Arbeidsnotater

STATISTISK SENTRALBYRÅ

Dronningensgt. 16, Oslo-Dep., Oslo 1. Tlf. 41 38 20

IO 74/22

22. februar 1974

BEREGNING AV BEFOLKNINGSRATER

av Jan M. Hoem[x)]

INNHOLD

x) Kommentarer fra Helge Brunborg, Gerd Skoe Lettenstrøm og Halvard Skiri har vært til god nytte under utarbeidelsen av dette notatet. Erling Berge har organisert de numeriske eksemplene og lest manuskriptkorrektur.

Den utforming av Lexis skjema som finnes i figur 1, skyldes Statistiska Centralbyrån i Stockholm. Også andre elementer i fremstillingen her skyldes impulser fra vår svenske søster-organisasjon.

1. Innledning.

1A. I Norge ble det registrert 168.3 levendefødte barn pr. 1 000 kvinner i alderen 20-24 år i 1971. Samme år ble det registrert 12.8 døde under 1 år pr. 1 000 levendefødte. I 1969-70 var det 170.3 vigsler pr. 1 000 menn i alder 25-29 i middelfolkemengden. (Se tabellene 26, 29 og 22 i Statistisk årbok 1973.) Disse tallene er eksempler på befolkningsrater. Slike rater kan beregnes på en del forskjellige måter, og valg av beregningsmåte kan ha ganske stor betydning for utfallet. Det er derfor viktig å være oppmerksom på de mulighetene som finnes. Bare da kan en velge den metoden som er mest adekvat i det aktuelle tilfellet. Da blir en også klar over nødvendigheten av å bringe på det rene hvilken metode som har vært brukt når en skal anvende tall beregnet av andre. Det er også viktig at de beregninger som utføres, merkes tilstrekkelig tydelig slik at en kan spare seg selv og andre for detektivarbeid i ettertid.

1B. I dette notatet skal vi gi en oversikt over en del beregningsmetoder for befolkningsrater. Vi skal også gi retningslinjer for merking av beregninger. Retningslinjene er ment å være normative, slik at de er å oppfatte som en instruks som skal følges av medarbeidere ved Byråets 1. kontor og Sosiodemografisk forskningsgruppe. En håper at også andre kontorer i Byrået vil se retningslinjene som nyttige i sitt eget arbeid.

I beskrivelsen av beregningsmetodene brukes et system av matematiske symboler. Også dette er ment å være normgivende. Hittil har hver enkelt saksarbeider stort sett nyttet sitt eget symbolsystem. Erfaringene har vist at kommunikasjonen lettes ved standardisering også på dette området.

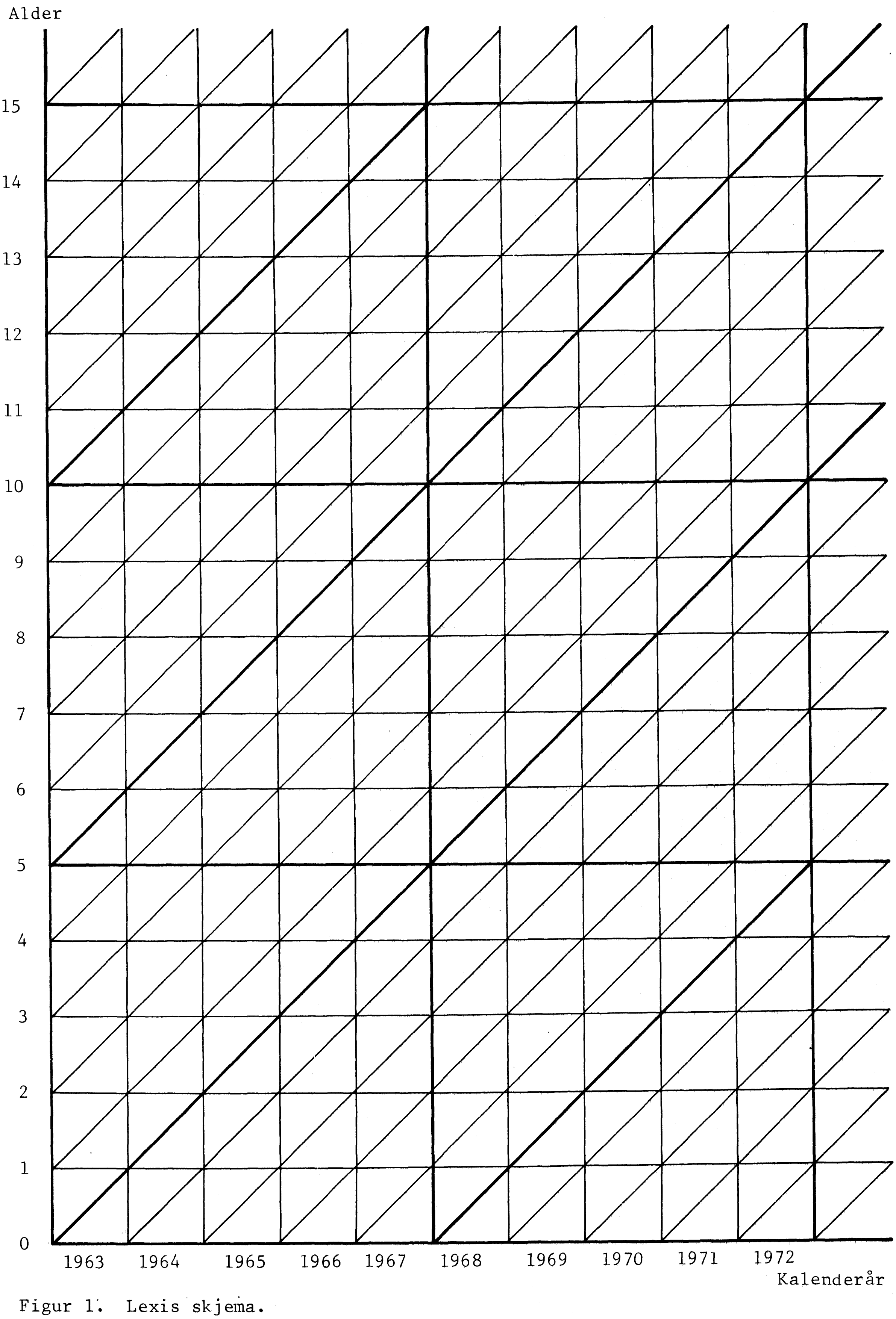
1C. I notatet gir vi enkelte numeriske eksempler og sporadiske kommentarer som antyder når en ratetype bør foretrekkes fremfor en annen. Det er neppe slik at én slags rate for alle formål bør foretrekkes fremfor alle andre, selv om forfatteren personlig har en forkjærlighet for rater beregnet etter alder ved utgangen av året og med en adekvat risikosum i nevneren. Det vil imidlertid føre for langt å gi noen prinsipiell diskusjon her, så meget mer som denne egentlig må baseres på resonnementer fra matematisk statistikk for å ha noen dybde. En vil imidlertid forsøke å gi mer utstrakte illustrasjoner i senere notater, og derigjennom belyse valg av ratetype med empirisk materiale.

2. Lexis skjema.

2A. Når man skal studere hvordan en befolkning utvikler seg over tiden, kan det ofte være vanskelig å beholde oversikten over alle faktorene som skal være med. Et nyttig virkemiddel i denne forbindelse er Lexis skjema, som gir en fremstilling i diagramform av de demografiske hendelsene i en befolkning, slik som fødsler, giftermål, dødsfall, osv. Lexis skjema er det viktigste instrument en har i beskrivelsen av de ulike metodene for beregning av befolkningsrater.

En fremstilling av grunnskjemaet er gitt i figur 1. Det har fått sitt navn etter den tyske befolkningsstatistikeren Wilhelm von Lexis, som beskrev det i en lærebok for ca. 100 år siden (Lexis, 1875).

2B. Grunnskjemaet består av et koordinatsystem der man avsetter kalendertid langs abscisse-aksen og alder langs ordinat-aksen. For hvert enkelt tidspunkt som avsettes, kan man trekke en dato-linje loddrett i koordinatsystemet. I figur 1 har vi trukket opp datolinjene for den 31/12 for hvert av de år som er tatt med.

