Arbeidsnotater

STATISTISK SENTRALBYRÅ
Dronningensgt. 16, Oslo-Dep., Oslo 1. Tlf. 41 38 20, 41 36 60

IO 74/6 30. januar 1974

METODEHEFTE NR. 10

Notater om dataterminologi, tidsnyttingsundersøkelsen 1971/1972, ferieundersøkelsene i 1968 og 1970, og konvergens i en klasse av matriseprosesser.

INNHOLD

FORORD

Metodehefter i serien Arbeidsnotater

I tilknytning til mange prosjekter i Statistisk Sentralbyrå utarbeides det mindre, utpretensiøse notater for avklaring av spørsmål av metodisk interesse. Det kan dreie seg om utvalgsteknikk, alternative spørsmålsformuleringer, presentasjonsmetoder, begrepsavklaringer, diskusjon av "funn" i data, systemidéer eller andre temaer. Selv om mange slike notater bare har begrenset interesse i ettertid, vil det blant dem være noen som kunne fortjene å bli alminnelig tilgjengelig. Det kan også være nyttig å ha dem registrert sentralt slik at det blir lettere å få oversikt over det stoffet som foreligger, og lettere å referere tilbake til det. Byrået publiserer derfor leilighetsvis et passende antall notater av dette slaget samlet i metodehefter i serien Arbeidsnotater.

Kontorlederne bes holde øynene åpne for denne nye publiseringsmuligheten.

Forsker Jan M. Hoem er redaktør av metodeheftene. Fullmektig Liv Hansen er redaksjonssekretær. Medarbeidere i Byrået som lager stoff som kan være aktuelt, bes sende dette til redaksjonen etter hvert som det blir ferdig. Retningslinjer for utformingen av inserater i metodeheftene finnes på side 46 til side 47 , Metodehefte nr. 9 (ANO IO 73/36).

NOEN BEGREPER VEDRØRENDE STATISTIKK OG PRIMÆRDATA
EN DISKUSJON AV TERMINOLOGI

Av Jan M. Hoem

Innhold

1. Innledning

Dette notatet tar i første rekke sikte på å bidra til en presisering og standardisering av gjeldende språkbruk i Statistisk Sentralbyrå omkring en del sentrale begreper vedrørende statistiske primærmaterialer og bearbeidet statistikk. Bare i liten utstrekning foreslås terminologi som ikke alt er i bruk. Notatet er også ment som et bidrag til den kodifisering av terminologi som er aktualisert av arbeidet med et system for sosiodemografisk statistikk.

Jeg har valgt å presentere definisjonene verbalt i en løpende tekst. Det har ikke vært lett å foreta en avveining mellom hensynene til generalitet, abstraksjon, leselighet og presisjon.

En vil derfor se at begrepene presenteres med ulikt presisjonsnivå, og at deti stor utstrekning appelleres til leserens egen innsikt og intuisjon. Begreper som det etter min oppfatning har vært særlig vanskelig å beskrive,er søkt belyst med.eksempler.

Et stikkordregister og et eksempelregister er tatt inn bakerst i notatet for å gjøre det mer egnet til oppslag.

Av tidligere arbeider som har noenlunde samme siktemål som dette, og som jeghar hentet impulser fra, kan jeg nevne Hellevik (1971, side 35-42), Øberg (1972), Hoffmann (1973), og Fastbom (1973). En annen fremstilling av lignende begreper i form av en ordliste med forklaringer finnes i Forsberg (1972). De nordiske statistiske foreninger har under arbeid en flerspråklig ordliste, som tar utgangspunkt i siste utgave av Kendall og Bucklands velkjente statistiske ordbok. Ingen av disse synes imidlertid å dekke akkurat samme behov som inneværende notat tar sikte på.

Diskusjoner med Erik Aurbakken, Svein Brenna, Roger Hansen, og særlig Erik Botheim og Eivind Hoffmann har vært til hjelp under utarbeidelsen av notatet.

2. Betegnelsene "bestand", "variabel", "kjennemerke", osv.

Når en statistisk undersøkelse gjennomføres av Statistisk Sentralbyrå, vil interessen som regel være konsentrert om en endelig samling (en populasjon, et univers) av elementer (observasjonsenheter, telleenheter, objekter), f.eks. bedrifter, husholdninger, eller personer. Denne samlingen vil vi omtale som undersøkelsesbestanden. (I eldre statistisk litteratur omtales ofte en bestand som en "masse". I nyere litteratur unngår en den siste betegnelsen, formodentlig fordi ordet gir assosiasjoner til massebegrepet i fysikken.)

Undersøkelsesbestanden kan være en samling av aktiviteter eller hendelser, f.eks. tyverier eller trafikkulykker, og det vi har å si, er ment å dekke også slike bestander. [Hoffmann (1973, kap. III) diskuterer slike muligheter.] For enkelthets skyld skal vi imidlertid ordlegge oss som om elementene er fysiske eller juridiske personer eller objekter.

For hvert element (medlem) i bestanden tenker vi oss utført en måling, slik at en får registrert en måleverdi. (En registrerer bedriftens brannforsikringsverdi, intervjuobjektets svar på et spørsmål, husholdningens månedlige husleie, osv.) I statistisk teori representerer vi en slik måling med en variabel definert over bestanden. Hvis vi kaller variabelen for v, vil det da til hvert medlem av bestanden svare en verdi av v, slik at medlem nr. k har variabelverdien v(k). Hvis bestanden har N medlemmer, kaller vi V = {v(1), v(2), ..., v(N)} for verdimengden til v. V er naturligvis endelig. I mange tilfeller vil V være en samling punkter innenfor et nærmere angitt område V' på tallinjen. Det kan da være nyttig å ha betegnelsen potensiell verdimengde på V'.

Som eksempel kan en tenke seg at elementene er norske skattytere i et gitt år og at v er skattbar inntekt i året. Mens V er samlingen av alle skattbare inntekter som forekommer dét året, kan det være nyttig å bruke V' = $[0,\infty>$, dvs. å "tillate alle ikkenegative skattbare inntekter å forekomme potensielt".

Vi har her knyttet variabelbegrepet i datasystemene helt til matematikkens variabelbegrep*). For å kunne arbeide med en variabel, er det således nødvendig å kjenne dens definisjonsområde (her: undersøkelsesbestanden), dens verdimengde (eller dens potensielle verdimengde), og det tilordningssystem som knytter disse sammen, dvs. de prinsipper målingen utføres etter. Det er viktig å holde disse tre aspektene fra hverandre. Til sammen definerer de variabelen. Hvordan de så symboliseres, enten det er med tall, bokstaver, eller på annen måte, er en sentral opplysning for datadokumentasjonen, men den hører ikke med i variabeldefinisjonen.

De symbolene en bruker for å representere verdimengden til en variabel, vil vi kalle for representasjonens "koder". Hvert symbol omtales altså som én kode, og en oppstilling over dem kalles en "kodeliste". For variabeldefinisjonen er det likegyldig hvilken kodeliste som benyttes, så lenge den er hensiktsmessig og kjent. Fra definisjonens synspunkt vil enhver éntydig transformasjon av kodelisten være like god.

---

*) En variabel skal her være en reell avbildning av {1, 2, ...,N}.

Dersom en endrer på definisjonen av bestanden, eller på den potensielle verdimengden (f.eks. ved å redusere antall mulige koder), går en imidlertid i prinsippet over fra én variabel til en annen, la oss si fra v_1 til v_2. Imidlertid vil en ofte beholde samme betegnelse på de to variablene, slik at v_1 og v_2 får samme navn. For eksempel snakker vi om "ekteskapelig status" både når kodelisten er {ugift, gift, separert, skilt, enke/enkemann} og når den er {ugift, gift, før gift}. Selv om navnet er det samme, er det allikevel viktig at en holder klart for seg at det fortsatt dreier seg om to variable, ikke bare én. En må ikke blande sammen betegnelse og begrep. Tenk bare på hvilken forskjell det er mellom variable som betegnes "arbeidsløshet" eller "flytting" når de er definert henholdsvis på individnivået (v_1) og i statistikk for kommuner (v_2).

Betegnelsen "kjennemerke" er synonym med "variabel". Tilsvarende brukes "kjennemerkeverdi" synonymt med "variabelverdi", osv.

I avsnittene nedenfor skal vi utdype vår omtale av disse begrepene.

3. Betegnelsene "klassifisering", "gruppering", osv.

La oss tenke oss verdimengden V delt opp i et antall disjunkte delmengder $V_1,\ldots,V_J$. (Disse kan eksempelvis være avledet av en oppdeling av V' i delintervaller.) Dette gir opphav til en klassifisering av elementene i bestanden ved at element nr. k sies å tilhøre klasse j dersom $v(k)\varepsilon V_j$. Vi har her nummerert klassene fra 1 til J. I mange sammenhenger, f. eks. i offisielle standarder, vil en utvikle en systematisk klassebetegnelse ved hjelp av sifre og eventuelt andre symboler. En slik betegnelse kalles klassekoden. Hvis hver klasse får et pregnant navn, genereres det en tilsvarende klassebetegnelse for elementene i klassen. (Eksempel: kommunetypebetegnelsene.)

En klassifisering av elementene i en bestand gir umiddelbart opphav til en ny variabel, avledet av den som ligger til grunn for klassifiseringen. Hvis vi kaller den opprinnelige variabelen for v, får vi en avledet variabel v^* ved å sette $v^*(k) = j$ hvis element nr. k tilhører klasse nr. j, slik at

$$v^*(k) = j \text{ hvis og bare hvis } v(k)\varepsilon V_j.$$

Igjen vil v og v^* ofte få samme navn. Ikke desto mindre dreier det seg om to ulike variable (bortsett fra i det trivielle tilfelle da hver V_j bare har ett element). (Eksempel: ett- og femårige aldersklasser.)

La oss tenke oss at det er definert et avstandsmål d over bestanden, slik at avstanden mellom elementene k og ℓ er $d(k,\ell)$.

