

Demografi kompendium

Department of Public Health, University of Copenhagen

Forord

Dette kompendium byder på baggrund og illustration af kurset 'Demografi' i studiet Folkesundhedsvidenskab ved 2. semester, Københavns Universitet. Vi siger tusind tak til Anna-Vera Jørring Pallesen for hendes bidrag til teksten og til Per Kragh Andersen og Johan Sebastian Ohlendorff for grundig gennemlæsning og masser af gode kommentarer.

Laust Hvas Mortensen og Thomas Alexander Gerds

Kapitel 1: Grundbegreber

1 Demografi

Demografi er studiet af humane befolkningers størrelse, sammensætning og udvikling over tid og sted. Demografi beskæftiger sig særligt med tre kernebegivenheder: Fødsel, migration og død. Kendskab til befolkningens sammensætning og struktur er vigtig for forståelsen af befolkningens udvikling. Derfor er demografi også mere end metoder til at holde regnskab over befolkninger, og beskæftiger sig med aldersfordeling, kønsfordeling, etnicitet, familiemandelse, fertilitet, sprog, religion, uddannelse, erhverv og andre relevante faktorer som påvirker befolkningens udvikling.

Demografer anvender forskellige metoder, såsom dataindsamling og statistisk analyse, til at undersøge og beskrive befolkningstendenser- og mønstre. Demografi spiller en vigtig rolle i samfundsplanlægning og i politikudvikling, da det kan give vigtig information om en befolknings sundhedstilstand, uddannelsesbehov, arbejdsstyrke og økonomisk udvikling. Nogle af de områder, hvor demografi er relevant, inkluderer folkesundhed, økonomi, byplanlægning, arbejdsmarked og socialpolitik. Metoderne, som vi bruger til demografi, er tæt beslægtet med metoder fra epidemiologi og statistik.

2 Befolkning

I demografi refererer udtrykket *befolkning* til den gruppe af mennesker, der bor inden for et geografisk område eller en juridisk enhed ved et bestemt tidspunkt. Dermed kan man tale om Danmarks befolkning den 26. juni 1992, og om Nordeuropas befolkning den 1. april 2000. En befolkning kan opdeles og analyseres efter kriterier som for eksempel bopæl, køn, alder, civilstand, etnisk gruppe, uddannelse, erhverv, fag og stilling. Tilsammen er alle disse grupper udtryk for befolkningsstrukturen. En *befolkningsgruppe* er en del af en befolkning som opfylder yderlige kriterier. For eksempel danner alle kvinder, som modtager SU, er yngre end 25 år, og bor i Region Sjælland 1. januar 2023, en befolkninggruppe.

Befolkning i demografi kan beskrives og analyseres med statistikker som befolkningsstørrelse, befolkningstæthed, aldersfordeling, kønsfordeling, etnicitet og mange andre faktorer, der kan give indsigt i befolkningens dynamik og sammensætning. Disse statistikker bruges til at forudsige og planlægge behovet for

offentlige tjenester på eksempeltvis sundheds-, uddannelses- og beskæftigelsesområdet.

2.1 Lukkede og åbne befolkninger

Hvis befolkningen ikke er genstand for ind- og udvandring, kaldes den en lukket befolkning. Hvis befolkningen ikke er lukket, er det en åben befolkning. De allerfleste befolkninger er åbne. Lukkede befolkninger er typisk af hypotetisk/tænkt natur og et *demografisk værktøj*, der bruges til at forklare komplekse principper og regnestykker i en simplificeret situation.

2.2 Kohorte

En kohorte er en gruppe personer som oplever en bestemt (demografisk) begivenhed i en bestemt periode og følges derefter, for eksempel alle kvinder født i Danmark i året 1900 (fødselskohorte) eller alle mænd, som påbegyndte FSV studiet mellem 2017 og 2020. En kohorte er defineret hvis der angives begivenheden (hvad), befolkningen (hvem) og perioden (hvornår).

2.3 Befolkningens størrelse

Lad os tage udgangspunkt i den danske befolkning og starte med at se på dens størrelse. Folketal $N(t)$ angiver befolkningens størrelse til tid t , hvor t er en dato i kalenderen. Ifølge statistikbanken¹ var der 5.806.081 personer i Danmarks befolkning den 1. januar 2019:

$$\text{Befolkning} = \text{Hele Danmark}$$

$$t_1 = 2019-01-01$$

$$N(t_1) = 5.806.081$$

På et senere tidspunkt vil befolkningens størrelse have ændret sig som følge af fødsler, dødsfal og ind- og udvandring. Den 1. januar 2023 var der 5.932.654 personer i Danmarks befolkning. Dermed er den danske befolkning vokset med 126.573 personer i de 1461 dage mellem den 1. januar 2019 og den 1. januar 2023. Nedenfor er givet nogle linjer R-koder, som kan bruges til at beregne antallet af dage mellem to datoer og ændringen i folketallet i perioden.

```
# R-code
t_1= as.Date("2019-01-01")
t_2= as.Date("2023-01-01")
N_1 = 5806081
N_2 = 5932654
data.frame("dato"=c(t_1,t_2), "Folketal"=c(N_1,N_2),
"Antal dage"=c(NA,t_2-t_1), "Vækst"=c(NA,N_2-N_1))
```

¹<https://statistikbanken.dk/>

		dato	Folketal	Antal.dage	Vækst
1	2019-01-01	5806081		NA	NA
2	2023-01-01	5932654		1461	126573

Folketallet ændrer sig hver dag, men det er svært at registrere og tælle nøjagtigt, hvor mange personer der er i befolkning for hver enkelt dag. Heldigvis betyder små unøjagtigheder i tallene typisk ikke særlig meget for demografiske konklusioner. Statistikbanken angiver folketal i starten af hvert kvartal siden 2008 (register FOLK1A), den 1. januar per år siden 1971 (register BEFOLK1), den 1. januar per år siden 1901 (register HISB3). Desuden er der summariske tal fra folketaellinger helt tilbage fra 1769 (register FT).

3 Instruks til brugen af R koder

Vi bruger en række R pakker og også hjemmelavede R funktioner, som gør det nemt arbejde med data fra statistikbanken og implementere specifikke demografiske metoder. For at afprøve R koderne, er det nødvendigt at først køre den følgende R-kode:

```
# load kursets funktioner direkte fra Internet
source("https://raw.githubusercontent.com/tagteam/demogRafi/main/R-
scripts/demofunk.R")
```

En vigtig funktion hedder `hent_data`. Den henter data fra statistikbankens register. For at se hvilke register der er og hvad de hedder, skal man besøge internetsiden <https://statistikbanken.dk/> hvor man finder dem under EMNE Borgere.

```
# Brug funktionen til at hente folketal fra 2023 1. kvartal
hent_data(register = "folk1a",tid = "2023K1")
```

```
# A tibble: 1 × 2
  TID     INDHOLD
  <chr>   <dbl>
1 2023Q1 5932654
```

I register FOLK1a kan man hente data per kvartal. I register BEFOLK1 kun fra første kvartal:

```
hent_data(register = "befolk1",tid = 2023)
```

```
# A tibble: 1 × 2
  TID     INDHOLD
  <dbl>   <dbl>
1 2023  5932654
```

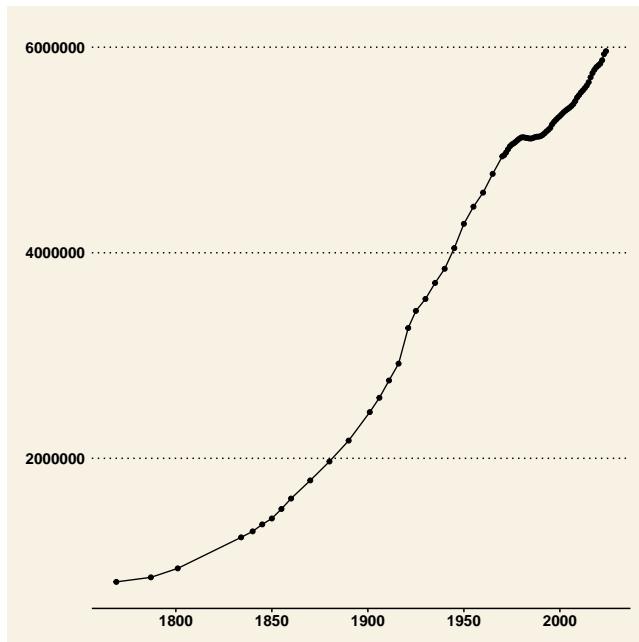
I register HISB3 er folketal rundet af og angivet som antal 1000 personer:

```
hent_data(register = "HISB3",tid = 2023,bevægelse = "M+K")
```

```
# A tibble: 1 × 3
  TID BEVÆGELSE      INDHOLD
  <dbl> <chr>          <dbl>
1 2023 Population Janua...     5933
```

Folketal fra register (FT) bliver vist i Figur 1. Her er R-koder som producerer denne figur:

```
# R-koder der producerer figur 1
dt=hent_data(register = "FT",tid = "all")
ggplot(dt,aes(TID,INDHOLD))+geom_line()+geom_point()+
  theme_wsj() + scale_colour_wsj("colors6")
```



Figur 1: Figuren viser udviklingen af det danske folketal siden 1769 fra folketællinger, statistikbankens register FT.

Alle R-koder som indgår i kapitel 1 kan også hentes som R-markdown fil her:
https://github.com/tagteam/demogRafi/blob/main/R_funk/demofunk.R

4 Befolkningsstatistik i Danmark

Befolkningsstatistikken i Danmark har en lang historie. Siden den første folketælling i 1769 er der foretaget regelmæssig og systematisk optælling af befolkningen. Efter folkestyrets indførelse i 1849 blev opgaven med at gennemføre folketællinger varetaget af Det Statistiske Bureau, som blev etableret i 1850. Det Statistiske Bureau blev senere til Danmarks Statistik, der siden har udgivet de officielle demografiske statistikker for Danmark som den centrale myndighed for dansk statistik. Danmarks Statistik udarbejder alene officiel statistik for Danmark, mens Statistik for de andre lande i rigsfællesskabet varetages af Naatsorsueqqisaartarfik (Grønlands Statistik) og Hagstova Føyora (Færøernes Statistik).

4.1 Administrative registre i demografisk statistik

På det demografiske område har Danmarks Statistik været en foregangsinstitution på et særligt område: Registerbaseret statistik. I 1968 blev Det Centrale Personregister oprettet, og CPR-nummeret indført som identifikationsnummer for personer for at lette den offentlige administration. Sammen med en række andre administrative kilder som Det Centrale Erhvervsregister (1975) og Bygnings- og Boligregistret (1977) gjorde CPR det muligt at lave mange typer af demografisk statistik på baggrund af administrative registre. Som en konsekvens af de nye muligheder lavede Danmarks Statistik i 1981 en af de første folketællinger, der alene var baseret på administrative registre. I dag er CPR registeret rygraden i den demografiske statistik.

CPR er et administrativt register. Det betyder, at registeret er lavet for at kunne bruges til at understøtte den offentlige administration generelt gennem let identifikation, for eksempel for at sikre adgang til offentlige ydelser (for eksempel lægebehandlinger, biblioteksadgang) og korrekt beskatning (for eksempel kilde- og ejendomsbeskatning).

CPR administreres af kommuner og indeholder oplysninger om de centrale begivenheder, som er nødvendige for at belyse fænomener af særlig demografisk interesse. Fordi CPR indeholder dataerne for fødsler, indvandring, udvandring og død, så kan CPR bruges direkte til at opgøre komponenterne i den demografiske ligevægtsligning (se afsnit 6), som er en helt central størrelse i demografien. CPR indeholder også information om ægtefælle, link mellem børn og forældre, samt information om bopæl. De informationer kan bruges til for eksempel familiedemografi, der beskæftiger sig med formation, ændring, sammensætning og udvikling i familier og husholdninger.

Kvaliteten af oplysningerne i CPR og de forskellige andre administrative registre, der anvendes til demografisk statistik er ikke altid perfekt. Kvaliteten af oplysningerne afhænger i praksis af om borgerne er motiveret til at sikre, at oplysningerne er korrekte eller at oplysningerne er kvalitetssikret af de myndigheder, der skal bruge oplysningerne til administrative formål. En anden svaghed ved at anvende administrative registerdata fra for eksempel CPR er, at der ofte vil være oplysninger, som ikke er tilgængelige. Det kan for eksempel være oplysninger om kønsidentitet, familieformer eller religiøst tilhørsforhold,

som ikke er noget, der findes data på i CPR. Styrken ved registerbaseret demografisk statistisk er, at den opgøres meget ofte og uden ret store omkostninger, sammenlignet med for eksempel demografisk statistik baseret på spørgeskema eller besøgsinterview.

4.2 Definitioner af befolkning

For at kunne opgøre befolkningstallet i den officielle statistik er det nødvendigt at have en præcis definition, som kan implementeres i data. Ydermere, er det i praksis også sådan, at definitionerne er samordnet mellem lande, sådan at alle personer tælles en gang og kun en gang. I de nordiske lande foregår samordningen ved, at de administrative registre er koordinerede, mens man i EU/EØS generelt sikrer overensstemmelse ved at dele statistikoplysninger mellem landene. Særligt i små og åbne lande som Danmark er det ikke helt trivielt at sikre, at befolkningen tælles korrekt på baggrund af registerdata. Det gælder generelt, at migrationskomponenten er langt sværere både at registrere og at fremskrive end fødsler og dødsfald.

5 Middelfolketal

Vi indfører nu begrebet middelfolketal, som bruges til at estimere det gennemsnitlige folketal i en given tidsperiode. Middelfolketallet er en vigtig demografisk indikator, der indgår for eksempel i fertilitetsrater og dødelighedstavler.

Middelfolketallet er defineret som det gennemsnitlige folketal i en given tidsperiode. Vi betegner middelfolketallet med \tilde{N} .

For at beregne middelfolketallet helt korrekt, ville man for alle personer være nødt til at tælle, hvor mange dage de har levet i befolkningen i perioden. Så kunne man beregne middelfolketallet eksakt som den samlede gennemlevede tid divideret med periodens længde. Denne beregning giver et gennemsnitligt antal mennesker, der bor i området over en given tidsperiode.

Rent praktisk kender man desværre ikke de præcise tal, altså hvor mange dage alle personer fra en befolkning har levet i en given tidsperiode. Fødselsdage og dødsdage har man typisk registreret korrekt, men ind- og udvandringsdatoer har man typisk mindre nøjagtigt registreret. Flere metoder kan dog bruges til at beregne middelfolketallet approksimativt baseret på enkelte folketal i en given tidsperiode.

5.1 Metode 1

Her skal man kun kende et enkelt folketal, nemlig folketallet cirka i midten af perioden. Hvis perioden starter i tidspunkt t_1 og slutter i tidspunkt t_2 , så er tidspunktet i midten af perioden givet som $(t_1 + t_2)/2$, og folketallet i midten af perioden er givet ved formlen:

$$\tilde{N}[t_1, t_2] = N((t_1 + t_2)/2). \quad (\text{K1.1})$$

Dette tal bruges som et estimat for middelfolketallet. Metoden er simpel og tilstrækkelig for mange formål, især når folketallet ikke ændrer sig særlig meget i den givne tidsperiode. For eksempel bruger Danmark Statistik folktal fra den 1. juli som årets middelfolketal i deres årlige rapporter om befolkningens udvikling.²

5.2 Metode 2

For at bruge denne metode skal man kende folketallet i starten og i slutningen af perioden. Det estimerede middelfolketal er gennemsnittet af de to folktal:

$$\tilde{N}[t_1, t_2] = (N(t_1) + N(t_2))/2. \quad (\text{K1.2})$$

5.3 Metode 3

Her skal man kende folketallet i starten, i slutningen og ved mindst et tidspunkt mere i perioden. Lad os antage, at vi kender folktal til J forskellige tidspunkter $t_1 < t_2 < \dots < t_J$. Det estimerede middelfolketal for perioden $[t_1, t_J]$ kan beregnes med følgende formel:

$$\begin{aligned} \tilde{N}[t_1, t_J] &= \frac{1}{(t_J - t_1)} \sum_{j=2}^J (t_j - t_{j-1}) \frac{(N(t_{j-1}) + N(t_j))}{2} \\ &= \frac{(t_2 - t_1)}{(t_J - t_1)} \frac{(N(t_2) + N(t_1))}{2} + \dots + \frac{(t_J - t_{J-1})}{(t_J - t_1)} \frac{(N(t_J) + N(t_{J-1}))}{2} \end{aligned} \quad (\text{K1.3})$$

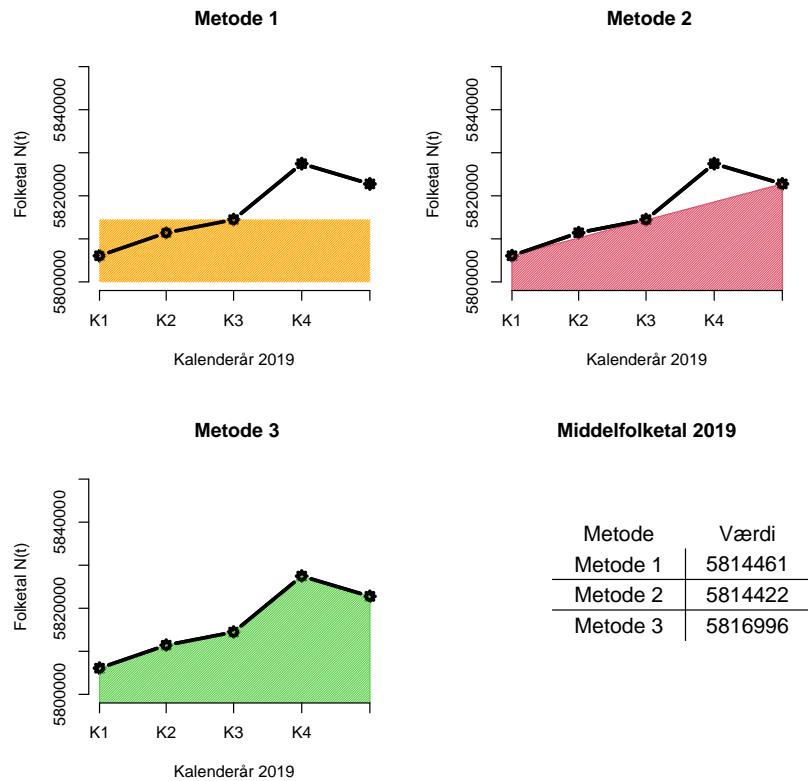
Alle tre metoder er lige gode, hvis folketallet er relativt stabil i perioden. Ændrer folketallet sig meget i perioden, er metode 2 en bedre tilnærmelse til det sande ukendte middelfolketal end metode 1, og ligeledes er metode 3 bedre end metode 2. Figur 2 visualiserer forskellen mellem de 3 metoder.

5.3.1 Eksempel

Vi beregner middelfolketal for den danske befolkning i en periode, som starter den 1. januar 2009 og slutter den 1. januar 2023. Fra statistikbankens BEFOLK1 henter vi folktal i starten, midten og slutningen af perioden:

```
dt=hent_data("BEFOLK1",tid=c(2009,2016,2023))
dt
```

²<https://www.dst.dk/da/Statistik/nyheder-analyser-publ/Publikationer/>



Figur 2: Figuren viser de 3 metoder for at beregne middelfolketal baseret på 5 folketal: 1. januar 2019, 1. april 2019, 1. juli 2019, 1. oktober 2019, 1. januar 2020.

```
# A tibble: 3 × 2
  TID INDHOLD
  <dbl> <dbl>
1 2009 5511451
2 2016 5707251
3 2023 5932654
```

Det er nemmest at anvende metode 1. Med metode 1 er det estimerede middelfolketal $\tilde{N}[1. januar 2016] = 5.707.251$ personer. For at benytte metode 2 bruger vi R som lommerregner:

```
# Metode 2
(5511451 + 5932654)/2
```

[1] 5722052.5

Med metode 2 estimerer vi middelfolketallet i perioden mellem 1. januar 2009 og 1. januar 2023 dermed til 5.722.053 personer. For metode 3 er beregningen i R den følgende:

```
# Metode 3
(2016-2009)/(2023-2009)*(5511451+5707251)/2 + (2023-2016)/(2023-2009)
*(5707251+5932654)/2
```

[1] 5714652

Med metode 3 estimerer vi middelfolketallet i perioden 1. januar 2009 og 1. januar 2023 til 5.714.652 personer baseret på de 3 folketal fra perioden. Vi ser i dette eksempel, at der er mindre end 10.000 personers forskel mellem metode 3 og metode 2. Om denne forskel er vigtig eller ej ville afhænge af formålet med den konkrete demografiske undersøgelse. Er det vigtigt, ville man prøve at estimere middelfolketallet så godt som muligt. For at gøre det ville man hente så mange folketal som muligt fra perioden, og så beregne middelfolketallet med metode 3 på alle disse tal. Følgende R-koder henter alle folketal mellem 1. januar 2009 og 1. januar 2023 fra statistikbankens register FOLK1a og anvender metode 3.

```
# Metode 3 baseret på 57 folketal mellem 2009 og 2023
# konstruere vektor 2009K1, 2009K2, ..., 2022K4
kvartal_years <- paste0(rep(2009:2022,rep(4,14)), "K", 1:4)
# tilføj 2023K1
kvartal_years <- c(kvartal_years, "2023K1")
# hent data fra FOLK1a
dt <- hent_data(register = "FOLK1a", tid=kvartal_years)
# transform årstal + kvartal til dato
dt <- format_dato(dt, variable = "TID")
# anvend middelfolketal metode 3
summarise(dt, {
```

```

len <- length(TID)
len_periode_total <- as.numeric(TID[length(TID)]-TID[1])
len_periode <- as.numeric(TID[-1]-TID[-len])
ft_gennemsnit <-(INDHOLD[-len]+INDHOLD[-1])/2
sum(len_periode*ft_gennemsnit)/len_periode_total
})%>% pull()

```

[1] 5717974

Baseret på 57 folketal i perioden mellem den 1. januar 2009 og den 1. januar 2023 estimerer vi middelfolketal for perioden til at være 5.717.974 personer.

6 Den demografiske ligevægtsligning

Den demografiske ligevægtsligning er en formel, der bruges i demografisk analyse til at beskrive forholdet mellem antallet af fødsler, dødsfald og migration i en periode i en befolkning. Jo flere dødsfald og jo flere personer, der udvandrer, jo mindre er folketallet i slutningen af perioden sammenlignet med starten af perioden. Ligeledes er folketallet i slutningen af perioden højere jo flere personer der bliver født og indvandrer til befolkningen. Den demografiske ligevægtsligning for en periode $[t_1, t_2]$ er:

$$N(t_2) = N(t_1) + (F[t_1, t_2] - D[t_1, t_2]) + (I[t_1, t_2] - U[t_1, t_2]). \quad (\text{K1.4})$$

hvor vi har brugt følgende notation:

- $N[t_1]$ er folketal på tidspunkt t_1 .
- $N[t_2]$ er folketal på tidspunkt t_2 .
- $F[t_1, t_2]$ er antallet af fødsler i perioden.
- $D[t_1, t_2]$ er antallet af dødsfald i perioden.
- $I[t_1, t_2]$ er antallet af indvandrere i perioden.
- $U[t_1, t_2]$ er antallet af udvandrere i perioden.