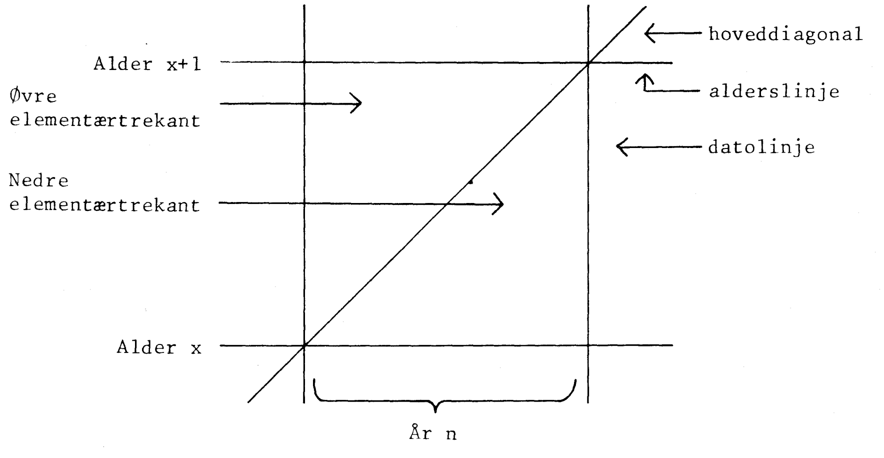


Figur 1. Lexis skjema.

Analogt kan man for hver enkelt alder trekke en alders-linje vannrett i diagrammet. Alderen regnes kontinuerlig, og man kan trekke alderslinjer både for hele aldre og for brudne aldre om man ønsker det. Vi har trukket alderslinjene for de hele aldrene i figur 1.

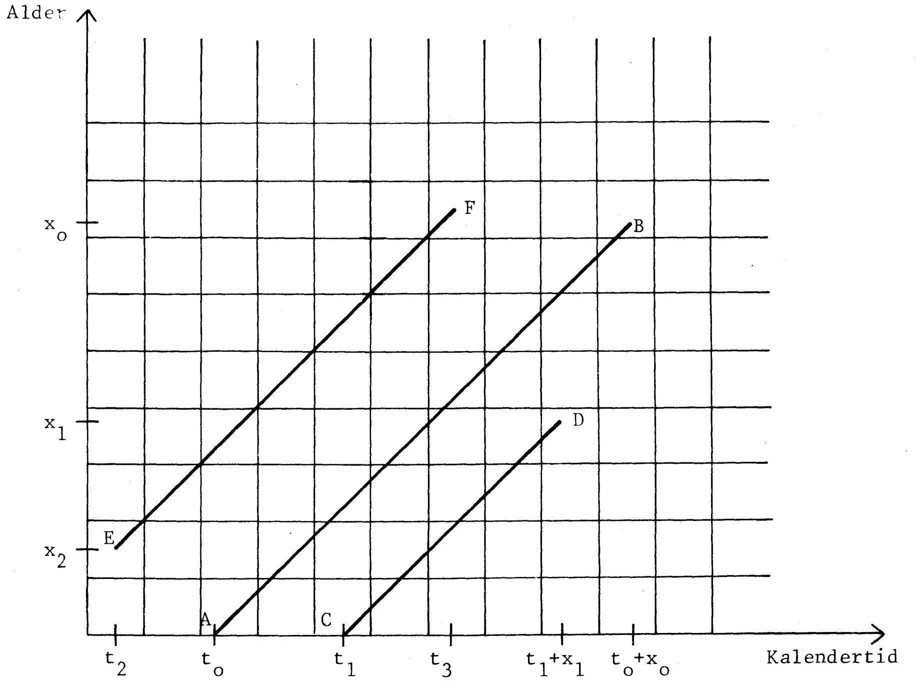
På dette vis får man et rutenett i koordinatsystemet. Skjæringspunktene mellom de datolinjene vi har tegnet, kaller vi for knutepunkter. Vi har også trukket opp rette linjer gjennom knutepunktene, i 45^{o} vinkel med dato-aksen. Disse rette linjene kaller vi for hoveddiagonaler.

Når vi har trukket alle disse linjene, er planet blitt delt opp i en rekke trekanter, som vi vil kalle elementærtrekanter. Hvert kvadrat består av to trekanter. Den øverste av disse har formen ◸. Slike trekanter vil vi kalle øvre elementærtrekanter. Den nederste trekanten i hvert kvadrat har formen ◿. Slike trekanter vil vi kalle nedre elementærtrekanter. Sml. figur 2.



Figur 2. Noen begreper i Lexis skjema

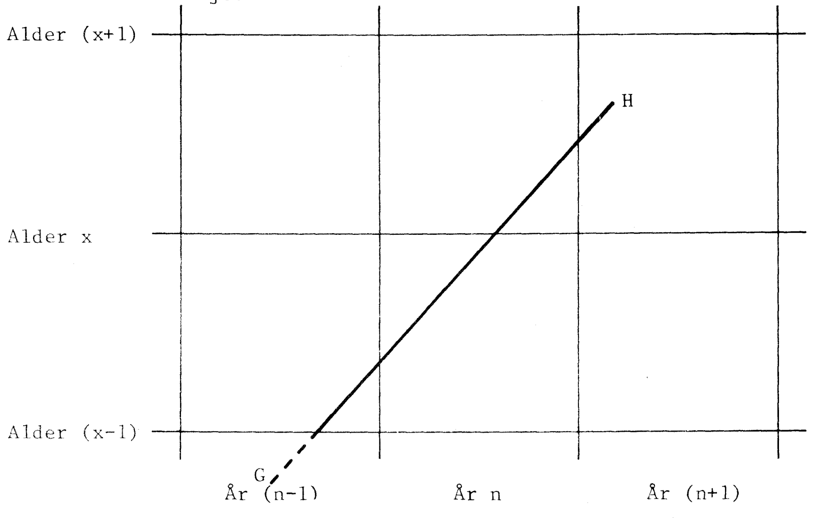
2C. La oss anta at en kvinne blir født på et tidspunkt t_o, og at hun er i befolkningen til hun blir x_o år, men at hun så dør på tidspunktet $t_o + x_o$. (x_o regnes kontinuerlig.) I Lexis skjema kan vi da trekke en rett linje fra punktet $(t_o, 0)$ til punktet (t_o+x_o, x_o). Denne rette linjen, som jo vil gå parallelt med hoveddiagonalene, kaller vi for hennes livslinje. I figur 3 er linjen AB livslinjen for en slik kvinne. Hvis hun får barn, kan tidspunktene for fødslene avmerkes på livslinjen hennes. Tilsvarende om hun gifter seg, osv.



Figur 3. Livslinjer i Lexis skjema.

En kvinne som blir født på tidspunkt t_1, og som blir i befolkningen inntil hun utvandrer i alder x_1 på tidspunkt t_1+x_1, vil ha en livslinje i skjemaet for denne befolkningen som ser akkurat tilsvarende ut. I figur 3 er CD en slik livslinje.

En kvinne som innvandrer på tidspunkt t_2, vil da være over 0 år gammel. La oss kalle alderen hennes for x_2. Hennes livslinje starter i punktet (t_2,x_2) i skjemaet, og den slutter i punktet (t_3,x_3) ved at hun dør eller utvandrer i alder x_3 på tidspunkt t_3. Her blir $x_3 = x_2 + (t_3-t_2)$. I figur 3 er EF en slik livslinje.



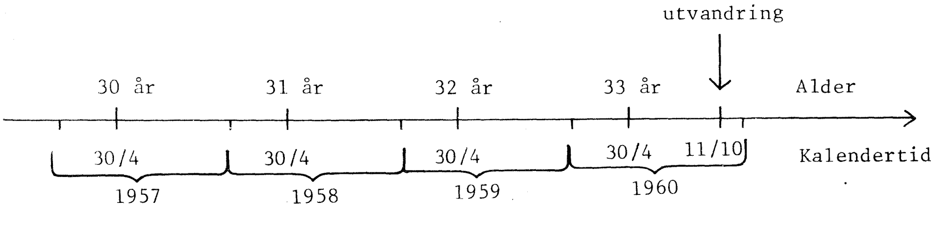
Figur 4. Hendelser langs en livslinje i Lexis skjema.

