{.2}, N_{.3}, N_{.4}, N_{.F.})$$

De stokastiske variable er her definert som de forskjellige befolkningsklassers størrelse, og det er ikke gjort noen forutsetninger om de enkelte personers atferd eller helseutvikling.

Modell A:

Vi forutsetter at sannsynlighetsfordelingen har en slik karakter at vi kan skrive:

$$(A. 1-4) \quad N_{.k} = a_k + \sum_h \beta_{hk} N_h. + \beta_{Fk} \cdot N_{F.} + \varepsilon_{kt} \quad (k = 1 \dots 4)$$

hvor a og β -ene er strukturparametre og ε et stokastisk restledd som bl. a. ivaretar virkningen av andre variable. Relasjonene sier at størrelsen på de enkelte helseklasser i befolkningen ved utgangen av periode t bestemmes av størrelsene på vedkommende klasse og de andre klasser ved inngangen av perioden, tallet på fødsler i perioden og den stokastiske variabel. Det kan være fristende, men neppe fruktbart å prøve å tolke β -ene som relative fordelingsandeler fordi summen vanligvis ikke vil være lik 1.

Vi kan gjøre oss opp visse meninger om størrelsesordenen på en del av parametrene. F. eks. er det rimelig at $\beta_{41} = 0$, dvs. tallet på varig syke vil neppe influere særlig på tallet på friske personer. På den annen side vil vi kunne vente at størrelsen på β_{34} er betydelig, etter som de varig syke vanligvis har vært behandlet på sykehus.

Med et tilstrekkelig statistisk materiale og under visse videre forutsetninger om fordelingsens egenskaper kan α - og β -er estimeres etter velkjente metoder.

Vi har ikke sagt noe om modellens femte endogene variabel, $N_{.D}$. Den blir bestemt i definisjonsrelasjonen:

$$(A.5) \quad \sum_h N_{h.} + N_{F.} = \sum_k N_{.k} + N_{.D}$$

En av modellens svakheter er at den i spesielle situasjoner kan tenkes å gi verdier på $N_{.1}, \dots, N_{.4}$, som til sammen er større enn verdiene på venstre side av A.5, og derfor fører til negativ verdi på $N_{.D}$.

Modell A kan bl. a. nyttes til å studere hvordan alternative fødselstall påvirker befolkningens helsetilstand, eller hvordan alternative helsetilstander kan føre til ulik utvikling over tiden, og hvordan den f. eks. gjenspeiler seg i helseindikatoren I . Fødselstallene som forutsettes forklart av en annen modell, kan selvsagt tenkes å bli påvirket av helsetilstandens utvikling. Bortsett fra å gi grunnlag for nokså grove slutninger om utviklingen, har modelltypen trolig liten verdi, selv om den har vært nyttet som del i mange makro-økonomiske modeller til forklaring av befolknings- og arbeidskraftutvikling.

For å få numeriske estimater på parametrene kreves følgende datasett:

$$(N_{.1}, N_{.2}, N_{.3}, N_{.4}, N_{.D}, N_{1.}, N_{2.}, N_{3.}, N_{4.}, N_{F.})$$

for seks eller flere perioder. Under visse forutsetninger vil vi da ved hjelp

av f. eks. minste kvadraters metode kunne estimere verdiene på α - og β -ene i modellen. For å få brukbare estimater er det imidlertid rimelig å anta at det trengs datasett for et større antall perioder. Når modellens parametre er tallfestet, vil bruk av modellen til analytiske formål kreve empiriske eller hypotetiske data om en initialsituasjon beskrevet ved $(N_{1.}, N_{2.}, N_{3.}, N_{4.})$ og en serie med fødselstall for de perioder en vil analysere eller lage prognoser for.

Modell B:

Fra en samfunnsplanleggers synsvinkel, er Modell A lite interessant, idet den ikke gir noen forklaring på sammenhengen mellom helsetiltak og helse-tilstand. I Modell B, som er en videre spesifisering av Modell A, lar vi \bar{X}_h , $h = 1 \dots 4$, og F , representere de ressurser som nyttes på helsefrem-mende tiltak pr. person i de respektive klasser i perioden $t - 1$ til t .

Vi kan forestille oss mange alternativer i sammenhenger mellom tiltak og helsetilstand. En særlig enkel utforming er:

$$(B.1-4) \quad N_{.k} = \alpha_k + \sum_h \beta_{hk} \cdot N_h + \beta_{Fk} \cdot N_F + \sum_h \lambda_{hk} \cdot \bar{X}_h + \lambda_{Fk} \cdot X_F + \varepsilon_{kt} \\ (k = 1 \dots 4)$$

λ -ene representerer her effekten på befolkningsstørrelsen $N_{.k}$ av hver helse-tiltakskrone pr. person i de forskjellige klasser. Ser vi f. eks. på λ_{12} , må vi anta at denne parameter er negativ, fordi økte helseressurser til forebyg-gende tiltak hos friske personer må antas å redusere tallet på personer som har behov for legehjelp. På den annen side er det ikke urimelig å anta at $\lambda_{21} > 0$ fordi økte ressurser til legebehandling kan føre til forkortet syketid og derfor, på et gitt tidspunkt, flere friske personer.

Den siste relasjonen i Modell B er som i den foregående også

$$(B.5) \quad \sum N_h + N_F = \sum N_{.k} + N_D$$

som fastlegger N_D .

Modell B's struktur kan estimeres under forutsetninger om tilstrekkelige data og om fordelingsregnskaper. Den numeriske modell kan i prinsippet nyttes til å studere virkninger på helsetilstanden av alternative størrelser og sammensetning på helsetiltak. Vil en gitt helsepolitikk føre til lavere dødelighet, men høyere sykelighet enn den en har på referansetidspunktet? Hvordan virker alternative ressursdisponeringer inn på helseindikatoren I ? Det datasett Modell B krever, er

$$(N_{.1}, N_{.2}, N_{.3}, N_{.4}, N_{.D}, N_{1.}, N_{2.}, N_{3.}, N_{4.}, N_F, \bar{X}_1, \bar{X}_2, \bar{X}_3, \bar{X}_4, \bar{X}_F)$$

numerisk tallfesting av parametrene vil kreve minst 11 «årganger», mens bruk av den estimerte modellen vil kreve mindre når formålet er å gi prediksjonsverdier på de endogene variable.