Formlen siger kort sagt, at den samlede befolkning på et tidspunkt t_2 er lig den samlede befolkning på tidspunktet t_1 plus en stigning i befolkningen på grund af fødsler og indvandring og en reduktion i befolkningen på grund af dødsfald og udvandring. Vi kalder forskellen mellem fødsler og dødsfald ($F[t_1, t_2] - D[t_1, t_2]$) for *naturlig vækst* (som kan være negativ) og forskellen mellem ind- og udvandring ($I[t_1, t_2] - U[t_1, t_2]$) for *nettovandring*. Det giver følgende version af den demografiske ligevægtsligning (formel (K1.4)):

$$\underbrace{N(t_2) - N(t_1)}_{\text{Vækst}} = \underbrace{(F[t_1, t_2] - D[t_1, t_2])}_{\text{Naturlig vækst}} + \underbrace{(I[t_1, t_2] - U[t_1, t_2])}_{\text{Nettovandring}}$$

6.0.1 Eksempel

Vi henter tal fra den danske befolkning i 2022 fra statistikbankens register FOLK1a, DOD, INDVAN og UDVAN.

```
N <- hent_data("FOLK1a", tid = c("2022K1", "2023K1"))[["INDHOLD"]]
D <- hent_data("DOD", tid=2022)[["INDHOLD"]]
F <- hent_data("FOD", tid = 2022)[["INDHOLD"]]
I <- hent_data("INDVAN", tid=2022)[["INDHOLD"]]
U <- hent_data("UDVAN", tid=2022)[["INDHOLD"]]
# data for ligevægtsligningen
tibble(X=c("Folketal jan 2022",
          "Folketal jan 2023",
          "Fødsler 2022",
          "Dødsfald 2022",
          "Indvandring 2022",
          "Udvandre 2022"),
       Antal=c(N[1], N[2], F, D, I, U))
```

```
# A tibble: 6 × 2
  X           Antal
  <chr>        <dbl>
1 Folketal jan 2022 5873420
2 Folketal jan 2023 5932654
3 Fødsler 2022     58430
4 Dødsfald 2022    59435
5 Indvandring 2022  121183
6 Udvandre 2022    62927
```

Baseret på disse tal beregner vi væksten i den danske befolkning i perioden til $(5.932.654 - 5.873.420) = 59.234$ personer. Den naturlige vækst i perioden er negativ: $(58.430 - 59.345) = -915$ personer og nettovandring i perioden positiv: $(121.183 - 62.927) = 58.256$ personer. Vi ser, at ligevægtsligningen (formel (K1.4)) ikke går op, da der mangler 1893 personer:

$$\underbrace{59.234}_{\text{Vækst}} = \underbrace{-915}_{\text{Naturlig vækst}} + \underbrace{58.256}_{\text{Nettvandring}} + \underbrace{1893}_{\text{fejl}}.$$

Det vil sige, at de forskellige registre, som statistikbanken internt bogfører, ikke er konsistente. Det kan der være mange grunde til. En vigtig grund er, at det er svært at registrere de præcise datoer, hvornår ind- og udvandringer sker.

Figur 3 viser vækst, fødsler, dødsfald, ind- og udvandring mellem 1980 og 2023 i den danske befolkning. Det er tydeligt, at indvandring er den dominerede faktor for ændringer af folketallet i denne periode, hvormod fødsler og dødsfald er på et rimeligt konstant niveau. Man kan også se, at udvandring er stigende helt op til 2019 men knækker i 2020 på grund af coronakrisen.

```
# R-koder der producerer figur 3
```

```

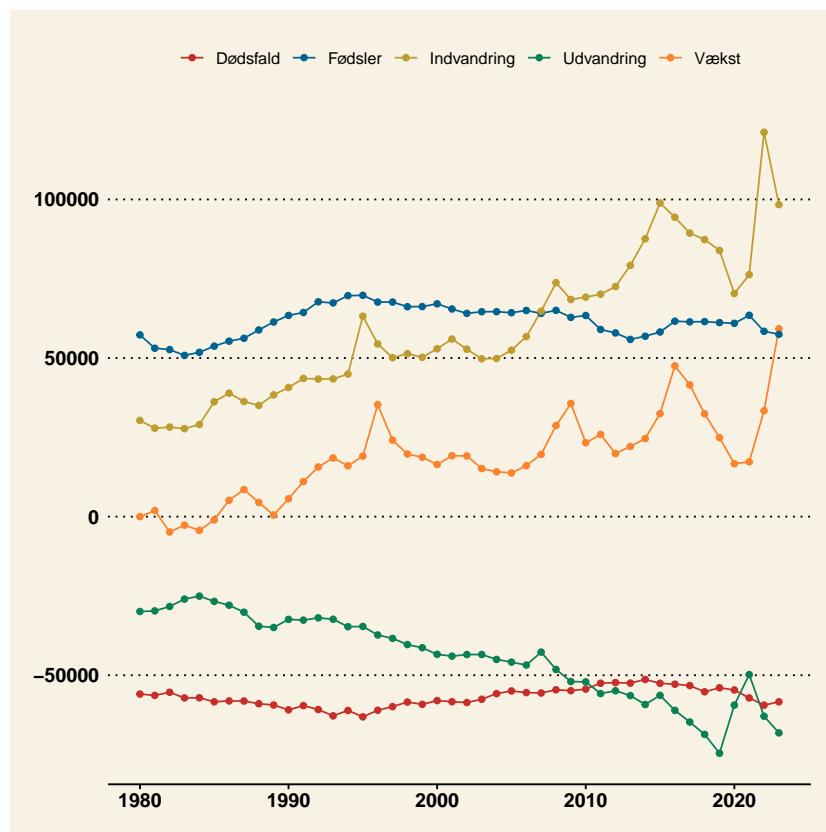
V = hent_data("BEFOLK1",tid=1980:2023)
V = V %>% mutate(INDHOLD = INDHOLD - c(INDHOLD[1],INDHOLD[-length(INDHOLD)]))
D = hent_data("dod",tid=1980:2023)
D = D %>% mutate(INDHOLD = -INDHOLD)
F = hent_data("FOD",tid=1980:2023)
I = hent_data("INDVAN",tid=1980:2023)
U = hent_data("UDVAN",tid=1980:2023)
U = U %>% mutate(INDHOLD = -INDHOLD)
# samle data
dat <- tibble(rbind(cbind(X="Vækst",V),
                     cbind(X="Dødsfald",D),
                     cbind(X="Fødsler",F),
                     cbind(X="Indvandring",I),
                     cbind(X="Udvandring",U)))
dat <- dat %>% mutate(X = factor(X))
ggplot(dat,aes(TID,INDHOLD,color=X,group=X))+geom_line()+
  geom_point()+theme_wsj() + scale_colour_wsj("colors6") +
  theme(legend.title=element_blank())

```

7 Rater

I demografi bruger vi rater til at beskrive befolkningens relative ændringer for at sammenligne forskellige befolkninger og for at sammenligne befolkningsgrupper indenfor en befolkning. For eksempel beskriver dødsraten antal døde relativt til befolkningens størrelse. Det er som udgangspunkt typisk ikke meningsfyldt at sammenligne absolut antal døde mellem befolkninger. For eksempel døde 569 personer på Bornholm og 2 personer på Christiansø i 2022. Her kan man næppe konkludere, at dødeligheden var højere på Bornholm end på Christiansø. Brugen af rater frem for absolut antal er yderst relevant, når formålet er at sammenligne befolkninger, som har forskellig storrelse. For eksempel var mortalitetsraten på Bornholm i 2022 lig med $569/39817 = 14,3$ per 1000 personår og på Christiansø $2/91 = 22,0$ per 1000 personår i samme tidsperiode.

Som enhed for dødsraten bruges ofte *antal døde per personår*. Her dividerer man antal døde i en periode med antal personår, som personer fra befolkningen har levet i samme periode. Mere generelt har en rate som kendetegn, at den er defineret som kvotient af to størrelser i forskellige måleenheder. Ved beskrivelse af en rates enheder bruges ordet "per" til at adskille enhederne for de to målinger, der bruges til at beregne raten. For eksempel er hastighed af en cykel en rate, som kan beskrives med enheden *kilometer per time*. De fleste demografiske rater bruger *risikotid* i nævneren og antal begivenheder i tælleren og har dermed en enhed *antal begivenheder per personår*.



Figur 3: Figuren viser ændringen i folketal (vækst), antal fødsler, dødsfald, ind- og udvandring siden 1980.

7.1 Risikotid

Vi betegner med $R[t_1, t_2]$ den samlede gennemlevede tid i perioden $[t_1, t_2]$ for alle personer i en befolkning og kalder den også for *risikotid*. Udtrykket *risikotid* giver egentlig kun mening når man studerer en risikabel hændelse, som for eksempel død blandt personer, som er eksponeret for risikoen for denne hændelse i perioden. Det er især i epidemiologi, hvor man for eksempel kan interesser sig for sygdomsrater, hvor nævneren er risikotid for personer, som var eksponeret for sygdomsrisiko. I demografi bruger vi udtrykket *risikotid* også i andre sammenhænge. Enheden for risikotid er antal personår. For at beskrive risikotid i små befolkninger kan den regnes om til antal personuger eller antal persondage. For store befolkninger vil man typisk regne om til enheder som *10.000 personår*, *100.000 personår* eller *1.000.000 personår*. Kender man det eksakte antal dage, som alle personer i en befolkning har levet i en given periode, beregner man risikotiden eksakt som sum af alle persondage. Det kræver dog, at man kender eksakte datoer for alle fødsler, dødsfald samt ind- og udvandringer i perioden. Det gør man sjældent. Man kan dog estimere risikotid baseret på registerdata. For at estimere risikotid i en befolkning baseret på registerdata ganger vi typisk periodens middelfolketal med periodens længde:

$$R[t_1, t_2] = \tilde{N}[t_1, t_2] \cdot (t_2 - t_1) \quad (\text{K1.5})$$

For eksempel var middelfolketallet i 2022 på Bornholm 39.817 personer (tal fra statistikbankens register FOLK1a, metode 1 for middelfolketallet). Vi estimerer dermed risikotid af Bornholms befolkning i året 2022 til 39.817 personår.

7.2 Perioderater

Mange demografiske rater er defineret som antal begivenheder i en periode, for eksempel antal dødsfald eller antal indvandringer, relativt til risikotid, altså antal gennemlevede personår i samme periode i en befolkning:

$$\text{Rate}_X[t_1, t_2] = \frac{\text{Antal begivenheder } X \text{ i perioden } [t_1, t_2]}{R[t_1, t_2]} \quad (\text{K1.6})$$

Denne formel kan anvendes rimeligt generelt. Man skal dog være opmærksom på, at en korrekt fortolkning af perioderater kræver kendskab til begivenheden (hvad), befolkningen (hvem) og perioden (hvornår). Desuden skal man huske at angive enheder, når man rapporterer perioderater.

Vi beregner mortalitetsraten for Bornholm i året 2022.

```
# folketal den 1. jan 2022 og 1 jan 2023
N_bornholm=hent_data(register = "folk1a",
                      tid = c("2022K1", "2023K1"),
                      område ="bornholm")
# middelfolketal metode 2
N_bornholm_metode2 = mean(N_bornholm$INDHOLD)
# antal døde
```

```
D_bornholm = hent_data(register = "fod207",tid = "2022",område ="bornholm")$INDHOLD
# mortalitetsrate per 1000 personår
1000*D_bornholm/N_bornholm_metode2
```

[1] 14.37831

og konkluderer:

Mortalitetsraten på Bornholm i året 2022 var 14,4 per 1000 personår.

Bemærkning til terminologi

Rater som tæller hændelser og risikotid i hele populationen kalder vi for *summariske rater*. I det her kapitel omtaler vi kun *summariske rater*. I næste kapitel, diskuterer vi også *aldersspecifikke rater* og *standardiserede rater*. I det her kapitel udelader vi at bruge prædikatet “summarisk”.

7.2.1 Eksempel

Vi illustrerer beregningen af perioderater og bruger flyttingsrater for flytninger indenfor Danmark i perioden fra 1. januar 2020 til 1. januar 2023 som eksempel. Først henter vi antal flytninger fra statistikbankens register FLY.

```
# Antal flytninger indenfor Danmark i årene 2020, 2021, 2022
FL <- hent_data("FLY",tid=2020:2022)
# Antal flytninger i perioden [2020,2022]
X <- pull(summarise(FL,sum(INDHOLD)))
X
```

[1] 2773056

Der er registreret 2.773.056 flytninger indenfor Danmark i perioden fra 1. januar 2020 til 1. januar 2023. Bagefter henter vi folketal fra statistikbankens register FOLK1a og beregner middelfolketal med metode 2. Vi beregner også risikotid.

```
# Folketal for den danske befolkning i perioden
N <- hent_data("FOLK1a",tid = c("2020K1","2023K1"))
# Middelfolketal metode 2
NN <- summarise(N,middelfolketal=mean(INDHOLD))
# Risikotid
Risikotid <- summarise(NN,R= middelfolketal * as.numeric(as.Date("2023-01-01")-as.Date("2020-01-01"))/365.25)
R <- pull(Risikotid)
R
```

[1] 17637149

Riskotiden af den danske befolkning i perioden fra 1. januar 2020 til 1. januar 2023 er estimeret til 17.637.149 personår.

Til sidst beregner vi flytningsraten i perioden.

```
# Flytningsrate per personår
X/R
# Flytningsrate per 1000 personår
1000*X/R
```

[1] 0.1572281

[1] 157.2281

Flytningsraten for flytninger internt i Danmark var 157.2 flytninger per 1000 personår i perioden fra 1. januar 2020 til 1. januar 2023.

7.3 Demografiske vækstrater

Vi kan anvende formlen for perioderater (formel (K1.6)) til mortalitetsrater (begivenhed X er et dødsfald), fødselsrater (begivenhed X er en fødsel), indvandringsrater (begivenhed X er en indvandring) og udvandringsrater (begivenhed X er en udvandring). På den måde kan vi beskrive en dekomposition af demografiske vækstrater.

Vi trækker $N(t_1)$ fra begge sidder af den demografiske ligevægtsligning (formel (K1.4)) og dividerer på begge sidder af lighedstegnet med $R[t_1, t_2]$. Det giver følgende dekomposition af befolkningens vækstrate i perioden $[t_1, t_2]$:

$$\underbrace{\frac{N(t_2) - N(t_1)}{R[t_1, t_2]}}_{\text{Vækstrate}} = \underbrace{\frac{F[t_1, t_2]}{R[t_1, t_2]}}_{\text{Fødselsrate}} - \underbrace{\frac{D[t_1, t_2]}{R[t_1, t_2]}}_{\text{Mortalitetsrate}} + \underbrace{\frac{I[t_1, t_2]}{R[t_1, t_2]}}_{\text{Immigrationsrate}} - \underbrace{\frac{U[t_1, t_2]}{R[t_1, t_2]}}_{\text{Emigrationsrate}} \quad (\text{K1.7})$$

Vi bemærker at for mortalitetsrater og udvandringsrater giver udtrykket *risikotid* mening, fordi $R[t_1, t_2]$ stammer fra de personer som faktisk var under risiko for hændelsen.

7.3.1 Eksempel

Vi beregner vækstrater, som defineret i ligning (K1.7), for den danske befolkning i perioden 1. januar 2022 til 1. januar 2023. Ud over data, som vi allerede har brugt i eksemplet for den demografiske ligevægtsligning, har vi nu brug for risikotid for den danske befolkning i perioden. Vi anvender metode 2 for middelfolketallet og beregner risikotid ved at gange med 1 år:

```
N <- hent_data("FOLK1a",tid=c("2022K1","2023K1"))[["INDHOLD"]]
vækst <- N[2]-N[1]
middelfolketal <- mean(N)
risikotid <- middelfolketal*1
risikotid
```

[1] 5903037

Risikotiden i den danske befolkning var således 5.903.037 personår mellem 1. januar 2022 og 1. januar 2023. Vi henter antal begivenheder og beregner raterne.

```
# mortalitetsrate
D <- hent_data("DOD",tid=2022)[["INDHOLD"]]
Drate <- 1000*D/risikotid
# fødselsrate
F <- hent_data("FOD",tid=2022)[["INDHOLD"]]
Frate <- 1000*F/risikotid
# indvandringsrate
I <- hent_data("INDVAN",tid=2022)[["INDHOLD"]]
Irate <- 1000*I/risikotid
# udvandringsrate
U <- hent_data("UDVAN",tid=2022)[["INDHOLD"]]
Urate <- 1000*U/risikotid
# væksrate
Vrate <- 1000*vækst/risikotid
# naturlige væksrate
NaturVrate <- Frate-Drate
# nettovandringsrate
NettoVrate <- Irate-Urate
x=tibble(X=c("Vækstrate",
            "Mortalitetsrate",
            "Fødselsrate",
            "Indvandringsrate",
            "Udvandringsrate",
            "Naturlige_vækst_rate",
            "Netto_vandrings_rate"),
          Rate=c(Vrate,Drate,Frate,Irate,Urate,NaturVrate,NettoVrate))
x
```

```
# A tibble: 7 × 2
  X                  Rate
  <chr>              <dbl>
1 Vækstrate         10.0 
2 Mortalitetsrate   10.1 
3 Fødselsrate       9.90 
4 Indvandringsrate  20.5
```

- 5 Udvandringsrate 10.7
6 Naturlige_vækst_rate -0.170
7 Netto_vandrings_rate 9.87

I 2022 voksede den danske befolkning med 10,0 personer per 1000 personår. Den naturlige vækstrate var -0,2 personer per 1000 personår og nettvandringsraten var 9,9 personer per 1000 personår. Ligesom den demografiske ligevægtsligning ikke går op, gør denne formel heller ikke, og der er en fejlrate:

$$\underbrace{10,0}_{\text{Vækstrate}} = \underbrace{-0,2}_{\text{Naturlig vækstrate}} + \underbrace{9,9}_{\text{Nettvandringsrate}} + \underbrace{0,3}_{\text{fejlrate}}.$$

Kapitel 2: Mortalitet og standardisering

1 Mortalitet

Mortalitet er et synonym til dødelighed og refererer til hændelsen død. I demografi betragtes dødsfald som en måde, hvorpå personer kan forlade en befolkning og mortalitet beskriver således en vigtig del af en befolknings dynamik og udvikling. I forrige kapitel blev I præsenteret for den demografiske ligevægtsligning, som beskriver befolkningstilvæksten ud fra fire demografiske komponenter: fødsler, dødsfald, indvandringer og udvandringer. Mortalitet er altså en af de grundlæggende emner inden for demografi.

Fra et samfundsperspektiv spiller demografiske undersøgelser af befolkningens mortalitet en vigtig rolle i at forstå forandringer i befolkningssammensætningen og befolkningens sundhedstilstand. Et eksempel er den faldende dødelighed på tværs af alle aldersgrupper, som har resulteret i, at vi lever længere. Dette medfører en stigende andel af ældre i befolkningen. Desuden giver analyser af mortalitet indblik i befolkningens generelle sundhedstilstand, og de kan informere om ulighed i sundhed/dødelighed på tværs af befolkningsgrupper. Disse informationer er væsentlige for det politiske arbejde, som former sundhedspolitikker og påvirker ressourceallokering i sundhedsvæsenet.

1.1 Den demografiske transition

Den demografiske transition beskriver et historisk skift i befolkningsmønstre fra et traditionelt samfund med høj fertilitet og høj dødelighed til et samfund præget af lav fertilitet og lav dødelighed. Den faldende tendens i dødelighed er det element af transitionen, der kommer først, hvorefter et fald i fertiliteten følger (Kapitel 4). Den demografiske transition er ofte drevet af økonomisk vækst, forbedret sundhed og sociale forandringer, som påvirker beslutninger om familiemandelse. Som konsekvens af den demografiske transition ændres befolkningssammensætningen.

1.2 Mål for mortalitet

Demografien anvender forskellige mål til at kvantificere og opgøre mortalitet. En summarisk mortalitetsrate angiver antallet af dødsfald per personår i en given periode. Beregningen af summariske mortalitetsrater blev forklaret i Kapitel 1. Den summariske mortalitetsrate er for eksempel nyttig, når formålet er at

beskrive dødeligheden i en befolkning i en given periode. Summariske mortalitetsrater kan udregnes per år i en given periode og således informere om ændringer i dødeligheden over tid. Figur 2 viser hvordan dødelighed i Danmark har udviklet sig mellem 1974 og 2022.

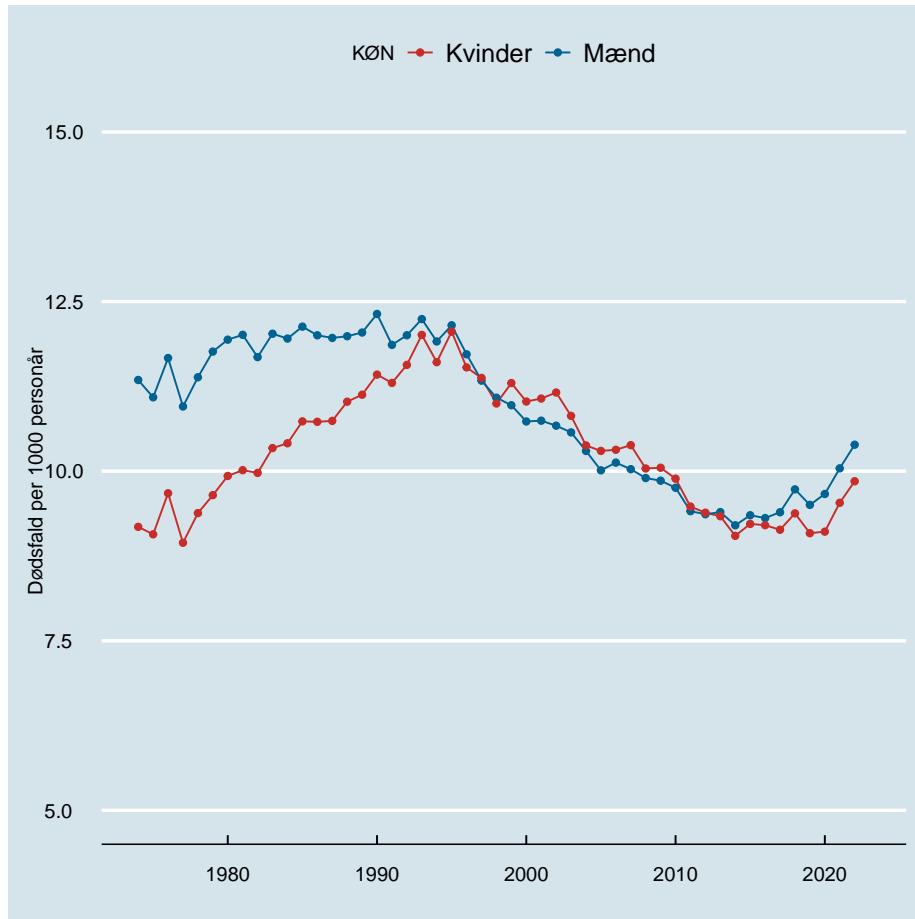


Figure 1: Udvikling i den summariske mortalitetsrate i perioden 1974-2022 i Danmark. Kilde: statistikbankens register DOD, BEFOLK2

Mortalitet kan også beskrives ud fra *aldersspecifikke mortalitetsrater*. Disse beregnes på samme vis som den summariske mortalitetsrate, men separat for hver aldersgruppe. Det bliver således muligt at undersøge forskelle i dødeligheden på tværs af aldersgrupper. Ligeledes kan aldersspecifikke mortalitetsrater udregnes per år over en given periode og derved give information om ændringer i aldersspecifik dødelighed over tid. Altså: Oplever nogle aldersgrupper et fald i mortaliteten, hvor andre oplever en stigning? Det kan vi se i nedenstående Figur 2, som viser køns- og aldersspecifikke mortalitetsrater. Her ser vi, at

dødeligheden er faldet i samtlige aldersgrupper og for begge køn i løbet af årene 1974–2022.