Elementene i bestanden kan da grupperes slik at elementene kommer i samme gruppe dersom $d(k,\ell)$ er liten. Hvis en krever $d(k,\ell) = 0$ for at elementene k og ℓ skal komme i samme gruppe, får en i prinsippet ingen problemer med å angi grensene mellom ulike grupper. Dersom en tillater d- avstanden mellom elementene i en gruppe å bli positiv, må det fastlegges kriterier for hvor stor d- avstanden skal kunne bli innen en gruppe, og for hvor grensene mellom gruppene skal trekkes.

Formelt sett er en gruppering og en klassifisering akkurat det samme, i den forstand at enhver klassifisering er en gruppering og enhver gruppering er en klassifisering.

La nemlig bestanden være klassifisert etter variabelen v, og la

$$d(k,\ell) = |j_k - j_\ell|,$$

der j_k og j_ℓ er numrene på de klassene elementene k og ℓ tilhører. Da er klassifiseringen en gruppering av elementene generert av kravet $d = 0$.

La det omvendt foreligge en gruppering av bestanden, og la en variabel v være definert ved at $v(k) = g_k$, der g_k er nummerert på den gruppen elementet k tilhører. Da er grupperingen samtidig en klassifisering etter v.

Det er allikevel nyttig å ha begge betegnelsene, både gruppering og klassifisering. Betegnelsen gruppering leder nemlig tanken i retning av at det skal foreligge en innledning av bestanden i grupper som i en meningsfylt forstand er homogene. En har bruk for en terminologi som kan ta med dette substansielle aspektet, selv om det formelt sett er overflødig.

Ovenfor har vi definert en klassifisering med utgangspunkt i en enkelt variabel. Formelt sett er dette tilstrekkelig, siden en med utgangspunkt i flere variable definert over den endelige bestanden alltid en-entydig kan definere en enkelt avledet variabel som alene representerer dem alle.

Det er allikevel nyttig å kunne omtale en fler-dimensjonal klassifisering.

La variablene $v_1, v_2, \ldots, v_n$ være definert over bestanden, og la $\underset{\sim}{v} = (v_1, \ldots, v_n)$. Verdimengden til $\underset{\sim}{v}$ er $\underset{\sim}{V} = \{\underset{\sim}{v}(k) : k = 1,2,\ldots,N\}$. Det foreligger en fler-dimensjonal klassifisering av bestanden tilsvarende hver oppdeling av $\underset{\sim}{V}$ i disjunkte delmengder $\underset{\sim}{V}_1, \ldots, \underset{\sim}{V}_J$. Den fler-dimensjonale klassifiseringen kan kalles en kryss-klassifisering etter $v_1, v_2, \ldots, v_n$ hvis den er fremkommet ved at verdimengdene $V_1, \ldots, V_n$ til de enkelte variable er delt opp i disjunkte undermengder V_{ij} ($i = 1,2,\ldots,n$; $j = 1,2,\ldots,J_i$) og hver $\underset{\sim}{V}_j$ er fremkommet som et kryssprodukt av typen

$$\underset{\sim}{V}_j = V_{1j_1} \times V_{2j_2} \times \ldots \times V_{nj_n}$$

$$= \{\underset{\sim}{v}(k) : v_i(k) \in V_{ij_i} \text{ for } i = 1,2,\ldots,n\}.$$

I praksis kan en gruppering fremkomme f.eks. ved at klasser i en opprinnelig, mer finmasket klassifisering (eventuelt en kryssklassifisering) slås sammen til større grupper.

En kan også få en gruppering av personer med utgamgspunkt i en opprinnelig klassifisering av sosiale systemer som disse personene tilhører. For eksempel kan det foreligge en klassifisering av bedrifter, la oss si i næringer, og dette gir opphav til en tilsvarende gruppering av arbeidstakere etter næring.

Stadig grovere grupperinger av elementene i en gitt bestand kan gi opphav til nye bestander av elementer på stadig høyere aggregeringsnivåer når det som representerer en gruppe elementer på ett nivå, oppfattes som et element på et høyere nivå. Eksempelvis kan personer grupperes i husholdninger, husholdninger i kommunale befolkninger, disse i fylkesbefolkninger, osv.

Variable kan være definert over bestander på hvert av disse nivåene. Noen ganger kan en variabel på et mer aggregert nivå fremkomme ved summering over variabelverdier fra et lavere nivå. Slik fremkommer f.eks. husholdningsinntekt, samlet personlig inntekt i en kommune, osv. Andre slike transformasjoner er gjennomsnitt, ulike spredningsmål, empiriske regresjonskoeffisienter, osv. Variable fremkommet på dette viset, kan omtales som aggregatvariable. Det er igjen viktig å være oppmerksom på at de variable en arbeider med "før" og "etter" en slik aggregering, kan ha samme betegnelse til tross for at det dreier seg om ulike variable.

Andre ganger defineres en variabel direkte for elementene i en aggregater uten at den selv fremkommer ved summering over variabelverdier. Antall folkeskoler i en kommune er et eksempel. Slike variable omtales ofte som globale variable (i relasjon til elementene i aggregatene).

4. Tidsdimensjonen.

I det ovenstående har vi sett bort fra ett aspekt ved statistiske undersøkelser, nemlig det at de alltid vil referere seg til et tidspunkt eller en periode. Vi skal nå ta dette med i betraktning.

Hvis en følger en gruppe av statistiske undersøkelsesenheter over tid, vil medlemsstokken kunne endre seg ved at noen forlater gruppen

(bedrifter nedlegges, personer skifter egenskaper eller dør) mens andre kan tre inn i den. En slik gruppe av undersøkelsesenheter der det kan være tilgang og avgang av elementer, kan vi kalle en dynamisk bestand.

Siden elementene i en slik bestand kan skifte egenskaper, er det ofte problematisk å avgjøre hva som skal få betraktes som "samme element" av bestanden på to ulike tidspunkter. Hvis elementene er personer, er vel dette nokså likeframt, men det er ikke alltid så lett å bestemme seg for hva som skal få anses som samme husholdning over tid, for eksempel. Øberg (1972, avsnitt 3.3) har gitt en grei oversikt over slike spørsmål, og han innfører et interessant skille mellom identiske og gjenidentiske enheter. Slik disse begrepene beskrives av Hoffmann (1973, side 8), er medlemmer av en dynamisk bestand identiske bare hvis de på alle tenkelige variable har de samme verdier, også på tidsvariabelen. Et element er derfor bare identisk med seg selv. For å få sammenheng ved bevegelse langs tidsvariabelen innføres så begrepet gjenidentitet, som svarer til det vi i dagligtale mener når vi sier at en ting vi har observert på ett tidspunkt er den samme som den vi observerer på et annet tidspunkt. Personen med personnummer 121043-36165 i skattemanntallet pr. 31/12 1972 er identisk med personen med samme personnummer i Det sentrale personregister pr. samme dato, og det er slike sammenknytninger mellom ulike datakilder som gjør identitetsbegrepet interessant. Den ovennevnte personen er "bare" gjenidentisk med personen med samme personnummer 26/2 1973. Mer relevant er dette skillet naturligvis ved grupper av personer eller ved institusjoner, der identifiserende variable kan skifte verdi over tid. (Eksempel: Familier med samme familienummer i personregisteret på to ulike tidspunkter kan ha helt ulik sammensetning. Det eneste som trenger å være felles, er den referansepersonen som familienummeret referer seg til.) En må ha et sett av regler for å bestemme hvor store endringer som skal tillates i hvilke variable før en slutter å snakke om gjenidentiske elementer.

Spørsmål vedrørende gjenidentitet er betydningsfulle for en diskusjon av hva identifikasjonsnumre for elementer i en dynamisk bestand eventuelt skal inneholde av informasjon. I Norge kan vi jo lese en persons kjønn og fødselsdato rett ut av personnummeret og fylket ut av kommunenummeret. Men hvordan skal en bygge opp identifikasjonsnumre for bedrifter (i bedrifts- og foretaksregisteret) eller helseinstitusjoner (i Det økonomisk-medisinske informasjonssytem)? Statistisk Sentralbyrås holdning har vært at det er uheldig å ta inn informasjon om enheten i identifikasjonsnummeret, og en begrunnelse har vært at enheten ville skifte identitet når et innebygget kjennetegn skiftet verdi.

Hvis kjennetegnene i identifikasjonsnummeret var sentrale nok, kunne det imidlertid være fordelaktig med et slikt automatisk identitetsskifte. Om en legger opp et system uten slik automatikk, kan gjenidentitet lett bli knyttet til mindre vesentlige (kanskje vilkårlige) kriterier, som f.eks. eksakt lokalisering. [Poengene i dette avsnittet skyldes Eivind Hoffmann.]

La oss imidlertid nå forutsette at en har løst de praktiske problemene knyttet til definisjonen av gjenidentitet, og la elementene i en dynamisk bestand entydig være tildelt permanente identifikasjonsnumre. Enkelte variable vil da være definert over de elementene som er medlemmer av bestanden på et gitt tidspunkt, nemlig tilstandsvariable og beholdningsvariable. Hvis samme formelle definisjon brukes til å definere slike variable over en dynamisk bestand for flere tidspunkter, slik at en får frem en serie variable med samsvarende definisjoner, kan en si at variabeldefinisjonene er integrert over tid og at de tilsvarende datamengdene er sammenlignbare over tid.