Selvsagt må modellene spesifiseres sterkere om de skal være til større praktisk nytte. Den virkelige vanskelighet med denne type modeller vil være å få tilstrekkelig mange årganger med ensartede data for å estimere parametrene. Fordi en har relativt små variasjoner i størrelsen på de forskjellige helseklasser og de ressurser som nyttes for helsetiltak i de respektive klasser fra periode til periode, vil selv lange tidsrekker sannsynligvis gi estimater som har liten statistisk presisjon.

2.3 Aggregerte mikro-modeller

I de følgende modeller bygger vi på at det er personenes atferd som må forklares og at klassenes størrelser og endringer følger som naturlig konsekvens av enkeltpersonenes atferd.

Vi tenker oss først at personenes helsehistorie kan forklares som en stokastisk prosess hvor den enkelte persons registrerte historie betraktes som en realisering fra

$$P(e_{.i} = H_{.k} \mid e_{i.} = H_h) = P_{hk}$$

hvor $e_{i.}$ er en variabel som beskriver person i 's helsetilstand ved utgangen ($t - 1$), mens $e_{.i}$ beskriver individets helsetilstand ved utgangen av t . P_{hk} er derfor sannsynligheten for at en person som på tidspunkt $t - 1$ er i tilstand H_h vil være i tilstand H_k ved utgangen av t , hvor t kan være en hvilken som helst periode.

Sannsynlighetsfordelingen er her beskrevet ved matrisen:

$$p = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{1D} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{2D} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{3D} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{4D} \\ p_{F1} & p_{F2} & p_{F3} & p_{F4} & p_{FD} \end{bmatrix}$$

Matrisen er karakterisert ved at alle sannsynligheter tilfredsstillter $0 \leq p_{hk} \leq 1$ og at $\sum_k p_{hk} = 1$. Den prosess vi her har, vil være en Markov-kjede-prosess.

Modell C:

Mikro-modellen ovenfor kan uten videre summeres eller aggregeres for flere personer, og vi får en modell som forklarer hvordan personer fra forskjellige helseklasser går over i andre klasser på et senere tidspunkt.

Modell C kan vi skrive ut slik :

$$(C.1-25) \quad N_{hk} = p_{hk} \cdot N_h. \quad \left(\begin{array}{l} h = 1, 2, 3, 4 \text{ og } F \\ k = 1, 2, 3, 4 \text{ og } D \end{array} \right)$$

som sier at de N_{hk} personer som går fra klasse h til klasse k , gjør dette som en tilfeldig trekning med sannsynlighet p_{hk} . p -ene er parametrene i denne modellen. De kan estimeres dersom det foreligger et materiale hvor et tilstrekkelig antall personer er karakterisert med sin helsetilstand ved utgangen av minst to på hverandre følgende perioder.

I tillegg omfatter modellen følgende definisjonsrelasjoner :

$$(C.26-30) \quad N_{.k} = \sum_h N_{hk} \quad (k = 1, 2, 3, 4 \text{ og } D)$$

som fastlegger verdien på de fem siste endogene variable. Relasjonen A.5—B.5 gjelder også i denne modellen, men er allerede implisitt med ved at summen av hver linje i P -matrisen er lik 1. Setter vi inn relasjonene for N_{hk} gitt ved (C.1—24) i relasjonene ovenfor, får vi :

$$N_{.k} = \sum_h p_{hk} \cdot N_h + p_{Fk} \cdot N_F. \quad (k = 1, 2, 3, 4)$$

Disse relasjonene har en viss likhet med (A.1—4). Mens relasjonene i modell A har et stokastisk restledd som blant annet kan tenkes å representere uspesifiserte eksogene variable, er relasjonene ovenfor fullstendig spesifisert.

I tillegg til det Modell A kan nyttes til, kan en ved hjelp av Modell C bl. a. beregne forventet gjenstående levetid i perioder for personer i de forskjellige helseklasser, gjennomsnittlig oppholdstid i de forskjellige tilstandsklasser, sannsynligheter for gitte helsehistorier spesifisert med tidspunktene for overganger mellom klasser, etc. Denne modell-typen er nyttet bl. a. for å lage befolkningsprognoser, hvor de forskjellige tilstander er bostedsområde, fødselsår, kjønn og ekteskapelig status. Den vanligste innvendig mot denne type modeller er at det er urealistisk å anta at sannsynlighetsmatrisen forblir konstant over en lengre periode.

Estimering av sannsynlighetsmatrisen som Modell C bygger på, forutsetter at det kan gis data for et tilstrekkelig stort og representativt utvalg av personer, hvor hver enkelt person er samtidig klassifiserbar etter tilstand på to tidspunkter. For estimering av P -matrisens elementer, kreves et data-sett

$$\left[\begin{array}{l} N_{11}, \dots\dots\dots N_{1D} \\ N_{21}, \dots\dots\dots N_{2D} \\ \dots\dots\dots\dots\dots\dots \\ N_{F1}, \dots\dots\dots N_{FD} \end{array} \right]$$

hvor tallene på personer i de forskjellige klasser bør være så store at estimatene får de nødvendige fordelingssegenskaper. Kjenner en sannsynlighetsmatrisen, vil bruk av modellen ikke sette andre krav til data enn Modell A.

Modell D:

La oss anta at helsetilstand og yrkesstatus gjensidig og simultant påvirker hverandre. Nytter vi modeller av typen Modell C til separat forklaring av helsetilstand og yrkesstatus, vil vi sannsynligvis få forskjellig forklaring på tallet på døde $N_{,D}$ som må inngå som endogen variabel i begge modeller. Med samme teknikk som den som er nyttet i Modell C, kan vi imidlertid utvikle en simultan modell til forklaring av både H og Y som gir konsistente resultater. Ved å kryssklassifisere etter H og Y , får vi en avledet klassifisering Z med i alt (5.A) klasser. Dersom Y har eksempelvis 10 klasser, får modellen i alt 2 550 relasjoner:

$$(D.1—2500) \quad N_{ij} = p_{ij} \cdot N_i. \quad (i = 1 \dots 50)$$

$$(j = 1 \dots 50)$$

$$(D.2501—2550) \quad N_{\cdot j} = \sum_i N_{ij} \quad (i = 1 \dots 50)$$

med en matrise av orden (50 . 50) med overgangssannsynligheter.

Som allerede nevnt vil den viktigste fordel ved Modell D sammenliknet med den foregående, være at den tar hensyn til eventuelle simultane vekselvirkninger mellom flere endogene variable, og at resultatene blir konsistente.

Datakravet fra Modell D vil være større, men av samme type som fra Modell C med datasettet

$$\begin{bmatrix} N_{11} & \dots & N_{1,50} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ N_{50,1} & \dots & N_{50,50} \end{bmatrix}$$

for estimering av sannsynlighetsmatrisen.