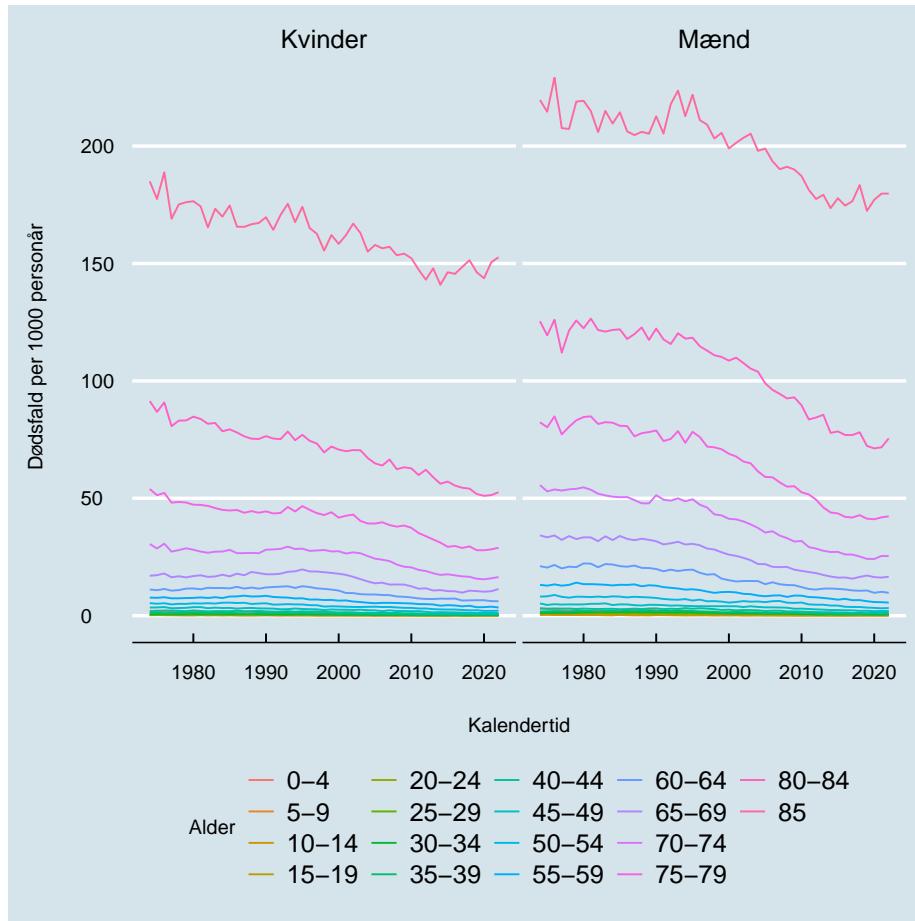


Figure 2: Udvikling i de køns- og aldersspecifikke mortalitetsrater i perioden 1974–2022 i Danmark. Kilde: statistikbankens register DOD, BEFOLK2.

Aldersspecifikke mortalitetsrater kan også angives for et enkelt år, hvilket giver et indtryk af forskelle i dødeligheden på tværs af aldersspektrum. Køns- og aldersspecifikke mortalitetsrater for 2022 kan ses i Figur 3. Her ser vi, at mortaliteten er lav og forbliver sådan frem til efter 60-årsalderen, hvorefter mortaliteten stiger - først en lille stigning og senere en mere markant stigning. Det er også tydeligt, at mortaliteten stiger hurtigere for mænd end for kvinder.

For at sammenfatte dødelighed i et tal beregner man ofte middellevetid inden for demografi. Dette er et mål for, hvor længe en nyfødt kan forvente at leve, givet at de aldersspecifikke mortalitetsrater, der er gældende det år, barnet er født i, holder sig konstant i al fremtid. Middellevetidens udregning bygger

på konstruktionen af såkaldte overlevellestavler, som vi beskrivelser i Kapitel 3. I dette kapitel vil vi fokusere på, hvordan summariske og aldersspecifikke rater kan standardiseres for at gøre det muligt at sammenligne rater - herunder mortalitetsrater - på tværs af befolkninger.

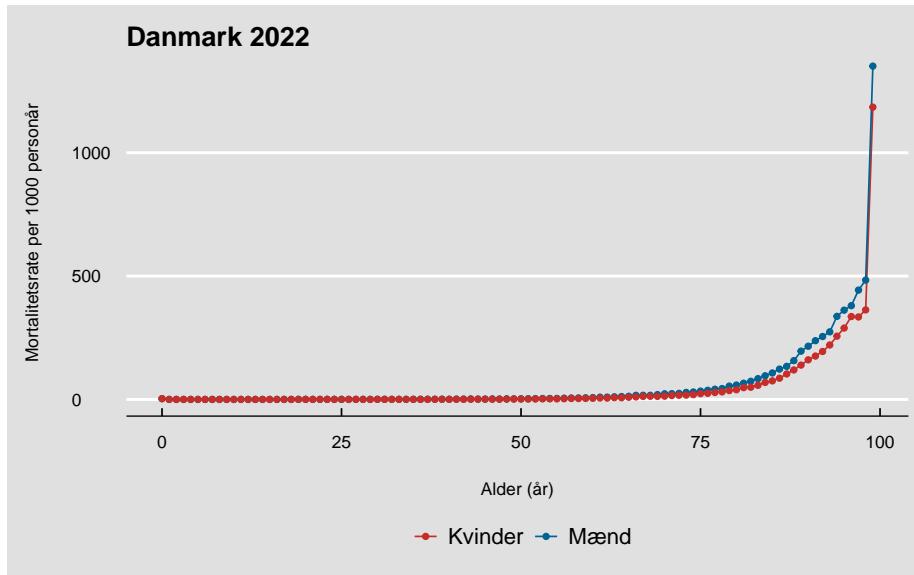


Figure 3: Aldersspecifikke mortalitetsrater fra hele den danske befolkning i 2022. Vi ser at dødeligheden blandt de ældre var højre for mænd end for kvinder.

2 Standardiserede rater

Vi vil nu gå i dybden med begreberne *aldersspecifikke rater* og *standardiserede rater*. Disse skal ikke forveksles med de summariske rater som Kapitel 1 har indført. I det følgende betegner vi derfor alle rater, som Kapitel 1 har indført, som *summariske rater*. *Summarisk* betyder, at raterne tæller hændelser og risikotid i hele befolkningen, altså uanset alder og uden standardisering. For at motivere standardiserede rater starter vi med at forklare begrænsningerne med de summariske rater når det kommer til sammenligning af forskellige befolkninger.

2.1 Sammenligning af summariske rater

Som udgangspunkt har det begrænset interesse at sammenligne forskellige befolkningers summariske rater. Det er især problematisk, når befolkningerne, som man ønsker at sammenligne, har forskellige aldersfordelinger. Afhængig af formålet med undersøgelsen kan det alligevel godt være, at man vil sammenligne summariske rater, men det er vigtigt, at man er klar over, at resultatet

afhænger af aldersfordelingerne. Problemet som opstår ved sammenligning af summariske rater er ret nemt at indse ved følgende eksempel. En matematisk forklaring (Kitagawas dekomposition) følger i afsnit 5.1.

2.1.1 Eksempel

Vi beregner de summariske mortalitetsrater for året 2011 i den kvindelige og i den mandlige befolkning.

I det følgende viser vi de R-koder som vi har brugt til at generere figurer og eksempler. For at afprøve R-koderne og for at se mellemresultaterne kan du hente filen: <https://github.com/tagteam/demogRafi/blob/main/R-scripts/kapitel2.Rmd>

```
# R-koder til beregning af summariske mortalitetsrater
x <- hent_data("FOLK1a",tid = "2011K3",køn = c(2,1))
# fjern TID fordi den er konstant
x$TID <- NULL
# Risikotid i 2011 baseret på middelfolketal metode 1
# ganget med 1 år
x <- rename(x,"risiko_tid"="INDHOLD")
# Antal dødsfald i 2011
d <- hent_data("DOD",tid="2011","køn" = c("K","M"))
# Fjern TID fordi den er konstant
d$TID <- NULL
# Endre variable navn
d <- rename(d,"antal_dod"="INDHOLD")
# join folktal og antal døde
dat <- left_join(x,d,by="KØN")
# Summariske mortalitetsrater per 1000 personår
dat <- mutate(dat,
              "Summariske mortalitetsrate"=1000*antal_dod/risiko_tid)
dat
```

```
# A tibble: 2 × 4
  KØN      risiko_tid antal_dod `Summariske mortalitetsrate`
  <chr>     <dbl>    <dbl>                <dbl>
1 Mænd      2760140   25939                 9.40
2 Kvinder    2806716   26577                 9.47
```

Vi ser at den summariske mortalitetsrate i året 2011 var 9,47 døde per 1000 personår for danske kvinder og 9,40 døde per 1000 personår for danske mænd. Ved første øjekast strider dette resultat imod den gængse viden, at danske kvinder lever længere end danske mænd. Fordi dødeligheden stiger med alderen og fordi der er flere kvinder med en høj alder end mænd med en høj alder, er den summariske mortalitetsrate højere for kvinder end for mænd. Den summariske

mortalitetsrate afspejler nemlig ikke kun dødeligheden men også aldersfordelingen i befolkningen. Resultatet er dog helt korrekt; kvinderne havde en højere summarisk mortalitetsrate end mænd i 2011. Den skyldes ikke deres køn men deres alder.

Hvordan skal disse rater fortolkes? En rate er jo ikke en sandsynlighed, og det ville ikke være helt korrekt at konkludere, at der døde 9,47 kvinder blandt 1000 kvinder, som man følger igennem 2011, fordi de kvinder, som dør midt i eller i starten af 2011, jo ikke bidrager med et helt personår til risikotiden. En bedre fortolkning opstår, når man sammenligner mortalitetsraten med hastigheden af en cykel. Hastigheden er raten cyklen bevæger sig med og kan for eksempel være 20 kilometer per time. Mortalitetsraten betegner hastigheden hvormed befolkningen dør; den kan for eksempel være 9,40 døde per 1000 personår. Denne hastighed, altså mortalitetsraten, betegner vi også med *dødelighed*. Det vil sige, at resultatet kan fortolkes på følgende måde: Danske kvinder har haft en lidt højere dødelighed i 2011 end danske mænd (fordi de var ældre).

3 Aldersfordeling

En aldersfordeling af folketal angiver hvor mange personer har en bestemt alder relativt til hvor mange personer, der er i befolkningen. Man kan også beregne aldersfordeling af andre størrelser end folketal. For eksempel skal vi bruge aldersfordeling af risikotid i beregningen af aldersstandardiserede mortalitetsrater.

3.1 Alderspyramide

For at sammenligne aldersfordelinger af kvinder og mænd kan man tegne en alderspyramide. Figur 4 viser alderspyramiden for den danske befolkning baseret på data fra 1. juli 2023. I toppen af pyramiden, kan man tydeligt se forskellen mellem mænd og kvinder: Der er flere ældre kvinder end ældre mænd. Pyramiden afspejler også historiske begivenheder som anden verdenskrig og den faldende tendens i dødeligheden og fertiliteten som følge af den demografiske transition. En mere sofistikeret og dynamisk version af den danske alderspyramide findes her <https://extranet.dst.dk/pyramide/pyramide.htm>.

```
# R-koder der genererer alderspyramiden for Danmark den 1. juli 2023
folk <- hent_data("FOLK1a", "alder"=0:125, "køn"=1:2, tid="2023K3")
# Fjern ældere, der har folketal 0
folk <- subset(folk, alder<106)
# Mænd skal vises på venstresiden, derfor bliver INDHOLD negativt
folk_m <- subset(folk, KØN=="Mænd") %>% mutate(INDHOLD=-INDHOLD)
# For kvinder på højre siden er INDHOLD positivt
folk_k <- subset(folk, KØN=="Kvinder")
# Grafik
g <- ggplot(folk, aes(x = alder, y = INDHOLD, fill = KØN)) +
  geom_bar(data=folk_m, stat = "identity") +
  geom_bar(data=folk_k, stat = "identity") +
  coord_flip() + theme_economist() +
```

```

ylab("Folketal N(t)")+xlab("Alder (år)") +
theme(legend.title=element_blank())
g <- g+ggtitle("Alderspyramide Danmark 1 juli 2023")
g <- g+theme(axis.title.y=element_text(margin=margin(t=0,r=20,b=0,l=0)))
g <- g+theme(axis.title.x=element_text(margin=margin(t=20,r=0,b=0,l=0)))
g

```

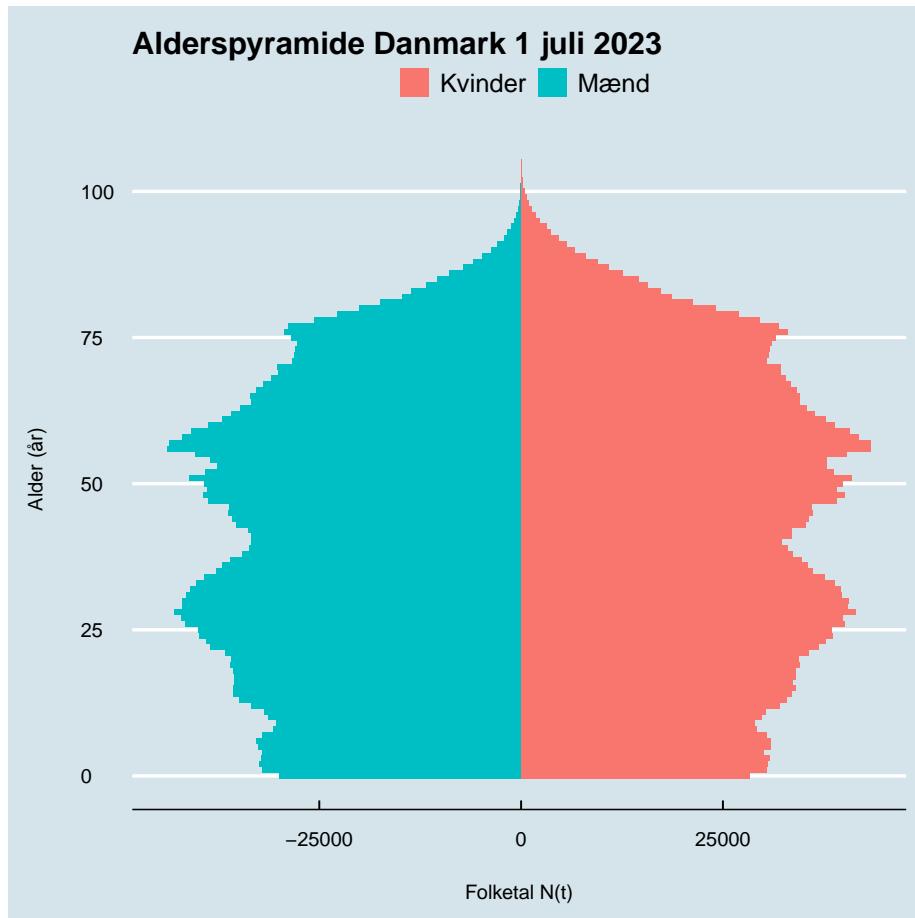


Figure 4: Data fra statistikbankens FOLK1a

3.2 Folketal i aldersgrupper

Aldersfordelingen af folketallet angiver hvor mange personer i en befolkning, der har en bestemt alder for alle aldre. Det kan den enten gøre i det abso-

lutte antal eller som procent i forhold til antal personer i hele befolkningen. For at beskrive aldersfordelinger vil man typisk vælge et passende antal aldersintervaller (passende til opgaven man sidder med) og fordele befolkningen på intervallerne. Intervallerne behøver ikke at være lige store. Da alle personers aldre ændrer sig hele tiden, skal man angive den dato, som folketallet refererer til. For eksempel kan vi tale om aldersfordelingen af kvinder i Danmark den 8. marts 1910 og om aldersfordelingen af Fyns befolkning den 1. juli 1989.

3.2.1 Eksempel

Vi finder aldersfordelingen af folketallet for hele den danske befolkning 1. januar 2023 og inddeler den i 4 intervaller: $[0, 25]$, $(25, 50]$, $(50, 75]$, $(75, 125]$. Bemærk, at vores notation for intervaller betyder, at intervalgrænsen er ekskluderet hvis parentesen er rund og inkluderet hvis parentesen er firkantet. Det vil sige, at personer, som er præcis 25 år gamle, falder i intervallet $[0, 25]$, og personer som er 50 falder ikke i intervallet $(50, 75]$, men i intervallet $(25, 50]$. Vi beregner nu andelen, som de enkelte aldersgrupper udgør og angiver den i procent (per hundrede). De fire procenttal er netop aldersfordelingen af folketal den 1. januar 2023 med hensyn til de fire intervaller.

```
# R-koder til beregning af aldersfordeling af folketal
folk <- hent_data("FOLK1a", "alder"=0:125, tid="2023K3")
# Aldersintervaller
folk <- mutate(folk, Aldersinterval=cut(alder,
                                         breaks=c(0,25,50,75,125),
                                         include.lowest = TRUE))
# antal person i de 4 aldersintervaller
af <- folk %>% group_by(Aldersinterval) %>%
  summarise(Antal=sum(INDHOLD))
# beregne procenter
af <- af %>% mutate(Procent=100*Antal/sum(Antal))
af
```

```
# A tibble: 4 × 3
  Aldersinterval    Antal   Procent
  <fct>           <dbl>    <dbl>
1 [0,25]        1742979  29.3
2 (25,50]       1882860  31.7
3 (50,75]       1778084  29.9
4 (75,125]      540222   9.09
```

3.2.2 Aldersfordeling af folketal i formler

Enhver definition af aldersintervaller opdeler en befolkning i aldersgrupper. For $x = 1, \dots, m$ aldersgrupper betegner vi med $N_x(t)$ folketallet i aldersgruppen x til kalendertid t . Vi betegner fortsat med $N(t)$ folketallet i hele befolkningen til

kalendertid t og udtrykker det som summen af folketallene i aldersgrupperne:

$$N(t) = N_1(t) + \cdots + N_m(t) = \sum_{x=1}^m N_x(t).$$

I eksemplet fra afsnit 3.2.1 er der $m = 4$ aldersgrupper og når vi indsætter tal i formlen finder vi folketallet som sum af de aldersspecifikke folketal:

$$N(1 \text{ jan } 2023) = 1742979 + 1882860 + 1778084 + 540222 = 5944145.$$

Vi beregner andelen af befolkningen i aldersgruppe x ved at dividere folketal i aldersgruppen med folketallet i hele befolkningen til tid t :

$$\frac{N_x(t)}{N(t)} = \{\text{Andel af befolkningen i aldersgruppe } x \text{ til tid } t\}.$$

Aldersfordelingen af folketal er lig med de aldersspecifikke andele af folketal, altså for en given opdeling i aldersintervaller givet ved:

$$\text{Aldersfordeling af folketal} = \left(\frac{N_1(t)}{N(t)}, \dots, \frac{N_m(t)}{N(t)} \right). \quad (\text{K2.1})$$

I eksemplet fra afsnit 3.2.1 har vi allerede beregnet aldersfordeling den 1. januar 2023 og angivet den som procent.

3.2.3 Sammenligning af aldersfordelinger

Vi sammenligner aldersfordelingen i hovedstadsområdet med aldersfordelingen i landdistrikter i Danmark i 2023. For at gøre det enkelt bruger vi inddelingen af befolkningen i de 4 aldersgrupper fra afsnit 3.2.1. Vi henter folketal fra statistikbankens register BY2, hvor man kan angive bystørrelse.

```
# R-koder til aldersfordeling af folketal for given bystørrelse
info <- register_info("BY2")
names(info)
b2 <- hent_data(register = "BY2", alder=0:125,
                  BYST=c("HOVEDS", "LAND"), tid="2023")
# Aldersintervaller
b2 <- mutate(b2, Aldersinterval=cut(alder,
                                         breaks=c(0,25,50,75,125),
                                         include.lowest = TRUE))
# Antal person i de 4 aldersintervaller
af <- b2 %>% group_by(BYST, Aldersinterval) %>%
      summarise(Antal=sum(INDHOLD))
# Procent
af <- af %>% mutate(Procent=100*Antal/sum(Antal))
af
```

# A tibble: 8 x 4	# Groups: BYST [2]		
BYST	Aldersinterval	Antal	Procent
<chr>	<fct>	<dbl>	<dbl>
1 Greater Copenhagen Region	[0,25]	424524	31.1
2 Greater Copenhagen Region	(25,50]	520217	38.2
3 Greater Copenhagen Region	(50,75]	329994	24.2
4 Greater Copenhagen Region	(75,125]	88561	6.50
5 Rural areas	[0,25]	184556	26.8
6 Rural areas	(25,50]	198151	28.8
7 Rural areas	(50,75]	258161	37.5
8 Rural areas	(75,125]	46720	6.79

En sammenligning af de to aldersfordelinger viser, at andelen af mennesker, der er over 75 år gamle, er cirka det samme, men at andelen af mennesker under 50 år er højest i hovedstadsområdet og andelen af mennesker mellem 50 og 75 år er højest i landdistrikterne.

3.3 Risikotid i aldersgrupper

For beregningen af aldersspecifikke mortalitetsrater har vi brug for aldersfordelingen af risikotiden i en bestemt kalenderperiode. Vi betegner med $R_x[t_1, t_2]$ den samlede gennemlevede tid i perioden $[t_1, t_2]$ af alle personer i aldersgruppe x . Vi bemærker, at en person, som har levet i befolkningen i perioden $[t_1, t_2]$, kan bidrage med risikotid til et eller flere aldersintervaller. Det sker for personer som har fødselsdag mellem dato t_1 og dato t_2 , hvis de den dag skifter fra aldersgruppe x til aldersgruppe $x + 1$. Vi betegner fortsæt med $R[t_1, t_2]$ risikotiden for hele befolkningen og kan nu udtrykke den som en sum af de aldersspecifikke risikotider:

$$R[t_1, t_2] = R_1[t_1, t_2] + \cdots + R_m[t_1, t_2] = \sum_{x=1}^m R_x[t_1, t_2].$$

Vi beregner andelen af risikotid i aldersgruppe x ved at dividere risikotiden i aldersgruppen med risikotiden i hele befolkningen i perioden $[t_1, t_2]$ og betegner den med V_x :

$$V_x[t_1, t_2] = \frac{R_x[t_1, t_2]}{R[t_1, t_2]} = \{\text{Andel af risikotid i aldersgruppe } x \text{ i perioden } [t_1, t_2]\}. \quad (\text{K2.2})$$

Risikotid beregnes ofte ved at gange middelfolketal med periodens længde (se Kapitel 1). I den særlige situation, hvor perioden er 1 år langt, altså når $t_2 - t_1 = 1$ år, har middelfolketallet (antal) og risikotiden (personår) den samme værdi men forskellige enheder. Vi vil senere bruge V_x som vægte i definitionen af aldersstandardiserede rater (afsnit 5).

3.3.1 Eksempel

Vi finder aldersfordelingen af risikotiden for hele den danske befolkning i perioden mellem 1. januar 2022 og 1. januar 2023 og inddeler den i fire aldersintervaller: [0, 25], (25, 50], (50, 75], (75, 125].

```
# Beregningen af aldersfordelingen af risikotiden
folk <- hent_data("FOLK1a", alder=0:125, tid=c("2022K1", "2023K1"))
# Risikotid= 1* Middelfolketal metode 2
folk <- folk %>% group_by(alder) %>%
  summarise(Risikotid=1*mean(INDHOLD))
# Aldersintervaller
folk <- mutate(folk, Aldersinterval=cut(alder,
                                         breaks=c(0,25,50,75,125),
                                         include.lowest = TRUE))
# Antal personår i de 4 aldersintervaller
af <- folk %>% group_by(Aldersinterval) %>%
  summarise(Personår=sum(Risikotid))
# Aldersfordelingen af risikotiden i procent
af <- af %>% mutate(Procent=100*Personår/sum(Personår))
af
```

```
# A tibble: 4 × 3
Aldersinterval Personår Procent
<dbl>      <dbl>    <dbl>
1 [0,25]     1747687   29.6
2 (25,50]    1867838.  31.6
3 (50,75]    1773568   30.0
4 (75,125]   513944.   8.71
```

3.4 Lexisdiagram

Et Lexisdiagram visualiserer sammenhængen mellem kalendertid (vertikal) og alder (horisontal). Hver person er repræsenteret af sin livslinje (Figur 5). I en lukket befolkning (hvor ind- og udvandring ikke forekommer) starter alle livslinjer ved fødslen, hvor personen er 0 år gammel og ender i dødsdatoen - den alder personen har levet til. I en åben befolkning starter livslinjer for immigranter den dag, de immigrerer og slutter for emigranter den dag, de emigrerer.