2D. La oss nå se litt nærmere på de hendelser kvinnen opplever når vi følger henne langs livslinjen hennes. I figur 4 har vi tegnet inn slutten på en slik livslinje GH, der vi skal anta at H angir et dødsfall. Vi ser at livslinjen skjærer datolinjene for 31/12 år n-1 og 31/12 år n. Dette betyr at hun opplever begge disse årsskiftene. Livslinjen skjærer derimot ikke datolinjen for 31/12 år (n+1), for kvinnen avgikk ved døden i første kvartal av året (n+1).

Likeledes skjærer livslinjen alderslinjene for aldrene (x-1) og x (der x er heltallig), så hun opplever sin (x-1)-te og sin x-te fødselsdag. Vi ser også straks at den x-te fødselsdagen ble feiret i året n. Kvinnen avgår så ved døden i en alder $x+\theta$ i året (n+1), der $0 < \theta < 1$ i dette eksempelet.

Det blir tilsvarende for en mann, naturligvis. Vi omtaler kvinner her først og fremst fordi det er knyttet flere interessante demografiske hendelser til dem enn til menn (fødsler).

2E. Vi kan også tenke oss livslinjen i Lexis skjema projisert ned på tidsaksen, slik at vi får en kombinert alders- og dato-linje som i eksempelet i figur 5.

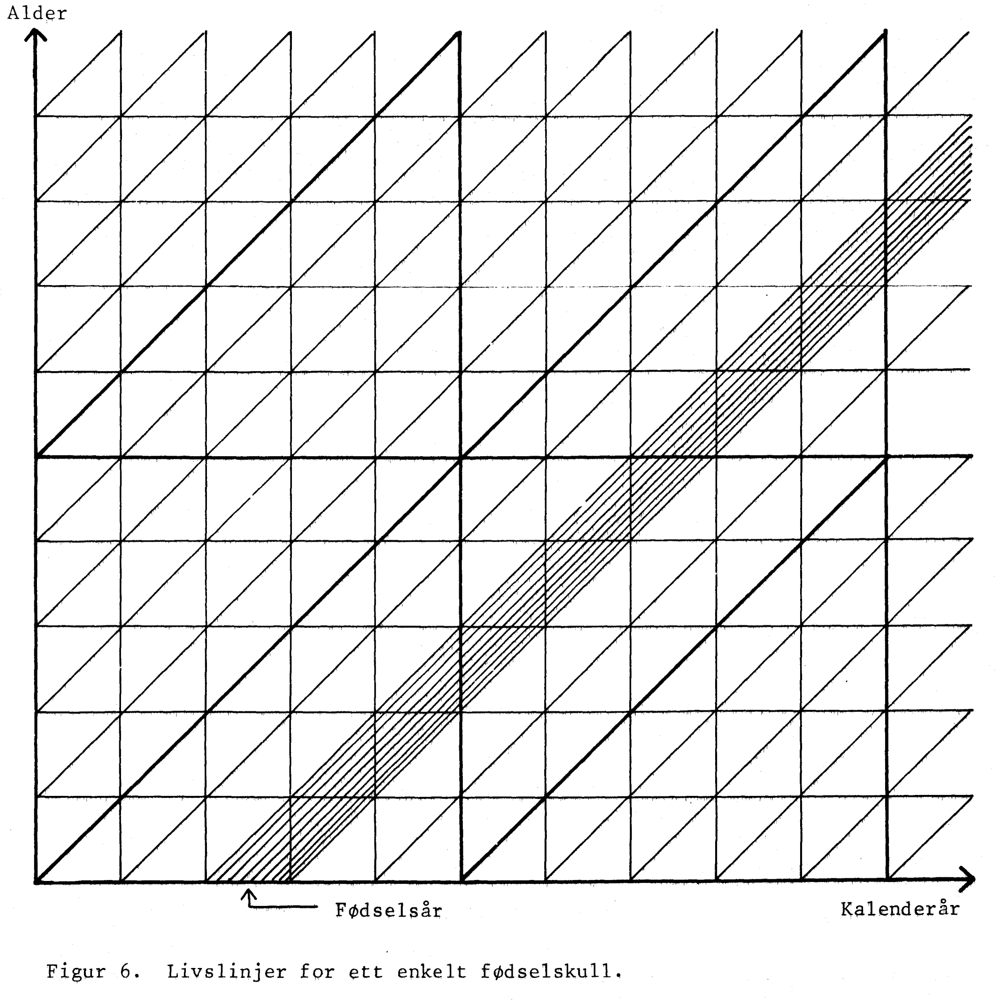


Figur 5. Alders- og dato-linje for en kvinne som fyller 30 år 30/4 1957, og som utvandrer fra befolkningen 11/10 1960.

Noen ganger ønsker vi å følge med i de demografiske hendelser en person opplever mellom en dato t_1 og en annen dato t_2. Andre ganger ønsker vi å holde vedkommende observasjon fra en alder x_1 og til en annen alder x_2. For enklere å kunne omtale slike observasjonsperioder, bruker vi betegnelsen kalenderåret om perioden fra 1/1 til 31/12 i et år. Analogt vil vi kalle perioden mellom en fødselsdag og den neste hos en person, for alders-året.

2F. Vi tenker oss nå at man tegner inn i skjemaet livslinjene for alle personene i befolkningen, eller f.eks. for alle kvinnene hvis det er opplysninger om dem man vil studere. Alle livslinjene for personer født i året n vil da ligge mellom de hoveddiagonalene som starter i datopunktene 1/1 og 31/12 år n. Sml. figur 6. Hvis man ønsker å studere de begivenheter som angår dette ene fødselskullet, kan man derfor konsentrere oppmerksomheten om livslinjene mellom de to hoveddiagonalene.

Analyser av de demografiske begivenhetene i de enkelte fødselskull kalles fødselskullsanalyse eller kohort-analyse.



Figur 6. Livslinjer for ett enkelt fødselskull.

3. Begrepet alder.

3A. Det er mange måter å angi en persons alder på. Dagligtalens språkbruk er ikke alltid presis nok for våre formål, og det er viktig å ha klart for seg de ulike mulighetene som foreligger.

Vi skal si at en person har eksakt alder x hvis vi måler alderen i år og deler av år, slik at x blir en kontinuerlig variabel. Når vi f.eks. oppgir at en slik x-åring dør innen han blir x+t år gammel, vil vi mene at han dør i en alder i intervallet <x,x+t], der også t regnes kontinuerlig.

På den annen side vil vi si at en person har fylt x år hvis x er et helt tall og han har opplevet sin x-te, men ennå ikke sin (x+1)-te fødselsdag. For å tilkjennegi at alderen regnes på denne måten, kan vi også si at x er personens alder i fylte år. Ved beregning av befolkningsrater bruker vi vanligvis alder i fylte år. Her er x naturligvis et helt tall.

Hvis vi oppgir at en person er x år gammel og omtaler ham som en x-åring, er det altså ikke alltid klart hva vi mener før vi har angitt om x måles kontinuerlig eller heltallig.

Når vi sier at den fødedyktige periode hos kvinnene er aldrene 15-49 år, mener vi at stort sett bare kvinner som har fylt 15,16,...,48 eller 49 år, får barn. Barna fødes da mens morens eksakte alder ligger i intervallet <15,50].

3B. Det finnes imidlertid en enda viktigere distinksjon enn denne når det gjelder angivelse av en persons alder. Hvilken alder som skal oppgis, avhenger nemlig naturligvis av når alderen måles. I sammenheng med demografiske hendelser (altså fødsler, giftermål, dødsfall, osv.) skal vi her sondre mellom

(i) alder ved utgangen av kalenderåret (eller alder pr. 31/12, eller ved neste årsskifte),

(ii) alder ved begynnelsen av kalenderåret (eller alder pr. 1/1, eller ved forrige årsskifte),

(iii) alder ved nærmeste årsskifte, og

(iv) alder ved hendelsestidspunktet (morens alder ved nedkomsten, brudens alder ved giftermålet, osv.).