Modell E:

Vi kan tenke oss en situasjon hvor befolkningen foruten etter helsetilstand også kan klassifiseres etter verdien av de helsetjenester de mottar. La oss anta at de helsetjenester som ytes den enkelte person ikke helt bestemmes av vedkommende persons helsetilstand. Vi tenker oss at verdien av de helsetjenester person j mottar, X_j i perioden $t - 1$ til t kan transformeres til en klassevariabel Z som har $G + 1$ klasser, bestemt på følgende måte

$$\begin{array}{ll}
 Z_j = 0 \text{ n\aa}r \ 0 \leq X_j < \bar{X}^{(1)} & \text{hvor } \bar{X}^{(1)}, \bar{X}^{(2)}, \dots, \bar{X}^{(G)} \\
 Z_j = 1 \text{ n\aa}r \ \bar{X}^{(1)} \leq X_j < \bar{X}^{(2)} & \text{er gitte grenseverdier} \\
 \dots\dots\dots & \\
 Z_j = G \text{ n\aa}r \ \bar{X}^{(G)} \leq X_j & (h = 1, 2, 3, 4 \text{ og } F) \\
 & (j = 1, 2, 3, 4 \text{ og } N \dots)
 \end{array}$$

$Z_j = G$ betyr med andre ord at personen j tilh\oraerer den klasse av personer som mottar de st\oraeste helsetjenester. Kryssklassifiserer vi befolkningen etter helsetilstand p\aa tidspunkt $t - 1$ og verdi av mottatte helsetjenester i den etterf\oralgende periode, f\aa r vi n\aa en avledet klassifisering med $5 \cdot (G + 1)$ klasser. For hver av disse klassene tenker vi oss en vektor med overgangssannsynligheter til de 5 helsetilstander p\aa tidspunkt t , dvs. en rektangul\aa r matrise av orden $[5 \cdot (G + 1)] \cdot 5$ med overgangssannsynligheter.

Med 5 klasser for mottatte helsetjenester blir relasjonene i Modell E:

$$\begin{array}{lll}
 \text{(E.1—125)} & N_{ik} = p_{ik} \cdot N_i & (i = 1, \dots, 25) \\
 & & (k = 1, 2, 3, 4 \text{ og } D) \\
 \text{(E.126—130)} & N_{.k} = \sum_i N_{ik} & (k = 1, 2, 3, 4 \text{ og } D)
 \end{array}$$

Modell E tillater at en studerer virkningen av alternative ressursanvendelser for helsetiltak. Modellen kan besvare sp\ora rsm\aa l av typen: Hva er virkningen p\aa forventet levetid, frisk levetid, sykdomsperiodisitet n\aa r en \aa ker ressursinnsatsen til forebyggende helsetiltak samtidig som en reduserer innsatsen for tiltak til personer som har behov for behandling i sykehus? Hvilken betydning har fordeling av helsetiltak innen den enkelte helseklasse?

Estimering av P -matrisen forutsetter datasett av typen:

$$\left[\begin{array}{cccc}
 N_{11}, \dots, \dots, \dots, \dots & \dots & \dots & N_{1D} \\
 \dots & \dots & \dots & \dots \\
 \dots & \dots & \dots & \dots \\
 N_{25,1} & \dots & \dots & N_{25,D}
 \end{array} \right]$$

Dette inneb\aa rer at en har observasjoner av den totale befolkning eller et tilstrekkelig utvalg klassifisert b\aa de etter helsetilstand og verdien av mottatte helsetjenester.

Modell F:

La oss anta at den helsemessige utvikling kan oppfattes som generert av en prosess med en sannsynlighetsfordeling

$$P(e_i = H_k \mid e_i = H_h; x_i = X) = p_{hk}(X)$$

hvor X er en kontinuierlig variabel som representerer verdien av mottatte helsetjenester. $p_{hk}(X)$ er sannsynligheten for at en person som p\aa tidspunkt $t - 1$ befant seg i helsetilstand h , skal befinne seg i tilstand k p\aa

tidspunkt t når vedkommende i den mellomliggende periode mottar helse-tjenester til en verdi av X .

Modellen krever løsning av to problemer. For det første må en finne funksjonsformer som for alle tillatte verdier av X tilfredsstillere identiteten:

$$\sum_k p_{hk}(X) \equiv 1 \quad \begin{array}{l} (h = 1, 2, 3, 4 \text{ og } F) \\ (k = 1, 2, 3, 4 \text{ og } D) \end{array}$$

og betingelsene:

$$0 \leq p_{hk}(X) \leq 1$$

For det andre må en finne estimeringsmetoder som tillater å estimere verdiene på de parametre som inngår i funksjonene.

Mens de foregående mikro-modeller lett lot seg aggregere til grupper av personer, er ikke dette tilfelle for modeller av type F . La $i = 1, \dots, N$.. være en nummerering av de enkelte personer i befolkningen, d_{ih} en variabel som har verdi 1 dersom person i på tidspunkt $t - 1$ er i helseklasse h , og verdi 0 ellers, og la x_i være en variabel som angir verdien av de helse-tjenester person i mottar mellom $t - 1$ og t . Ved aggregering kan vi nå formulere Modell F.

$$(F.1-25) \quad N_{hk} = \sum_i p_{hk}(x_i) \cdot d_{ih} \quad \begin{array}{l} (h = 1, 2, 3, 4 \text{ og } F) \\ (k = 1, 2, 3, 4 \text{ og } D) \end{array}$$

og

$$(F.26-30) \quad N_{.k} = \sum_h N_{hk} \quad (k = 1, 2, 3, 4 \text{ og } D)$$

Dersom X bare tillates å anta $G + 1$ forskjellige verdier er Modell F den samme som Modell E.

For å tallfeste parametrene i de betingede sannsynligheter $p_{hk}(X)$, kreves det for hele befolkningen, eller et tilstrekkelig utvalg, individualdata av typen:

$$(d_{ih}, d_{ik}, x_i) \quad \begin{array}{l} (h = 1, 2, 3, 4 \text{ og } F) \\ (k = 1, 2, 3, 4 \text{ og } D) \end{array}$$

hvor d_{ik} er en variabel som har verdi 1 dersom person i på tidspunkt t er i helseklasse k , og verdi 0 ellers. Bruk av modellen for å undersøke virkningen av alternative fordelinger av helsetjenestene vil også kreve individualdata som uttrykker fordelingen av helsetjenester på befolkningene innen de enkelte helseklasser.