Det kan her være tale om atskilte tidspunkter t_1, t_2,..., eller om observasjon som i prinsippet er kontinuerlig over et gitt tidsrom (slik som ved et løpende personregister), eller en annen tidsmengde. Vi vil i alle fall kalle samlingen av tidspunkter som kommer i betraktning, for T. For hvert tidspunkt $t\varepsilon T$ tenker vi oss da angitt en variabel v_t definert over elementene i bestanden. Vi antar at $\{v_t : t\varepsilon T\}$ er en samling av ulike variable med samsvarende definisjoner. I dagligtalen vil en gjerne omtale disse som "samme variabel v målt på ulike tidspunkter t i T", og det er ingen grunn til at man skulle forsøke å "forby" en slik språkbruk i fagspråket heller. Den er enkel, presis, og fullt tillatt etter den matematiske definisjonen av begrepet variabel. [Fra matematisk synspunkt vil det her dreie seg om en variabel v med verdimengde $\{v_t(k) : t\varepsilon T, k\varepsilon B_t\}$, der B_t er samlingen av elementer i bestanden på tidspunkt t.]

Tilsvarende betraktninger gjelder for strømningsvariable, men det er vel nødvendig med en viss utdypning. [Vi bygger her delvis på Haavelmo, 1951.]

Vi skal sondre mellom to typer av strømningsvariable, nemlig
(i) slike som refererer seg til et gitt tidsrom, som f.eks. en persons skattbare inntekt i et gitt år, og en bedrifts investeringer i et gitt kvartal, og
(ii) slike som refererer seg til et gitt tidspunkt, som f.eks. en persons inntekt på årsbasis regnet etter inntekten pr. 1/7 1973.

Den første typen vil vi kalle periodiserte strømningsvariable. Den andre typen vil vi kalle intensitetsvariable. La oss betegne en strømningsvariabel definert for perioden $[t,t+s\rangle$ med $v(t,s)$. For å konkretisere, kan vi tenke på en persons inntekt i dette tidsrommet. Hans inntekt pr. tidsenhet i $[t,t+s)$ er da $\bar{v}(t,s) = v(t,s)/s$, og hans inntekt pr. tidsenhet på tidspunkt t er

$$w(t) = \lim_{s \to 0} v(t,s)/s.$$

Mens v og $\bar{v}$ er periodiserte strømningsvariable, er w en intensitetsvariabel.

Språkbruken vedrørende integrasjon og sammenlignbarhet over tid osv. kan nå overføres direkte til intensitetsvariable. For periodiserte strømningsvariable går det like greit når en bare skifter ut ord som "tidspunkt" med ord som "periode" overalt. Eksempelvis blir det tale om "samme periodiserte strømningsvariabel v målt for ulike perioder t i T", der T nå blir en samling tidsintervaller. Hvis disse utgjør ikke-overlappende delintervaller av en lengere periode, er det naturlig å nummerere dem fortløpende og la t betegne periodenummeret eller noe lignende.

For en strømningsvariabel definert for en gitt periode blir definisjonsområdet vanligvis samlingen av de elementer som er medlemmer av den dynamiske bestanden en eller annen gang i løpet av perioden.

5. Mer om betegnelsene "sammenlignbarhet", "integrasjon", osv.

I avsnitt 4 ovenfor omtalte vi bl.a. variabeldefinisjoner som er integrert over tid og datamengder som er sammenlignbare over tid. Tilsvarende skal vi si at det foreligger integrasjon innen et statistikkområde dersom en har samsvar mellom begreper, enheter og definisjoner for all statistikk innen området. (Analogt for deler av et statistikkområde.) Dette innebærer f.eks. at en har samsvar mellom periodiserte strømningsvariable målt på kvartalbasis og korresponderende variable målt på årsbasis. Kvartals- og årsstatistikkene blir da sammenlignbare. Integrert statistikkproduksjon gir altså opphav til sammenlignbare data (sammenlignbar statistikk). Der hvor samsvaret er ivaretatt når det gjennomføres aggregering (over tid eller over elementene i en bestand eller begge deler), snakker en om vertikal integrasjon.

Analogt sier en at en har sammenlignbarhet mellom statistikken fra to statistikkområder (eller deler av dem) hvis det foreligger tilsvarende samsvar områdene imellom, og det sies å foreligge horisontal integrasjon mellom områdene. Skal en eksempelvis ha sammenlignbarhet mellom næringsstatistikk og lønnsstatistikk, må bl.a. definisjonen av

"bedrift" være den samme i begge områder.

Sammenligning mellom datamengder (tabeller, statistikk) kan naturligvis foregå på ulike aggregeringsnivåer. Noen ganger kan en sammenligne individuelle variabelverdier for medlemmene i to bestander. Andre ganger foretar en sammenligninger mellom summer eller gjennomsnitt av variabelverdier (aggregattall) for grupper av elementer, eller mellom totalsummer for to eller flere bestander. Ordet "sammenlignbarhet" brukes til dels upresist om alle slags muligheter til å holde ulike tallmaterialer opp mot hverandre.

I neste kapittel skal vi ta for oss to begreper som er relevante når det foreligger sammenlignbarhet.

6. Samkobling og sammenbinding.

Verdier for to ulike variable, la oss si v_1 og v_2, definert over én og samme bestand, vil kunne foreligge i to forskjellige datakilder. Når disse variabelverdiene hentes ut fra de to datakildene og føres sammen på ett datamedium, sier vi at de to materialene samkobles. Skal samkobling kunne finne sted, må en altså ha observasjoner på det som i den gitte situasjon betraktes som elementnivået (individnivået). Det er også en forutsetning at definisjonen av bestandens enheter og omfang er felles for de to datakildene. (Eksempel: Levekårsundersøkelsen og Forbruksundersøkelsen 1973.)

La oss nå tenke oss at det foreligger en klassifisering K av elementene i en bestand, og at en kjenner fordelingen av elementene i hver klasse i K etter to variable, w_1 og w_2, uten at en har tilgjengelig individobservasjonene av verdiene for begge disse variable.

En kan da lage aggregattabeller kryssgruppert etter K og w_1, og separat etter K og w_2, men ikke aggregattabeller kryssgruppert etter w_1 og w_2 (eller K,w_1 og w_2 simultant). Vi vil foreslå at en i et slikt tilfelle omtaler K som et bindeledd mellom w_1 og w_2. Dette er et forsøk på oversettelse av det svenske ordet "länk" for samme begrep. (Se Fastbom, 1973, punkt 2.) Vi foreslår dessuten at w_1 og w_2 sies å være bundet sammen ved hjelp av K.

Sammenbundne materialer inviterer til visse typer av feilslutninger som det er grunn til å advare mot. Selv om en får vite at 33 % av norske 19-åringer var under utdanning i 1971, og at 2 % av norske 19-åringer var siktet for forbrytelser det året, kan en ikke slutte noe eksakt om hvor stor del av 19-åringene under utdanning som var siktet. Hellevik (1971, side 276 ff) gir en interessant diskusjon av feilslutninger av denne typen.

Stikkordregister med sidehenvisning.

Eksempelregister med sidehenvisning.

Referanser.

[1] Fastbom, L. (1973): "Några kommentarer avseende Eivind Hoffmann: Linkages within the System of Social and Demographic Statistics: A Framework for Discussion." Nordiska utskottet för ett system för socio-demografisk statistik, Kommentarer 1973-02-23.

[2] Forsberg, Robert (1972): "Standard avseende vissa statistiska grunnbegrepp." Statistiska Centralbyrån, Stockholm, SCB metodinformation Nr. 72:4.

[3] Hellevik, Ottar (1971): "Forskningsmetode i sosiologi og statsvitenskap." Universitetsforlaget.

[4] Hoffmann, Eivind (1973): "Noen informasjonsteoretiske grunnbegreper og deres relevans for arbeidet med et system for sosiodemografisk statistikk." Notat EH/EH, 24/2-73. Gjengitt på side 16 til 24 i dette metodehefte.

[5] Haavelmo, Trygve (1951): "Noen alminnelige merknader om tidsrekker og tidsrekkeanalyse." Avsnitt III.7 i "Innføring i teoretisk statistikk. Del II" av T. Haavelmo, redigert av Arne Amundsen; Memorandum av 15/2 1951 fra Sosialøkonomisk institutt, Universitetet i Oslo. Også gjengitt på side 54-59 i "Utdrag av forelesninger i teoretisk statistikk fra kompendier av Ragnar Frisch og Trygve Haavelmo", redigert av Herdis Thorén Amundsen; Universitetsforlaget, 1964.

[6] Kendall, Maurice G. and William R. Buckland: "A Dictionary of Statistical Terms." Oliver and Boyd, for the International Statistical Institute. Flere utgaver.

[7] Øberg, S. (1972): "Ett system för socio-demografisk statistik (SSDS), grunnläggande begrepp." Statistiska Centralbyrån, Stockholm, Rapport 1972.09.18.

NOEN INFORMASJONSTEORETISKE GRUNNBEGREPER OG DERES RELEVANS FOR ARBEIDET MED ET SYSTEM FOR SOSIODEMOGRAFISK STATISTIKK[1)]

av EIVIND HOFFMANN

INNHOLD

1) Notat til møtet i Nordisk utvalg for et system for sosiodemografisk statistikk. Oslo 26/2 - 28/2-73. Litt bearbeidet.

NOEN INFORMASJONSTEORETISKE GRUNNBEGREPER OG DERES RELEVANS FOR ARBEIDET MED ET SYSTEM FOR SOSIODEMOGRAFISK STATISTIKK av EIVIND HOFFMANN

I INNLEDNING

1. De følgende merknader har som viktigste utgangspunkt et notat av Svante Øberg [5] samt den diskusjon av notatets første versjon [4] som fant sted på møtet i Nordisk utvalg for et system for sosiodemografisk statistikk (NUSD) i København, januar 1972. Dette notatet er en redegjørelse for noen av de spørsmål som har reist seg i forsøket på å komme fram til slike konkretiseringer av de generelle begrep i [4] og [5] som bør følge av at vi skal fram til et system for sosiodemografisk statistikk; og et forsøk på å vurdere hvor nyttige de vil være for forståelsen og oppbyggingen av et SSDS.

II HVA MENER VI MED ET SSDS?