3C. Alderen (i fylte år) ved utgangen av kalenderåret er det samme som alderen på fødselsdagen det året. Den er også lik differensen mellom kalenderåret og fødselsåret. Er kalenderåret 1972 og fødselsåret 1941, blir alderen i fylte år pr. 31/12 1972 lik 31. Dette er en aldersdefinisjon som passer f.eks. i tilknytning til befolkningsprognoser, der en jo regner seg frem fra årsskifte til årsskifte. Den kan også brukes ved kohortanalyser, og har da bl.a. den fordelen at perioderater og kohortrater faller sammen.

3D. Alderen i fylte år ved begynnelsen av ett kalenderår n er naturligvis det samme som alderen i fylte år ved utløpet av foregående kalenderår n-1 og én mindre enn alderen i fylte år ved utgangen av år n. For å unngå sammenblanding skal en i Statistisk Sentralbyrå vanligvis unngå å bruke alder definert ved begynnelsen av et kalenderår ved beregning av befolkningsrater.

3E. Alderen i fylte år ved nærmeste årsskifte kan være et aktuelt alternativ når en studerer atferd som er nær knyttet til skoleåret. (Se f.eks. Hoem, 1972.) Såvidt vites, er denne muligheten ikke brukt i Byrået.

3F. Alder i fylte år ved hendelsestidspunktet er den aldersdefinisjon som er mest brukt ved beregning av befolkningsrater for vanlige statistikkformål. De ratene som publiseres i Statistisk årbok i en tabell med overskriften "Fruktbarheten etter morens alder. Reproduksjonskvotienter.", er beregnet med utgangspunkt i denne definisjonen for årene siden 1961. (For tidligere år gjelder ratene for alderen ved utgangen av året. I Statistisk årbok 1964 kan en finne ratene for 1961 beregnet etter begge aldersdefinisjoner.)

4. Noen betegnelser.

4A. Det er hensiktsmessig å innføre en del symboler for størrelser en får å arbeide med.

La n være nummeret på et kalenderår. Eksempelvis kan vi la n = 1974. La x være heltallig. Da skal

$L_x(n)$ betegne antall x-årige personer (eventuelt antall x-årige kvinner, osv.) i befolkningen ved utgangen av år n, og

$M_x(n)$ betegne det antall som opplever sin x-te fødselsdag i år n.

Spesielt blir $M_0(n)$ antall levendefødte i perioden n. Vi vil også bruke betegnelsen F(n) om dette antall, slik at altså $F(n) = M_0(n)$.

Det er grenser for hvor gammelt et menneske kan bli. Vi vil betegne høyest mulig (heltallig) levealder med ω. Her kan altså $M_\omega(n)$ være positiv, mens vi må ha $M_{\omega+1}(n) = 0$ for alle n.

I norske dødelighetstabeller o.l. regner en vanligvis med ω = 105 år.

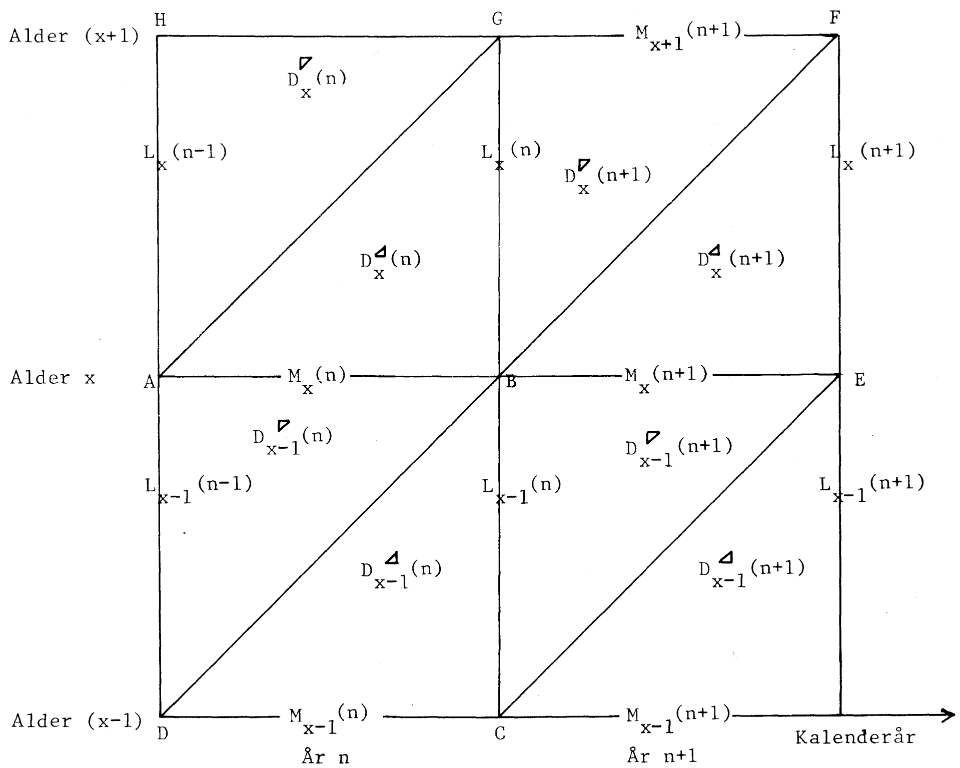
4B. Vi innfører også

$D_x^{\text{◿}}(n)$ som antall personer som fyller x år i året n og som dør i året <u>etter</u> fødselsdagen sin, og

$D_x^{\text{◸}}(n)$ som antall personer som dør i året n <u>før</u> sin fødselsdag (nummer x+1) i det året, og som var fylt x år ved dødsfallet.

Vi merker oss at vi kan ha både $D_0^{\text{◿}}(n) > 0$ og $D_0^{\text{◸}}(n) > 0$. Videre kan vi ha $D_\omega^{\text{◿}}(n) > 0$ og $D_\omega^{\text{◸}}(n) > 0$, men vi må ha $D_{\omega+1}^{\text{◸}}(n) = D_{\omega+1}^{\text{◿}}(n) = 0$.

I figur 7 viser vi hvordan disse størrelsene kan settes inn i Lexis skjema.



Figur 7. Utsnitt av Lexis skjema.

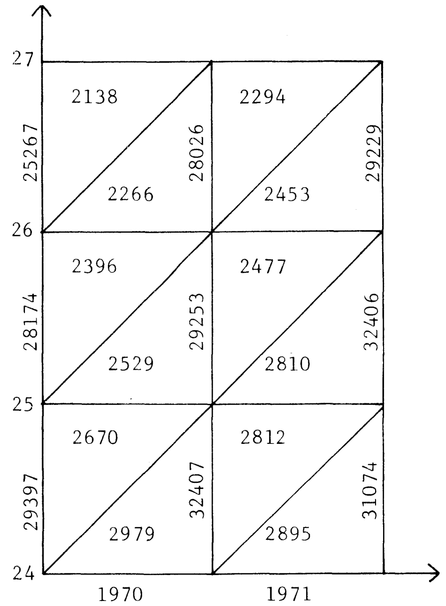
Som vi ser av figuren, blir $L_x(n)$ lik antall livslinjer som skjærer linjestykket BG i diagrammet. Likeledes blir $M_x(n)$ lik antall livslinjer som skjærer AB, osv.

$D_x^{\triangle}(n)$ blir antall dødsfall avmerket i elementærtrekanten ABG, og $D_x^{\triangledown}(n)$ blir antall dødsfall avmerket i elementærtrekanten AGH. Vi skjønner lett hvorfor symbolene for disse størrelsene har fått den utforming de har.