Figur 5 viser 5 personers livslinjer fra en åben befolkning. Den mørkeblå linje repræsenterer en person, som bliver født i foråret 2015 og forbliver i befolkningen indtil foråret 2020 hvor Lexisdiagrammet slutter. Lexisdiagrammet kan også bruges til at forklare forskellen mellem kohorteprincippet (man følger en fødselskohorte) og kalenderårsprincippet (man studerer en befolkning i en kort periode). Figur 6 viser et Lexisdiagram med skematisk forklaring til, hvordan man kan studere en befolkning i en kort kalenderperiode, følge en aldergruppe igennem kalendertid, og en fødselskohorte igennem både kalendertid og alder.

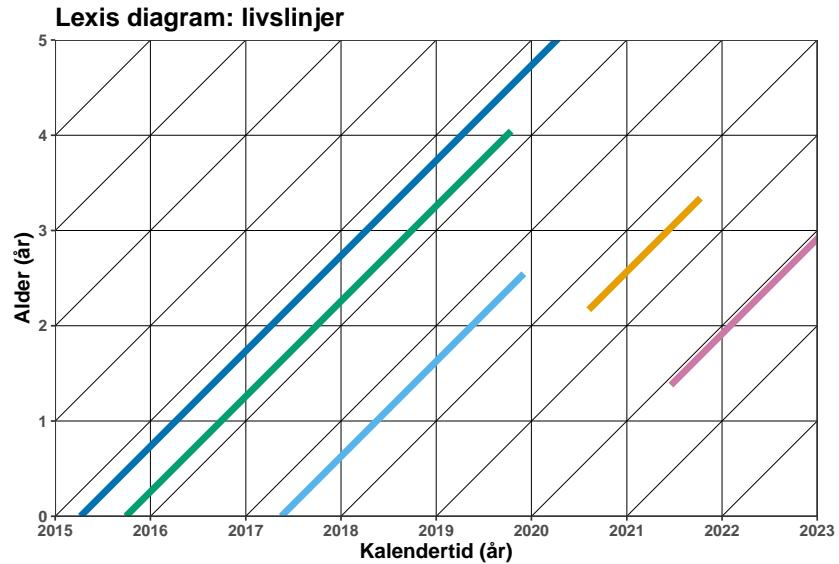


Figure 5: Figuren viser 5 personers livslinjer i (den nederste del af) et Lexis diagram. Livslinjer der ikke starter i alder '0' representerer immigranter og livslinjer som stopper representerer enten dødsfald eller emigranter.

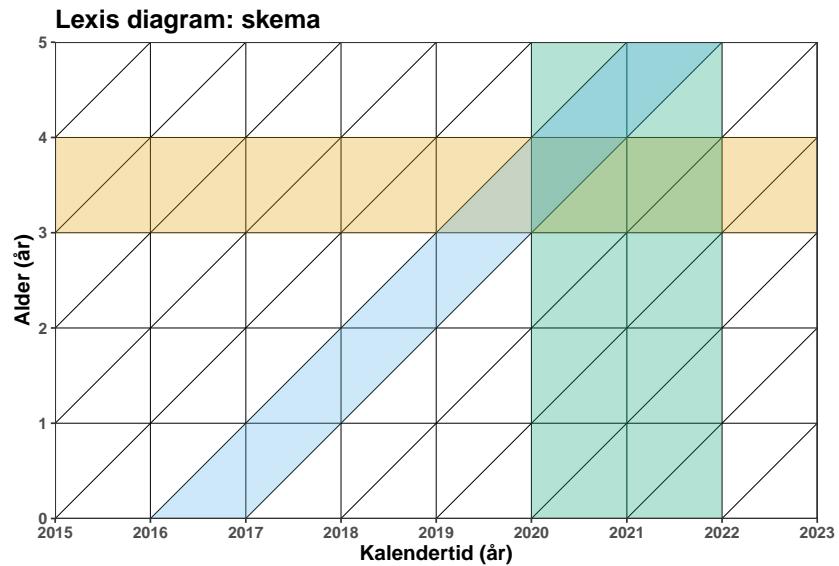


Figure 6: I et Lexis diagram kan man følge en aldersgruppe igennem kalendertid (gul) eller en fødselskohorte igennem både alder og kalendertid (blå). Det grønne område viser en kort kalenderperiode.

4 Aldersspecifikke mortalitetsrater

Vi ser på en befolkning i en kalenderperiode $[t_1, t_2]$ og inddeler den i $x = 1, \dots, m$ aldersgrupper. Vi betegner med $D_x[t_1, t_2]$ antal dødsfald i perioden hvor personens alder ved dødsdatoen falder i aldersgruppe x . For at lette notationsbyrden dropper vi kalenderperioden og forkorter $D_x[t_1, t_2]$ til D_x , og ligeledes skriver vi R_x for den aldersspecifikke risikotid $R_x[t_1, t_2]$ i samme periode. De aldersspecifikke mortalitetsrater er defineret som kvotienten mellem antal dødsfald og risikotid:

$$\text{Aldersspecifikke mortalitetsrate: } M_x = \frac{D_x}{R_x}, \quad x = 1, \dots, m. \quad (\text{K2.3})$$

Bemærk at den aldersspecifikke mortalitetsrate M_x afhænger kalenderperioden: $M_x = M_x[t_1, t_2]$.

4.1 Eksempel

Vi finder antal dødsfald for hele den danske befolkning i perioden mellem 1. januar 2022 og 1. januar 2023 og beregner det samlede antal døde i samme 4 aldersintervaller ($[0, 25], (25, 50], (50, 75], (75, 125]$) som vi har brugt i eksemplet i afsnit 3.3.1. Vi finder tal i statistikbankens DOD og bemærker at det sidste aldersinterval hedder "99 år og derover".

```
# R-koder der henter aldersfordelingen af antal døde
dd <- hent_data("dod", alder="all_no_total", tid="2022")
# Aldersintervaller
dd <- mutate(dd, Aldersinterval=cut(alder,
                                         breaks=c(0,25,50,75,125),
                                         include.lowest = TRUE))
# Antal døde i de 4 aldersintervaller
group_dd <- dd %>% group_by(Aldersinterval) %>%
  summarise(antal_døde=sum(INDHOLD))
group_dd
```

```
# A tibble: 4 × 2
  Aldersinterval antal_døde
  <fct>           <dbl>
1 [0,25]            461
2 (25,50]           1621
3 (50,75]          18194
4 (75,125]         39159
```

For at beregne de aldersspecifikke mortalitetsrater skal vi samle personår (afsnit 3.3.1) og antal døde i aldersgrupper i et **tibble**. Det gør vi med funktionen `left_join`:

```
# Fortsættelse af sidste R-chunk
x <- left_join(af,group_dd,by="Aldersinterval")
# Aldersspecifikke mortalitetsrater
x <- x %>% mutate(mrate=1000*antal_døde/Personår)
x
```

```
# A tibble: 4 × 5
  Aldersinterval Personår Procent antal_døde mrate
  <fct>          <dbl>   <dbl>      <dbl>   <dbl>
1 [0,25]         1747687  29.6       461    0.264
2 (25,50]        1867838. 31.6       1621   0.868
3 (50,75]        1773568  30.0       18194  10.3
4 (75,125]       513944.  8.71      39159  76.2
```

4.2 Sammenligning af aldersspecifikke mortalitetsrater

For at sammenligne mortalitet i to befolkninger (studiebefolkning A mod studiebefolkning B) kan man sammenligne de aldersspecifikke mortalitetsrater mellem dem (M_x^A mod M_x^B). Det giver lige så mange resultater, som der er aldersintervaller, altså et resultat for hver aldersgruppe. Hvis der er blot 4 aldersgrupper kan man på en overskulig måde vise resultater i en tabel. Men med mange aldersgrupper er det nemmere at se forskellen i en figur, som viser de aldersspecifikke mortalitetsrater af de to befolkninger ved siden af hinanden. Figur 7 viser mortalitetsrater for kvinder og mænd i Danmark i 2011 for 100 aldersgrupper hvor den sidste samler personer som er 99 år og derover. Forskellen mellem Figur 3 som viser mortalitetsrater for 2022 og Figur 7 er at vi nu bruger en logaritmisk skala.

4.2.1 Eksempel

Vi beregner aldersspecifikke mortalitetsrater for mænd og kvinder i 2011 og visualiserer forskellen.

```
# R-koder der genererer grafik som viser køns- og aldersspecifikke
# mortalitetsrater
folk <- hent_data("FOLK1a",alder=0:125,køn=c(2,1),tid="2011K3")
# Endre variable navn
folk <- folk %>% rename("risikotid"="INDHOLD")
# Samle antal personer over 99 (fordi register DOD gør det samme)
folk <- samle_alder(folk,variable = "risikotid",value = "99plus",by =
  "køn")
# Hent dødstal fra 2011
dd <- hent_data(register="dod",alder=0:99,køn=c("K","M"),tid="2011")
# Endre variable navn
dd <- dd %>% rename("antal_døde"="INDHOLD")
# Join folketal og antal dødsfald
x <- left_join(folk,dd,by=c("alder","KØN"))
```

```
# Aldersspecifikke mortalitetsrater
x <- x %>% group_by(KØN) %>% mutate(mrate=1000*antal_døde/risikotid)
# Grafik
g <- ggplot(x,aes(x=alder,y=mrate,color= KØN))+geom_line()+geom_point()
g <- g+theme_economist()+scale_colour_wsj("colors6") +theme(legend.title=element_blank())
g <- g+ylab("Mortalitetsrate per 1000 personår")+xlab("Alder (år)")+
  ggtitle("Danmark 2011")
g <- g+theme(axis.title.y=element_text(margin=margin(t=0,r=20,b=0,l=0)))
g <- g+theme(axis.title.x=element_text(margin=margin(t=20,r=0,b=0,l=0)))
g <- g+theme(plot.background=element_rect(fill="gray88",colour=NA))
g <- g+theme(legend.title=element_blank())+theme(legend.position="bottom")
g + scale_y_log10()
```

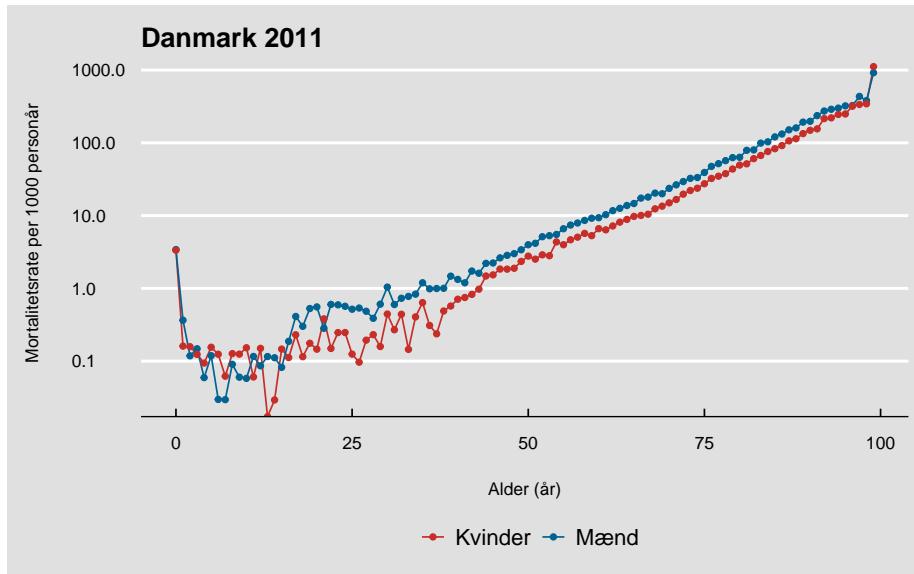


Figure 7: Aldersspecifikke mortalitetsrater for hele den danske befolkning i 2011. Vi ser at dødeligheden var højere for mænd i de fleste aldersgrupper.

5 Aldersstandardisering

Formålet med aldersstandardisering er at sammenligne mortalitetsrater (og andre rater) mellem to eller flere befolkninger, som har forskellige aldersfordelinger. Den overordnede idé er at udskifte den rigtige aldersfordeling med en anden

aldersfordeling, og at beregne mortalitetsraten som den ville have været, hvis befolkningen havde haft den anden aldersfordeling. På den måde kan man sammenligne dødelighed mellem to eller flere befolkninger uanset aldersfordeling. Her er det vigtigt, at man vælger den samme aldersfordeling for alle befolkninger, som skal sammenlignes, men det er typisk ikke så vigtigt, hvilken aldersfordeling man vælger. For eksempel kan vi spørge, hvor meget højere mortalitetsraten er blandt danske mænd sammenlignet med danske kvinder, hvis aldersfordelingen havde været den samme for mænd og kvinder. Vi mangler kun at specificere den aldersfordeling, som de standardiserede rater skal have til fælles. Her er der umiddelbart flere forskellige muligheder: aldersfordeling blandt mænd, aldersfordeling blandt kvinder, aldersfordeling blandt alle danskere uanset køn eller en helt anden aldersfordeling.

Vi beskriver to standardiseringsformer, *direkte standardisering* (afsnit 5.2) og *indirekte standardisering* (afsnit 5.3). Vi starter med en matematisk forklaring af resultatet fra afsnit 2.1 (afsnit 5.1) og slutter med en sammenligning af direkte og indirekte standardisering (afsnit 5.4).

5.1 Kitagawas dekomposition

For en given inddeling af en befolkning i aldersgrupper i en periode $[t_1, t_2]$ er dens summariske mortalitetsrate et vægtet gennemsnit af de aldersspecifikke mortalitetsrater. For at indse dette skal vi bruge aldersfordelingen af risikotid, som vi har indført i afsnit 3.3. For aldersgruppe x er andelen af risikotid givet som

$$V_x = \frac{R_x}{R}, \quad (\text{K2.4})$$

hvor R betegner befolkningens totale risikotid i perioden, lige som i formel (K2.2). Vi omskriver formlen for den aldersspecifikke mortalitetsrate (K2.3) sådan, at antal dødsfald i aldersgruppen står isoleret:

$$D_x = M_x R_x. \quad (\text{K2.5})$$

Vi betegner fortsat M som befolkningens summariske mortalitetsrate og D som antal dødsfald i hele befolkningen i perioden. Det følgende regnestykke viser, at den summariske mortalitetsrate er et vægtet gennemsnit af de alder-

sspecifikke mortalitetsrater, hvor vægtene er lig med aldersfordelingen af risikotid:

$$\begin{aligned}
 M &= \frac{D}{R} \\
 &= \frac{D_1 + D_2 + \dots + D_m}{R} \\
 &= \frac{M_1 R_1 + M_2 R_2 + \dots + M_m R_m}{R} \\
 &= M_1 \frac{R_1}{R} + M_2 \frac{R_2}{R} + \dots + M_m \frac{R_m}{R}, \\
 &= M_1 V_1 + M_2 V_2 + \dots + M_m V_m \\
 &= \sum_{x=1}^m M_x V_x.
 \end{aligned} \tag{K2.6}$$

I afsnit 2.1 har vi diskuteret, at forskellen mellem kvinders og mænds summariske mortalitetsrater skyldes ikke kun kønsforskellen af mortalitetsrater men også kønsforskellen af aldersfordelinger. Kitagawas dekomposition viser dette klart og mere generelt som matematisk formel. I stedet for det specifikke valg, kvinder og mænd, skal vi skrive formlen i abstrakt form for en *studiebefolkning A* og en *studiebefolkning B*. Vi kan anvende formel (K2.6) og skrive de to summariske mortalitetsrater som

$$M^A = \sum_{x=1}^m M_x^A V_x^A \text{ og } M^B = \sum_{x=1}^m M_x^B V_x^B$$

hvor V_x^A og V_x^B er aldersfordelinger af risikotid fra henholdsvis studiebefolkning A og studiebefolkning B. Kitagawas dekomposition beskriver forskellen mellem to summariske mortalitetsrater:

$$\begin{aligned}
 M^A - M^B &= \sum_{x=1}^m M_x^A V_x^A - \sum_{x=1}^m M_x^B V_x^B \\
 &= \sum_{x=1}^m (M_x^A V_x^A - M_x^B V_x^B) \\
 &= \underbrace{\sum_{x=1}^m (M_x^A - M_x^B) \frac{V_x^A + V_x^B}{2}}_{\text{Komponent 1}} + \underbrace{\sum_{x=1}^m (V_x^A - V_x^B) \frac{M_x^A + M_x^B}{2}}_{\text{Komponent 2}}
 \end{aligned} \tag{K2.7}$$

Her beskriver komponent 1 forskellen mellem de aldersspecifikke mortalitetsrater vægtet med de gennemsnitlige andele af risikotid og komponent 2 forskellen mellem aldersfordelingerne vægtet med de gennemsnitlige mortalitetsrater. Det kræver lidt algebra, hvis man vil indse, hvorfor formel (K2.7) (Kitagawas dekomposition) holder: For hvert aldersinterval x gælder:

$$\begin{aligned}
(M_x^A V_x^A - M_x^B V_x^B) &= \frac{(M_x^A V_x^A - M_x^B V_x^B) + (M_x^A V_x^A - M_x^B V_x^B)}{2} \\
&= \frac{M_x^A V_x^A}{2} - \frac{M_x^B V_x^B}{2} + \frac{M_x^A V_x^A}{2} - \frac{M_x^B V_x^B}{2} \\
&= \frac{M_x^A V_x^A}{2} - \frac{M_x^B V_x^B}{2} + \frac{M_x^A V_x^A}{2} - \frac{M_x^B V_x^B}{2} \\
&\quad + \left(\frac{M_x^A V_x^B}{2} - \frac{M_x^A V_x^B}{2} \right) + \left(\frac{M_x^B V_x^A}{2} - \frac{M_x^B V_x^A}{2} \right) \\
&= \frac{M_x^A V_x^A}{2} + \frac{M_x^A V_x^B}{2} - \frac{M_x^B V_x^A}{2} - \frac{M_x^B V_x^B}{2} \\
&\quad + \frac{M_x^A V_x^A}{2} + \frac{M_x^B V_x^A}{2} - \frac{M_x^A V_x^B}{2} - \frac{M_x^B V_x^B}{2} \\
&= (M_x^A - M_x^B) \frac{V_x^A + V_x^B}{2} + (V_x^A - V_x^B) \frac{M_x^A + M_x^B}{2}.
\end{aligned} \tag{K2.8}$$

Da ligning (K2.8) holder for hver aldersgruppe x , holder den også for summen over alle aldersgrupper, og det giver formel (K2.7).

5.2 Direkte standardisering

Formålet med den såkaldte *direkte* standardisering er at sammenligne mortalitetsrater mellem to befolkninger uanset forskelle i aldersfordeling. Direkte standardisering er en metode, som sammenfatter forskellen af mortalitetsrater med et tal der benævnes *standardiserede rate ratio*. For at definere direkte standardisering skal vi bruge to studiebefolkninger (A mod B) og en standardbefolkning (S). Ideen er at se, om studiebefolkning A har højere mortalitet end studiebefolkning B , hvis (hypotetisk) begge havde samme aldersfordeling som standardbefolkning S . Vi fortolker den direkte standardiserede mortalitetsrate

$$\tilde{M}_S^A = \sum_{x=1}^m M_x^A V_x^S, \tag{K2.9}$$

som den mortalitetsrate vi ville have set i studiebefolkning A , hvis aldersfordeling af risikotid havde været den samme som i standardbefolkning S . Ideen er, at vi nu kan direkte sammenligne de standardiserede mortalitetsrater fra studiebefolkninger A og B . Den hyppigste form, der afrapporterer denne sammenligning mellem to standardiserede mortalitetsrater, er den *standardiserede rate ratio* (SRR):

$$\text{SRR}(A, B, S) = \frac{\sum_{x=1}^m M_x^A V_x^S}{\sum_{x=1}^m M_x^B V_x^S}. \tag{K2.10}$$

5.2.1 Eksempel

For at illustrere idéen med at bruge den direkte standardisering fortsætter vi med eksemplet fra afsnit 2.1.1, som viste, at mænd havde en lavere summarisk mortalitetsrate end kvinder i 2011. Vi anvender direkte standardisering, hvor vi bruger alle danske kvinder i 2011 som studiebefolkning A , alle danske mænd i 2011 som studiebefolkning B , og vælger alle danskere i 2011 (uanset køn) som standardbefolkning S .

```
# R-koder der beregner direkte standardisering
af <- hent_data("FOLK1a","alder"=0:125,tid="2011K3",
                 køn = c(1,2,"TOT"))
# Fordeling af risikotid i aldersintervaller
af <- rename(af,R = INDHOLD)
af <- intervAlder(af,breaks=c(0,25,50,75,125),by="KØN",vars="R")
# Antal døde i aldersintervaller
dd <- hent_data("DOD","alder"="all_no_total",tid="2011",køn=c("M","K"
    ))
dd <- rename(dd,D = INDHOLD)
dd <- intervAlder(dd,breaks=c(0,25,50,75,125),by="KØN",vars="D")
# Aldersspecifikke mortalitetsrater
# Kvinder
A <- left_join(filter(af,KØN == "Kvinder"),
                  filter(dd, KØN == "Kvinder"),
                  by = c("KØN","aldersinterval"))
A <- mutate(A,M = 1000*D/R)
A <- select(A,aldersinterval,M)
# Mænd
B <- left_join(filter(af,KØN == "Mænd"),
                  filter(dd, KØN == "Mænd"),
                  by = c("KØN","aldersinterval"))
B <- mutate(B,M = 1000*D/R)
B <- select(B,aldersinterval,M)
# Aldersfordeling i standardbefolkning
S <- select(filter(af,KØN == "I alt"),!KØN)
S <- mutate(S,V=(R/sum(R)))
# Join
A <- left_join(A,S,by = "aldersinterval")
B <- left_join(B,S,by = "aldersinterval")
# Direkte standardisering
tibble("srate_kvinder" = pull(summarise(A,sum(M * V))),
       "srate_maend" = pull(summarise(B,sum(M * V))))
```

```
# A tibble: 1 × 2
  srate_kvinder srate_maend
  <dbl>        <dbl>
1     8.35       10.7
```

Resultatet fortolkes på følgende måde: Den standardiserede mortalitetsrate i Danmark i 2011 var 8,4 døde per 1000 personår for kvinder og 10,7 døde per

1000 personår for mænd, hvis aldersfordelingen havde været den samme som den var blandt alle danskere i 2011. Dette bekræfter, at grunden til, at den summariske mortalitetsrate var højere for mænd end for kvinder i 2011, var forskelle i aldersfordeling.

5.3 Indirekte standardisering

Formålet med den såkaldte *indirekte* standardisering er at sammenligne mortalitetsrater i studiebefolkning A med mortalitetsrater i en standardbefolkning S . Konkret sammenligner man det totale antal dødsfald i studiebefolkning A med det forventede antal døde i studiebefolkning A , hvis (hypotetisk) de aldersspecifikke mortalitetsrater havde de samme som i standardbefolkning S . Er det forventede antal dødsfald højere, kan man konkludere, at den samlede dødelighed (det vil sige de aldersspecifikke mortalitetsrater samlet set) har været højere i standardbefolkningen end i studiebefolkning A .