2. Jeg vil ta utgangspunkt i at SSDS skal være et system for å organisere statistiske opplysninger om den del av samfunnet som er definert ved

a forhold som gjelder enkeltpersoner og/eller grupper av personer; og
b de forbindelser mellom enheter som fremgår av følgende skjema;[1)]

	Personer	Grupper av personer	Institusjoner
Personer	1	2	4
Grupper av personer		3	5

1) Begrepene i skjemaet er nærmere omtalt i avsnitt III nedenfor.

på en slik måte at data om personer (og grupper av personer) som er innsamlet på ulike tidspunkt og/eller over ulike sett av variable kan samordnes så langt det er mulig utfra de prinsippielle beskrankninger og sammenhenger og de administrative og analytiske behov som ligger til grunn for systemet.[2)]

At et SSDS bør omfatte statistisk informasjon som knytter seg til forbindelsene 4 og 5 i tabellen ovenfor er kanskje ikke like innlysende som at informasjon om 1 - 3 må omfattes av systemet. Poenget er at den statistiske informasjon som et SSDS skal organisere, vil ha som formål å gi en beskrivelse og bidra til en forståelse av hvordan menneskene i

1) Begrepene i skjemaet er nærmere omtalt i avsnitt III nedenfor.
2) Noen synspunkter på hvilke beskrankninger og sammenhenger vi vil ha i et SSDS er presentert i [2].

et samfunn har det, og hvordan utviklingen i deres vilkår er. En slik beskrivelse og forståelse vil være ufullstendig hvis ikke informasjon om menneskenes og institusjonelle omgivelser trekkes inn. Vi må derfor i et SSDS få med informasjon om forbindelsene mellom personer og institusjoner - selv on informasjonene om institusjonene qua institusjoner organiseres utenfor systemet, f.eks. i et nasjonalregnskap.

Å gi en mer presis avgrensing av området for et SSDS er ellers vanskelig. Når strukturen i SSDS er blitt klarere og utprøvet vil det vise seg om avgrensingen også er vanskelig i praksis.

3. Denne forståelsen av et SSDS svarer antakelig ganske godt til Øbergs, når han på side 4 i [5] aier at "SSDS skal vara et system med tre delsystem:

- et system av definitioner
- et system av data
- et produktionssystem".

selv om jeg vil foretrekke å si at SSDS skal være et system for organisering av data som forutsetter at det finnes et system av definisjoner og klassifikasjoner og bl.a. derfor vil stille bestemte krav til vårt produksjonssystem for statistikk. Jeg er ikke helt sikker på at min forståelse av hva som er et system for data, slik det er skissert ovenfor, samsvarer med Øbergs, men den diskusjonen vil jeg utsette til et senere avsnitt.

III OBJEKT, EGENSKAP, RELASJON

4. Før Øberg diskuterer nærmere hva han forstår med henholdsvis et system av definisjoner, et system av data og et produksjonssystem redegjør han for noen grunnleggende informatologiske begreper, hvorav tre fremgår av overskriften til dette avsnittet. Det er ikke klart for meg hvilken rolle disse begrep vil spille i utformingen av et SSDS, og i dette avsnittet vil jeg søke å klargjøre hva min uforstand grunner seg i.

Øberg har sagt følgende om begrepenes funksjon [5], s. 5: "När man praktiskt försöker exempelvis dokumentera data måste dessa begrepp konkretiseras gjenom att man anger vad som skall betraktas som objekt, som egenskaper, som relationer och som variabler." Dette eksemplet gjelder alså dokumentasjon og knytter seg således til produksjonssystemet, og det har ledet meg til å søke etter svar på følgende spørsmål:

1. Er det i en SSDS-sammenheng ønskelig å konkretisere hva som skal betraktes som objekter, egenskaper og relasjoner?

2. Er disse begrepene - med eller uten relevante konkretiseringer - nyttige ved oppbyggingen av systemet for data (som jeg oppfatter som den sentrale delen av et SSDS)? (For dokumentasjonsformal - dvs. i tilknytting til produksjonssystemet - er disse begrepene meget rentable såvidt jeg kan forstå.)

5. La meg først se litt på spørsmål 2 da jeg tror at diskusjonen av det også kan bidra til å finne svar på spørsmål 1. Jeg vil ta utgangspunkt i objekt, da det som Øberg påpeker i [4] er nær sammenheng mellom avgrensingen av objekter og hva som blir egenskaper og relasjoner i systemet. Øberg definerer, [5] s. 5, et objekt som det "om hvilket vi vill registrera eller hämta inn uppgifter i det fysiska systemet, samhället, som SSDS skal ge information om . ..". Definisjonene av egenskaper og relasjoner tar så utgangspunkt i denne definisjonen. Med en slik definisjon knytter han forbindelsen direkte til prosuksjonssystemet og dokumentasjonen av det. Ved overføring til systemet for data må vi velge litt andre formuleringer: Et objekt er det 'noe' som det skal gis informasjon om. Om objektet skal det gis informasjon vedrørende egenskaper ved det og de relasjoner det har til andre objekter. En egenskap ved objektet A gis det informasjon om når vi sier noe om A, uten samtidig eksplisitt eller implisitt å si noe om andre objekter. En relasjon med tilknytting til A sier vi noe om hvis vi også trekker inn i utsagnet minst et annet objekt. ("A er ti år gammel" er et utsagn om en egenskap ved A."A er datter til B" sier noe om en relasjon mellom A og B.)

6. Fra den forståelsen av hva SSDS skal være som jeg redegjorde for i avsnitt II vil jeg finne det rimelig at vi f.eks. skjelnet mellom følgende grupper av objekter i et SSDS (dvs. i system av data):

a Personer. Nærmere definisjon antas å være overflødig. Ved anvendelse av informasjonen vil mengden av personer naturligvis avgrenses ved ett sett av variabelverdier - f.eks. ved "alle personer bosatt i Norge i hele perioden 1. januar 1974 - 31.desember 1975".

b Grupper av personer. I [3], s. 88, er en gruppe av personer definert som "en delmängd av samhällets totala population ... (der) ... samtliga enheter i den har vissa betämda tilstånd eller egenskaper, står i betämda relationer til andra enheter av olika slag och/eller är (har varit) utsatta för betämda förändringar eller händelser". En slik gruppe kan være definert for et bestemt tidspunkt eller en tidsperiode, men i denne sammenheng må det være rimelig å la definisjonen være uten tidsangivelse hvis det ikke spesielt gjelder en kohort. (En annen sak er

at i en analyse der gruppen inngår, må vi presisere hvilken tidsperiode analysen gjelder.) Dette fordi vi i første rekke er interesserte i slik informasjon som gjelder gruppen som sådan - f.eks. om og hvordan den fungerer som en enhet - og endringer i den. Hvis vi f.eks. i en analyse i første rekke er interesserte i de enkelte personer i gruppen i en periode og egenskaper/relasjoner ved dem, så er det personene som er objekter i analysen.

c Institusjoner. Det må være rimelig å tenke seg en forholdsvis vid forståelse av hva som er institusjoner. Foruten offentlige organer som departementer, sykehus, skoler og bedrifter, foreninger og organisasjoner, bør man antagelig inkludere omgivelsene - nabolaget, kirkesognet, kommunen - ikke som administrative enheter, men som miljøer i vid forstand. Det vil kunne være vanskelig å avgrense f.eks. foreninger og organisasjoner qua institusjoner fra at de er grupper av personer. Eksistensen av et byråkrati i problemstillingen (og i foreningen) kan være et kriterium. I stor utstrekning må altså avgrensingen framgå av sammenhengen. Institusjonene får rang av en slags annenrangs objekter i system, siden det som nevnt ovenfor, ikke skal organisere annen informasjon om dem enn den som knytter seg til deres relasjoner til personer eller grupper av personer.

7. Med denne avgrensingen av objekter vil egenskaper og relasjoner nok omtrent svare til hva jeg i avsnitt II har betegnet som de 'forhold' og 'forbindelser' som SSDS skal organisere informasjon om. Men kanskje det da er like naturlig å velge dem som objekter i systemet?

De egenskaper og relasjoner (forhold og forbindelser) som vi har særlig interesse av for SSDS, er slike som har relevans for å belyse personers og gruppers sosiale og demografiske tilstander (situasjoner), aktiviteter og begivenheter. Disse vil derfor også kunne være aktuelle som objekter i SSDS.

Det er ingen grunn til å tro at dette uttømmer listen over mulige og rimelige avgrensinger av hva som bør kunne være objekter i SSDS.

8. Konklusjonen på disse betraktningene kan såvidt jeg ser være

a Den struktur vi etterhvert kommer fram til for SSDS vil kunne føre til at settet av objekter i SSDS får en helt bestemt avgrensing. Eventuelt kan det være slik at en bestemt avgrensing - med de avgrensinger av hva som er egenskaper og relasjoner som følger av denne - hjelper oss i særlig grad til å finne en tilfredsstillende struktur på SSDS.
eller

b Vi finner intet avgrenset sett av objekter som vurderes som tilfredsstillende utfra målsettingene med SSDS. Siden de ulike mulige sett har svært ulike konsekvenser for hva som blir egenskaper og relasjoner, synes det lite rimelig at vi i dette tilfelle klarer å finne fram til en hensiktsmessige struktur på datasystemet i SSDS med utgangspunkt i disse begrepene. Jeg vil altså tro at i dette tilfelle vil de begrepene som er diskutert ovenfor spille en begrenset rolle ved utformingen av struktur datasystem SSDS.

Ut fra den struktur for et SSDS som vi nå etterhvert kan skimte, blant annet i Øbergs to notater [5] og [6]og i mitt eget [2], kan jeg ikke se at konklusjonen a virker rimelig. I denne strukturen inngår foruten "regnskapselementene" (se [2]) personer, tiden, areal og ressurser, også aktiviteter, resultater, tilstander og hendelser som sentrale elementer. I denne strukturen vil vi sikkert kunne identifisere enkelte elementer som henholdsvis objekter, egenskaper og relasjoner, men det vil være ex post og denne inndelingen vil ikke ha noen grunnleggende rolle ex ante ved utformingen av systemet.