4C. Vi tenker oss at antall levendefødte barn også avmerkes på livslinjene. Antall avmerkinger for levendefødte barn i trekantene ABG og AGH betegner vi henholdsvis $F_x^{\triangle}(n)$ og $F_x^{\triangledown}(n)$. De tilsvarende antall jentebarn betegner vi $J_x^{\triangle}(n)$ og $J_x^{\triangledown}(n)$. For levendefødte guttebarn er betegnelsene $G_x^{\triangle}(n)$ og $G_x^{\triangledown}(n)$. Andre typer av barn (dødfødte, levendefødte pluss dødfødte, osv.) har vi såpass sjelden bruk for å beregne befolkningsrater for, at vi ikke skal innføre noen standardbetegnelser her, men overlater til den enkelte å improvisere. Der det passer, kan en jo også innføre flere topp- og fotskrifter på symbolene.

Tilsvarende antall vigsler (giftermål) betegner vi med $V_x^{\triangle}(n)$ og $V_x^{\triangledown}(n)$. Toppskrift kan benyttes hvis en skal sondre mellom bruder og brudgommer. (Vi kan ikke bruke bokstaven G for giftermål, fordi den alt er "opptatt" av antall levendefødte guttebarn.)

I figur 8 har vi oppgitt noen faktisk registrerte antall levendefødte barn (gutter pluss piker) og de tilsvarende antall kvinner ved årsskiftene.



Figur 8. Noen antall levendefødte barn i Norge 1970 og 1971 og tilsvarende antall kvinner ved utgangen av årene 1969 til 1971.

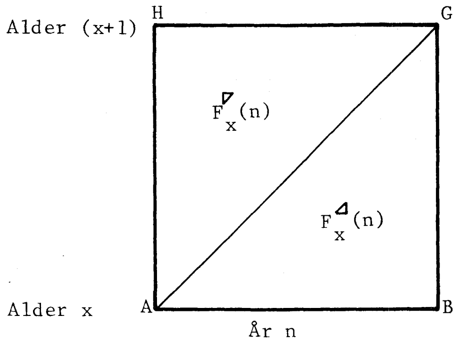
5. Censusmetoden, kalenderårsmetoden, og aldersårsmetoden.

5A. Censusmetoden. Vi vil nå sette

$$F_x^{\square}(n) = F_x^{\triangle}(n) + F_x^{\triangledown}(n),$$

og ser at $F_x^{\square}(n)$ blir antall levendefødte avmerket i firkanten ABGH i figur 9. (Betegnelsene på punktene A, B,... osv. er de samme som i figur 7.) Vi sier at $F_x^{\square}(n)$ er et antall beregnet ved censusmetoden, og skjønner at dette blir antall barn levendefødt i år n av kvinner som er fylt x år ved nedkomsten.

Av figur 8 ser vi at en i Norge f.eks. hadde $F^{\square}_{25}(1970) = 2529 + 2396 = 4925$.



Figur 9. Elementært censuskvadrat.

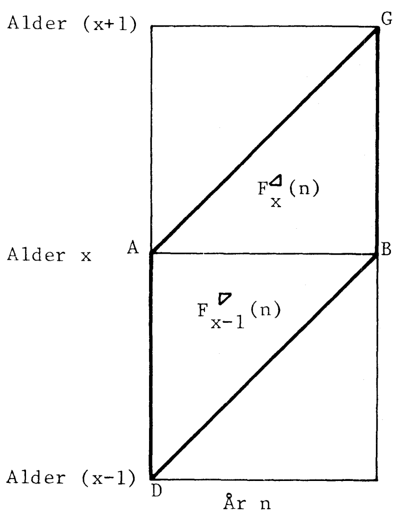
5B. <u>Kalenderårsmetoden</u>. Vi setter

$$F^{\text{▱}}_{x}(n) = F^{\triangledown}_{x-1}(n) + F^{\triangle}_{x}(n),$$

og ser at $F^{\text{▱}}_{x}(n)$ blir antall levendefødte avmerket i parallellogrammet DBGA i figur 10. Vi sier at $F^{\text{▱}}_{x}(n)$ er beregnet ved kalenderårsmetoden, og skjønner at dette blir antall barn levendefødt i år n av kvinner som er fylt x år ved utgangen av år n.

Av figur 8 ser vi at en i Norge f.eks. hadde $F^{\text{▱}}_{25}(1970) = 2670 + 2529 = 5199$.

Tilsvarende for dødsfall osv.



Figur 10. Elementært kalenderårsparallellogram.

5C. <u>Aldersårsmetoden</u>. Vi setter

$$F^{\text{▱}}_{x}(n) = F^{\triangle}_{x}(n) + F^{\triangledown}_{x}(n+1),$$

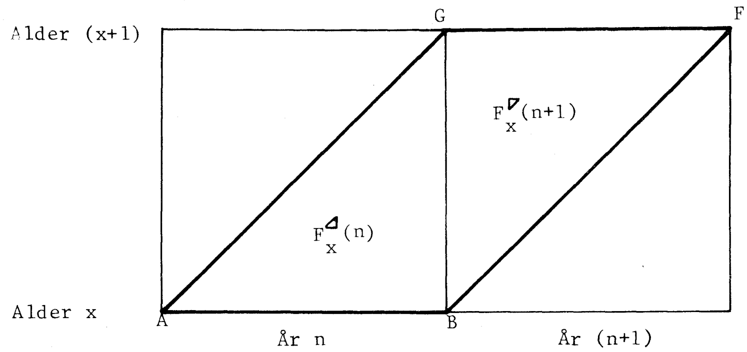
og ser at $F^{\text{▱}}_{x}(n)$ blir antall levendefødte avmerket i parallellogrammet ABFG i figur 11. Vi sier at $F^{\text{▱}}_{x}(n)$ er beregnet ved aldersårsmetoden, og skjønner at dette blir antall barn levendefødt i x-te aldersår av kvinner som fyller x år i året n.

Aldersårsmetoden kan brukes ved kohortundersøkelser som et alternativ til kalenderårsmetoden. Det kan være en ulempe ved metoden at en trenger data fra to kalenderår når en skal bruke den.

Av figur 8 ser vi at en i Norge hadde $F_{25}^{\unicode{x25B1}}(1970) = 2529 + 2477 = 5006$.

Analogt for dødsfall osv.

Siden et kvadrat er et spesielt slags parallellogram, vil vi omtale de tre elementære figurene for censusmetoden, kalenderårsmetoden, og aldersårsmetoden som elementærparallellogrammer. Gå tilbake til figur 7 og se hvordan de tre elementærparallellogrammene for samme x og n ligger i forhold til hverandre.



Figur 11. Elementært aldersårsparallellogram.

6. Risikotid og middelfolkemengde.

6A. I Lexis skjema vil vi nå la siden i hvert kvadrat ha lengden $1/\sqrt{2}$, slik at diagonalen i kvadratet får lengde 1.

La så $R_x^{\triangle}(n)$ være summen av lengdene av de delene av livslinjene i lexisskjemaet som faller i trekanten ABG i figur 7. Hvis en person fyller x år 20. februar i år n og dør 32 dager senere, er altså denne personens bidrag til $R_x^{\triangle}(n)$ lik 32/365. Analogt er bidraget fra en som fyller x år 28. november år n og som blir i befolkningen året ut, lik 34/365, for vedkommende bidrar med tre dager i november og 31 dager i desember. (La oss si at fødselsdagen selv også skal regnes med.) Alle bidrag til $R_x^{\triangle}(n)$ vil være mindre enn 1, unntatt for dem som fyller x år i befolkningen 1. januar år n og som blir der året ut. Vi vil kalle $R_x^{\triangle}(n)$ for <u>risikosummen</u> for den elementærtrekanten vi omtaler.

Analogt kan vi definere risikosummen $R_x^{\nabla}(n)$ for den øvre elementærtrekanten AGH som summen av lengdene av de delene av livslinjene i lexisskjemaet som faller i den trekanten.

En slik risikosum forteller oss hvor mange personår folk til sammen har levet under "risiko" for å oppleve demografiske hendelser i det området i Lexis skjema som den refererer seg til. (Mens en i dagligtalen mest tenker på noe ubehagelig en kan oppleve når en snakker om en risiko, har dette ordet ikke noen slik overtone i faglig språkbruk. Det en kan oppleve, kan godt være noe hyggelig som en ønsker seg.)