Andre modeller:

Vi har her bare gitt eksempler på relativt enkle personmodeller som forutsetter at vi finner det tilstrekkelig å betrakte befolkningen på bestemte

tidspunkter. For en rekke formål vil det imidlertid være nødvendig å forutsette at prosessene er tidskontinuerlige. Å bygge modeller som tar vare på en slik forutsetning, setter også krav til tilgang på individuelle data med presis datering av overganger mellom forskjellige tilstander.

3. Statistiske regnskapssystemer

3.1 Sosio-demografiske tabeller

Formålet med de statistiske regnskapssystemer er å framstille statistikk som ikke er innbyrdes selvmotsigende for bl. a. modellbyggere og -brukere. Systemene har ellers ofte som siktepunkt å være «general purpose», og en søker gjerne å holde begrepsdefinisjonene og klassifikasjonssystemene faste over lengre perioder. Det ligger en viss fare i dette, idet en lett kan komme i skade for å betrakte statistikk framkommet ved andre klassifikasjoner som vill-ledende eller mindreverdig. Det er lite rimelig å tro at det lar seg gjøre å utvikle et regnskapssystem som er så generelt at det dekker alle behov.

Vi kan tenke oss utviklet flere parallelle regnskapssystemer. Vi har for lengst i de fleste land økonomiske nasjonalregnskapssystemer, og i det følgende skal vi drøfte en type av sosio-demografiske regnskapssystemer. Som for modellene kan vi innenfor de sosio-demografiske regnskapssystemer foreta en hensiktsmessig oppdeling etter regnskapsenhet i personregnskaper, familieregnskaper og andre sosio-demografiske regnskaper. Mellom disse regnskaper kan det for visse poster eksistere relasjoner etc. som må være oppfylt. Tallet på familier kan f. eks. aldri bli større enn tallet på personer. Vi kunne også tenke oss fysiske regnskapssystemer med en eller annen fysisk mengdeenhet som regnskapsenhet og regnskapssystemer over tidsbruk med tiden som enhet.

I mange anvendelser, bl. a. i Modell B og Modell E i avsnitt 2.2 og 2.3, vil det være behov for å stille sammen tall som gir uttrykk for befolkningsgruppens størrelse og de økonomiske transaksjoner som vedrører denne gruppen. Dette krever at klassifiseringene til en viss grad er de samme i de forskjellige regnskapssystemer.

For å få økonomiske tall i samsvar med det sosio-demografiske systemet, kan en tenke seg nasjonalregnskapet utbygd med hva en kunne kalle sosio-økonomiske tabeller, som bl. a. gir økonomiske mål for innsatsen av eller kostnadene ved de helsetjenester som mottas av personer i de forskjellige helsetilstandsklasser.

3.2 Personregnskapstabeller

I et *personregnskapssystem* er det personene i befolkningen som klassifiseres. I regnskapet tillegges hver person lik størrelse, dvs. verdien 1. Regnskapet kan tenkes presentert i to sett av tabeller, *beholdningstabeller* og *overgangstabeller*. Hver beholdningstabell viser hvordan befolkningen på et bestemt tidspunkt fordeler seg på de enkelte og gjensidig utelukkende klasser i en sosio-demografisk klassifisering. Summene i beholdningstabellene vil være like og ofte også en rekke undersummer. Dette vil særlig bli tilfelle i tabeller hvor en systematisk har kryssklassifisert etter f. eks. fødselsår. Overgangstabeller oppnår en når en kryssklassifiserer den samme befolkning etter personenes tilstand på to forskjellige tidspunkter, f. eks. helsetilstand ved utgangen av to på hverandre følgende år.

Et utbygd personregnskap vil tenkes å kunne inneholde tabeller med en eller flere klassifiseringer av den samme befolkning innen følgende områder:

1. Fødselsår og kjønn
2. Helsetilstand og sykdom
3. Utdanningsnivå og type
4. Familiestatus
5. Sosial status
6. Kriminell status
7. Arbeidsaktivitet og yrke
8. Fritidsaktivitet
9. Annen tidsbruk
10. Inntekt og inntektstype
11. Forbruk
12. Meninger og holdninger

I det følgende skal vi fortsatt nytte helsetilstanden som eksempel i tabellene, men det må være underforstått at tilsvarende tabeller må etableres for andre klassifiseringer.

Visse tilstander er *permanente* som en persons kjønn og fødselsår. Vi kan dele befolkningen inn i *separable* befolkninger etter slike kriterier og behandle hver enkelt uavhengig av de andre, fordi det aldri skjer overganger mellom delbefolkningene. De følgende tabeller kan f. eks. alltid betraktes i to versjoner, en for menn og en for kvinner.

Tabell I viser en hensiktsmessig måte å stille opp beholdningstabeller for befolkningen etter helsetilstand på. Som det allerede er pekt på i avsnitt 2.1 og 2.2, er den dynamiske problemstilling sentral i personmodellbyggingen. Det er også hensiktsmessig av hensyn til senere konsolidering med overgangstabellene å ta med så vel siste som foregående års beholdningstall i tabellen.

Tabell I. Personer på tidspunkt t og $t - 1$ etter helsetilstand

Helsetilstand	Ved utgangen av	
	$t - 1$	t
H_1 : Frisk	$N_{1.}$	$N_{.1}$
H_2 : Behov for legehjelp u. sykehus	$N_{2.}$	$N_{.2}$
H_3 : Behov for legehjelp i sykehus	$N_{3.}$	$N_{.3}$
H_4 : Behov for varig pleie	$N_{4.}$	$N_{.4}$
H_F : Fødte i perioden t	$N_F.$	
H_D : Døde i perioden t		$N_{.D}$
Alle klassifiserte personer	$N_{..}$ (=)	$N_{..}$

Det kan synes merkelig at en i befolkningen på tidspunkt $t - 1$ medregner personer som ikke er født og på tidspunkt t personer som allerede er døde. Det er to begrunnelser for dette. For det første er dette en hensiktsmessig måte å få med to viktige størrelser i systemet. For det andre sikrer denne oppstilling en regnskapsmessig kontroll av de definisjonsmessige relasjoner A.5 og B.5 i Modell A og Modell B mot utelatelser eller dobbelttelling. $N_{..}$ kan oppfattes som et mål på en bruttobefolkning.

Tabeller av type Tabell I vil tilfredsstillende databehovet for estimering av parametrene i Modell A om de foreligger for tilstrekkelig mange år. Finnes det tabeller med tilsvarende klasseinndeling også i et økonomisk regnskapsystem, vil også Modell B's behov dekkes. Likeens vil beholdningstabeller av denne type gi de nødvendige data for bruk av modeller av typen Modell C.