Beregningen af indirekte standardisering kræver kendskab til de aldersspecifikke mortalitetsrater i standardbefolkning S , de aldersspecifikke risikotider i studiebefolkning A og det totale antal dødsfald i studiebefolkning A . For $x = 1, \dots, m$ aldersgrupper er det totale antal dødsfald i studiebefolkning A givet ved

$$D^A = D_1^A + \dots + D_m^A = \sum_{x=1}^m D_x^A = \sum_{x=1}^m M_x^A R_x^A.$$

Her har vi brugt formel (K2.5). Relativt til den totale risikotid R^A er det forventede antal døde (hvis dødeligheden havde været den samme som den var i standardbefolkning S) givet ved

$$\sum_{x=1}^m M_x^S V_x^A = \sum_{x=1}^m M_x^S \frac{R_x^A}{R^A} = \frac{1}{R^A} \sum_{x=1}^m M_x^S R_x^A.$$

En sammenligning af mortalitetsrater mellem studiebefolkning A og standardbefolkning S er den såkaldte *standardiserede mortalitetsrateratio*:

$$\begin{aligned} \text{SMR}(A, S) &= \frac{\text{Observeret antal døde}}{\text{Forventet antal døde}} \\ &= \frac{\sum_{x=1}^m M_x^A V_x^A}{\sum_{x=1}^m M_x^S V_x^A} \\ &= \frac{\sum_{x=1}^m M_x^A R_x^A}{\sum_{x=1}^m M_x^S R_x^A} \\ &= \frac{\sum_{x=1}^m D_x^A}{\sum_{x=1}^m M_x^S R_x^A} \end{aligned} \tag{K2.11}$$

Den indirekte standardiserede mortalitetsrate i befolkning A er givet ved

$$\bar{M}_S^A = \text{SMR}(A, S) \cdot M^S = \frac{\sum_{x=1}^m M_x^A V_x^A}{\sum_{x=1}^m M_x^S V_x^A} \sum_{x=1}^m M_x^S V_x^S. \tag{K2.12}$$

5.4 Direkte versus indirekte standardisering

Typisk ønsker man at sammenligne flere studiebefolkninger, og man vil derfor beregne en standardiseret rate for hver studiebefolkning, idet den samme standardbefolkning bruges i alle beregninger. Dødeligheden i de forskellige befolkninger kan så sammenlignes ved at sammenligne de standardiserede rater. Direkte standardisering kræver kendskab til aldersfordelingen af risikotiden i standardbefolkningen. Kender man ikke aldersfordelingen fra standardbefolknigen, kan man ikke anvende direkte standardisering og dermed ikke beregne SRR. Hvis man til gengæld kender de aldersspecifikke mortalitetsrater i standardbefolkningen, kan man beregne SMR. Man kan dog umiddelbart ikke direkte sammenligne SMR for to studiebefolkninger A og B , fordi $\text{SMR}(A, S)$ afhænger aldersfordeling af studiebefolkning A og $\text{SMR}(B, S)$ afhænger aldersfordeling af studiebefolkning B . Dette problem har direkte standardisering $\text{SRR}(A, B, S)$ ikke. Derfor vil man bruge direkte standardisering i en situation hvor man kan vælge mellem direkte og indirekte standardisering.

For at foretage en direkte eller indirekte standardisering skal man først have valgt en standardbefolkning. Valget vil typisk afspejle problemstillingen. Hvis vi skal sammenligne dødeligheden i forskellige lande i Europa, vil det være naturligt at vælge hele Europa som standardbefolkning. Hvis vi vil sammenligne dødeligheden i forskellige erhvervsgrupper, vil det være naturligt at bruge alle erhvervsaktive som standardbefolkning. Det kan også være rimeligt at vælge en af studiebefolkningerne eller de to studiebefolkninger sat sammen som standardbefolkning. I det specielle tilfælde, hvor man bruger en af to studiebefolkninger som standardbefolkning, vil SRR fra direkte standardisering og SMR fra en indirekte standardisering give det samme resultat:

$$\text{SRR}(A, B, A) = \frac{\sum_{x=1}^m M_x^A V_x^A}{\sum_{x=1}^m M_x^B V_x^A} = \text{SMR}(A, B).$$

Kapitel 3: Overlevelsestavler

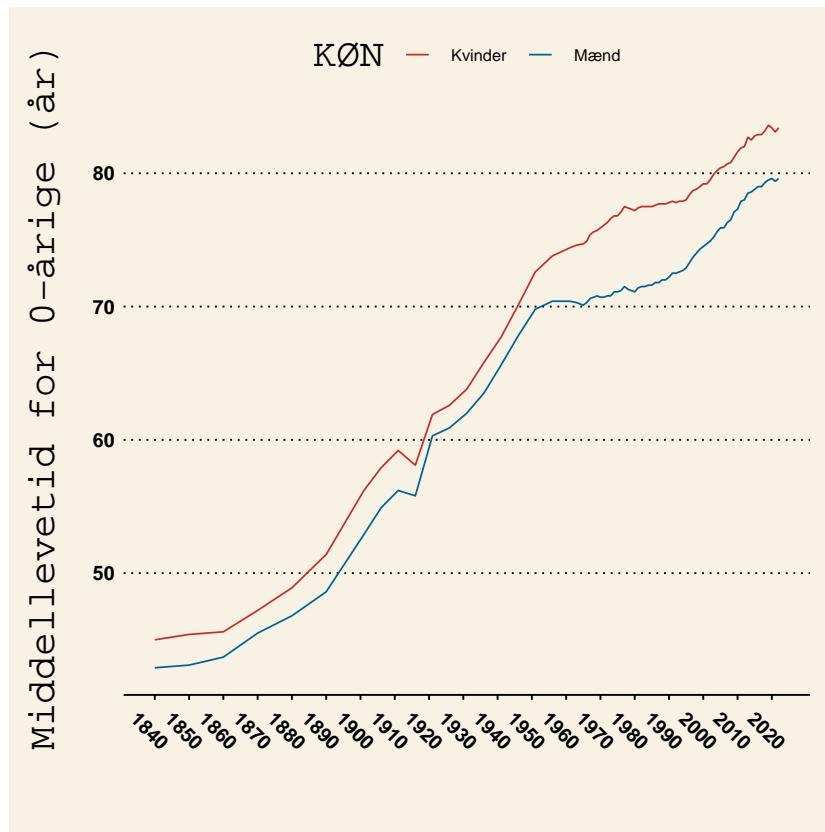
1 Introduktion

Overlevelsestavlen repræsenterer en matematisk model, der beskriver forskellige dødelighedsmål ved hjælp af konkrete demografiske data. Modellen genererer derefter en omfattende beskrivelse af dødelighedsforholdene i den specifikke befolkning. De forskellige mål for dødelighed konstrueres på baggrund af overlevelsestavlen. Lignende metoder bruges til målene vedrørende forekomsten af vielser, skilsmisses, folkevandringer og i vis grad forskellige fertilitets- og reproduktionsmål. Overlevelsestavlen kunne derfor gennemgås på en ret abstrakt måde og fortolkes forskelligt, afhængigt af om den skal anvendes til at beskrive dødelighed, vielser, skilsmisses eller fertilitet. I det følgende vil vi dog fokusere på at opbygge modellen omkring målingen af befolkningens dødelighed for at gøre det lettere at forstå modellens umiddelbare anvendelighed.

1.1 Middellevetid

Hvor mange år kan en nyfødt i dag forvente at leve? Dette spørgsmål er umuligt at besvare korrekt, fordi svaret umiddelbart afhænger af, hvad der sker i fremtiden. Alligevel er middellevetid, altså den forventede gennemsnitlige levetid af en nyfødt, et demografisk værktøj, som anvendes hyppigt til belysning af befolkningens nuværende dødelighedsniveau. Middellevetid bruges også som sammenligningsgrundlag på tværs af befolkninger og tid. Tallet angiver det antal år, som en nyfødt kan forvente at leve *under den forudsætning*, at de nuværende mortalitetsrater for alle grupperinger af køn og alderstrin holder sig på samme niveau i fremtiden. Med middellevetiden har man et relativt simpelt begreb, som gør det muligt at sammenligne forskellige befolkningers dødelighed. I praksis vil de nuværende dødshyppigheder formentlig ikke holde sig på samme niveau i fremtiden, så det skal man tage højde for når man fortolker middellevetiden.

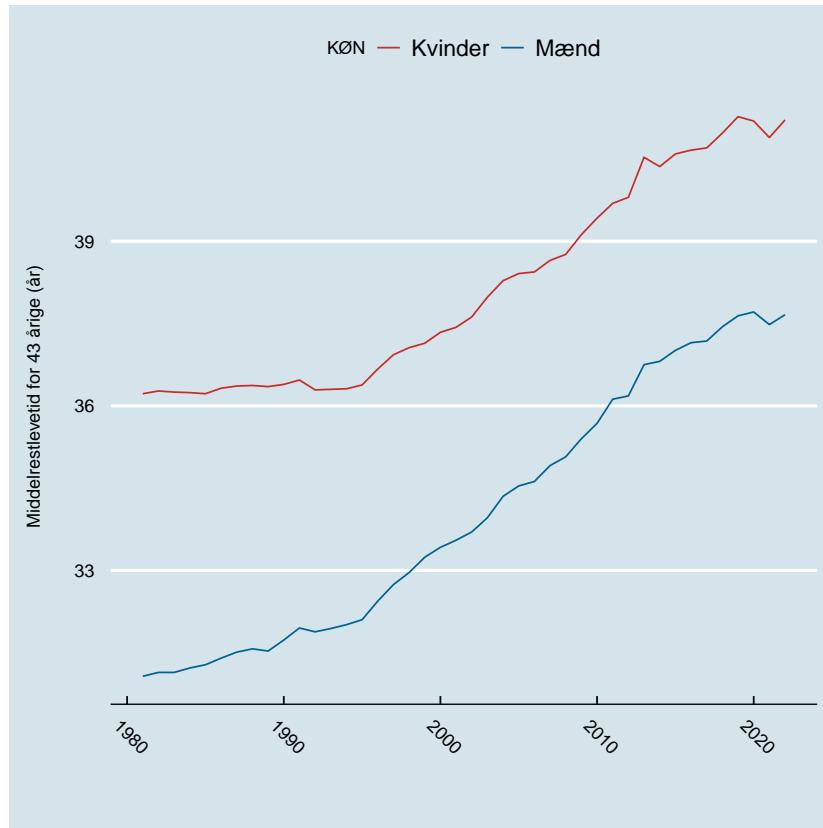
Igennem mange år har der været en tendens til faldende mortalitetsrater, og der er meget, som tyder på, at det er en udvikling som fortsætter. Den konkrete fortolkning af middellevetiden for 0-årige som det gennemsnitlige antal år, som en nyfødt kan forventes at leve, vil derfor formentlig undervurdere den faktiske middellevetid. Men formålet med middellevetiden er heller ikke at forudsige præcist, hvor længe en nyfødt vil leve. Formålet er at have et simpelt begreb, der kan sammenlignes på tværs af befolkninger og tid. Figur 1 viser udviklingen af middellevetid for 0-årige i Danmark siden 1840.



Figur 1: Udviklingen i middellevetid for 0-årigé. Kilde: statistikbankens HISB7.

1.2 Andre dødelighedsmål

Middellevetiden er måske det vigtigste mål, som resulterer af overlevelsestavlen. En overlevelsestavle beskriver også en række andre dødelighedsmål, såsom den forventede restlevetid fra alder x , sandsynligheden for at dø inden alder x og sandsynligheden for at være i live ved alder x . Figur 1 viser middelrestlevetiden for 43-årige i Danmark siden 1981.



Figur 2: Udviklingen i middelrestlevetid for 43-årige. Kilde: statistikbankens HISB8.

1.3 Eksempel

Vi henter data fra statistikbankens register FOLK1a og FOD207 og beregner aldersspecifikke mortalitetsrater for kvinder i Danmark i 2019. Vi inddeler i 12 aldersintervaller, hvor det første interval har længde 1 år, det andet interval har længde 9 år, resten af intervallerne har længde 10 år, og det sidste aldersinterval er fra 90 til 125 år.

```
x <- hent_mortalitetsrate_data(tid = 2019,
                                breaks = c(0,1,10,seq(20,90,10),Inf),
                                køn = "kvinder")
x <- mutate(x,M = Dod/R)
x
```

```
# A tibble: 11 × 6
  aldersinterval KØN     TID     R   Dod      M
  <fct>       <chr>    <dbl>  <dbl> <dbl>    <dbl>
1 0            Kvinder 2019 29448 74 0.00251
2 1-9          Kvinder 2019 270111 24 0.0000889
3 10-19         Kvinder 2019 332202 32 0.0000963
4 20-29         Kvinder 2019 383578 73 0.000190
5 30-39         Kvinder 2019 336414 128 0.000380
6 40-49         Kvinder 2019 378914 342 0.000903
7 50-59         Kvinder 2019 397594 1160 0.00292
8 60-69         Kvinder 2019 336747 2855 0.00848
9 70-79         Kvinder 2019 293474 6016 0.0205
10 80-89        Kvinder 2019 129929 8878 0.0683
11 90+          Kvinder 2019 32094 6921 0.216
```

Med disse tal fra den rigtige befolkning konstruerer vi overlevelsestavlen, som beskriver dødeligheden i en hypotetisk befolkning, der bliver født i 2019 og lever hele deres “liv” igennem alle alderstrin i 2019, hvor de bliver utsat for mortalitetsraterne fra 2019.

```
x <- mutate(x,a = c(0.1,4.5,rep(5,9)),k = c(1,9,rep(10,9)))
tavle_kvinder <- overlevelsestavle(x,
                                     mortalitet = "M",
                                     alder = "aldersinterval")
print(tavle_kvinder,digits = 2)
```

```
# A tibble: 11 × 9
  Alder     l     d     p     q     o     L     T     e
  <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 0      100000 251 0.997 0.00251 1 99774 8334430. 83.3
2 1-9    99749  80 0.999 0.000799 0.997 897385 8234656. 82.6
3 10-19   99670  96 0.999 0.000963 0.997 996216 7337271. 73.6
4 20-29   99574 189 0.998 0.00190 0.996 994789 6341056. 63.7
5 30-39   99384 377 0.996 0.00380 0.994 991955 5346266. 53.8
6 40-49   99007 890 0.991 0.00899 0.990 985620 4354311. 44.0
7 50-59   98117 2821 0.971 0.0288 0.981 967065 3368691. 34.3
8 60-69   95296 7751 0.919 0.0813 0.953 914204 2401626. 25.2
9 70-79   87545 16278 0.814 0.186 0.875 794062 1487422. 17.0
10 80-89  71267 36296 0.491 0.509 0.713 531192 693360. 9.73
11 90+    34971 34971 0     1     0.350 162168 162168. 4.64
```

Tabel 1: Forklaring af kolonner i en overlevelsestavle

Kolonne	Betydning
Alder	Aldersinterval
1	Dekrementfunktion: Antal tabelpersoner i starten af intervallet
d	Antal døde i intervallet
p	Sandsynlighed for at overleve i intervallet givet i live ved intervallets start
q	Dødshyppighed: sandsynlighed for at dø i intervallet givet i live ved intervallets start
o	Sandsynlighed for at overleve indtil starten af intervallet
L	Samlet risikotid i intervallet
T	Samlet levetid fra starten af intervallet
e	Middelrestlevetid (i første interval = middellevetid)

Fra overlevelsestavlen aflæser vi af kolonne **e**: under antagelsen af, at mortalitetsraterne i 2019 ikke ændrer sig i al fremtid, vil man forvente, at en nyfødt pige lever 83,3 år og at en kvinde som er 30 år gammel kan forvente at leve 53,8 år.

2 Konstruktion af overlevelsestavler

Overlevelsestavler beskriver, hvordan en tænkt lukket fødselskohorte reduceres med stigende alder alene på grund af dødsfald. Fordi kohorten er lukket, er død den eneste mulige afgang fra kohorten. Der tages udgangspunkt i en fiktiv tabelbefolkning bestående af ℓ_0 personer, som antages at være født på nøjagtig samme tidspunkt. Antallet af fiktive tabelpersoner ℓ_0 kaldes for ‘radix’, og radix sættes typisk til $\ell_0 = 100.000$.

2.1 Dekrementfunktionen

Funktionen ℓ_x angiver, hvor mange tabelpersoner som stadigvæk er i live ved alder x og beskriver, hvordan tabelbefolkningen reduceres på grund af dødsfald. Startværdien ℓ_0 angiver, hvor mange tabelpersoner, der er i tabelbefolkningen helt i begyndelsen, hvor alder er lig med 0, og ℓ_{30} angiver hvor mange tabelpersoner som er i live ved alder 30. Fordi ℓ_x er monoton faldende som funktion af alder, det vil sige, at der gælder $\ell_x \geq \ell_{x+1}$, kaldes man den for dekrementfunktionen. Af tabellen kan man aflæse, hvor mange personer som forventes at overleve til en bestemt alder. For eksempel betyder $\ell_{30} = 99.345$, at 99.345 personer ud af $\ell_0 = 100.000$ tabelpersoner stadigvæk er i live ved alder 30. I dette eksempel er overlevelsessandsynligheden i tabelbefolkningen ved alder 30 lig med

$$o(30) = \frac{\ell_{30}}{\ell_0} = \frac{99.345}{100.000} = 99,3\%,$$

eftersom overlevelsесfunktionen er defineret som

$$o_x = \frac{\ell_x}{\ell_0}.$$

Under konstruktionen af overlevelsestavler er opgaven at beregne dekrementfunktionens værdier ℓ_x for alle alderstrin $x = 0, 1, \dots, x^{max}$ hvor x^{max} er det sidste alderstrin. Per konstruktion dør alle resterende tabelpersoner i det sidste alderstrin - det vil sige $\ell_{x^{max}+1} = 0$ og dermed også $o_{x^{max}+1} = 0$. Vi vil forklare hvorfor senere.

2.2 Dødshyppigheder

Dødshyppigheden kq_x beskriver for en person med eksakt alder x sandsynligheden for at dø inden alderen $x + k$. Dødshyppigheder forbinder den ægte, åbne befolkning, som man interesserer sig for, med den tænkte, lukkede tabelbefolkning, der definerer overlevelsestavlen. Man beregner dødshyppigheder baseret på aldersspecifikke mortalitetsrater, og den underliggende idé er, at mortalitetsraterne er ens i den ægte befolkning og i tabelbefolkningen for begge køn og alle alderstrin.

Bemærkning til notation:

Det er standardnotation i demografi at have indeks på begge sider af symbolet ligesom i kq_x . Her er indeks til højre startalderen og indeks til venstre er antal år, som tælles med inklusive startalderen. Det er lidt forvirrende, fordi intervallet inkluderer startalderen x :

Symbol	Start	Længden	Slut	Betydning
$_1D_0$	0		1	Antal døde i alder 0
$_4D_1$	1		4	Antal døde i alder 1, 2, 3, 4
$_5D_5$	5		9	Antal døde i alder 5, 6, 7, 8, 9

Vi ændrer nu også notationen for de aldersspecifikke mortalitetsrater. I Kapitel 2 har vi brugt M_x for mortalitetsraten i det x -te aldersinterval. Fra nu af bruger vi den mere præcise betegnelse kM_x for mortalitetsraten i det aldersinterval, som starter i alderen x og slutter i alderen $x + k - 1$.

2.2.1 Approksimationsformlen

For at beregne dødssandsynligheder i den ægte befolkning vil man gerne dividere antal dødsfald i en kalenderperiode med antal personer i starten af perioden. Problemet er, at den ægte befolkning er *åben*: Dødsfald bliver ikke registreret for personer, som udvander i perioden, og både udvandrere og indvandrere i perioden bidrager ikke med risikotid til hele perioden. Ideen er derfor at tilnærme dødshyppighederne baseret på mortalitetsrater. Altersspecifikke mortalitetsrater kan beregnes på de registrerede data, ved at dividere antal dødsfald i befolkningen med risikotiden, hvor indvandrere og udvandrere kun bidrager med den tid de har været i befolkningen (se Kapitel 1 og 2). Nøglen til en tilnærmelse af

dødshyppighederne baseret på mortalitetsrater er følgende centrale formel for overlevelsestavlen:

$${}_k q_x = \frac{k \cdot {}_k M_x}{1 + (k - {}_k a_x) \cdot {}_k M_x}. \quad (\text{K3.1})$$

Formlen afhænger aldersspecifikke mortalitetsrater ${}_k M_x$, længden af aldersintervallet k og også en konstant ${}_k a_x$, som kaldes Chiang's a ¹. Konstanten ${}_k a_x$ beskriver den gennemsnitlige levetid i aldersintervallet for personer, der døde mellem alderen x og alderen $x + k - 1$. Dermed beskriver $(k - {}_k a_x)$ den gennemsnitlige tid, som en person, der døde i aldersintervallet, var død. Hvis vi for eksempel ser på et aldersinterval mellem 70 og 79 år, og en person døde i alder 74, så har den person været i live i 4 år (70, 71, 72, 73) og død i 6 år (74, 75, 76, 77, 78, 79). En person som døde i alder 78 har været i live i 8 år og død i 2 år, og så videre. Værdien af ${}_{10} a_{70}$ skal afspejle det gennemsnitlige antal år, som personer, der døde i denne aldersgruppe, var i live. For de fleste intervaller vil man antage, at gennemsnittet ligger i midten, altså i eksemplet vil man vælge ${}_{10} a_{70} = 5$.

2.2.2 Chiang's a

For at beregne dødshyppigheder med den centrale formel (K3.1) har vi brug for at specificere Chiang's a for alle aldersintervaller. Chiang's a skal tilnærme det forventede antal år levet i intervallet af en person, som dør i intervallet. Hvis Chiang's a opfylder dette, kan vi tilnærme den samlede dødstid, som alle personer, der døde i aldersintervallet, har været døde:

Samlede dødstid i aldersintervallet = $(k - {}_k a_x) \cdot {}_k D_x$,

k = Antal år i aldersintervallet

${}_k D_x$ = Antal døde i aldersintervallet

$k - {}_k a_x$ = Gennemsnitlige antal dødsår i intervallet

$\{x, x + 1, \dots, x + k - 1\}$ = Aldre i intervallet.

Hvis vi antager at dødstider er lige fordelt i aldersintervallet, altså at det er lige så sandsynligt at dø i starten, som det er at dø i slutningen af aldersintervallet, er det rimeligt at vælge

$${}_k a_x = \frac{k}{2}.$$

Det første og sidste aldersinterval vil dog altid kræve særlige værdier af ${}_k a_x$. I det første leveår er dødstiderne meget skævt fordelt over året - de fleste dødstider inden 1-års fødselsdagen ligger kort efter fødslen. Derfor sætter vi ${}_{10} a_0 = 0,1$. For det sidste interval x^{max} vælger vi

$${}_\infty a_{x^{max}} = \frac{1}{{}_\infty M_{x^{max}}}, \quad (\text{K3.2})$$

¹Chin Long Chiang (1984). The Life table and its applications. Malabar, Fla. : Krieger

så dødshyppigheden i det sidste interval bliver 1, og det betyder, at alle tabelpersoner dør i det sidste aldersinterval, dvs. $\infty q_{x^{max}} = 1$. ved formel (K3.1).

Tabel 2: Tabellen viser, hvordan vi vælger Chiang's a for 1-års, 5-års og 10-års aldersintervaller.