6B. På samme måte som for $F_x^{\square}(n)$, $F_x^{\lozenge}(n)$, og $F_x^{\unicode{x25B1}}(n)$ definerer vi risikosummene for de tre elementærparallellogrammene som

$$R_x^{\square}(n) = R_x^{\triangle}(n) + R_x^{\nabla}(n),$$

$$R_x^{\lozenge}(n) = R_{x-1}^{\nabla}(n) + R_x^{\triangle}(n),$$

og

$$R_x^{\unicode{x25B1}}(n) = R_x^{\triangle}(n) + R_x^{\nabla}(n+1).$$

Disse risikosummene omtales også som <u>de eksakte middelfolkemengdene</u> i de tre parallellogrammene.

Hver person som har alder x i fylte år pr. 31/12 år n, gir et bidrag til $R_x^{\lozenge}(n)$ som er lik den brøkdelen av år n vedkommende er medlem av befolkningen. De fleste vil vel være med hele året, og bidraget blir da lik 1, men folk som blir født, innvandrer, dør, eller utvandrer i løpet av året, vil ha et bidrag mindre enn 1.

Noe tilsvarende gjelder for bidragene til $R_x^{\unicode{x25B1}}(n)$, mens bidragene til $R_x^{-}(n)$ neppe kan regnes ut ved direkte betraktning. Censusmetoden har jo ikke noe enkelt referanseår i den enkeltes liv slik kalenderårsmetoden og aldersårsmetoden har.

6C. De eksakte verdiene av risikosummene kan bare beregnes med utgangspunkt i individdata. For hvert enkelt medlem av befolkningen må en avgjøre hvor stort bidrag som gis til de enkelte risikosummer, og så må disse bidragene legges sammen systematisk. Dette kan ofte være besværlig eller praktisk ugjennomførlig. I mange situasjoner bruker man derfor heller tilnærmede verdier for risikosummene.

Vi skal gjengi noen tilnærmelsesformler for risikosummene. (For en begrunnelse, se Sverdrup, 1967, avsnitt III. 2.C.)

(6.1) $$R_x^{\triangledown}(n) \simeq \frac{1}{3}L_x(n-1) + \frac{1}{6}L_{x+1}(n),$$

(6.2) $$R_x^{\triangle}(n) \simeq \frac{1}{6}L_{x-1}(n-1) + \frac{1}{3}L_x(n),$$

(6.3) $$R_x^{\square}(n) \simeq \frac{1}{6}L_{x-1}(n-1) + \frac{1}{3}L_x(n-1) + \frac{1}{3}L_x(n) + \frac{1}{6}L_{x+1}(n),$$

(6.4) $$R_x^{\lozenge}(n) \simeq \frac{1}{2}\{L_{x-1}(n-1) + L_x(n)\},$$

(6.5) $$R_x^{\unicode{x25B1}}(n) \simeq \frac{1}{6}L_{x-1}(n-1) + \frac{2}{3}L_x(n) + \frac{1}{6}L_{x+1}(n+1).$$

For $R_x^{\square}(n)$ brukes forøvrig vanligvis den noe mindre tilfredsstillende men trolig brukbare tilnærmelsesformelen

(6.6) $$R_x^{\square}(n) \simeq \frac{1}{2}\{L_x(n-1) + L_x(n)\}$$

istedenfor (6.3). Analogt kan en bruke

(6.7) $$R_x^{\unicode{x25B1}}(n) \simeq L_x(n)$$

istedenfor (6.5).

Av figur 8 får vi for Norge at

$$R_{25}^{\triangledown}(1970) \simeq \frac{1}{3}\cdot 28174 + \frac{1}{6}\cdot 28026 = 14062\tfrac{1}{3},$$

$$R_{25}^{\triangle}(1970) \simeq \frac{1}{6}\cdot 29397 + \frac{1}{3}\cdot 29253 = 14674\tfrac{1}{2},$$

$$R_{25}^{\square}(1970) \simeq \frac{1}{6}\cdot 29397 + \frac{1}{3}\cdot 28174 + \frac{1}{3}\cdot 29253 + \frac{1}{6}\cdot 28026 = 28712.8,$$

$$R_{25}^{\lozenge}(1970) \simeq \frac{1}{2}\cdot 29397 + \frac{1}{2}\cdot 29253 = 29325,$$

og $$R_{25}^{\unicode{x25B1}}(1970) \simeq \frac{1}{6}\cdot 29397 + \frac{2}{3}\cdot 29253 + \frac{1}{6}\cdot 29229 = 29273$$

når vi bruker (6.1)-(6.5). Tilsvarende gir (6.6) at

$$R_{25}^{\square}(1970) \simeq \frac{1}{2}\cdot 28174 + \frac{1}{2}\cdot 29253 = 28713.5,$$

som ikke ligger langt fra det (6.3) ga. Formel (6.7) gir

$$R_{25}^{\unicode{x25B1}}(1970) \simeq 29253,$$

som jo også "treffer" ganske bra.

7. Befolkningsrater.

7A. Fødselsrater, dødsfallsrater, vigselsrater osv. kan beregnes for ett enkelt kalenderår og én enkelt alder med utgangspunkt i hvert av de tre elementærparallellogrammene vi har presentert. Vi definerer fødselsratene for levendefødte

$$f_x^{\square}(n) = F_x^{\square}(n)/\,R_x^{\square}(n),$$
$$f_x^{\lozenge}(n) = F_x^{\lozenge}(n)/\,R_x^{\lozenge}(n),$$
og
$$f_x^{\text{▱}}(n) = F_x^{\text{▱}}(n)/\,R_x^{\text{▱}}(n),$$

og tilsvarende med d for dødsfallsrater, v for vigselsrater, j for jentefødselsrater, osv. Vi får tilnærmede rateverdier ved å bruke tilnærmede verdier for risikosummene. Således får vi for Norge at

$$f_{25}^{\square}(1970) \simeq \frac{4925}{28712.8} = 0.171526$$

ved (6.3), og

$$f_{25}^{\square}(1970) \simeq \frac{4925}{28713.5} = 0.171522$$

ved (6.6). Analogt får vi

$$f_{25}^{\text{▱}}(1970) \simeq \frac{5006}{29273} = 0.171011$$

ved (6.5), og

$$f_{25}^{\text{▱}}(1970) \simeq \frac{5006}{29253} = 0.171128$$

ved (6.7). Endelig blir

$$f_{25}^{\lozenge}(1970) \simeq \frac{5119}{29325} = 0.177289$$

ved (6.4). Vi ser at de ulike tilnærmelsene til samme eksakte rate bare atskiller seg ubetydelig fra hverandre i disse eksemplene, mens de tre typene av rater er noe forskjellige innbyrdes. Hvis vi hadde valgt en annen alder, f.eks. x=20 eller x=27, ville forskjellene kunne blitt større. Slike momenter diskuteres av Brunborg (1973, kap.4).

7B. Hvis vi betrakter kvinnene i en befolkning og ser bort fra inn- og utvandring, vil alle de $F_x^{\lozenge}(n)$ levendefødslene i år n der moren har alder x i fylte år ved årets utløp, skyldes de $L_{x-1}(n-1)$ kvinnene i befolkningen som ved foregående års utløp var x-1 år gamle. I en del sammenhenger vil en derfor beregne

$$\tilde{f}_x^{\lozenge}(n) = F_x^{\lozenge}(n)/L_{x-1}(n-1)$$

istedenfor $f_x^{\lozenge}(n)$. Tilsvarende vil en av og til beregne

$$\tilde{f}_x^{\text{▱}}(n) = F_x^{\text{▱}}(n)/M_x(n)$$

istedenfor $f_x^{\text{▱}}(n)$. Hvis $M_x(n)$ er ukjent, kan en bruke tilnærmelsesformelen

$$M_x(n) \simeq \frac{1}{2}\{L_{x-1}(n-1) + L_x(n)\}. \tag{7.1}$$

I Norge ble

$$\tilde{f}_{25}^{\lozenge}(1970) = \frac{5199}{29397} = 0.176885$$

og

$$\tilde{f}_{25}^{\text{▱}}(1970) \simeq \frac{5006}{\frac{1}{2}\cdot(29397 + 29253)} = 0.170708.$$

Vi ser at $f_{25}^{\square}(1970)$ og $\tilde{f}_{25}^{\square}(1970)$ er ganske like. Tilsvarende er det ikke stor forskjell mellom $f_{25}^{\diamond}(1970)$ og $\tilde{f}_{25}^{\diamond}(1970)$.