Dersom vi bare har beholdningstabeller i systemet, vil det være få regnskapsmessige relasjoner mellom tall i forskjellige tabeller. En vanlig antakelse er at de fleste endringer er avhengig av befolkningens alderssammensetning. Da fødselsår også er en permanent egenskap hos personen, kan det være formålstjenlig å kryssklassifisere etter alder i alle beholdningstabeller. En får da tabeller av typen Tabell II.

I tabell II neste side vil en for hver årgang ha tall som er avstemt mot hverandre. I realiteten betyr dette at en deler befolkningen inn i t separable befolkninger som det ikke kan foregå overganger mellom. Tabell II er derfor i prinsippet den samme som Tabell I.

Beholdningstabeller gir ikke de data som er nødvendige for å tallfeste sannsynlighetsmatrisene i Modell C, D og E. En naturlig videreutvikling er overgangstabellene som vil kunne gi de nødvendige tall.

Tabell III gir en illustrasjon på oppbygningen av en overgangstabell som gir overgangstall for helsedata.

Tabell II. Personer på t og $t-1$ etter fødselsår og helsetilstand

Fødselsår : helsetilstand		Ved utgangen	
		$t-1$	t
År 0:	H_1	$N_{1.}^{(0)}$	$N_{.2}^{(0)}$
	H_2	$N_{2.}^{(0)}$	$N_{.2}^{(0)}$
	H_3	$N_{3.}^{(0)}$	$N_{.3}^{(0)}$
	H_4	$N_{4.}^{(0)}$	$N_{.4}^{(0)}$
	H_D		$N_{.D}^{(0)}$
Alle t -åringer klassifisert		$N_{..}^{(0)} (= N_{..}^{(0)})$	
År t :	H_1		$N_{.1}^{(t)}$
	H_2		$N_{.2}^{(t)}$
	H_3		$N_{.3}^{(t)}$
	H_4		$N_{.4}^{(t)}$
	H_F	$N_{F.}^{(t)}$	
	H_D		$N_{.D}^{(t)}$
Alle 0-åringer klassifisert		$N_{..}^{(t)} (= N_{..}^{(t)})$	
Alle personer klassifisert		$N_{..} (= N_{..})$	

Tabell III. Personer etter helsetilstand på tidspunkt $t-1$ og t

Helsetilstand på tidspunkt $t-1$:	Helsetilstand på tidspunkt t :					Alle
	H_1	H_2	H_3	H_4	H_D	
H_1	N_{11}	N_{12}	N_{13}	N_{14}	N_{1D}	$N_{1.}$
H_2	N_{21}	N_{22}	N_{23}	N_{24}	N_{2D}	$N_{2.}$
H_3	N_{31}	N_{32}	N_{33}	N_{34}	N_{3D}	$N_{3.}$
H_4	N_{41}	N_{42}	N_{43}	N_{44}	N_{4D}	$N_{4.}$
H_F	N_{F1}	N_{F2}	N_{F3}	N_{F4}	N_{FD}	$N_{F.}$
Alle	$N_{.1}$	$N_{.2}$	$N_{.3}$	$N_{.4}$	$N_{.D}$	$N_{..}$

En merker seg at siste kolonne og siste linje er identiske med de to tallkolonner i beholdningstabellen Tabell I. Som den sistnevnte tabell kunne utvides ved kryssklassifisering med andre klassifiseringssystemer, vil vi også kunne lage større overgangstabeller. Vi tar for oss kryssklassifisering etter yrke, Y , og helsetilstand, H , på begge tidspunkter. La oss fortsatt anta at tallet på alternativer i Y er 10. Vi får da utvidede overgangstabeller som er av den type Tabell IV viser. Slike tabeller ville tilfredsstillende behovet for data til en simultan modell av typene D og E.

Tabell IV. Personer etter yrke og helsetilstand på $t - 1$ og t

Yrke : helsetil- stand på $t - 1$	Y_1					Y_n					Alle	
	H_1	H_2	H_3	H_4	H_D	\dots	H_1	H_2	H_3	H_4		H_D
Yrke 1: H_1	$N_{1 \cdot 1}$	$N_{1 \cdot 2}$	$N_{1 \cdot 3}$	$N_{1 \cdot 4}$	$N_{1 \cdot 5}$	\dots	$N_{1 \cdot 46}$	$N_{1 \cdot 47}$	$N_{1 \cdot 48}$	$N_{1 \cdot 49}$	$N_{1 \cdot 50}$	$N_{\cdot \cdot}$
H_2	$N_{2 \cdot 1}$	$N_{2 \cdot 2}$	$N_{2 \cdot 3}$	$N_{2 \cdot 4}$	$N_{2 \cdot 5}$	\dots	$N_{2 \cdot 46}$	$N_{2 \cdot 47}$	$N_{2 \cdot 48}$	$N_{2 \cdot 49}$	$N_{2 \cdot 50}$	$N_{\cdot 2}$
H_3	$N_{3 \cdot 1}$	$N_{3 \cdot 2}$	$N_{3 \cdot 3}$	$N_{3 \cdot 4}$	$N_{3 \cdot 5}$	\dots	$N_{3 \cdot 46}$	$N_{3 \cdot 47}$	$N_{3 \cdot 48}$	$N_{3 \cdot 49}$	$N_{3 \cdot 50}$	$N_{\cdot 3}$
H_4	$N_{4 \cdot 1}$	$N_{4 \cdot 2}$	$N_{4 \cdot 3}$	$N_{4 \cdot 4}$	$N_{4 \cdot 5}$	\dots	$N_{4 \cdot 46}$	$N_{4 \cdot 47}$	$N_{4 \cdot 48}$	$N_{4 \cdot 49}$	$N_{4 \cdot 50}$	$N_{\cdot 4}$
H_F	$N_{5 \cdot 1}$	$N_{5 \cdot 2}$	$N_{5 \cdot 3}$	$N_{5 \cdot 4}$	$N_{5 \cdot 5}$	\dots	$N_{5 \cdot 46}$	$N_{5 \cdot 47}$	$N_{5 \cdot 48}$	$N_{5 \cdot 49}$	$N_{5 \cdot 50}$	$N_{\cdot 5}$
Yrke n : H_1	$N_{46 \cdot 1}$	$N_{46 \cdot 2}$	$N_{46 \cdot 3}$	$N_{46 \cdot 4}$	$N_{46 \cdot 5}$	\dots	$N_{46 \cdot 46}$	$N_{46 \cdot 47}$	$N_{46 \cdot 48}$	$N_{46 \cdot 49}$	$N_{46 \cdot 50}$	$N_{46 \cdot}$
H_2	$N_{47 \cdot 1}$	$N_{47 \cdot 2}$	$N_{47 \cdot 3}$	$N_{47 \cdot 4}$	$N_{47 \cdot 5}$	\dots	$N_{47 \cdot 46}$	$N_{47 \cdot 47}$	$N_{47 \cdot 48}$	$N_{47 \cdot 49}$	$N_{47 \cdot 50}$	$N_{47 \cdot}$
H_3	$N_{48 \cdot 1}$	$N_{48 \cdot 2}$	$N_{48 \cdot 3}$	$N_{48 \cdot 4}$	$N_{48 \cdot 5}$	\dots	$N_{48 \cdot 46}$	$N_{48 \cdot 47}$	$N_{48 \cdot 48}$	$N_{48 \cdot 49}$	$N_{48 \cdot 50}$	$N_{48 \cdot}$
H_4	$N_{49 \cdot 1}$	$N_{49 \cdot 2}$	$N_{49 \cdot 3}$	$N_{49 \cdot 4}$	$N_{49 \cdot 5}$	\dots	$N_{49 \cdot 46}$	$N_{49 \cdot 47}$	$N_{49 \cdot 48}$	$N_{49 \cdot 49}$	$N_{49 \cdot 50}$	$N_{49 \cdot}$
H_F	$N_{50 \cdot 1}$	$N_{50 \cdot 2}$	$N_{50 \cdot 3}$	$N_{50 \cdot 4}$	$N_{50 \cdot 5}$	\dots	$N_{50 \cdot 46}$	$N_{50 \cdot 47}$	$N_{50 \cdot 48}$	$N_{50 \cdot 49}$	$N_{50 \cdot 50}$	$N_{50 \cdot}$
Alle	$N_{\cdot 1}$	$N_{\cdot 2}$	$N_{\cdot 3}$	$N_{\cdot 4}$	$N_{\cdot 5}$	\dots	$N_{\cdot 46}$	$N_{\cdot 47}$	$N_{\cdot 48}$	$N_{\cdot 49}$	$N_{\cdot 50}$	$N_{\cdot \cdot}$