9. I produksjonssystemet vil det såvidt jeg kan forstå ikke være spesielt ønskelig å avgrense et bestemt sett av objekter som gjeldende i eller tilhørende et SSDS. Ved de forskjellige statistiske undersøkelser vil hva som er objekt i undersøkelsen i stor grad avgjøres av hvordan oppgavene er samlet inn. Samme fenomen kan belyses ved data fra undersøkelser som på grunn av ulike opplegg har ulike ting som objekter. For dokumentasjonen av slike undersøkelser vil det være sentralt å få klarlagt hva som er objekter, hvilke sett av egenskaper som er observert om de enkelte objekter og hvilke sett av relasjoner som er observert mellom dem.

Mitt svar på det første av spørsmålene i begynnelsen av dette avsnittet vil derfor være negativt. Det synes ikke ønskelig å spesifisere generelt hva som er objekter, egenskaper og relasjoner i det produksjons- og dokumentasjonssystem som støtter opp under SSDS som et system av data. Øberg synes da heller ikke å ha gjort noe slikt forsøk på spesifikasjon, selv om han i det som er sitert fra [5], ovenfor antyder nokså sterkt at det må gjøres.

IV TIDEN

10. Tiden og tidsbegrepet vil være sentralt i et system for statistikk som skal kunne belyse dynamiske problemstillinger og utviklinger over tid. I SSDS gjelder dette både for systemet for data og for produksjons- og dokumentasjonssystemet. Øberg har i avsnitt 3.3 i (5) gitt en grei oversikt om hvordan tiden påvirker behandlingen av objekter. Oversikten har i første rekke relevans for dokumentasjonssystemet. Av særlig interesse er hans skille mellom identiske og gjenidentiske objekter. Objekter er bare identiske om de på alle tenkelige variable har de samme verdier - også tidsvariablen. Det betyr at objekt bare er identiske med seg selv. Med andre objekter kan de være identiske med hensyn til et sett av variable. For å få sammenheng ved bevegelse langs tidsvariablen innføres begrepet gjenidentitet, som svarer til det vi mener i dagligtalen når vi sier at en ting vi har observert på et tidspunkt er den samme som den vi observerer på et annet tidspunkt. Personen med personnummer 12104336165 i skattemantallet 31/12-72 er identisk med personen med samme personnummer i personregistret 31/12-72, og gjenidentisk med personen med dette personnummer 26/2-73. Mer relevant er dette skillet ved grupper av personer eller institusjoner, hvor identifiserende variable kan skifte verdi over tid. Vi må ha et sett av regler for å bestemme hvor store endringer i hvilke variable som kan tillates før vi slutter å snakke om gjenidentiske objekter. Slike regler må fastsettes under hensyntagen til både data og produksjonssystemet i SSDS, og til den anvendelse data vil få.

11. I systemet for data vil tiden være sentral på to måter. For det første vil tiden være en grunnleggende dimensjon som er med på å binde hele systemet sammen og danne dets struktur, gjennom at de fenomener vi er interesserte i opptrer samtidig eller sekvensielt. Dette allokeringsperspektivet på tiden/som også er berørt i mitt notat [2], er noe av det som gjør at f.eks. spørsmålet om når noe fant sted er så sentralt. For det andre vil varigheten (alderen) av tilstander etc. være en sentral variabel i mange av de problemstillinger SSDS-organiserte data skal nyttes til å belyse.

V VARIABEL

12. Det er naturligvis helt nødvendig å ha klart for seg hva vi skal mene med en variabel i SSDS - både i dokumentasjonssystemet og i systemet for data. Øberg har i avsnitt 3.4 [5] en forholdsvis klar framstilling som - såvidt jeg kan forstå - knytter variabelbegrepet i SSDS nær til matematikkens variabelbegrep. For å kunne arbeide med en variabel er det

således nødvendig å kjenne dens definisjonsområde, verdimengde, tilordningssystem og opprettelses- og opphørstidspunkt. Det er de to første elementene som definerer variablen (og de to neste vil ofte framgå av disse). (Dette er omtalt litt nærmere i [1].) I SSDS vil en variabels verdimengde gjerne bestå av ett sett av egenskaper. Hvordan vi så symboliserer disse egenskapene - med tall, bokstaver eller på annen måte - er en sentral opplysning for dokumentasjonen av data, men hører ikke med i definisjonen av variablen. Det er likegyldig hvilket av de mulige symbolsett for verdimengden som velges i sammenheng, så lenge det er hensiktsmessig og kjent, og det er en en-entydig sammenheng til andre mulige symbolsett på samme aggregeringsnivå. Øberg virker uklar i sin behandling av dette i [5], bilag 1, synes jeg; og det virker også som om han merner at det er en motsetning som jeg ikke ser, mellom det matematiske variabelbegrep og det vi skal anvende.

For variabler som er definert over samme sett av egenskaper, men med ulik aggregeringsgrad (spesifiseringsgrad) - dvs. med ulike tilordningssystemer - må det finnes entydige regler for overgangen fra et lavere til et høyere aggregeringsnivå, for at de skal kunne anvendes innenfor samme system.

I tillegg til denne form for aggregering - en sammenslåing av elementer i verdimengden for en variabel - som er reflektert f.eks. i den pyramidiske oppbygging av mange statistiske standarder, snakker vi gjerne om aggregering i form av sammenslåing av elementer i definisjonsmengden for en variabel - f.eks. sammenslåing av personer til grupper av personer. Slike sammenslåinger gir oss naturligvis nye objekter og også nye variable. Bl.a. fordi de variable med 'før' og 'etter' slike aggregeringer ofte har samme betegnelse, er det viktig at vi i arbeidet med SSDS er oppmerksom de problemer som lett følger med slike ulike analysenivåer.

VI SYSTEM AV DATA

13. Som nevnt ovenfor er jeg ikke sikker på at Øberg og jeg mener helt det samme når vi sier henholdsvis "et system av data" og "og et system for (organisering av) data". Øberg synes å definere data som "fysisk representasjon av informasjon" [5], side 9, mens jeg mener det samme foran er betegnet som "statistiske opplysninger". Et system for data slik jeg forstår det er en måte å organisere disse på slik at man får sammenheng og konsistens i dem - jfr. avsnitt II. Hva Øberg forstår med dette synes dels å være bestemte sett av tabeller; men det er også individdata, der data fra en miniatyrpopulasjon bør spille en sentral

rolle [5], s. 4 og 25. Slik jeg ser det er tabellsett en viktig presentasjonsform for data organisert i et SSDS og individdata er et viktig råmateriale for denne organisering, mens en miniatyrpopulasjon er en potensielt viktig kilde til slikt råmateriale. Jeg mistenker at mitt "system for data" i meningsinnhold ligger nær opp til Øbergs "system av definitioner" (jfr. framstillingen i [5], s. 4 og 11) uten at jeg tror at denne terminologiske uoverensstemmelse er vesentlig for diskusjonen i avsnittene foran.

VII KONKLUSJON

14. Det virker på meg som om den diskusjonen av sentrale informasjonsteoretiske begreper som Øberg gir, er viktig i oppbyggingen av et dokumentasjonssystem for data, som igjen er et viktig, men ikke sentralt, ledd i SSDS-arbeidet. For den sentrale delen av SSDS synes diskusjonen å ha mindre relevans, ved at den ikke synes å gi noe grunnlag å arbeide ut fra i det fortsatte arbeidet med å finne en struktur for SSDS. I diskusjonen av variabel synes han å gjøre dette begrepet mer komplisert i en SSDS-sammenheng enn det trenger være.

HENVISNINGER

[1] Hoem, J.M. og Hoffmann, E: Variabelkatalogen. Notat JMH/EH/ES, 16/6-72. Gjengitt på s. 26-29 i Metodehefte nr. 6, ANO IO 73/8.

[2] Hoffmann, E.: Linkages within the System of Social and Demographic statistics: A Framework for Discussion. Socio-Demographic Research Unit, Central Bureau of Statistics, Oslo, EH/GH, 14.II.73.

[3] Larsson, B.: "Identifikation av statistisk information." Statistisk tidsskrift 1969: 2 & 3.

[4] Øberg, S.: SSDS-grunnläggande begrepp. Utkast PM. Statistiska Centralbyrån, 1.11.71

[5] Øberg, S.: Ett system för socio-demografisk statistik (SSDS), grunnläggande begrepp. Rapport. Statistiska Centralbyrån. 1972.09.18.

[6] Øberg, S.: Innhållsstrukturering av SSDS. PM. Statistiska Centralbyrån, 1973-02-12.

Sig H/GH, 9/5-73.

Tidsnyttingsundersøkelsen 1971/1972

av Sigurd Høst.

INNHOLD

I. INNLEDNING

Gjennom Tidsnyttingsundersøkelsen 1971/1972 foretar Statistisk Sentralbyrå en prøve i full målestokk av en generell tidsnyttingsundersøkelse. Formålet med prøven er både å få testet om en gjennom det valgte opplegg kan komme fram til resultater av tilfredsstillende teknisk kvalitet, og om det er tilstrekkelig behov for den type opplysninger undersøkelsen gir til at tidsnyttingsundersøkelser kan gå inn som en del av Byråets ordinære undersøkelsesprogram.

Bakgrunnen for arbeidet med en generell tidsnyttingsundersøkelse har vært at en etter hvert har registrert behov for kunnskap om en lang rekke aktiviteter: kulturaktiviteter, friluftsliv, undervisning, arbeid og arbeidsreiser, husarbeid o.l. Ved en generell undersøkelse av tidsnytting håper en å dekke en god del slike behov som ellers ville kreve større eller mindre spesialundersøkelser.