Legg forøvrig merke til at det er samme høyreside i (7.1) som i (6.4), slik at vi regner med at $M_x(n) \simeq R_x^{\square}(n)$.

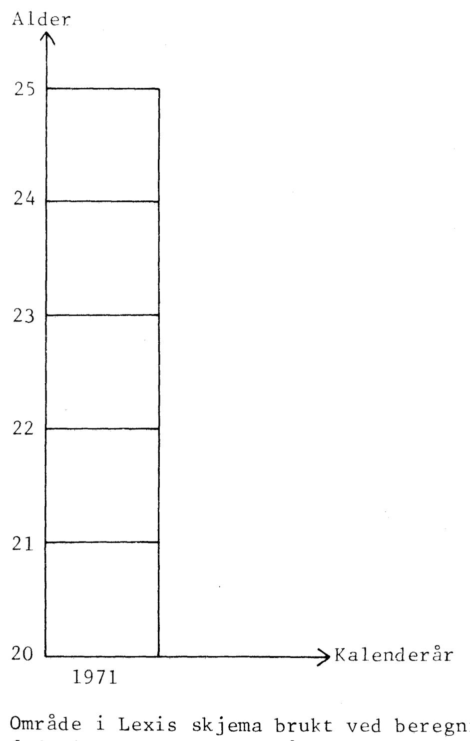
7C. I de to foregående underkapitler har vi diskutert befolkningsrater beregnet for en enkelt alder og et enkelt år. (Ved aldersårsmetoden bruker en riktignok data fra to kalenderår, men bare for ett aldersår.) I mange situasjoner vil en i stedet regne rater for flere år og/eller for aldersgrupper. Eksempelvis oppga vi i underkapittel 1A levendefødselsrater for norske kvinner i aldersgruppen 20-24 år i 1971 og vigselsraten for norske menn i aldersgruppen 25-29 år i 1969-70. Slike rater defineres ved en likefrem generalisering av de ettårige ratene vi har sett på hittil. En avgrenser først et passende område i lexisskjemaet, vanligvis et område som er sammensatt av flere elementærparallellogrammer av samme sort. For dette området defineres så befolkningsraten av den typen vi betrakter (f.eks. fødselsraten), som antall registrerte hendelser avmerket i området (f.eks. antall levendefødte) dividert med risikosummen for området.

Vi skal gi noen eksempler.

7D. Den fødselsraten på 168.6/1000 som ble oppgitt i begynnelsen av underkapittel 1A, er beregnet som en brøk, slik alle rater er. Telleren i brøken er antall levendefødte barn registrert i Norge i 1971 der mødrene var fylt 20-24 år ved nedkomsten. Nevneren er gjennomsnittet av antall kvinner i alder 20-24 registrert som bosatt her i landet 31/12 1970 og det tilsvarende antall 31/12 1971. Formelen raten er regnet ut etter, er altså

$$\frac{\sum_{x=20}^{24} F_x^{\square}(1971)}{\frac{1}{2}\sum_{x=20}^{24}\{L_x(1970)+L_x(1971)\}}, \tag{7.2}$$

der hver $L_x(n)$ representerer et antall kvinner. Området i lexisskjemaet fremgår av figur 12. Nevneren i (7.2) er å betrakte som en tilnærmelse til middelfolkemengden (risikosummen) for området.



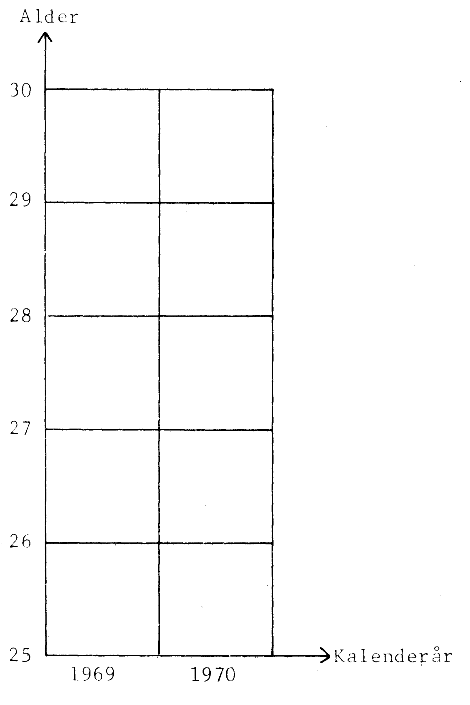
Figur 12. Område i Lexis skjema brukt ved beregning av vanlige fødselsrater for en femårig aldersgruppe i ett kalenderår.

7E. Den vigselsraten på 170.3/1000 som også ble nevnt i begynnelsen av underkapittel 1A, er fremkommet ved at antall giftermål registrert her i landet i 1969 og 1970 der brudgommen var fylt 25-29 år ved giftermålet, er dividert med det dobbelte av antall ugifte menn i alder 25-29 pr. 31/12 1969.

Formelen blir nå

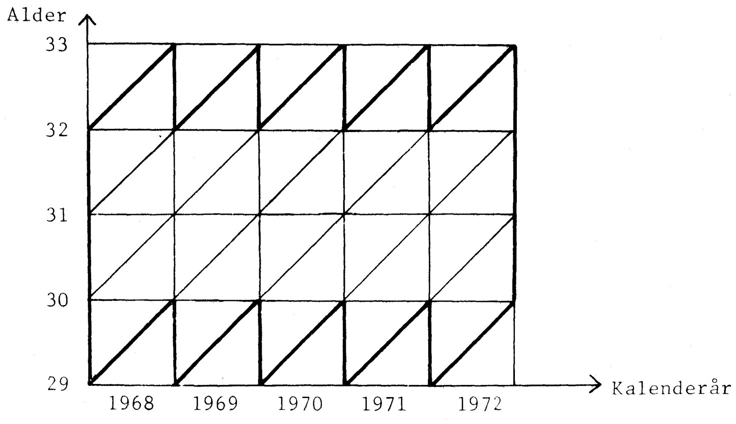
$$\frac{\sum_{x=25}^{29}\{V_x^{\square}(1969) + V_x^{\square}(1970)\}}{2\sum_{x=25}^{29} L_x(1969)}, \tag{7.3}$$

der hver $L_x(n)$ nå representerer et antall <u>ugifte menn</u>. Området i lexisskjemaet fremgår av figur 13. Nevneren i figur (7.3) må igjen betraktes som en tilnærmelse til risikosummen for ugifte menn i området. En multipliserer med 2 under brøkstreken fordi dataene er for to kalenderår.



Figur 13. Område i Lexis skjema brukt ved beregning av vigselsrater for en femårig aldersgruppe og to kalenderår.

7F. I de to foregående eksemplene har vi vist hvordan censusmetoden kan generaliseres. De andre to metodene kan utvides tilsvarende. Hvis en eksempelvis skulle beregne en fødselsrate for aldersgruppen 30-32 etter kalenderårsmetoden på grunnlag av data for årene 1968-1972, ville en basere seg på det området som er innrammet med tykk strek i figur 14.



Figur 14. Område i Lexis skjema for beregning av fødselsraten for aldersgruppen 30-32 ved kalenderårsmetoden på grunnlag av data for årene 1968 til 1972.

Beregningsformelen for denne fødselsraten vil se slik ut:

(7.4)
$$\frac{\sum\limits_{n=1968}^{1972}\sum\limits_{x=30}^{32} F_x^{\square}(n)}{\frac{1}{2}\sum\limits_{n=1968}^{1972}\sum\limits_{x=30}^{32}\{L_{x-1}(n-1)+L_x(n)\}}.$$

7G. Når en skal beregne f.eks. dødsfallsrater ved kalenderårsmetoden, får en problemer med alderen x=0. I figur 10 vil jo ikke trekanten DBA være med i Lexis skjema og en vil ikke ha noen personer i alder -1 med livslinjer som skjærer linjestykket DA, så formelverket for denne metoden blir tilsynelatende meningsløst for x=0. En kan riktignok "redde" det meste formelt ved å sette $L_{-1}(n)=0$, $D_{-1}^{\square}(n)=0$ osv. for alle n, men det er like greit å se på dette tilfellet spesielt, så la oss gjøre det.