Dersom denne tabelltypen nyttes i et personstatistisk tabellverk bygd på f. eks. 10 forskjellige klassifikasjoner, vil det få et betydelig omfang. Nyttet det klassifiseringer med gjennomsnittlig 15 klasser, vil vi for hvert kjønn få overgangstabeller av størrelsesorden 225 linjer og 225 kolonner for hver 2-veis kryssklassifisering. Dette svarer til om lag 100 publikasjonssider pr. tabell. I alt vil det være 45 forskjellige 2-veis kryssklassifiseringer pr. kjønn og totalt 9 000 sider. Dersom noen forskere ville arbeide med hypoteser som bygde på at 3 eller flere egenskaper påvirket hverandre simultant, ville dette i tillegg kreve 3-veis kryssklassifiseringstabeller eller mer. Dette ville føre til tabellverk av fullstendig utenkelig omfang.

Det kan således lages personregnskapstabeller som tilfredsstillter datakrav for estimering av strukturen i modeller av typen A, B og C. Når det gjelder modeller av typene D og E, ville imidlertid eventuelle tabeller få et omfang som det ikke vil være praktisk å presentere i tabellform. Vil det være hensiktsmessig å utvikle alle overgangstabeller basert på 1-veis, 2-veis, 3-veis, osv. klassifiseringer, og lagre disse på et maskinlesbart medium for kopiering etter behov?

Det viser seg at allerede tallet på celler i de overgangstabeller som bygger på 2-veis klassifikasjoner overstiger størrelsen på den norske befolkning. Det vil da være andre lagringsformer som er mer hensiktsmessige.

Til slutt må vi fastslå at et tabellsystem ikke vil gi modellbyggerne de data som mikro-modeller av typen F forutsetter.

4. Persondataarkiver

4.1 Arkivstatistisk system

Overgangstabellene kan utvikles på to forskjellige måter:

1. De kan bygges på periodiske *retrospektive* undersøkelser hvor en spør hver enkelt person både om den aktuelle situasjon og om situasjonen ved utgangen av foregående periode. Vanskeligheten ved denne framgangsmåte er at det ikke er praktisk mulig å dekke alle spørsmål i en undersøkelse. Dessuten er det en viss fare for at svarene om tilstanden ved utgangen av forrige periode ikke blir sikre på grunn av hukommelsessvikt.
2. Alternativt kan de bygge på et *arkivstatistisk* system som arkiverer løpende identifiserte persondata fra forskjellige kilder på en måte som muliggjør integrering av individuelle data.

Har en et operativt arkivstatistisk system, vil spørsmålet om et personregnskap i stor utstrekning reduseres til en hensiktsmessig presentasjon av hovedtallene i personstatistikk for alminnelige informasjonsformål. Spesielle data for modellbyggere vil systemet kunne gi på andre og mer hensiktsmessige former.

De deler av det arkivstatistiske systemet som er av interesse i denne sammenheng er:

- a. Identifiserings- og kodifiseringsrutiner.
- b. Integreringsrutiner.

Identifiserings- og *kodifiseringsrutiner* beskriver hvordan det innsamlede materialet transformeres til systemets dataform. Den viktigste komponenten i identifiseringsrutinen er et løpende vedlikeholdt register over alle personer som både inneholder nødvendige opplysninger, f. eks. navn og adresse, for å identifisere den enkelte person i samfunnet, og et entydig og permanent identifikasjonssymbol, f. eks. et identifikasjonsnummer for hver person.

De oppgaver som samles løpende inn fra forskjellige kilder, er individuelle verdier på to typer av kjennemerker. De *kvantitative* kjennemerkene kan anta et endelig eller uendelig antall forskjellige verdier som gir et »størrelsesmål» på vedkommende person. Vi skal her ta for oss *kvalitative* kjennemerker, dvs. kjennemerker som bare kan anta alternative klasseverdier, og hvor verdiene ikke har noen kvantitativ betydning i forhold til hverandre. Kjønn, utdanning, helsetilstand, etc. er eksempler. Ved hjelp av kodelister utføres *kodifisering* hvor opplysninger om de kvalitative kjennemerker transformeres til symboler. På samme måte som identifikasjonsnummeret er en entydig og permanent identifikasjon av personen, må koden være en entydig og permanent identifikasjon av kjennemerke og -klasse.