En generell undersøkelse gir videre muligheter for å se de ulike aktivitetene i sammenheng, i og med at de alle foregår innenfor rammen av døgnets 24 timer. På denne måten kan tidsnyttingsundersøkelsen også fungere som en rammeundersøkelse i forhold til eventuelle undersøkelser av spesielle aktiviteter.

En sammenlikning mellom forskjellige tidsbruksundersøkelser vil gjøre det klart at de metodiske og teoretiske problemer en slik undersøkelse reiser er forsøkt løst på til dels svært forskjellige måter. Det dreier seg altså ikke om en forskningstradisjon med klart definerte teoretiske problemer og metodiske retningslinjer. Blant de undersøkelsene en hadde kjennskap til var det heller ingen som umiddelbart kunne danne mønster for Byråets undersøkelse. Byrået har derfor vært nødt til å komme fram til sitt eget undersøkelsesopplegg. I dette arbeidet ble en stilt overfor en rekke valg. Noen av disse valgene ble foretatt i forbindelse med opplegget av en prøveundersøkelse i mindre målestokk, mens andre ble foretatt på grunnlag av resultatene fra den prøveundersøkelsen.

Nedenfor skal vi prøve å karakterisere undersøkelsen ved å gå gjennom noen av de viktigste avgjørelsene som ble truffet i løpet av planleggingsperioden.

II. UTVALG

Som populasjon for undersøkelsen er valgt personer i alderen 15-74 år. Av praktiske årsaker er personer bosatt i felleshusholdninger (sykehus, militærleire o.l.) holdt utenfor populasjonen. Den nedre grensen på 15 år er vanlig for Intervjukontorets undersøkelser, og skyldes at en ikke vil hente opplysninger direkte fra mindreårige personer. Årsaken til at de eldste ble holdt utenfor populasjonen, var en antakelse av at svært mange av dem ville ha vansker med å delta i undersøkelsen p.g.a. sykdom o.l. Også denne avgjørelsen er i tråd med Byråets praksis.

For å få en tilfredsstillende representasjon av forskjellige undergrupper i befolkningen, var det nødvendig med et såpass stort utvalg som 5 000 personer. Utvalget er trukket i to trinn i henhold til Intervjukontorets vanlige utvalgsplan [4].

III. DATAINNSAMLINGSMETIDE: BRUK AV SKJEMA OG DAGBOK

Som i de fleste andre tidsnyttingsundersøkelser er hovedvekten lagt på innsamling av tidsopplysninger gjennom dagbøker, der intervjupersonene selv noterer hva de har gjort i løpet av dagbokperioden. Oppbyggingen av dagboken vil bli nærmere beskrevet senere.

I tillegg til dagbokføringen ble det også samlet inn opplysninger gjennom et intervju foretatt i tilknytning til utdelingen av dagboken. Opplysninger som samles inn gjennom intervjuet faller i tre hovedkategorier:

1. Rene bakgrunnskjennetegn og opplysninger som først og fremst har verdi som bakgrunnskjennetegn. (Eks.: yrke, fullført utdanning, ekteskapelig status.)

2. Opplysninger om sjeldne aktiviteter. Slike aktiviteter vil bli mer tilfredsstillende registrert gjennom intervju enn gjennom dagboken. (Eks.: teaterbesøk, organisasjonsdeltaking, opplæring i yrket.)

3. Opplysninger som kan supplere den informasjonen om bestemte aktiviteter som samles inn gjennom dagboken. (Eks.: avstand til arbeidsstedet, tittel på boken IO leser på intervjutidspunktet, forekomsten av hjelpemidler i husholdningen.)

IV. UNDERSØKELSESPERIODE

Befolkningens tidsnytting vil til en viss grad variere over året. Fritiden vil således være mer uteorientert om våren og sommeren, mer inneorientert om vinteren og senhøsten. Arbeidstidens lengde kan også til en viss grad variere etter årstid. Variasjonene vil trolig være sterkest for yrker i primærnæringene, men også på arbeidsplasser som f.eks. statsadministrasjonen varierer arbeidstiden etter årstid. I tillegg kommer at friperioder som juleferie, påskeferie og sommerferie skaper store endringer i folks aktivitetsmønster. Dette betyr at en alltid må være forsiktig med å trekke generelle slutninger på grunnlag av undersøkelser med kortere undersøkelsesperiode enn ett år. For å få et mest mulig dekkende uttrykk for folks samlede tidsdisponering, og for å få muligheter til å studere årstidsvariasjoner i tidsbruken, har en for Byråets undersøkelse valgt året som undersøkelsesperiode. Av praktiske grunner valgte en å avvike fra kalenderåret; undersøkelsesperioden ble i stedet lagt fra 1. september 1971 til 31. august 1972.

Ved opplegget av undersøkelsen tok en sikte på at hver person skulle føre dagbok i to eller tre dager. Det vil senere bli gjort rede for hvorfor en valgte denne dagbokperioden.

Ved fordelingen av intervjupersoner over undersøkelsesperioden var det nødvendig å komme fram til et opplegg som var enkelt å administrere og som kunne tillate analyser av uke- og årstidsvariasjoner i tidsdisponeringen. For å oppnå dette ble året delt inn i 156 tidsperioder, og det ble trukket et delutvalg av personer for hver av disse periodene. En periode dekket enten ukedagene tirsdag og onsdag, torsdag og fredag, eller lørdag, søndag og mandag. Tre sammenhengende delperioder vil altså utgjøre en uke. For å gjøre det hele mer oversiktlig ble det utarbeidet tre dagboktyper, en for bruk tirsdag og onsdag, en for bruk torsdag og fredag, og en for bruk lørdag - mandag. Ulempen ved dette stramme opplegget er at midlertidige fravær fører til økt frafall i undersøkelsen. For å rette litt på dette tillot en at dagbokperiodene kunne forskyves en dag i hver retning.

V. NÆRMERE OM OPPLEGGET AV DAGBOKEN

a. Valg av minste dataenhet

Registrering av opplysninger om den enkeltes tidsbruk kan enten skje ved at det registreres hva personen har gjort i på forhånd bestemte tidsperioder, eller ved at den enkelte aktivitet brukes som minste dataenhet.

Valget på dette punktet har store metodiske konsekvenser, men setter samtidig visse grenser for hvilke problemstillinger som kan tas opp.

Det som taler for å velge tidsperiode som minste dataenhet er først og fremst hensynet til databehandlingen. Et opplegg med faste perioder og dermed faste recordlengder vil være betydelig enklere å håndtere enn et opplegg med aktivitet som minste dataenhet og variabel recordlengde. Siden tidsnyttingsundersøkelser gjerne er preget av store datamengder og krever en komplisert databehandling, vil det være naturlig å legge stor vekt på et slikt teknisk hensyn.

Det viktigste argument mot et opplegg med faste tidspunkter vil også være av teknisk art, nemlig at det gir større muligheter for at en del kortvarige aktiviteter ikke blir registrert. I denne forbindelse er det viktig å være klar over at heller ikke ved valg av aktivitet som dataenhet vil alle aktiviteter bli oppgitt. Her kan f.eks. nevnes at ved en undersøkelse i Ammerud fikk Dagfinn Ås oppgitt et gjennomsnitt på 30 aktiviteter pr. dag. Hvis vi holder tiden brukt til søvn utenfor, vil dette si et gjennomsnitt på om lag en halvtime pr. aktivitet. En oppdeling av aktivitetene i hovedkategorier viser på tilsvarende måte at ingen kategori hadde en gjennomsnittlig varighet pr. aktivitet på under 25 minutter. Det ser altså ut til at et opplegg med registrering av aktiviteter for hvert kvarter vil gi et vel så finregistrerende måleinstrument som et opplegg med registrering av hver aktivitet.

Selv om de to metodene stort sett vil gi de samme resultater og dekke de samme aktiviteter, kan det argumenteres for at et opplegg med faste tidspunkter vil føre til at kortvarige, men viktige, aktiviteter ikke registreres. Med viktige aktiviteter tenker en i denne forbindelse først og fremst på aktiviteter som inngår som et nødvendig element i en aktivitetssekvens. For å ta et eksempel: En registrering av aktiviteter i forbindelse med et besøk hos naboen blir ikke fullstendig dersom de fem minuttene som blir brukt til gange hver vei ikke kommer med. Ut fra et ønske om å forstå

sekvenser av individuell atferd vil denne gangtiden være viktigere enn f.eks. fem minutter som blir brukt til å bla gjennom et gammelt ukeblad.

Med den hovedmålsetting som er nevnt tidligere, nemlig å kartlegge hvorledes befolkningen fordeler sin tid på ulike aktiviteter, vil imidlertid aktivitetene være likeverdige. Ut fra det som tidligere er nevnt om de tekniske aspekter ved de to opplegg, følger dermed at det bare er aktuelt å bruke aktivitet som enhet dersom ønsket om å foreta en uttømmende analyse av atferdssekvenser utgjør en viktig del av målsettingen for undersøkelsen. Dette var ikke tilfelle for Byråets undersøkelse, slik at et opplegg med faste tidsperioder ble valgt.

Lengden av periodene ble satt til ½ time i tiden 00.00 - 02.00, 1 time i tiden 02.00 - 05.00, ½ time i tiden 05.00 - 06.00, og 1/4 time fra kl. 06.00 til kl. 24.00.

b. Opplysninger pr. dataenhet

Datamassen ved undersøkelsen vil være tilnærmet proporsjonal med tallet på opplysninger pr. dataenhet. I og med at undersøkelsen nødvendigvis vil bli nokså omfangsrik, må det være en forutsetning for opplegget at bare de opplysninger som er viktige og som det er praktisk mulig å utnytte ved en maskinell behandling blir samlet inn.

Ved planleggingen av undersøkelsen ble disse typene opplysninger vurdert:

1. Primæraktivitet
2. Sekundæraktivitet
3. Lokalisering
4. Geografisk sted
5. Tilstedeværende

1. Primæraktivitet

Registreringen av primæraktivitet er naturlig nok det sentrale ved en generell tidsnyttingsundersøkelse. Det eneste problemet med dette kjennetegnet er at det av og til kan være vanskelig å avgjøre hva som er primær- og hva som er sekundæraktivitet.