Vi lar

$$d_0^{\triangle}(n) = d_0^{\lrcorner}(n) = D_0^{\triangle}(n)/R_0^{\lrcorner}(n)$$

og

$$\tilde{d}_0^{\lrcorner}(n) = \tilde{d}_0^{\lrcorner}(n) = D_0^{\lrcorner}(n)/F(n)$$

i analogi med det foregående. Her er F(n) fortsatt antall levendefødte barn (av det kjønn som dødsfallsraten beregnes for) i år n. En grov tilnærmelsesformel for $R_0^{\triangle}(n)$ er

(7.5)
$$R_0^{\triangle}(n) \simeq \frac{1}{4}\{F(n) + L_0(n)\}.$$

Følgende tankegang ligger bak (7.5): Det fødes F(n) levendefødte barn i året n, og av disse lever $L_0(n)$ ved årets utløp. Gjennomsnittlig antall under risiko for å dø i året n er derfor $\frac{1}{2}\{F(n) + L_0(n)\}$. Siden fødslene er spredt over hele året, vil de levendefødte barna i gjennomsnitt leve i et halvt år hver før de når årsskiftet. Risikosummen fremkommer derfor, grovt regnet, ved å multiplisere $\frac{1}{2}\{F(n) + L_0(n)\}$ med $\frac{1}{2}$. (En bør antakelig søke å utvikle en bedre formel enn denne ved anledning.)

7H. Alle de ratene vi har diskutert hittil, er bygget opp etter et prinsipp der en dividerer et registrert antall hendelser med (en tilnærmelse til) et antall personer som direkte er under "risiko" for å oppleve disse hendelsene, eller vanligvis med et tilsvarende antall personår. Teller og nevner i hver rate refererer seg således til samme delbefolkning.

På ett punkt avviker en imidlertid fra dette prinsippet, antakelig først og fremst på grunn av bekvemmelighetshensyn. Når en regner ut den raten som omtales som antall døde under 1 år pr. 1000 levendefødte for et kalenderår, dividerer en nemlig antall barn under 1 år som dør i kalenderåret med antall levendefødte <u>i samme kalenderår</u> (og regner om til promille), til tross for at noen av disse dødsfallene skyldes barn født året før, og til tross for at noen av dem som fødes ett år, dør året etter før sin ett-års fødselsdag. Formelen som brukes, er altså

$$\tilde{d}_0^{\square}(n) = \frac{D_0^{\square}(n)}{F(n)} .$$

Denne må ikke blandes sammen med $d_0^{\square}(n)$, eller med

$$d_0^{\square}(n) = \frac{D_0^{\square}(n)}{R_0^{\square}(n)} = \frac{D_0^{\square}(n)}{R_0^{\square}(n) + R_0^{\square}(n)}$$

$$\simeq \frac{D_0^{\square}(n)}{\frac{1}{4}\{F(n) + L_0(n)\} + \{\frac{1}{3} L_0(n-1) + \frac{1}{6} L_1(n)\}} ,$$

og heller ikke med

$$\tilde{d}_0^{\square}(n) = \frac{D_0^{\square}(n)}{F(n)} .$$

8. <u>Retningslinjer for merking av beregninger over befolkningsrater.</u>

8A. Enhver tabellarisk oppstilling av befolkningsrater bør merkes slik at en tydelig ser hvordan tallene er beregnet, hvem som har gjort det, og når det ble gjort. Det bør også angis om tallene har vært kontrollert. Dette gjelder både for håndskrevne beregninger og for maskinutskrifter.

8B. Enhver beregning av befolkningsrater bør i tillegg til vanlig tabelltekst og de numeriske resultatene også inneholde følgende elementer:

(i) Aldersdefinisjon.
(ii) Karakteren av telleren i ratene.
(iii) Karakteren av nevneren i ratene.
(iv) Datakilde.
(v) Den dato beregningen er avsluttet.
(vi) Beregnerens signatur.
(vii) Angivelse av hvilke tall som eventuelt er kontrollert, med kontrollørens daterte signatur.

Vi skal kommentere noen av disse punktene.

8C. <u>Aldersdefinisjon.</u> Det er ikke nok å kalle en av kolonnene i beregningen for "alder". Skriv "alder ved årets utgang", "morens alder ved nedkomsten", eller den aldersdefinisjon som ellers er brukt. Dette gjelder også for maskinskrevne og trykte tabeller.

8D. <u>Karakteren av ratenes teller og nevner</u>. Sørg for at en senere bruker klart får angitt hva for slags rate det dreier seg om. Dette kan en oppnå gjennom tilstrekkelig teksting, gjennom oppstilling av beregningsformelen, eller på annen måte. For en del beregningsmetoder som er særlig mye anvendt, kan en bruke en enkel liten symbolsk tegning etter følgende system.

La oss bruke bokstaven H med passende topp- og fotskrifter til å betegne en veldefinert men uspesifisert type hendelser, og la h betegne en tilsvarende befolkningsrate.

Hvis en har brukt beregningsformelen

$$h= \frac{H_x^{\square}(n)}{\frac{1}{2}\{ L_x(n-1) + L_x(n)\}},$$

kan en angi dette ved å tegne følgende figur på beregningen:  . Det er da unødvendig å skrive beregningsformelen på beregningsarket.

Dersom en i stedet har brukt formelen

$$h= \frac{H_x(n)}{\frac{1}{6} L_{x-1}(n-1) + \frac{1}{3} L_x(n-1) + \frac{1}{3} L_x(n) + \frac{1}{6} L_{x+1}(n)},$$

bruker en i stedet følgende figur:  .

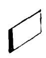
Har en brukt formelen

$$h= \frac{H_x(n)}{\frac{1}{2}\{L_{x-1}(n-1) + L_x(n)\}},$$

blir figuren  .

Har en brukt formelen

$$h= \frac{H_x(n)}{L_{x-1}(n-1)},$$

blir figuren  .

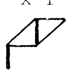
Har en brukt formelen

$$h= \frac{H_x(n)}{L_x(n)},$$

får en figuren  .

Har en brukt formelen

$$h= \frac{H_x(n)}{\frac{1}{2}\{L_{x-1}(n-1) + L_x(n)\}},$$

blir figuren  . Har en brukt (7.2), blir figuren  . Har en brukt (7.3), blir figuren  . Har en beregnet $\tilde{d}_0(n)$ i underkapittel 7H, blir figuren  . Tilsvarende for andre beregningsformler. Hvis tegning av en slik figur blir for besværlig, eller hvis figuren blir for komplisert, bør beregningsformelen angis, eller en bør påføre en henvisning til et notat der den er oppgitt.

8E. <u>Beregningskontroll</u>. Kontroll av beregnede tall utføres av andre enn beregneren selv. Tall som er kontrollert, bør merkes med en hake.

REFERANSER

[1] Brunborg, Helge (1973): "Statistisk Sentralbyrås befolkningsprognosemodell IX. Framskrivinga 1971-2000. Teknisk dokumentasjon." Arbeidsnotat IO 73/15.

[2] Hoem, Jan M. (1972): "Aldersbegrepet i arbeidskraftundersøkelsene." Side 26-37 i Arbeidsnotat IO 72/19 (Metodehefte nr. 1).

[3] Lexis, Wilhelm von (1875): "Einleitung in die Theorie der Bevölkerungsstatistik." Strassburg.

[4] Statistisk Sentralbyrå (1973): "Statistisk årbok 1973", NOS XII 276.

[5] Sverdrup, Erling (1967): "Statistiske metoder ved dødelighetsundersøkelser." Stensiltrykk. Matematisk institutt, Universitetet i Oslo.