I de statistiske presentasjonssystemer søker en å integrere, dvs. stille sammen, kontrollere og gjøre fullstendig sammenliknbart, statistikk for befolkningsklasser. På det individuelle nivå er den tilsvarende prosessen å stille sammen, kontrollere og gjøre sammenliknbart alle de kodede opplysninger vi har med samme identifikasjonsnummer, dvs. for samme person. I et arkivstatistisk system kan det finnes hundrevis av kodede opplysninger om den enkelte person, og det kan bli spørsmål om hvor langt i retning av en fullstendig livshistorie en bør gjennomføre integreringen.¹ Vi skal i det følgende se litt nærmere på hva modeller av den type vi drøftet i avsnitt 2 vil kreve av individuell dataintegrering.

4.2 Integrerte individualdata

La H være et av de kvalitative kjennemerker som inngår i det arkivstatistiske system. Når det hos en person inntreffer en endring i H som betinger at personen går over i en annen klasse, forutsetter vi at det arkivstatistiske systemet mottar en datert melding som identifiseres og kodifiseres. Ved å ordne meldingene etter identifikasjonsnummer, kjennemerke og endringsdato, kan en ved utgangen av hver periode, f. eks. ved utgangen av t , lage et datasett som beskriver hver enkelt i landet levende persons situasjon på dette tidspunkt.

En slik situasjonsrecord vil kunne tenkes å ha en struktur som vist i Figur b.

Felt 0:	Personens identifikasjonsnummer:	i
Felt 1:	Kjennemerke A 's verdi	A_i
Felt h :	Kjennemerke H 's verdi	H_i
Felt x :	Kjennemerke X 's verdi	X_i

Figur b: *Situasjonsrecordens oppbygning*

I dette settet vil ikke kodene A_F , A_D , H_F , H_D , osv. forekomme, fordi det bare omfatter eksisterende personer.

Ved systematisk å kople sammen situasjonssett fra to forskjellige tidspunkter, vil en for den mellomliggende periode kunne lage et datasett hvor den individuelle record vil kunne ha en form som vist i Figur c.

Note: ¹ For en nærmere drøfting av de muligheter det arkivstatistiske system gir samfunnsforskerne, vises til AUKRUST, O. og NORDBOTTEN, S.: Dataregistrering, dataarkiver og samfunnsforskning, Artikler fra Statistisk Sentralbyrå, Nr. 34, Oslo 1969.

Felt 1:	Kjennemerke A 's verdi på $t - 1$:	A_i .
Felt 2:	Kjennemerke A 's verdi på t	: A_i
Felt $2 \cdot h$:	Kjennemerke H 's verdi på $t - 1$:	H_i .
Felt $2 \cdot h + 1$:	Kjennemerke H 's verdi på t	: H_i
Felt $2 \cdot x$:	Kjennemerke X 's verdi på $t - 1$:	X_j .
Felt $2 \cdot x + 1$:	Kjennemerke X 's verdi på t	: X_i

Figur c: Den integrerte datarecords oppbygning

I dette datasettet har en ikke lenger behov for de individuelle identifikasjonsnummer som er utelatt. Fordi situasjonssettet for tidspunkt $t - 1$ vil inneholde personer som ikke forekommer i situasjonssettet for tidspunkt t på grunn av avgang, og omvendt på grunn av tilgang i den mellomliggende periode, vil det integrerte datasettet inneholde personrecords hvor også kjennemerkealternativene H_F , H_D , etc. vil forekomme.

Det integrerte datasettet vil tillate utkjøring av en hvilken som helst tabell innenfor det tabellverk som kan tenkes med de gitte sosio-demografiske kjennemerker.

Er X en kvantitativ variabel, vil datasettet også tilfredsstillte datakravene fra Modell F.

5. Sammenfatning

Personmodellene vil i stor utstrekning være mikro-modeller i den forstand at de bygger på forutsetninger om enkeltmenneskenes atferd. Estimering av strukturen i disse modeller stiller slike datakrav at det neppe vil kunne tilfredsstilltes ved et personregnskapstabellverk.

Personregnskapstabellene vil sannsynligvis ha største verdi som et hjelpemiddel for å presentere de mest sentrale hovedtall innenfor et oversiktlig og integrert system.

Et utbygd arkivstatistisk system vil tillate individuell integrering av data om personer fra en lang rekke forskjellige kilder og tidspunkter. Det vil være det beste grunnlag for framstilling av personregnskapstabeller, og for betjening av de spesielle databehov forskerne vil ha for å tallfeste strukturen i sine modeller.

English summary

The paper concentrates on socio-demographic aspects which can be described by the population distribution by socio-demographic classes, e.g. by health classes. Two macro- and four micro-model types are introduced and discussed with respect to their relevance in explaining socio-demographic situations and development and to their data requirements.

A set of population accounting tables is introduced. The simple stock tables will satisfy the data requirements for estimation of the macro-models and for the analytical use of micro-models with pre-estimated structures. The more detailed and extensive transition tables may supply necessary data also for estimation of transition probabilities in some types of models.

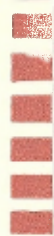
An individual data bank system, based on identified and dated information, permits an integration of individual data from different sources and times. A well organized data bank system of this type may be an efficient basis for developing socio-demographic accounting tables and a superior alternative to such tables for satisfying the social researcher with data for model structure estimation. The main tasks are the development of meaningful concept and classifications.