	5-års aldersintervaller	10-års aldersintervaller
Første leveår	${}_1a_0 = 0,1$	${}_1a_0 = 0,1$
Aldersinterval 1-5 år	${}_4a_1 = 4 \cdot 0,5 = 2$	${}_9a_1 = 9 \cdot 0,5 = 4,5$
Alle andre intervaller	${}_ka_5 = 5 \cdot 0,5 = 2,5$	${}_ka_{10} = 10 \cdot 0,5 = 5$
Sidste aldersinterval	$a_{x^{max}} = \frac{1}{\infty M_{x^{max}}}$	$a_{x^{max}} = \frac{1}{\infty M_{x^{max}}}$

2.2.3 Forklaring af den centrale formel

I det følgende skal vi på en uformel måde forklare formel (K3.1). Hvis den ægte befolkning var lukket, altså uden forekomst af ind- og udvandring, ville man kunne beregne dødshyppighederne simpelt som antal dødsfald i aldersintervallet divideret med antal personer i starten af aldersintervallet:

$$\text{Dødshyppighed} = \frac{\text{Antal dødsfald i aldersintervallet}}{\text{Antal personer i starten}}.$$

Hvis aldersintervallet er over k år gælder

$$\text{Antal personer i starten} = \frac{\text{Risikotid} + \text{Dødstid}}{k}.$$

Her er risikotiden det samlede antal år, som befolkningens personer har levet (i aldersintervallet), og dødstiden er tilsvarende det samlede antal år, som befolkningens personer var døde. Med denne formel kan dødshyppigheden skrives som

$$\text{Dødshyppighed} = \frac{k \cdot \text{Antal dødsfald i aldersinterval}}{\text{Risikotid} + \text{Dødstid}}. \quad (\text{K3.3})$$

Vi sætter Chiang's a sådan, at

$$\text{Dødstid i aldersinterval} = (k - {}_ka_x) \cdot {}_kD_x$$

er en god tilnærming af den samlede dødstid, som alle personer der døde i aldersintervallet, har været døde (cf., afsnit 2.2.2). Hvis vi nu anvender formlen for den aldersspecifikke mortalitetsrate fra Kapitel 2,

$${}_kM_x = \frac{{}_kD_x}{{}_kR_x},$$

ser vi at den centrale formel (K3.1) faktisk er lig med formel (K3.3):

$$\begin{aligned} \frac{k \cdot {}_k M_x}{1 + (k - {}_k a_x) \cdot {}_k M_x} &= \frac{k \cdot \frac{{}_k D_x}{{}_k R_x}}{1 + (k - {}_k a_x) \cdot \frac{{}_k D_x}{{}_k R_x}} \\ &= \frac{k \cdot {}_k D_x}{{}_k R_x \cdot (1 + (k - {}_k a_x) \cdot \frac{{}_k D_x}{{}_k R_x})} \\ &= \frac{k \cdot {}_k D_x}{{}_k R_x + (k - {}_k a_x) \cdot {}_k D_x}. \end{aligned}$$

2.2.4 Beregningen af antal dødsfald og overlevelser

Vi fortsætter nu konstruktionen af overlevelsestavlen. Vi starter med en radix af ℓ_0 tabelpersoner. For at beregne antal tabelpersoner som overlever indtil det første alderstrin, $x = 1$, skal vi beregne, hvor mange tabelpersoner som dør mellem alder $x = 0$ og alder $x = 1$. For at beregne hvor mange tabelpersoner, der overlever indtil alder $x + k$, skal vi beregne, hvor mange af de resterende ℓ_x tabelpersoner der dør i aldersintervallet. Vi betegner med ${}_k d_x$ antallet af tabelpersoner, som dør mellem alder x og alder $x + k - 1$. Dermed er ${}_1 d_x$ antallet af tabelpersoner, som dør ved alder x . Sandsynligheden for at dø mellem to alderstrin (dødshyppighederne) er det centrale element ved konstruktionen af overlevelsestavlen. Vi beregner antal dødsfald i aldersintervallet ved at gange antal tabelpersoner i starten af intervallet med dødshyppigheden:

$${}_k d_x = {}_k q_x \cdot \ell_x. \quad (\text{K3.4})$$

Det er vigtigt at skelne mellem antal døde ${}_k D_x$ i den ægte befolkning og antal døde ${}_k d_x$ i tabelbefolkningen. Med formel (K3.4) er det en enkel sag at finde antallet af tabelpersoner der er i live i starten af det næste aldersinterval:

$$\ell_{x+k} = \ell_x - {}_k d_x.$$

Alternativt kan vi starte med at beregne dekrementfunktionen baseret på døds-hyppigheden

$$\ell_{x+k} = \ell_x \cdot (1 - {}_k q_x).$$

Bagefter er det simpelt at beregne antal dødsfald som

$${}_k d_x = l_x - l_{x+k}.$$

Med disse formler kan vi konstruere overlevelsestavlens vigtigste kolonner (ℓ_0 og ${}_k d_x$). Vi beskriver nu de vigtigste dødelighedsmål, som overlevelsestavlen viser.

2.2.5 Beregning af middelrestlevetid og middellevetid

Vi betegner med ${}_k L_x$ den samlede gennemlevede tid i tabelbefolkningen i alderen mellem x og $x + k - 1$. Da dødsfald er eneste afgangssag i tabelbefolkningen,

har vi

$$\begin{aligned} {}_k L_x &= \text{bidrag fra overlevende} + \text{bidrag fra døde} \\ &= k \cdot \ell_{x+k} + {}_k a_x \cdot {}_k d_x \\ &= {}_k a_x \cdot \ell_x + (k - {}_k a_x) \cdot \ell_{x+k}. \end{aligned}$$

Vi skal nu beregne den *forventede restlevetid* for en x -årig tabelperson. For en nyfødt er $x = 0$ og dermed bliver den forventede restlevetid til den forventede levetid, som betegnes med *middellevetid*. Lad T_x angive den samlede levetid i tabelbefolkningen efter x -års fødselsdagen, specielt er T_0 den samlede levetid i tabelbefolkningen. Vi beregner

$$\begin{aligned} T_x &= {}_k L_x + \dots + {}_k L_{x^{\max}} \\ &= {}_k a_x \cdot \ell_x + (k - {}_k a_x) \cdot \ell_{x+k} + \dots + {}_\infty a_{x^{\max}} \cdot \ell_{x^{\max}}. \end{aligned}$$

I tabelbefolkningen overlever ℓ_x personer til deres x -års fødselsdag, så den gennemsnitlige levetid efter x -års fødselsdagen bliver

$$e_x = \frac{T_x}{\ell_x} = \text{gennemsnitlige restlevetid.} \quad (\text{K3.5})$$

Denne kvotient kaldes den forventede restlevetid eller middelrestlevetid for en x -årig tabelperson. På tilsvarende vis bliver middellevetid beregnet som

$$e_0 = \frac{T_0}{\ell_0} = \text{middellevetid.} \quad (\text{K3.6})$$

2.2.6 Fortolkning

Når man fortolker middellevetid og middelrestlevetid, er det vigtigt at huske at fremhæve at beregningen bygger på en hypotetisk tabelbefolkning, som lever hele deres liv i en bestemt kalenderperiode. Danmarks Statistik forklarer middelrestlevetiden sådan²:

Middelrestlevetiden er det gennemsnitlige antal år, som personer på en given fødselsdag har tilbage at leve i, *hvis deres dødelighed fremover (alder for alder) svarer til det niveau, som er konstateret i den aktuelle periode*.

2.3 Overlevelsestabler med 5-års intervaller

```
x5 <- hent_mortalitetsrate_data(tid = 2019,
                                    breaks = c(0,1,seq(5,95,5),Inf),
                                    køn = "kvinder")
x5 <- mutate(x5,M = Dod/R)
x5 <- mutate(x5,a = c(0.1,2,rep(2.5,19)),k = c(1,4,rep(5,19)))
```

²<https://www.dst.dk/da/Statistik/emner/borgere/befolkning/middellevetid>

```
tavle_kvinder_5 <- overlevelsestable(x5,
                                      mortalitet = "M",
                                      alder = "aldersinterval")
print(tavle_kvinder_5,digits = 2,n = 100)
```

```
# A tibble: 21 × 9
#>   Alder     l     d     p     q     o     L     T     e
#>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#> 1 0      100000 251 0.997 0.00251 1 99774 8338941. 83.4
#> 2 1-4    99749  40 1.00  0.000402 0.997 398917 8239167. 82.6
#> 3 5-9    99709  40 1.00  0.000398 0.997 498447 7840250. 78.6
#> 4 10-14   99669  30 1.00  0.000303 0.997 498272 7341804. 73.7
#> 5 15-19   99639  66 0.999 0.000658 0.996 498033 6843532. 68.7
#> 6 20-24   99574  82 0.999 0.000823 0.996 497664 6345499. 63.7
#> 7 25-29   99492  107 0.999 0.00108 0.995 497192 5847835. 58.8
#> 8 30-34   99385  146 0.999 0.00147 0.994 496558 5350643. 53.8
#> 9 35-39   99238  233 0.998 0.00235 0.992 495609 4854085. 48.9
#> 10 40-44  99005  322 0.997 0.00325 0.990 494219 4358476. 44.0
#> 11 45-49  98683  561 0.994 0.00568 0.987 492012 3864257. 39.2
#> 12 50-54  98122  1135 0.988 0.0116 0.981 487773 3372245. 34.4
#> 13 55-59  96987  1709 0.982 0.0176 0.970 480661 2884472. 29.7
#> 14 60-64  95278  3081 0.968 0.0323 0.953 468685 2403810. 25.2
#> 15 65-69  92196  4715 0.949 0.0511 0.922 449195 1935126. 21.0
#> 16 70-74  87482  6638 0.924 0.0759 0.875 420815 1485930. 17.0
#> 17 75-79  80844  10209 0.874 0.126 0.808 378699 1065115. 13.2
#> 18 80-84  70635  15912 0.775 0.225 0.706 313396 686417. 9.72
#> 19 85-89  54723  21608 0.605 0.395 0.547 219597 373021. 6.82
#> 20 90-94  33116  20382 0.385 0.615 0.331 114623 153424. 4.63
#> 21 95+    12734  12734 0       1 0.127 38801 38801. 3.05
```

2.4 Overlevelsestable med 1-års intervaller

```
x1 <- hent_mortalitetsrate_data(tid = 2019,
                                    breaks = c(0:99,Inf),
                                    køn = "kvinder",
                                    right = FALSE)

x1 <- mutate(x1,M = Dod/R)
x1 <- mutate(x1,a = c(0.1,rep(0.5,99)),k = rep(1,100))
tavle_kvinder_1 <- overlevelsestable(x1,
                                       mortalitet = "M",
                                       alder = "aldersinterval")
print(tavle_kvinder_1,digits = 2,n = 100)
```

```
# A tibble: 100 × 9
```

Alder	l	d	p	q	o	L	T	e
<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 0	100000	251	0.997	0.00251	1	99774	8340603.	83.4
2 1-1	99749	16	1.00	0.000164	0.997	99741	8240828.	82.6
3 2-2	99733	7	1.00	0.0000658	0.997	99730	8141087.	81.6
4 3-3	99726	7	1.00	0.0000671	0.997	99723	8041357.	80.6
5 4-4	99720	10	1.00	0.000105	0.997	99714	7941634.	79.6
6 5-5	99709	14	1.00	0.000142	0.997	99702	7841920.	78.6
7 6-6	99695	3	1.00	0.0000343	0.997	99693	7742218.	77.7
8 7-7	99692	10	1.00	0.000101	0.997	99687	7642524.	76.7
9 8-8	99682	6	1.00	0.0000638	0.997	99678	7542838.	75.7
10 9-9	99675	6	1.00	0.0000619	0.997	99672	7443159.	74.7
11 10-10	99669	3	1.00	0.0000304	0.997	99668	7343487.	73.7
12 11-11	99666	3	1.00	0.0000300	0.997	99665	7243820.	72.7
13 12-12	99663	9	1.00	0.0000915	0.997	99659	7144155.	71.7
14 13-13	99654	9	1.00	0.0000906	0.997	99649	7044496.	70.7
15 14-14	99645	6	1.00	0.0000607	0.996	99642	6944847.	69.7
16 15-15	99639	24	1.00	0.000241	0.996	99627	6845205.	68.7
17 16-16	99615	3	1.00	0.0000308	0.996	99613	6745578.	67.7
18 17-17	99612	9	1.00	0.0000907	0.996	99607	6645965.	66.7
19 18-18	99603	9	1.00	0.0000886	0.996	99598	6546358.	65.7
20 19-19	99594	20	1.00	0.000203	0.996	99584	6446759.	64.7
21 20-20	99574	17	1.00	0.000169	0.996	99565	6347176.	63.7
22 21-21	99557	14	1.00	0.000139	0.996	99550	6247610.	62.8
23 22-22	99543	16	1.00	0.000158	0.995	99535	6148060.	61.8
24 23-23	99527	16	1.00	0.000156	0.995	99520	6048525.	60.8
25 24-24	99512	20	1.00	0.000198	0.995	99502	5949005.	59.8
26 25-25	99492	20	1.00	0.000201	0.995	99482	5849503.	58.8
27 26-26	99472	28	1.00	0.000277	0.995	99458	5750021.	57.8
28 27-27	99445	15	1.00	0.000154	0.994	99437	5650563.	56.8
29 28-28	99429	23	1.00	0.000232	0.994	99418	5551126.	55.8
30 29-29	99406	21	1.00	0.000210	0.994	99396	5451708.	54.8
31 30-30	99385	16	1.00	0.000163	0.994	99377	5352312.	53.9
32 31-31	99369	39	1.00	0.000397	0.994	99349	5252935.	52.9
33 32-32	99330	29	1.00	0.000290	0.993	99315	5153586.	51.9
34 33-33	99301	26	1.00	0.000266	0.993	99288	5054271.	50.9
35 34-34	99274	36	1.00	0.000367	0.993	99256	4954983.	49.9
36 35-35	99238	46	1.00	0.000467	0.992	99215	4855727.	48.9
37 36-36	99192	41	1.00	0.000413	0.992	99171	4756512.	48.0
38 37-37	99151	21	1.00	0.000214	0.992	99140	4657341.	47.0
39 38-38	99129	88	0.999	0.000887	0.991	99085	4558201.	46.0
40 39-39	99041	37	1.00	0.000377	0.990	99023	4459116.	45.0
41 40-40	99004	51	0.999	0.000513	0.990	98979	4360093.	44.0
42 41-41	98953	39	1.00	0.000395	0.990	98934	4261114.	43.1
43 42-42	98914	73	0.999	0.000733	0.989	98878	4162180.	42.1
44 43-43	98842	79	0.999	0.000804	0.988	98802	4063302.	41.1
45 44-44	98762	77	0.999	0.000784	0.988	98724	3964500.	40.1
46 45-45	98685	115	0.999	0.00117	0.987	98627	3865777.	39.2
47 46-46	98570	60	0.999	0.000609	0.986	98540	3767150.	38.2
48 47-47	98510	131	0.999	0.00133	0.985	98444	3668610.	37.2
49 48-48	98379	123	0.999	0.00125	0.984	98318	3570165.	36.3
50 49-49	98257	133	0.999	0.00135	0.983	98190	3471848.	35.3
51 50-50	98124	169	0.998	0.00172	0.981	98039	3373658.	34.4
52 51-51	97955	163	0.998	0.00167	0.980	97873	3275618.	33.4
53 52-52	97792	251	0.997	0.00256	0.978	97666	3177745.	32.5
54 53-53	97541	262	0.997	0.00269	0.975	97410	3080078.	31.6
55 54-54	97279	280	0.997	0.00287	0.973	97139	2982668.	30.7

56	55-55	96999	248	0.997	0.00256	0.970	96875	2885530.	29.7
57	56-56	96751	296	0.997	0.00306	0.968	96602	2788655.	28.8
58	57-57	96454	348	0.996	0.00361	0.965	96280	2692052.	27.9
59	58-58	96106	373	0.996	0.00388	0.961	95919	2595772.	27.0
60	59-59	95733	460	0.995	0.00480	0.957	95503	2499853.	26.1
61	60-60	95273	471	0.995	0.00494	0.953	95038	2404350.	25.2
62	61-61	94802	599	0.994	0.00632	0.948	94503	2309312.	24.4
63	62-62	94203	611	0.994	0.00649	0.942	93898	2214809.	23.5
64	63-63	93592	660	0.993	0.00705	0.936	93262	2120911.	22.7
65	64-64	92932	746	0.992	0.00803	0.929	92559	2027649.	21.8
66	65-65	92186	823	0.991	0.00893	0.922	91775	1935090.	21.0
67	66-66	91363	873	0.990	0.00956	0.914	90927	1843315.	20.2
68	67-67	90490	985	0.989	0.0109	0.905	89997	1752389.	19.4
69	68-68	89505	1030	0.988	0.0115	0.895	88990	1662391.	18.6
70	69-69	88475	1009	0.989	0.0114	0.885	87970	1573402.	17.8
71	70-70	87466	1113	0.987	0.0127	0.875	86909	1485432.	17.0
72	71-71	86353	1137	0.987	0.0132	0.864	85784	1398522.	16.2
73	72-72	85215	1329	0.984	0.0156	0.852	84551	1312738.	15.4
74	73-73	83887	1454	0.983	0.0173	0.839	83159	1228187.	14.6
75	74-74	82432	1600	0.981	0.0194	0.824	81632	1145028.	13.9
76	75-75	80832	1857	0.977	0.0230	0.808	79904	1063396.	13.2
77	76-76	78975	1928	0.976	0.0244	0.790	78011	983492.	12.5
78	77-77	77047	1968	0.974	0.0255	0.770	76063	905481.	11.8
79	78-78	75078	2224	0.970	0.0296	0.751	73967	829419.	11.0
80	79-79	72855	2390	0.967	0.0328	0.729	71660	755452.	10.4
81	80-80	70465	2698	0.962	0.0383	0.705	69116	683792.	9.70
82	81-81	67767	2878	0.958	0.0425	0.678	66328	614675.	9.07
83	82-82	64889	3227	0.950	0.0497	0.649	63275	548347.	8.45
84	83-83	61662	3529	0.943	0.0572	0.617	59897	485072.	7.87
85	84-84	58133	3837	0.934	0.0660	0.581	56215	425174.	7.31
86	85-85	54296	3922	0.928	0.0722	0.543	52335	368960.	6.80
87	86-86	50374	4430	0.912	0.0879	0.504	48159	316625.	6.29
88	87-87	45944	4262	0.907	0.0928	0.459	43813	268465.	5.84
89	88-88	41682	4616	0.889	0.111	0.417	39375	224652.	5.39
90	89-89	37067	4390	0.882	0.118	0.371	34872	185278.	5.00
91	90-90	32677	4427	0.865	0.135	0.327	30464	150406.	4.60
92	91-91	28250	4264	0.849	0.151	0.282	26118	119942.	4.25
93	92-92	23986	4081	0.830	0.170	0.240	21945	93824.	3.91
94	93-93	19905	3563	0.821	0.179	0.199	18124	71879.	3.61
95	94-94	16342	3425	0.790	0.210	0.163	14630	53756.	3.29
96	95-95	12918	3040	0.765	0.235	0.129	11398	39125.	3.03
97	96-96	9877	2494	0.748	0.252	0.0988	8631	27728.	2.81
98	97-97	7384	1939	0.737	0.263	0.0738	6414	19097.	2.59
99	98-98	5444	1690	0.690	0.310	0.0544	4599	12683.	2.33
100	99+	3754	3754	0	1	0.0375	8084	8084.	2.15

2.5 Danmark Statistik

Danmark Statistik offentliggør egne beregninger af middellevetiden og middelrelstlevetiden.³ I dette afsnit forklarer vi, hvordan Danmark Statistikks beregninger bliver mere præcise, fordi de bruger datoer for fødsler, dødsfald og folkevandringer.⁴

³<https://www.dst.dk/da/Statistik/emner/borgere/befolkning/middellevetid>.

⁴<https://www.dst.dk/ext/36380110073/0/befolkning/Hvordan-beregner-vi-middellevetid?--pdf>

Med etableringen af den personstatistiske database har Danmarks Statistik fået nye muligheder for at beregne dødshyppighederne mere korrekt, idet databasen for alle personer i Danmark indeholder eksakt information om eventuel dødsdato og ind- og udvandringsdatoer. Der kan således for hver enkelt person udregnes nøjagtigt, hvor mange dage personen i en årsperiode har været i Danmark og hvor mange af dagene i årsperioden, personen har været død. Den søgte dødshyppighed skal præcist angive sandsynligheden for at dø i et bestemt alderstrin, dvs. mellem to fødselsdage. For at opnå denne hyppighed laves der en særlig beregning for hver enkelt person fra fødselsdag til fødselsdag i en periode, der omfatter to kalenderår. I offentliggørelsen af middellevetid fra 19. marts 2010 er det kalenderårene 2008 og 2009, der ligger til grund for beregningerne. For alle personer, der var i den danske befolkning på et eller andet tidspunkt mellem deres fødselsdag i 2008 og i 2009, er der lavet en beregning for antallet af dage, personen var i Danmark og antallet af dage, personen var død i perioden mellem de to fødselsdage. For personer, der ikke dør mellem to fødselsdage, vil antallet af dage som død naturligvis være 0. Efterfølgende laves der en sammenlægning for personer med samme køn og alderstrin for at få det samlede antal levedage og dødsdage. Personer vil placeres på det alderstrin, som svarer til det antal år, de fyldte i startåret, hvilket i eksemplet vil sige 2008. En person, som fyldte 60 år 1. januar 2008 vil f.eks. tilhøre de 60-årige. Det samme vil en person, der fyldte 60 år 31. december 2008. Der kan altså i yderste konsekvens være næsten et års forskel mellem den periode, som personer på samme alderstrin følges.

3 Dødsårsager

Menneskers død har forskellige årsager, som fortæller, hvad der er sket lige inden, for eksempel en trafikulykke, eller sammenfatter et kortere eller længere sygdomsforløb, før døden indtræffer. I Danmark har man siden 1875 samlet data om dødsårsager, som nu er verdens største digitaliserede dødsårsagsregister⁵. Dødsårsagsregisteret bygger på dødsattester fra personer med folkeregisteradresse i Danmark, der døde i Danmark. Den demografiske analyse af dødsårsager er formåls- og metodemæssigt relateret til de andre fag i dette semester (sygdomslære, epidemiologi og biostatistik). Vi bruger data fra dødsårsagsregisteret til at konstruere årsagsspecifikke dødelighedstavler (Afsnit 3.3) og til at beregne restlivstidsrisikoen (sandsynligheden for at en x -årig dør af en bestemt årsag).

3.1 Gruppering af dødsårsager

Dødsårsager i dødsårsagsregisteret er opdelt i 26 A grupper, og hver A gruppe har en eller flere B undergrupper. Der er 109 B-grupper. Statistikbankens register DODA1 har antal døde fordelt på 26 A-grupper:

```
da <- hent_data("doda1",årsag = "all_no_total",tid = 2022)
print(da,n = 26)
```

⁵<https://www.rigsarkivet.dk/udforsk/doedsaarsagsregister-1943-2019/>

# A tibble: 26 × 3	ÅRSAG	TID INDHOLD
	<chr>	<dbl> <dbl>
1	A-01 Infektiøse inkl. parasitære sygdomme	2022 1824
2	A-02 Kræft	2022 15777
3	A-03 Andre svulster (anden neoplasi)	2022 359
4	A-04 Sygdomme i blod (-dannende) organer, sygdomme,	2022 231
5	A-05 Endokrine og ernæringsbetingede sygdomme samt s	2022 2003
6	A-06 Psykiske lidelser og adfærdsmæssige forstyrrelser	2022 3954
7	A-07 Sygdomme i nervesystemet og sanseorganerne	2022 3207
8	A-08 Hjertesygdomme	2022 8019
9	A-09 Andre kredsløbssygdomme	2022 4117
10	A-10 Sygdomme i åndedrætsorganer	2022 6297
11	A-11 Sygdomme i fordøjelsesorganer	2022 2379
12	A-12 Sygdomme i hud og underhud	2022 72
13	A-13 Sygdomme i knogler, muskler og bindevæv	2022 354
14	A-14 Sygdomme i urin- og kønsorganer	2022 948
15	A-15 Komplikationer ved svangerskab, fødsel og barsel	2022 1
16	A-16 Visse sygdomme, der opstår i perinatalperioden	2022 100
17	A-17 Medfødte misdannelser og kromosomanomalier	2022 141
18	A-18 Symptomer og abnorme fund, dårligt definerede å	2022 2337
19	A-19 Ulykker	2022 1692
20	A-20 Selvmord og selvmordsforsøg	2022 572
21	A-21 Drab, overfald	2022 39
22	A-22 Hændelser med uvis omstændighed	2022 45
23	A-23 Legale interventioner inkl. krigshandlinger	2022 1
24	A-23x Covid-19 - Corona	2022 1590
25	A-24 Dødsfald uden medicinske oplysninger	2022 3062
26	Årsag ikke oplyst	2022 314

Vi ser, at dødsfald på grund af kræft har været den største A-gruppe i 2022. Der er i alt 109 B-grupper, hvor hver B-gruppe hører under en A-gruppe. Kræft er den dødsårsag med flest B-undergrupper. A-grupper er mere overordnede, mens B-grupper er mere specifikke. For eksempel er der tre B-grupper som opdeler gruppen A-08 HJERTESYGDOMME I ALT:

```
db <- hent_data("doddb1",årsag = "all_no_total",tid = 2022)
print(filter(db,str_detect(ÅRSAG,"A-08|B-057|B-058|B-059")))
```

# A tibble: 4 × 3	ÅRSAG	TID INDHOLD
	<chr>	<dbl> <dbl>
1	A-08 HJERTESYGDOMME I ALT	2022 8019
2	B-057 Iskæmiske hjertesygdomme	2022 3275
3	B-058 Blodtryksforhøjelse	2022 1462
4	B-059 Andre hjertesygdomme	2022 3282

3.2 Årsagsspecifikke mortalitetsrater

For en given dødsårsag Q beregner vi de summariske årsagsspecifikke mortalitetsrater på samme måde som de summariske mortalitetsrater som antal begivenheder per personår. For en kalenderperiode $[t_1, t_2]$ og risikotid $R[t_1, t_2]$ er formlen for den årsagsspecifikke mortalitetsrate:

$$\frac{D^Q[t_1, t_2]}{R[t_1, t_2]} = \frac{\text{Antal døde med årsag } Q \text{ i perioden } [t_1, t_2]}{\text{Risikotid i perioden } [t_1, t_2]}.$$

Vi kan også beregne årsagsspecifikke mortalitetsrater i aldersgrupper. Vi betegner med ${}_k D_x^Q$ antal dødsfald, hvor dødsårsagen var Q i aldersgruppen af personer, der var mellem x -år og $(x + k - 1)$ -år gamle i perioden. Det giver følgende notation for de aldersspecifikke rater:

$$\frac{{}_k D_x^Q[t_1, t_2]}{{}_k R_x[t_1, t_2]}.$$

For eksempel kan vi beregne rater af dødsulykker per 10.000 personår blandt unge mennesker (15-29 år) og se, hvordan de har udviklet sig siden 2007. Figur 3 viser at disse rater er faldet fra omkring 1,5 dødsulykker per 10.000 personår i 2007 til omkring 0,75 dødsulykker per 10.000 personår i 2022.