2. Sekundæraktivitet

Med sekundæraktivitet menes en aktivitet som utføres parallelt med primæraktiviteten. Resultater fra flere tidligere undersøkelsen, som den internasjonale tidsbudsjettundersøkelsen [2] eller undersøkelser av radio- og fjernsynsbruk [1], viser at en vil få et skjevt bilde av folks tidsnytting ved bare å se på primæraktiviteten.

3. Lokalisering

Med lokalisering menes her en generell opplysning om type sted, f.eks. i hjemmet, på arbeidsplassen, på offentlig transportmiddel osv. I svært mange tilfelle vil opplysninger om primæraktiviteten også bestemme personens lokalisering. I andre tilfelle vil den handlingssekvens som en aktivitet inngår i gi entydige opplysninger om hvor aktiviteten er lokalisert. Ut fra dette har en i Byråets tidsnyttingsundersøkelse valgt et opplegg der dagbokførerne ikke selv gir eksplisitte opplysninger om lokalisering. Disse opplysningene blir i stedet ført inn gjennom kodingen, ved at koderne bruker beskrivelsen av aktiviteten og den sekvens den inngår i for å bestemme hvor personer har oppholdt seg. Ved konstruksjonen av klassifikasjonssystem for lokalisering har en lagt spesiell vekt på å skille mellom forskjellige typer transportmidler.

4. Geografisk sted

Med geografisk sted menes en mer konkret plassering av aktiviteten, f.eks. foretatt ved avmerking på et kart. I og med at Byråets undersøkelse gjelder et landsomfattende utvalg trukket i to trinn, vil eventuelle opplysninger om geografisk sted bare kunne brukes til å konstruere generelle mål som f.eks. avstand fra hjemmet. Ut fra dette ble det nokså tidlig klart at merarbeidet med å trekke inn kjennetegnet geografisk sted ikke ville stå i noe rimelig forhold til den nytte en kunne vente å få av denne typen opplysninger. For tidsnyttingsundersøkelser blant personer bosatt i mindre geografiske områder kan det derimot ha mening å trekke kjennetegnet grografisk sted inn i en tidsnyttingsundersøkelse.

5. Tilstedeværende

Opplysninger om hvem personen er sammen med til de forskjellige tidspunkter kan brukes til å si noe både om personer og aktiviteter. Brukt til å karakterisere personer kan opplysninger om sosial kontakt brukes som en velferdsindikator. For en rekke aktiviteter kan det videre være nyttig å vite i hvilken grad de utføres i en sosial sammenheng eller i ensomhet osv. Ut fra slike hensyn tok en i første omgang sikte på å skaffe opplysninger om tilstedeværende gjennom dagboken. Ved prøveundersøkelsen viste det seg imidlertid at rubrikken for tilstedeværende var så dårlig utfylt at det ikke ville være forsvarlig å bruke disse opplysningene. En hadde da valget mellom å legge opp til en rutine med omfattende intervjuerkontroll og påfølgende oppretting av dagbøkene, og å sløyfe rubrikken for tilstede-

værende. For ikke å belaste forholdet mellom intervjuer og intervjuperson med en slik kontrollrutine, og for å holde omfanget av dagboken på et rimelig nivå, valgte en den siste løsningen.

c. Lengden av undersøkelsesperioden

I tidligere tidsnyttingsundersøkelser har en operert med forskjellige dagbokperioder. Mest vanlig er perioder på ett døgn; en slik periode er blant annet valgt i den internasjonale tidsbudsjettundersøkelsen [3]. Dersom en kan regne med at deltakerne i undersøkelsen er spesielt motivert for å føre dagbok, kan undersøkelsesperioden være så lang som en uke. Dette gjelder bl.a. den svenske undersøkelsen av læreres arbeidsforhold [2].

Valget av registreringsperiode vil være helt avgjørende for kostnadene og kvaliteten av undersøkelsen. I og med at intervjuerkostnadene er tilnærmet uavhengig av lengden av registreringsperioden, vil kostnadene ved registrering av tidsbruk pr. person pr. dag tilnærmet være omvendt proporsjonal med lengden av registreringsperioden.

Innsparingene i samlede kostnader vil ikke bli like omfattende; på grunn av regelmessigheten i den enkeltes tidsbruk vil et opplegg med lange perioder kreve et større antall registrerte persondager. For å ta et eksempel: Et opplegg med 715 personer som skal føre dagbok i en uke vil ikke gi like gode resultater som et opplegg med 5 000 personer som fører dagbok i en dag. Det er likevel ikke tvil om at ut fra et kostnadssynspunkt vil det være gunstig med så lange perioder som mulig.

Ulempene ved lange registreringsperioder vil være den reduserte kvaliteten av materialet. Vi må regne med at med lange perioder vil folk ha en tendens til å føre dagboken slurvet eller rett og slett glemme å føre den. Mulighetene for at eventuelle feil kan rettes opp når intervjueren henter dagboken vil også være mindre ved en lang dagbokperiode. Det er videre naturlig å regne med en viss sammenheng mellom lengden av dagbokperioden og frafallet, i og med at den arbeidsbyrden respondenten må si seg villig til å utføre avhenger av dagbokperioden.

I prøveundersøkelsen ble tre forskjellige undersøkelsesperioder testet: 1 dag, 3 dager og 7 dager. I tillegg ble det undersøkt om ekstra motivasjon av respondentene i form av betaling for dagbokføringen ville påvirke resultatene. Som testvariable ble brukt både frafallet p.g.a. nekting, tallet på personer som avbrøt dagbokføringen, antall tomme rubrikker i dagboken, antall rubrikker utfylt i kolonnen for sekundær-

aktiviteter, og rapportene fra intervjuerne. Hovedinntrykket av disse testene var at en periode på syv dager var for lang, også når intervjupersonene ble betalt for dagbokføringen. Derimot var det ingen vesentlige forskjeller mellom dagbokperioder på 1 og 3 dager. For en periode på tre dager så det ut til at betalingen hadde en viss virkning på kvaliteten av dagbøkene. Forbedringen var imidlertid ikke så stor at en fant det nødvendig å innføre betaling for det opplegget med to- og tre-dagers perioder som ble valgt for undersøkelsen.

d. Oppsummering

Ønsket om å komme fram til et noenlunde enkelt undersøkelsesopplegg har sammen med erfaringene fra den første prøveundersøkelsen ført til at en i undersøkelsen bruker dagbøker som dekker to eller tre undersøkelsesdager. Dagbøkene er bygd opp omkring faste tidsperioder, for det meste av ett kvarters varighet. For hver periode skal respondenten bare føre opp hovedaktivitet (primæraktivitet) og en eventuell biaktivitet. Ved kodingen av dagbøkene blir det videre innført en egen kode for lokalisering/transportmåte.

VI. KODING AV AKTIVITETER

Ved utarbeidingen av kodeliste for grupperingen av aktiviteter har en tatt utgangspunkt i den kodelisten som er utarbeidet for bruk i den internasjonale tidsbudsjettundersøkelsen [3]. En har imidlertid funnet det hensiktsmessig å foreta visse mindre endringer. Viktigst er det kanskje at en har gått bort fra prinsippet om at 1. siffer i den to-sifrede koden skulle brukes til å beskrive aktivitetsgruppe. En har også redusert tallet på kategorier som beskriver reiser fra 9 til 4. Dette har gjort det mulig å innføre mer detaljerte kategorier for husarbeid, vedlikeholdsarbeid og fritidsaktiviteter.

VII. REFERANSER

[1] P.-H. Kühl, I. Koch-Nielsen og K. Westergaard: Fritidsvaner i Danmark med særligt hensyn til radio og fjernsyn. København 1966.

[2] Sverige. Lärarnas arbete. Statens offentliga utredningar 1971: 53-55. Stockholm 1971.

[3] A. Szalai et al.: The Multinational Comparative Time Budget Research Project. American Behavioral Scientist. Vol. 10,No. 4, 1966. (Med Appendix.)

[4] S. Tamsfoss. Om bruk av stikkprøver ved Kontoret for intervjuundersøkelser, Statistisk Sentralbyrå. Artikler nr. 37, Statistisk Sentralbyrå, 1970.

JT/eh, 18/6-73

HVORDAN OPPLEGGET AV EN INTERVJUUNDERSØKELSE KAN PÅVIRKE RESULTATENE- EN SAMMENLIKNING AV FERIEUNDERSØKELSENE I 1968 OG 1970.

av Jon Teigland

Statistisk Sentralbyrå gjennomførte våren 1968 en intervjuundersøkelse for å registrere nordmenns ferievaner. Vel to år seinere, høsten 1970, ble en ny og liknende undersøkelse foretatt for å se på eventuelle endringer i feriemønsteret.

Da resultatene fra '70-undersøkelsen ble publisert, var det imidlertid ikke lagt noen vekt på å presentere resultatene i en form som var sammenliknbar med undersøkelsen i 1968. Forklaringen er at ferievanene syntes å ha forandret seg radikalt i løpet av disse to årene - hvis en skulle tro undersøkelsene. F.eks. økte antallet av yrkesaktive som hadde fridager (dvs. dager fri etter ferieloven) med omtrent 200 000 personer fra 1968 til 1970 - en økning på 20 prosent. Samtidig økte antallet av nordmenn som reiste på ferietur med godt over 200 000 personer - også dette en økning på 20 prosent i løpet av disse to årene. Denne siste økningen virker spesielt høy når en sammenlikner med resultatene fra de årlige ferieundersøkelsene i Storbritannia. Den tilsvarende prosentvise endring tok der over 15 år (British Tourist Authority, 1971: "The British on Holiday").