- Nr. 21 Petter Jakob Bjerve: Trends in Quantitative Economic Planning in Norway *Utviklingsstendensar i den kvantitative økonomiske planlegginga i Norge* 1968 29 s. kr. 5,00
- « 22 Kari Karlsen og Helge Skaug: Statistisk Sentralbyrås sentrale registre *Registers i the Central Bureau of Statistics* 1968 24 s. kr. 3,50
- « 23 Per Sevaldson: MODIS II A Macro-Economic Model for Short-Term Analysis and Planning *MODIS II En makroøkonomisk modell for korttidsanalyse og planlegging* 1968 40 s. kr. 4,50
- « 24 Olav Bjerkholt: A Precise Description of the System of Equations of the Economic Model MODIS III *Likningssystemet i den økonomiske modell MODIS III* 1968 30 s. kr. 4,50
- « 25 Eivind Hoffmann: Prinsipielt om måling av samfunnets utdanningskapital og et forsøk på å måle utdanningskapitalen i Norge i 1960 *On the Measurement of the Stock of Educational Capital and an Attempt to Measure Norway's Stock of Educational Capital in 1960* 1968 60 s. kr. 5,00
- « 26 Hallvard Borgenvik: Aktuelle skattetall 1968 *Current Tax Data* 1969 40 s. kr. 7,00
- « 27 Hallvard Borgenvik: Inntekts- og formuesskattlegging av norske kapitalplasseringer i utlandet *Income and Net Wealth Taxes of Norwegian Investment in Foreign Countries* 1969 40 s. kr. 7,00
- « 28 Petter Jakob Bjerve og Svein Nordbotten: Automasjon i statistikkproduksjonen *Automation of the Production of Statistics* 1969 30 s. kr. 7,00
- « 29 Tormod Andreassen: En analyse av industriens investeringsplaner *An Analysis of the Industries Investment Plans* 1969 26 s. kr. 5,00
- « 30 Bela Balassa og Odd Aukrust: To artikler om norsk industri *Two Articles on Norwegian Manufacturing Industries* 1969 40 s. kr. 5,00
- « 31 Hallvard Borgenvik og Hallvard Flø: Virkninger av skattereformen av 1969 *Effects of the Taxation Reform of 1969* 1969 35 s. kr. 7,00
- « 32 Per Sevaldson: The Stability of Input-Output Coefficients *Stabilitet i kryssløpskoeffisienter* 1969 40 s. kr. 7,00
- « 33 Odd Aukrust og Hallvard Borgenvik: Inntektsfordelingsvirkninger av skattereformen av 1969 *Income Distribution Effects of the Taxation Reform of 1969* 1969 29 s. kr. 7,00
- « 34 Odd Aukrust og Svein Nordbotten: Dataregistrering, dataarkiver og samfunnsforskning *Data Registration, Data Banks and Social Research* 43 s. kr. 7,00
- « 35 Odd Aukrust: PRIM I A Model of the Price and Income Distribution Mechanism of an Open Economy *PRIM I En modell av pris- og inntektsfordelingsmekanismen i en åpen økonomi.* 61 s. kr. 7,00
- « 36 Arne Amundsen: Konsumets og sparingens langsiktige utvikling *Consumption and Saving in the Process of Long-Term Growth* 18 s. kr. 5,00
- « 37 Steinar Tamsfoss: Om bruken av stikkprøver ved kontoret for intervjuundersøkelser, Statistisk Sentralbyrå *On the Use of Sampling Surveys by the Central Bureau of Statistics, Norway* 46 s. kr. 7,00

Utkommet i serien ART

Issued in the series Artikler fra Statistisk Sentralbyrå (ART)

- Nr. 1 Odd Aukrust: Investeringenes effekt på nasjonalproduktet *The Effects of Capital Formation on the National Product* 1957 28 s. Utsolgt
- « 2 Arne Amundsen: Vekst og sammenhenger i den norske økonomi 1920—1955 *Growth and Interdependence in Norwegian Economy* 1957 40 s. Utsolgt
- « 3 Statistisk Sentralbyrås forskningsavdeling: Skattlegging av personlige skattytere i årene 1947—1956 *Taxation of Personal Tax Payers* 1957 8 s. Utsolgt
- « 4 Odd Aukrust og Juul Bjerke: Realkapital og økonomisk vekst 1900—1956 *Real Capital and Economic Growth* 1958 32 s. Utsolgt
- « 5 Paul Barca: Utviklingen av den norske jordbruksstatistikk *Development of the Norwegian Agricultural Statistics* 1958 23 s. kr. 2,00
- « 6 Arne Amundsen: Metoder i analysen av forbruksdata *Methods in Family Budget Analyses* 1960 24 s. kr. 5,00
- « 7 Arne Amundsen: Konsumelastisiteter og konsumprognoser bygd på nasjonalregnskapet *Consumer Demand Elasticities and Consumer Expenditure Projections Based on National Accounts Data* 1963 44 s. kr. 5,00
- « 8 Arne Øien og Hallvard Borgenvik: Utviklingen i personlige inntektsskatter 1952—1964 *The Development of Personal Income Taxes* 1964 30 s. kr. 5,00
- « 9 Hallvard Borgenvik: Personlige inntektsskatter i sju vesteuropeiske land *Personal Income Taxes in Seven Countries in Western Europe* 1964 16 s. kr. 5,00
- « 10 Gerd Skoe Lettenstrøm og Gisle Skancke: De yrkesaktive i Norge 1875—1960 og prognoser for utviklingen fram til 1970 *The Economically Active Population in Norway and Forecasts up to 1970* 1964 56 s. kr. 6,00
- « 11 Hallvard Borgenvik: Aktuelle skattetall 1965 *Current Tax Data* 1965 38 s. kr. 6,00
- « 12 Idar Møglestue: Kriminalitet, årskull og økonomisk vekst *Crimes, Generations and Economic Growth* 1956 63 s. kr. 7,00
- « 13 Svein Nordbotten: Desisjonstabeller og generering av maskinprogrammer for granskning av statistisk primærmateriale *Decision Tables and Generation of Computer Programs for Editing of Statistical Data* 1965 11 s. kr. 4,00
- « 14 Gerd Skoe Lettenstrøm: Ekteskap og barnetall — En analyse av fruktbarhetsutviklingen i Norge *Marriages and Number of Children — An Analysis of Fertility Trend in Norway* 1965 29 s. kr. 6,00
- « 15 Odd Aukrust: Tjue års økonomisk politikk i Norge. Suksesser og mistak *Twenty Years of Norwegian Economic Policy: An Appraisal* 1965 38 s. kr. 6,00
- « 16 Svein Nordbotten: Long-Range Planning, Progress- and Cost-Reporting in the Central Bureau of Statistics of Norway *Langtidsprogrammering, framdrifts- og kostnadsrapportering i Statistisk Sentralbyrå* 1966 17 s. kr. 4,00
- « 17 Olav Bjerkholt: Økonomiske konsekvenser av nedrustning i Norge *Economic Consequences of Disarmament in Norway* 1966 25 s. kr. 4,00
- « 18 Petter Jakob Bjerve: Teknisk revolusjon i økonomisk analyse og politikk? *Technical Revolution in Economic Analysis and Policy?* 1966 23 s. kr. 4,00
- « 19 Harold W. Watts: An Analysis of the Effects of Transitory Income on Expenditure of Norwegian Households 1968 28 s. kr. 5,00
- « 20 Thomas Schiøtz: The Use of Computers in the National Accounts of Norway *Bruk av elektronregnemaskiner i nasjonalregnskapsarbeidet i Norge.* 1968 28 s. kr. 5,00



Publikasjonen utgis i kommisjon hos
H. Aschehoug & Co., Oslo, og er til salgs hos alle bokhandlere
Pris kr. 7,00

Omslag trykt hos Grøndahl & Søn, Oslo
For øvrig trykt hos Reklametrykk A/S, Bergen