3.3 Årsagsspecifikke dødelighedstavler

Vi konstruerer nu forskellige mål for, hvordan de specifikke dødsårsager bidrager til den samlede dødelighed. Disse mål er periodemål ligesom middellevetid og beregnes som udgangspunkt i en tænkt lukket tabelbefolkning. Vi beregner restlivstidsrisiko for at dø af en given årsag under antagelsen, at de årsagsspecifikke mortalitetsrater er som observeret i en given kalenderperiode.

3.3.1 Årsagsspecifikke dødshyppigheder

I den ægte befolkning finder vi andelen af dødsfald, som blev tilskrevet årsag Q i et givent aldersinterval i en given periode og betegner den med:

$${}_k h_x^Q = \frac{{}_k D_x^Q}{{}_k D_x}. \quad (\text{K3.7})$$

Andelen af dødsfald af alle andre årsager end Q i aldersintervallet bliver dermed

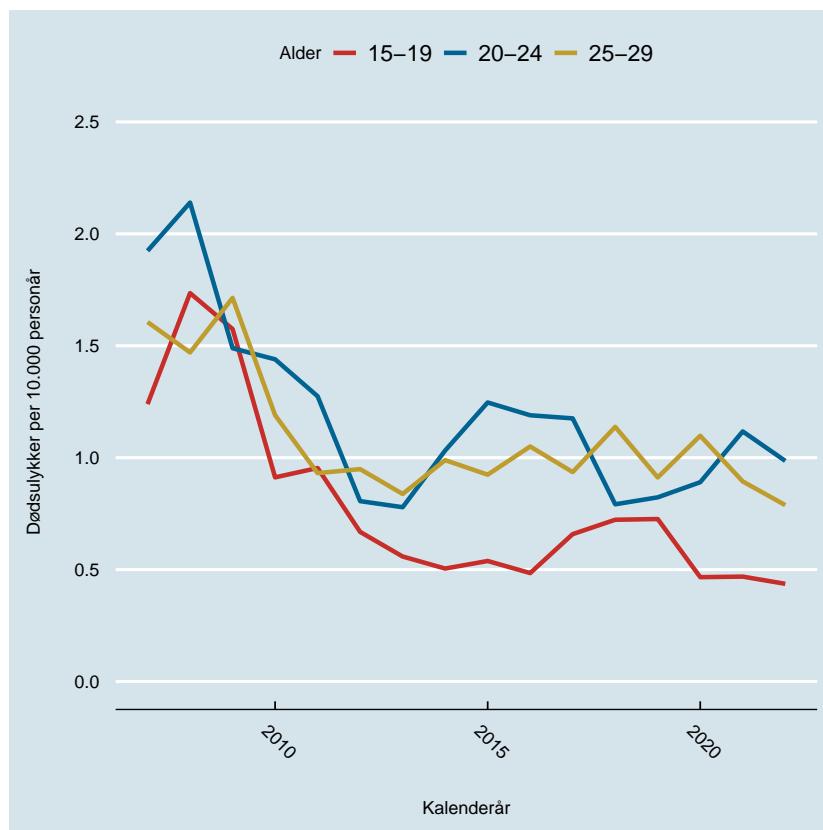
$${}_k h_x^{\bar{Q}} = 1 - {}_k h_x^Q = \frac{{}_k D_x - {}_k D_x^Q}{{}_k D_x}. \quad (\text{K3.8})$$

I en tænkt lukket tabelbefolkning kan vi nu beregne hyppigheden for at dø af årsag Q i alderen mellem x og $x + k - 1$:

$${}_k q_x^Q = {}_k q_x \cdot {}_k h_x^Q, \quad (\text{K3.9})$$

og tilsvarende er hyppigheden for at dø af en anden årsag:

$${}_k q_x^{\bar{Q}} = {}_k q_x \cdot (1 - {}_k h_x^Q). \quad (\text{K3.10})$$



Figur 3: Udviklingen i raten af dødsulykker blandt danskere i alderen mellem 15 og 29. Kilde: statistikbankens DOD1A og FOLK1A.

3.3.2 Konstruktion af en årsagsspecifik dødelighedstavle

Dekrementfunktionen, som genererer en årsagsspecifik dødelighedstavle, tager udgangspunkt i en radix af 100.000 tabelpersoner. I hvert alderstrin beregnes antal dødsfald ligesom i en almindelig overlevelsestavle. Dødsårsagerne bliver opdelt i to grupper sådan, at det samlede antal dødsfald er summen af dødsfald med årsag Q og dødsfald med andre årsager. Det samlede antal døde i tabelbefolkningen mellem alder x og alder $x + k - 1$ er

$${}_k d_x = \ell_x \cdot {}_k q_x. \quad (\text{K3.11})$$

Det kan nu opdeles i dødsfald med årsag Q

$${}_k d_x^Q = {}_k d_x \cdot {}_k h_x^Q \quad (\text{K3.12})$$

og dødsfald med andre årsager

$${}_k d_x^{\bar{Q}} = {}_k d_x \cdot (1 - {}_k h_x^Q). \quad (\text{K3.13})$$

Derfor gælder i tabelbefolkningen:

$$\ell_x = \underbrace{\ell_{x+k}}_{\text{i live}} + \underbrace{{}_k q_x \cdot {}_k h_x^Q}_{\text{årsag } Q} + \underbrace{{}_k q_x \cdot (1 - {}_k h_x^Q)}_{\text{andre årsager}}. \quad (\text{K3.14})$$

Fordi der nu er to muligheder (1: dødsårsag Q , 2: alle andre dødsårsager) for at forlade tabelbefolkningen hedder funktionen ℓ_x , der genererer en årsagsspecifik dødelighedstavle, double decrement function.

3.3.3 Den årsagsspecifikke restlivstidsrisiko

Det samlede antal dødsfald med årsag Q i tabelbefolkningen, hvor tabelpersonen er ældre end x -år, er (for 1-års intervaller) givet som:

$${}_{\infty} d_x^Q = {}_1 d_x^Q + {}_1 d_{x+1}^Q + \cdots + {}_{\infty} d_{x^{\max}}^Q, \quad (\text{K3.15})$$

og restlivstidsrisikoen blandt x -årige for at dø af årsag Q beregnes som

$$\text{LTR}_x^Q = \frac{({}_1 d_x^Q + {}_1 d_{x+1}^Q + \cdots + {}_{\infty} d_{x^{\max}}^Q)}{\ell_x}. \quad (\text{K3.16})$$

3.3.4 Risikoen for at dø af en bestemt årsag

I tabelbefolkningen er det samlede antal dødsfald, hvor årsagen var Q mellem alder 0 til alder x , givet ved (for 1-års intervaller):

$${}_x d_0^Q = {}_1 d_0^Q + \cdots + {}_1 d_{x-1}^Q. \quad (\text{K3.17})$$

Risikoen for, at en nyfødt dør af årsag Q inden alder x , bliver

$${}_x q_0^Q = {}_x d_0^Q / \ell_0. \quad (\text{K3.18})$$

Tilsvarende er risikoen for at dø på grund af andre årsager inden alder x :

$${}_x q_0^{\bar{Q}} = {}_x d_0^{\bar{Q}} / \ell_0. \quad (\text{K3.19})$$

Vi kan også beregne sandsynligheden for at overleve alle årsager inden alder x :

$$o_x = 1 - {}_x q_0 = 1 - {}_x q_0^Q - {}_x q_0^{\bar{Q}} \quad (\text{K3.20})$$

3.4 Eksempel

Vi beregner dodelighedstavlen for at dø af kræft blandt mænd i 2020 i Danmark. Vi henter folketal fra statistikbankens register FOLK1a og antal døde med kræft fra register doda1. Vi inddeler i 19 aldersintervaller, hvor det første interval har længde 1 år, det andet interval har længde 4 år, resten af intervallerne har længde 5 år og det sidste aldersinterval er fra 85 til 125 år. Vi beregner aldersspecifikke mortalitetsrater for mænd i 2020 og andelen af dødsfald med kræft.

```
x <- hent_dodsaarsag_data(tid = 2020, årsag = c("A02"), køn = "Mænd")
# mortalitetsrater
x <- mutate(x, M = Dod/R)
# andel kræftdødsfald
x <- mutate(x, hQ = QDod/Dod)
x
```

# A tibble: 19 × 9	ÅRSAG	aldersinterval	KØN	TID	R	Dod	QDod	M	hQ
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	A-02 Kræft 0		Mænd	2020	31512	109	0	0.00346	0
2	A-02 Kræft 1-4		Mænd	2020	127529	18	5	0.000141	0.278
3	A-02 Kræft 5-9		Mænd	2020	154685	15	3	0.0000970	0.2
4	A-02 Kræft 10-14		Mænd	2020	173860	18	2	0.000104	0.111
5	A-02 Kræft 15-19		Mænd	2020	174529	35	2	0.000201	0.0571
6	A-02 Kræft 20-24		Mænd	2020	192608	82	11	0.000426	0.134
7	A-02 Kræft 25-29		Mænd	2020	204302	112	7	0.000548	0.0625
8	A-02 Kræft 30-34		Mænd	2020	185281	88	10	0.000475	0.114
9	A-02 Kræft 35-39		Mænd	2020	165161	157	25	0.000951	0.159
10	A-02 Kræft 40-44		Mænd	2020	179809	219	38	0.00122	0.174
11	A-02 Kræft 45-49		Mænd	2020	196936	380	81	0.00193	0.213
12	A-02 Kræft 50-54		Mænd	2020	204696	690	193	0.00337	0.280
13	A-02 Kræft 55-59		Mænd	2020	197362	1132	385	0.00574	0.340
14	A-02 Kræft 60-64		Mænd	2020	171437	1663	630	0.00970	0.379
15	A-02 Kræft 65-69		Mænd	2020	155595	2556	1029	0.0164	0.403
16	A-02 Kræft 70-74		Mænd	2020	155082	3779	1548	0.0244	0.410
17	A-02 Kræft 75-79		Mænd	2020	115932	4584	1616	0.0395	0.353
18	A-02 Kræft 80-84		Mænd	2020	66656	4613	1338	0.0692	0.290
19	A-02 Kræft 85		Mænd	2020	44684	7745	1552	0.173	0.200

Med disse data beregner vi dødelighedstavlen.

```
# Chiang's \(\alpha\)
x <- mutate(x,a = c(0.1,2,rep(2.5,17)),k = c(1,4,rep(5,17)))
kraeftdodtavle_maend <- dodsaarsagtavle(data = x,
                                             mortalitet = "M",
                                             hQ = "hQ",
                                             alder = "aldersinterval",
                                             radix = 100000)
# fravælger kolonner som ikke er vigtige lige her
kraeftdodtavle_maend <- select(kraeftdodtavle_maend,-c(p,q,o,T))
print(kraeftdodtavle_maend,digits = 2)
```

	Alder	l	d	dQ	hQ	qQ	L	e	LTR_Q	risiko_Q
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0	100000	345	0	0	0	99690	79.6	0.280	0
2	1-4	99655	56	15.6	0.278	0.000157	398508	78.9	0.281	0.000156
3	5-9	99599	48	9.66	0.2	0.0000969	497874	75.0	0.281	0.000253
4	10-14	99551	52	5.72	0.111	0.0000575	497624	70.0	0.281	0.000310
5	15-19	99499	100	5.70	0.0571	0.0000573	497246	65.0	0.281	0.000367
6	20-24	99399	211	28.4	0.134	0.000285	496469	60.1	0.281	0.000651
7	25-29	99188	272	17.0	0.0625	0.000171	495261	55.2	0.282	0.000820
8	30-34	98917	235	26.7	0.114	0.000270	493996	50.4	0.282	0.00109
9	35-39	98682	468	74.5	0.159	0.000755	492240	45.5	0.283	0.00183
10	40-44	98214	596	103.	0.174	0.00105	489579	40.7	0.283	0.00287
11	45-49	97618	937	200.	0.213	0.00205	485745	35.9	0.284	0.00486
12	50-54	96680	1616	452.	0.280	0.00467	479363	31.2	0.285	0.00938
13	55-59	95065	2688	914.	0.340	0.00962	468603	26.7	0.285	0.0185
14	60-64	92377	4374	1657.	0.379	0.0179	450948	22.4	0.283	0.0351
15	65-69	88002	6943	2795.	0.403	0.0318	422655	18.4	0.278	0.0630
16	70-74	81059	9309	3813.	0.410	0.0470	382024	14.8	0.268	0.101
17	75-79	71750	12909	4551.	0.353	0.0634	326479	11.4	0.249	0.147
18	80-84	58841	17358	5035.	0.290	0.0856	250812	8.33	0.227	0.197
19	85	41484	41484	8313.	0.200	0.200	239335	5.77	0.200	0.280

Kolonnen LTR_Q i aldersintervallet fra x til $x + k - 1$ angiver restlivstidsrisikoen for at dø af kræft for en tabelperson med alder x , og $risiko_Q$ angiver risiko for, at en tabelperson dør af kræft inden alder $x + k$, se Tabel 3. Fra dødelighedstavlen kan vi for eksempel aflæse, at sandsynligheden for, at en nyfødt dreng i 2020 dør på grund af kræft inden alder 74, er 10,1%, hvis de årsagsspecifikke mortalitetsrater for mænd i fremtiden forbliver, som de var i 2020. Vi ser også, at restlivstidsrisikoen for kræftdød er 27,8% for en 65-årig mand under antagelsen, at de årsagsspecifikke mortalitetsrater for mænd forbliver i fremtiden, som de var i 2020.

Tabel 3: Forklaring af kolonner i en årsagsspecifikke dødelighedstavle produceret med R-funktionen **dodsaarsagtavle**.

Kolonne	Betydning
Alder	Aldersinterval
l	Dekrementfunktion: Antal tabelpersoner i starten af intervallet
d	Antal døde i intervallet
dQ	Antal dødsfald af årsag Q i aldersintervallet
hQ	Andelen af dødsfald af årsag Q i aldersintervallet
qQ	Dødshyppighed: sandsynlighed for at dø af årsag Q i intervallet bland tabelpersoner i live ved intervallets start
L	Samlet risikotid i intervallet
e	Middelrestlevetid (i første interval = middellevetid)
LTR_Q	Restlivstidsrisikoen for at dø af kræft blandt tabelpersoner som er i live ved intervallets start
risiko_Q	Risiko for kræftdød inden alder ved start af intervallet

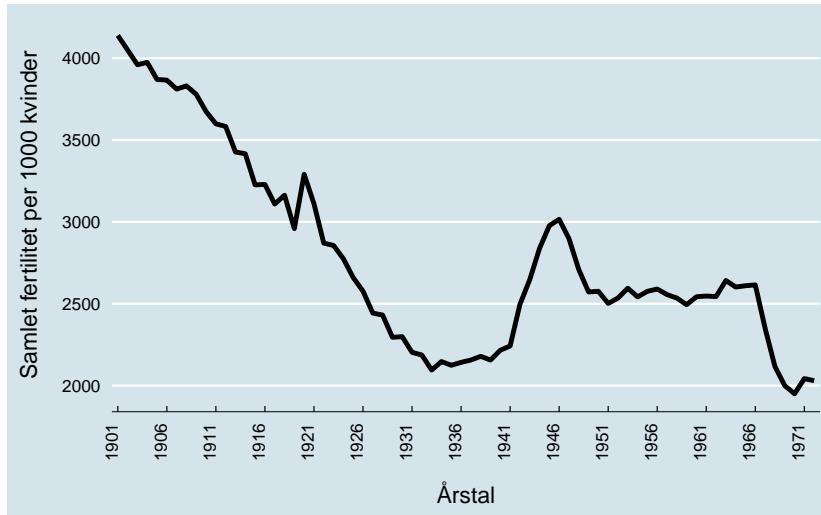
Kapitel 4: Fertilitet

1 Introduktion

Fertilitet er en vigtig komponent i demografi og beskriver sammen med mortalitet den naturlige vækst i befolkningen. Ændringer i fertiliteten har stor betydning for befolkningens størrelse, og det fylder derfor meget i den offentlige debat, når fertiliteten falder, hvilket netop er den tendens, der ses i de fleste vestlige lande. Fertilitet i demografi handler dog ikke om den biologiske evne til at få børn hos enkeltpersoner (frugtbarheden), men om hvor mange børn en befolkning samlet set får.

Over det sidste århundrede har fertiliteten gennemgået store forandringer i Danmark. Det skyldes blandt andet økonomiske kriser, krig, ændringer i kvinders rolle på arbejdsmarkedet, introduktionen af p-pillen, retten til abort, ændrede familiemønstre og holdninger til familieplanlægning. Da der i starten af 1900-tallet forekom en stigende urbanisering i Danmark, hvormed befolkningen rykkede ind til byerne, skete et markant fald i fertiliteten. Den *samlede fertilitet* er et fertilitetsmål, som bygger på fertilitetsrater fra en kort periode, typisk 1 år. Den samlede fertilitet angiver, hvor mange børn en (hypotetisk) kvinde i gennemsnit får, hvis hun gennemlever aldersperioden mellem 15 og 49 år, hvor de aldersspecifikke fertilitetsraterne fra perioden gælder. Den samlede fertilitet var over 4 børn per kvinde i 1901 og faldt derefter hurtigt. I 1933 nåede den samlede fertilitet under 2,1 børn per kvinde (Figur 1).

Under Anden Verdenskrig steg fertiliteten derimod. Den stigende fertilitet skete i alle aldersgrupper, og det medførte meget store fødselsårgange. På trods af udsving fortsatte en lille stigning i fertiliteten frem til slut 1960'erne. Denne periode omtales ofte som babyboomperioden. Der skete også en stigning i antallet af ægeskaber i samme periode. I 1967 blev p-pillen introduceret og i 1973 blev fri abort indført. Det gav kvinder – og familier – mulighed for i højere grad at planlægge, hvornår de ønskede at få børn. Det betød, at alderen for førstegangsfødende kvinder steg (se Figur 2). Samtidig opstod nye familieformer, og det blev mere normalt for kvinder og par at få børn uden for ægeskabet. Endeligt fik kvinder en større rolle på arbejdsmarkedet. Alle disse ændringer skete samtidig med, at fertiliteten faldt i denne periode og frem til starten af 1980'erne (Figur 6). Dette fald skete i alle aldersgrupper. Og den samlede fertilitet nåede igen ned under 2,1 børn per kvinde. Siden har fertiliteten været let stigende – særligt blandt de 30+ årige kvinder. Førstegangsforældre bliver stadigt ældre, som det ses af Figur 2. I 2022 var kvinder i gennemsnit 29,9 år og mænd 31,6 år



Figur 1: Udviklingen i den samlede fertilitet mellem 1901 og 1972 i Danmark.
Kilde: statistikbankens arkiverede register FOD3.

gamle, da de fik deres første barn. Vi finder disse tal i statistikbankens register FOD11:

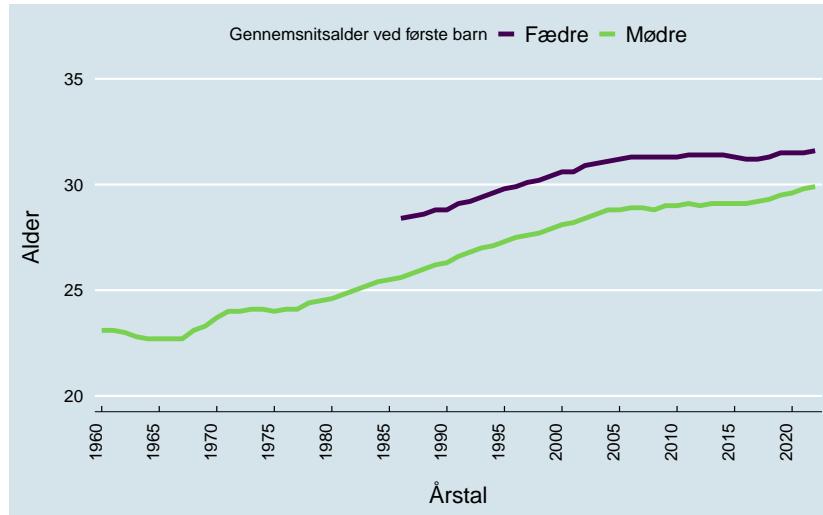
```
x <- hent_data("fod11", tid = 2022, alder = c(610, 617))
with(x, paste(ALDER, INDHOLD))
```

```
[1] "Gennemsnitsalder for førstegangsfødende kvinder 29.9"
[2] "Gennemsnitsalder for førstegangs fædre 31.6"
```

2 Fertilitetsmål

Når fertilitet undersøges i demografiske analyser, er det med fokus på fødsler af levendefødte børn. Ifølge Sundhedsstyrelsen anses et barn for levendefødt, hvis barnet kommer til verden med tydelige tegn på liv uanset graviditetens længde. Er dette ikke tilfældet, beskrives barnet som dødfødt og indgår ikke i demografiske analyser. Begrundelsen for dette er, at disse fødsler ikke bidrager til befolkningsudviklingen eller -sammensætningen.

Forskellige mål kan anvendes til at beskrive fertiliteten og udviklingen heraf. En simpel fremgangsmåde er en opgørelse af *antallet af levendefødte*. Et eksempel på sådan en opgørelse fremgår af Figur 3, som viser udviklingen i antallet af levendefødte børn siden 1973. Det absolute antal levendefødte beskriver størrelsen på en fødselskohorte, hvilket i demografisk sammenhæng har betydning for befolkningssammensætningen og dens udvikling over tid. For eksempel har størrelsen på en fødselskohorte betydning for, hvor mange pladser der er brug



Figur 2: Udvikling i forældres gennemsnitsalder ved første barn i perioden 1960-2022 i Danmark. Kilde: Statistikbankens register FOD11.