Forklaringene på de urimelige sterke "endringene" ligger nok i noen tilsynelatende små, men likevel vesentlige endringer i opplegget av undersøkelsene:

1. I 1970 var utvalgsenheten en person og personene redegjorde selv for feriene sine.

I 1968 derimot var husholdningen utvalgsenheten, og en person - ofte husmoren - gav opplysninger om alle feriene for alle medlemmene av husholdningen, hvis da disse andre ikke var til stede selv. Den eller de som svarte kan ha glemt ferieturer og feriedager som de andre husholdningsmedlemmene hadde. Både antall ferieturer og friperioder ble på den måten for lave i 1968.

2. I 1968 ble intervjuobjektet (IO) forelagt en kalender over årets dager og spurt generelt om husholdningens medlemmer hadde hatt fridager og ferieturer det siste året.

Også i 1970 fikk intervjuobjektet utlevert en kalender over årets dager, men i dette tilfellet var kalenderen delt inn i årstider eller ferieperioder. Det ble deretter spurt systematisk; hadde de fri sommeren 1970, var de på ferietur da? Hadde de fri våren 1970, var de på ferietur da? Hadde de fri i påsken, var de på ferietur da osv. til hele året forut for intervjuet var kartlagt.

Den systematiske gjennomgåelsen kan ha ført til at ferier som ellers ville vært glemt - og som ble glemt i 1968 - er blitt registrert i '70-undersøkelsen.

3. En del av forklaringen kan dessuten være at IO'ene hadde problemer med å huske og/eller å gi sikre opplysninger om ferier de hadde relativt kort tid tilbake i tiden. I det tilfelle er det vesentlig at 1970-undersøkelsen ble foretatt om høsten like etter at den viktigste ferietiden om sommeren var over, mens '68-undersøkelsen ble foretatt i mai rett etter at ferieåret var over, og nesten ett helt år etter de siste sommerferiene. Hvis folk glemmer hendelser litt tilbake i tiden eller tidsforskyver dem, dvs. glemmer når noe egentlig skjedde, da hadde de som ble intervjuet i '68 større mulighet for å huske feil enn de som ble spurt i 1970.

Nå er sannsynligvis slike erindrings-eller tidsforskyvningstendenser viktige feilkilder ved intervjuundersøkelser. Dalenius nevner f.eks. at den slags "telescoping"-effekter var den viktigste ikke-tilfeldige feilen som ble funnet i en metode-undersøkelse foretatt av U.S. Bureau of the Census (Dalenius, 1971, "Information for survey design"). Andre amerikanske undersøkelser viser at denne feilkilden også kan ha meget store virkninger ved ferieundersøkelser (Lansing, Blood, 1964: "The changing travel marked").

Enkelte resultater fra Ferieundersøkelsen 1970 ser ut til å bekrefte dette. I 1970 ble det nemlig spurt systematisk om ferieturer de to siste årene, dvs. det ble spurt om IO hadde vært på ferietur sommeren 1970 og sommeren 1969, påsken 1970 og påsken 1969, julen 1969 og julen 1968. Dette for å se på endringer i feriemønsteret fra et år til et annet ved hjelp av en og samme undersøkelsen. Denne registreringen viste at det tilsynelatende var en generell nedgang i feriehyppigheten fra 1969 til 1970 - på tross av at sammenlikninger av '68- og '70-undersøkelsen viste en drastisk økning.

Det var riktig nok "bare" registrert en 5 prosents nedgang i feriehyppigheten fra sommeren 1969 til sommeren 1970 - noe som tilsvarer at nesten 100 000 færre nordmenn reiste på sommerferie i 1970 enn i 1969. Men det var 13 prosent færre som hadde vært på ferietur påsken 1970 enn

i 1969, og det var 40 (!) prosent færre som sa de hadde vært på juleferie i 1969 enn i 1968. Dette må være feil og resultatene ble da heller ikke tatt med i feriepublikasjonen fra 1970.

Det vesentligste her er imidlertid at nedgangen økte jo lengre tilbake i tiden en forsøkte å spørre om. Det samsvarer med at jo lengre tid tilbake en forsøker å huske jo lettere er det å glemme når noe ikke alt for viktig har foregått. IO vet f.eks. at han/hun har vært på juleferie tidligere, men er ikke bevisst når ferien var og forskyver juleferien 1966 eller 1967 opp til 1968, og så svarer de; jo vi var på juleferie i fjor.

Nå er det sannsynlig at denne tidsforskyvningen eller glemselseffekten ikke bare forekommer ved ferier som intervjuobjektene hadde mer enn et år forut for intervjuingen. Feilkilden kan være og er nok til stede også ved kortere kartleggingsperioder - om enn i mindre grad. I praksis betyr det f.eks. at de publiserte juleferietallene fra 1970-undersøkelsen er noe usikre fordi intervjuene her foregikk om høsten omtrent 9 måneder etter jul.

Hvor vesentlig denne feilen er kan en forøvrig teste ved å foreta to ferieundersøkelser samme år, en liten undersøkelse (f.eks. tilknyttet AKU) på vårparten for å se nærmere på jul-, vinter- og påskeferiene, og så om høsten ha en større undersøkelse som kartlegger feriene hele det foregående år analogt 1970-undersøkelsen. Siden begge undersøkelsene dekker vinterhalvåret, skulle en derved få en sjekk av hvordan glemselseffekten påvirker resultatene i hovedundersøkelsen om høsten. Samtidig er hovedundersøkelsen sammenliknbar med '70-undersøkelsen, og den antyder endringene fra tidligere år. Dessuten vil den mindre undersøkelsen på vårparten gi gode muligheter til en grundigere kartlegging av feriene i vinterhalvåret.

DW/eh, 16/10-73

Convergence of a Class of Matrix Processes

by David Walker

In the disaggregated economic model MODIS IV a question has arisen concerning the convergence of the following matrix process:

$$x_n = B^{-1}(b - C\hat{m}Ax_{n-1}), \tag{1}$$

where vectors x_n, x_{n-1} and $b \in R^p$, m is a positive real vector which may be of dimension other than p, and A, B and C are real matrices. The indices n and n-1 represent the $(n)^{th}$ and $(n-1)^{th}$ time periods, respectively, and $\hat{m}$ is a diagonal matrix with the elements of m along the diagonal. Therelation (1) may be rewritten as

$$x_n = - Wx_{n-1} + c, \tag{2}$$

where $W = B^{-1}C\hat{m}A$ and $c = B^{-1}b$.

Matrices A, B, and C, and the vector m remain fixed during the process. However, $\hat{m}$ will be different in parallel calculations carried out at the same time. The process in (2) will converge if and only if the eigenvalues of -W are all less than one in modulus. This result is well known [1].

The present memorandum is concerned with drawing conclusions from the literature on bounds on eigenvalues, to enable testing of W in order to see whether or not the eigenvalue condition above holds, and thus whether or not the process (1) will converge.

Our results are equally applicable to both (1) and other processes which can be expressed in the form (2). We consider only real general matrices W.

The strongest results in the literature for our purpose appear to be the conditions derived by Parker (1937) and Farnell (1944). Other related results are due to Frobenius (1908 - for non-negative matrices only), Schur (1909), Gerŝgorin (1931), Browne (1930), and Brauer (1946). References to these are given in a 1964 survey by Marcus and Minc [2].

We begin by defining the row and column sums of the absolute values of the elements of W:

$$R_i = \sum_j | W_{ij} |, \; i = 1, \ldots, p;$$

and (3)

$$T_j = \sum | W_{ij} |, \; j = 1, \ldots, p.$$

Then (Parker, 1937)

$$| \lambda_t | < \max_i \{ \tfrac{1}{2} (R_i + T_i) \} \qquad (4)$$

for $t = 1, \ldots, p$, where $| \lambda_t |$ is the modulus of the t^{th} eigenvalue of W.

Again, define

$$R = \max_i R_i$$

and (5)

$$T = \max_j T_j$$

Then (Farnell, 1944)

$$| \lambda_t | < (RT)^{1/2} \text{ for } t = 1, \ldots, p. \qquad (6)$$

We define a constant K equal to the minimum of the right hand sides of (4) and (6).

Three cases may be distinguished:

(i) If $K \geq 1$, then we cannot say for certain whether or not process (1) will converge.

(ii) If $K < 1$, we know that the process will converge.

(iii) If $K << 1$, we can say that the process will converge rapidly.

In Case (i), I recommend an experiment to see whether process (1) converges or not.

The value for K will also indicate approximately the amount of variation which can be tolerated in the original matrices A, B and C, if these should be altered in later runs of the model, and in the vector m. This vector is to be given different values in parallel calculations, as stated above. If K is near 1, alterations which increase the R_i and T_i may prevent convergence.

The vector m is the most relevant here. If any element of m is increased, the possibility that K will be increased correspondingly must be considered. A change in an element of m may or may not alter K, depending on the way in which the change alters R, T, and the expression $\max_i (R_i + T_i)/2$. It may alter none of these. If the elements of m are systematically increased, the process will ultimately diverge.

This leads to a further possibility, namely that of calculating a value for K using a vector m in which every element is given the highest value which it is ever expected to be given in the model. It should be

pointed out in this respect that increasing all the elements of m may have a considerable effect on the value of K, whereas relatively larger increases in only a few elements of m may be tolerable.

Furthermore, it can be remarked that the model MODIS IV would gain in stature as a representation of the actual economy in Norway if process (1) could be shown to converge with the data used, because that process can be interpreted as a dynamic adjustment to a disturbance in the activity levels of the economy, caused by a change in the vector m, which is an exogenous vector representing proportional changes in the market shares of imported goods. We assume here something which is not obvious, but which is not likely to be questioned, and that is the stability of the actual economy with respect to such disturbances.

Bibliography

[1] K. Lancaster, Mathematical Economics, MacMillan, N.Y., 1968.

[2] M. Marcus & H. Minc, A Survey of Matrix Theory and Matrix Inequalities, Allyn & Bacon, Boston, 1964.