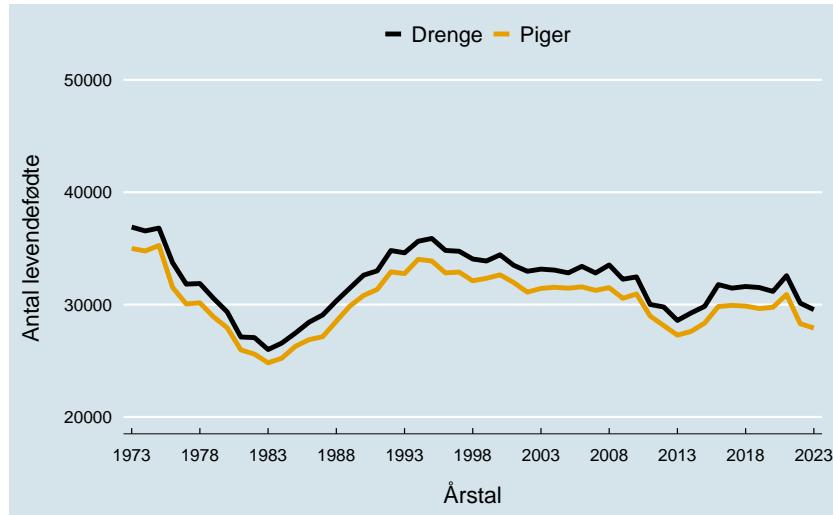
for i daginstitutionerne. Ligeledes har det senere betydning for, hvor mange der vil være på arbejdsmarkedet og endnu senere, hvor mangle ældre der vil være i samfundet. Alt dette har betydning for politiske beslutninger og planlægning. Det absolute antal levendefødte har dog også dets begrænsing i og med, det ikke kan bruges til at sammenligne fertiliteten over tid eller på tværs af befolkningsgrupper/lande. Hertil er der brug for (aldersspecifikke) fertilitetsrater.

2.1 Fertilitetsrater

Den summariske fertilitetsrate defineres som antal levendefødte per 1.000 indbyggere, hvor moren er mellem 15 og 49 år gammel og beregnes ved at dividere antallet af levendefødte børn i en given periode med risikotid for samme periode. Fremgangsmåden er altså den samme som for den summariske mortalitetsrate (Kapitler 1 og 2). Vi bruger følgende notation, som er standard i demografinen (Kapitel 3), for antal levendefødte børn i perioden $[t_1, t_2]$, hvor kvinden var mellem 15 og 49 år gammel ved fødslen:

$${}_{35}F_{15} = {}_{35}F_{15}[t_1, t_2]. \quad (\text{K4.1})$$

Her er indeks til højre startalderen og indeks til venstre er antal år som tælles med inklusiv startalderen. Dermed tæller formel (K4.1) alle levendefødte børn hvor moren var x -år gammel for alle $x \in \{15, 16, \dots, 48, 49\}$, og der er 35 tal i denne mængde. Det er lidt uklart, hvorfor man ikke tæller alle levendefødte børn, altså også dem hvor moren var yngre end 15 eller ældre end 49 år gammel.



Figur 3: Udvikling i antal levendefødte i perioden 1973-2023 i Danmark. Kilde: Statistikbankens register FOD.

Formlen for den summariske fødselsrate ser således ud:

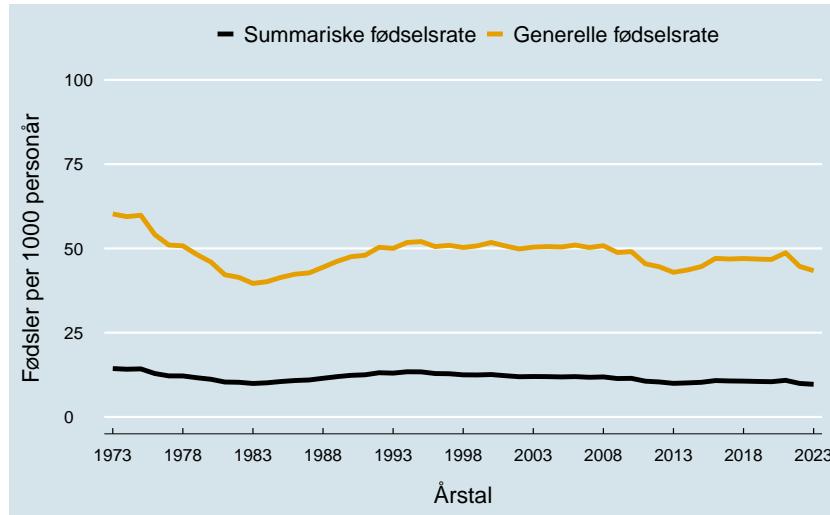
$$f_{sum} = \frac{35F_{15}[t_1, t_2]}{R[t_1, t_2]} = \frac{\text{Antal fødsler: Kvinder mellem 15 og 49 år}}{\text{Risikotid: hele befolkningen}} \quad (\text{K4.2})$$

Den summariske fødselsrate f_{sum} angiver dermed fertilitetsraten i tidsintervallet fra kalenderdato t_1 til kalenderdato t_2 . Det er vigtigt at notere sig, at der i risikotiden for den givne periode, $R[t_1, t_2]$ indgår personer, som ikke kan føde børn – herunder mænd, børn og ældre kvinder. Den *generelle fertilitetsrate* erstatter risikotid af alle personer med risikotid fra kvinder i den fertile alder (15-49 år). Den generelle fertilitetsrate defineres således som antal levendefødte børn per 1.000 kvinder i den fertile alder:

$$f_{gen} = \frac{35F_{15}[t_1, t_2]}{35R_{15}^{\text{Kvinder}}[t_1, t_2]} = \frac{\text{Antal fødsler: kvinder mellem 15 og 49 år}}{\text{Risikotid: kvinder mellem 15 og 49 år}} \quad (\text{K4.3})$$

Her står notationen $35R_{15}^{\text{Kvinder}}$ for risikotid fra alle kvinder, som var mellem 15 og 49 år gamle i perioden. Sammenlignet med summariske fertilitetsrater har generelle fertilitetsrater højere værdier, da risikotiden fra den kvindelige befolkning mellem 15 og 49 er mindre end risikotiden fra hele befolkningen. Udviklingen i både den generelle og summariske fertilitetsrate siden 1973 fremgår af Figur 4. Figuren viser, at den summariske fertilitetsrate har ligget ret stabilt over årene, og at den generelle fertilitetsrate var utsat for udsving, som reflekterer udsving i størrelsen på den kvindelige befolkning i aldersgruppen 15 til 49 år.

Aldersspecifikke fertilitetsrater udregnes på samme vis som aldersspecifikke mortalitetsrater (Kapitel 2). For en given periode og en given aldersgruppe, for



Figur 4: Udviklingen i både den generelle og summariske fertilitetsrate siden 1973 i Danmark. Kilde: statistikbankens register FOD, BEFOLK2.

eksempel 30-34 årlige, beregnes aldersspecifikke fertilitetsrater ved at dividere antallet af levendefødte børn, der fødes af kvinder, der er mellem 30 og 34 år gammel ved barnets fødsel, med risikotiden af 30-34 årlige kvinder i perioden:

$${}_5f_{30} = \frac{{}_5F_{30}}{{}_5R_{30}^{\text{Kvinder}}} = \frac{\text{Antal fødsler: Kvinder mellem 30 og 34 år}}{\text{Risikotid: Kvinder mellem 30 og 34 år}}.$$

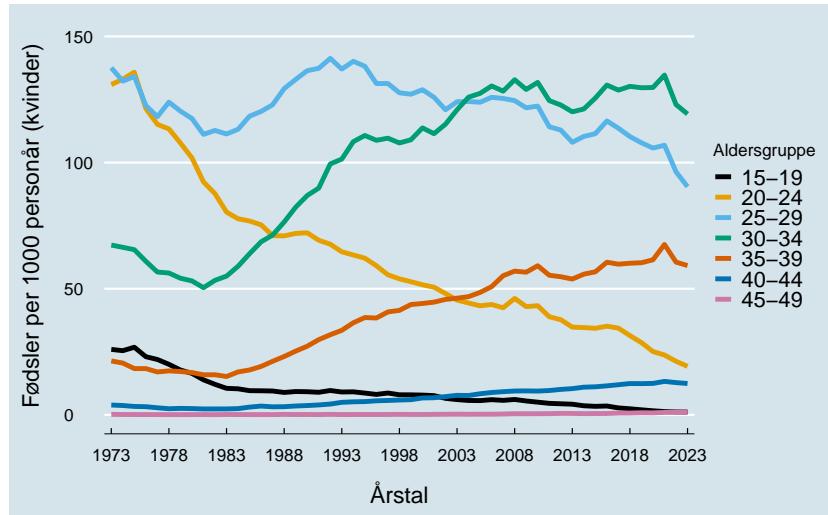
Mere generelt ser formlen for den aldersspecifikke fertilitetsrate fra alder x til og med alder $x + k - 1$ således ud:

$${}_k f_x = \frac{{}_k F_x}{{}_k R_x^{\text{Kvinder}}} = \frac{\text{Antal fødsler: Kvinder mellem } x \text{ og } x + k - 1 \text{ år}}{\text{Risikotid: Kvinder mellem } x \text{ og } x + k - 1 \text{ år}}. \quad (\text{K4.4})$$

Aldersspecifikke fertilitetsrater gør det muligt at undersøge fertilitetsudviklingen for enkelte aldersgrupper. Når kvinder og par vælger at få børn senere, vil det således vise sig ved, at fertilitetsraten falder i yngre aldersgrupper og ligeledes stiger i ældre aldersgrupper. Figur 5 viser udviklingen af de aldersspecifikke fertilitetsrater i Danmark for 5-års aldersgrupper mellem 1973 og 2023. Her fremgår det, at fertiliteten blandt kvinder i de yngre aldersgrupper (15-19 år og 20-24 år) er faldet markant. Samtidig er fertiliteten blandt de ældre aldersgrupper (30-44 årlige) steget betydeligt. Dette afspejler tendensen, at kvinder og par udsætter tidspunktet for familiestiftelse.

2.2 Den samlede fertilitet

Hvis man ønsker at beskrive fertiliteten i en befolkning med et enkelt tal, benyttes ofte *den samlede fertilitet*. Denne beregnes ud fra de aldersspecifikke fertilitetsrater. Fortolkningen af den samlede fertilitet er det antal levendefødte børn,

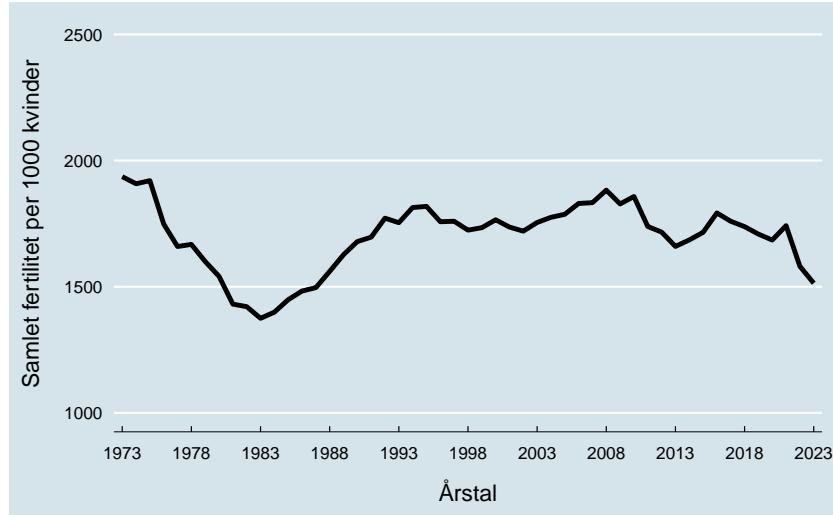


Figur 5: Udviklingen i aldersspecifikke fertilitetsrater siden 1973 i Danmark.
Kilde: statistikbankens register FOD, BEFOLK2.

som 1.000 kvinder vil føde i løbet af deres fertile alder (15-49 år), hvis hypotetisk ingen af kvinderne dør før de fylder 50 år, og under antagelsen at de 1.000 kvinder i hvert aldersinterval føder, som det var tilfældet i den givne kalenderperiode. Den samlede fertilitet er således et mål for fertiliteten i en hypotetisk kohorte i stil med dødelighedsmål for hypotetiske tabelbefolkninger (Kapitel 3). For 5-års aldersintervaller er den samlede fertilitet given ved følgende formel:

$$f_{samlet} = 5 \cdot f_{15} + 5 \cdot f_{20} + \dots + 5 \cdot f_{45}. \quad (\text{K4.5})$$

Formel (K4.5) viser, at den samlede fertilitet beregnes som summen af de aldersspecifikke fertilitetsrater multipliceret med aldersintervallets længde. Der multipliceres med aldersintervallets længde, da kvinderne kan føde børn i alle årene i et givent aldersinterval. Udviklingen i den samlede fertilitet siden 1973 kan ses i Figur 6. Tal for den samlede fertilitet nævnes ofte i den offentlige debat om fertilitet i Danmark og andre steder. Det er nemlig tal for den samlede fertilitet, der angiver, at kvinder skal føde 2,1 børn i gennemsnit for, at den danske befolkning kan reproducere sig selv. Tanken bag tærsklen 2,1 børn per kvinde er enkel: Hvis hver kvinde føder mindst en pige, som senere i livet føder mindst en pige, og så videre, kan befolkningen reproducere sig selv. To børn i gennemsnit per kvinde er dog ikke helt nok, fordi sandsynligheden for at få en pige er en smule mindre end for at få en dreng, og på grund af død (Figur 7).



Figur 6: Udviklingen i den samlede fertilitet siden 1973 i Danmark. Kilde: statistikbankens register FOD, BEFOLK2.

2.3 Produktionstal

Om fertiliteten er tilstrækkelig til, at befolkningen kan reproducere sig selv, kan undersøges nærmere med hjælp af *bruttoreproduktionstal*¹ og *nettoreproduktionstal*². Bruttoreproduktionstallet beregnes på samme vis, som den samlede fertilitet, men kun med nyfødte piger i tælleren. Definitionen er derfor antallet af levendefødte piger, som vil blive født af 1.000 kvinder i løbet af deres fertile alder (15-49 år), givet at ingen af kvinderne dør før de fylder 50 år, og at de i hvert aldersinterval føder børn per periodens aldersspecifikke fertilitetsrater. For 5-års aldersintervaller beregnes bruttoreproduktionstallet (BRT) med følgende formel:

$$BRT = 5 \cdot {}_5f_{15}^{\text{piger}} + 5 \cdot {}_5f_{20}^{\text{piger}} + \dots + 5 \cdot {}_5f_{45}^{\text{piger}}. \quad (\text{K4.6})$$

Her betegner symbolet ${}_5f_x^{\text{piger}}$ raten af pigefødsler hvor moren var mellem x og $x + 4$ år gammel. Kender man ikke antal pigefødsler, men kun det samlede antal fødsler, altså

$${}_5f_x = {}_5f_x^{\text{piger}} + {}_5f_x^{\text{drenge}}, \quad (\text{K4.7})$$

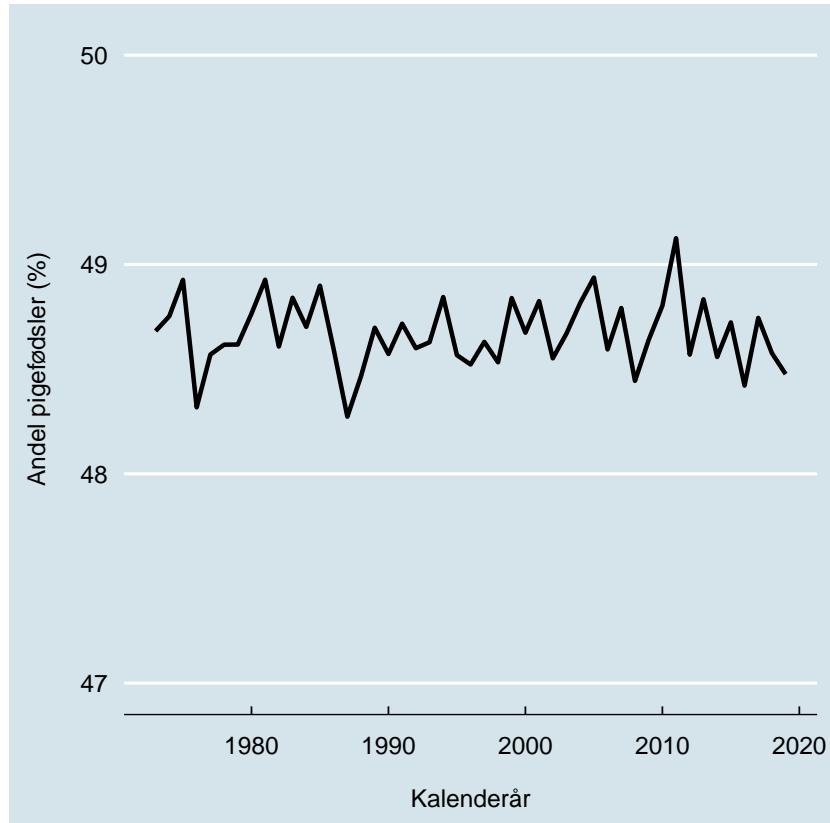
kan man bruge følgende approksimationsformel for forholdet mellem antal pigefødsler og samlet antal fødsler:

$$c = \frac{{}_5f_x^{\text{piger}}}{{}_5f_x} \approx 0,485. \quad (\text{K4.8})$$

¹Engelsk: Gross Reproduction Rate

²Engelsk: Net Reproduction Rate

Denne approksimationsformel bygger på, at forholdet mellem antal pige- og drengefødsler er det samme uanset moderens alder. Figur 7 viser at 48,5% er den gennemsnitlige andel pigerfødsler blandt alle fødsler.



Figur 7: Udviklingen i andelen af pigefødsler siden 1973 i Danmark. Kilde: statistikbankens register FOD, BEFOLK2.

Bruger man formel (K4.8), ses følgende relation mellem den samlede fertilitet og bruttoreproduktionstallet:

$$\text{BRT} = c \cdot f_{\text{samlet}}.$$

Nettoreproduktionstallet er et mere komplekst mål. Her tages der højde for, at kvinderne godt kan dø, inden de fylder 50 år, ligesom det er tilfældet i den virkelige verden. Nettoreproduktionstallet defineres således som antallet af levendefødte piger, som vil blive født af 1.000 kvinder i løbet af deres fertile alder (15-49 år), givet at kvinderne dør fra alder 0 i overensstemmelse med overlevelsestavlen (for kvinder) for den givne periode, og at kvinderne i hvert aldersinterval føder børn per periodens aldersspecifikke fertilitetsrater. For 5-års

aldersintervaller er nettoreproduktionstallet (NRT) givet ved følgende formel:

$$NRT = 5f_{15}^{\text{piger}} \frac{5L_{15}}{\ell_0} + 5f_{20}^{\text{piger}} \frac{5L_{20}}{\ell_0} + \dots + 5f_{45}^{\text{piger}} \frac{5L_{45}}{\ell_0}. \quad (\text{K4.9})$$

Her er ℓ_0 overlevelsestavlens radix og kL_x den samlede gennemlevede tid beregnet i overlevelsestavlens tabelbefolning (Kapitel 3). Fortolkningen af NRT er antal piger, en kvinde i gennemsnit vil føde i løbet af de fødedygtige aldre, hvis fødsels- og dødsraterne er, som observeret i perioden.

2.4 Eksempel

Vi henter antal fødsler fra statistikbankens register FODIE og risikotid fra mødrene mellem 15 og 49 år fra FOLK1A fra 2020 og grupperer dem i 5-års intervaller efter moderens alder:

```
f2020 <- hent_fertilitetsrate(2020)
f2020
```

```
# A tibble: 7 × 4
  aldersinterval    TID      R Fødsler
  <fct>           <dbl>    <dbl>    <dbl>
1 15-19            2020 166521     274
2 20-24            2020 184282     4694
3 25-29            2020 196845    20771
4 30-34            2020 177779    22773
5 35-39            2020 161606     9987
6 40-44            2020 179062    2268
7 45-49            2020 196756     157
```

Med disse data beregner vi de aldersspecifikke fertilitetsrater:

```
options(pillar.sigfig = 5)
f2020 <- mutate(f2020,frate = 1000*Fødsler/R)
f2020
```

```
# A tibble: 7 × 5
  aldersinterval    TID      R Fødsler     frate
  <fct>           <dbl>    <dbl>    <dbl>    <dbl>
1 15-19            2020 166521     274  1.6454
2 20-24            2020 184282     4694 25.472
3 25-29            2020 196845    20771 105.52
4 30-34            2020 177779    22773 128.10
5 35-39            2020 161606     9987  61.798
6 40-44            2020 179062    2268  12.666
7 45-49            2020 196756     157   0.79794
```

Vi ser for eksempel, at fertilitetsraten i Danmark i 2020 var 106 fødsler per 1000 personår blandt kvinder mellem 25 og 29 år. Vi bruger formel (K4.5) og beregner den samlede fertilitet:

```
summarize(f2020,samlet_fertilitet = sum(frate*5))
```

```
# A tibble: 1 × 1
  samlet_fertilitet
  <dbl>
1 1680.0
```

Den samlede fertilitet var altså 1680,0 fødsler per 1000 kvinder i Danmark i 2020. Det betyder, at i en hypotetisk befolkning, hvor fødselsraterne er, som de var blandt danske kvinder mellem 15 og 49 år i 2020 i Danmark, og hvor ingen dør, får hver kvinde i gennemsnit cirka 1,68 børn i løbet af sine fødedygtige aldre.

For at beregne bruttoreproduktionstallet, gennemfører vi de samme R-koder, men bruger aldersspecifikke pigefødselsrater i stedet for aldersspecifikke fødselsrater:

```
pige2020 <- hent_fertiliteratsrate_data(2020,barnkon = "Piger")
pige2020 <- mutate(pige2020,frate_piger = 1000*Fødsler/R)
summarize(pige2020,BRT = sum(frate_piger*5))
```

```
# A tibble: 1 × 1
  BRT
  <dbl>
1 821.
```

Vi ser at bruttoreproduktionstallet var 0,82 pigefødsler per kvinde i 2020. I en hypotetisk befolkning, hvor pigefødselsrater var ligesom i 2020, og ingen kvinde dør i aldersperioden fra 15 år til 49 år, vil en kvinde i gennemsnit føde 0,82 piger. Dette ligger allerede tydeligt under 1 og hvis man tager højde for at kvinder kan dør, bliver tallet endnu mindre. For at beregne nettoreproduktionstallet henter vi også data fra statistikbankens register DOD, beregner overlevelsestavlen (Kapitel 3) og samlar de aldersspecifikke risikotider (kolonne L) og pigefødselsrater (kolonne frate_piger):

```
fx <- fertilitets_tavle(tid = 2020)
select(fx,aldersinterval,L,frate_piger)
```

```
# A tibble: 7 × 3
  aldersinterval      L frate_piger
  <fct>            <dbl>      <dbl>
1 15-19             498117    0.000805
2 20-24             497681    0.0127
```

3 25-29	497203	0.0512
4 30-34	496478	0.0627
5 35-39	495585	0.0303
6 40-44	494071	0.00613
7 45-49	491847	0.000381

Nu er det enkelt at beregne nettoreproduktionstallet med formel (K4.9)

```
fx <- mutate(fx,bidrag_NRT=frate_piger*L/100000)
summarize(fx,NRT = sum(bidrag_NRT))
```

```
# A tibble: 1 × 1
  NRT
  <dbl>
1 0.815
```

Nettoreproduktionstallet var dermed 0,82 pigefødsler per kvinde i 2020. I en hypotetisk befolkning, hvor pigefødselsrater og mortalitetsrater var lige som i 2020, vil en kvinde i gennemsnit føde 0,815 piger.