

# Kapitel 2: Mortalitet og standardisering

Anna-Vera Jørring Pallesen, Johan Sebastian Ohlendorff, Laust  
Hvas Mortensen and Thomas Alexander Gerds

## 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 ligevægtssigning, som beskriver befolkningstilvæksten ud fra fire demografiske komponenter: fødsler, dødsfald, indvandring og udvandring. 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 familiedannelse. 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.

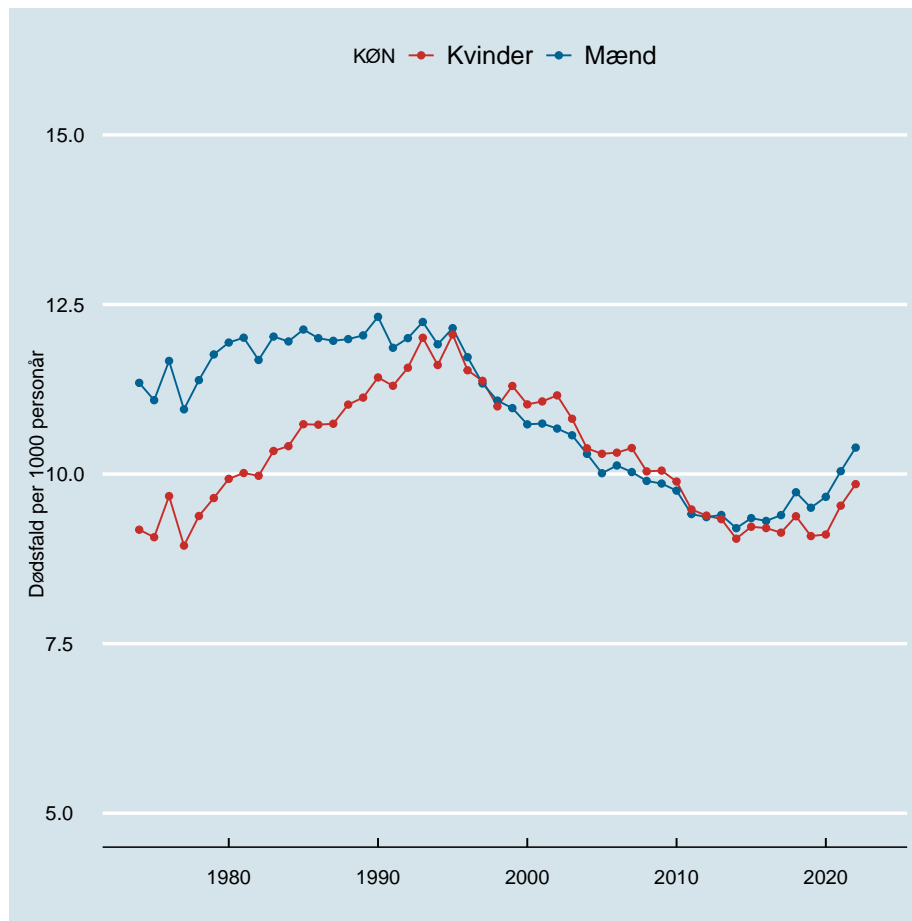


Figure 1: Udvikling i den summariske mortalitetsrate i perioden 1974-2022 i Danmark. Kilde: statistikbankens register DOD, BEFOLK2

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.

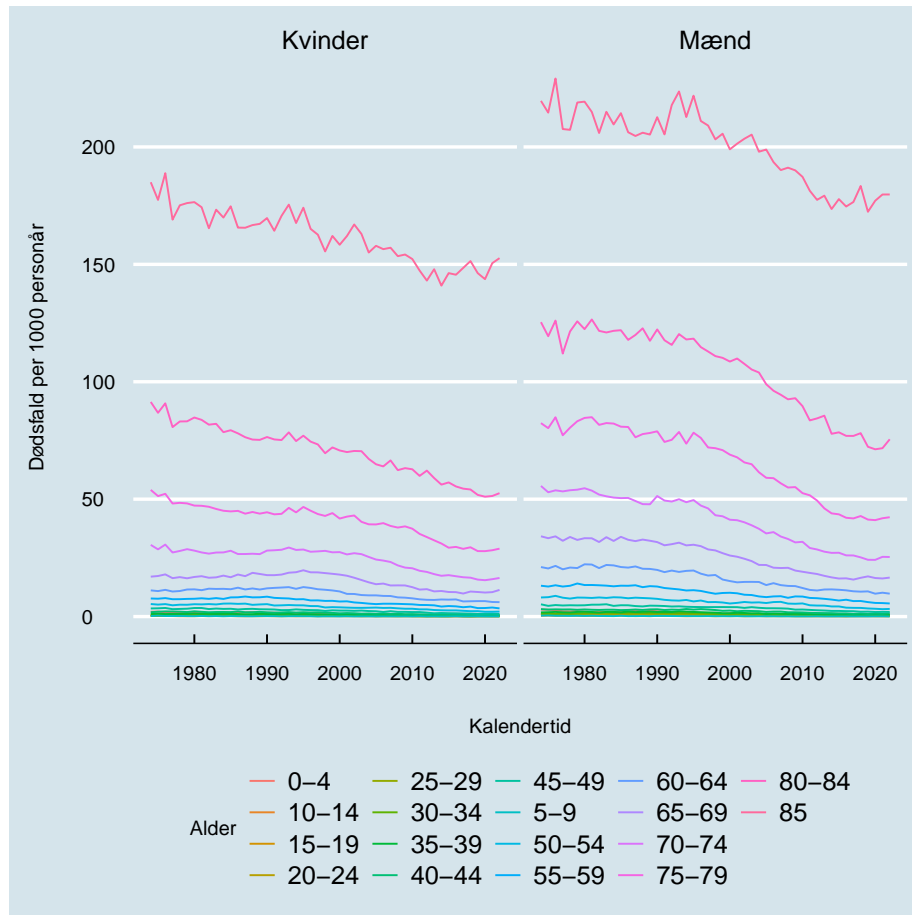


Figure 2: Udvikling i de køns- og aldersspecifikke mortalitetsrater i perioden 1974-2022 i Danmark. Kilde: statistikbankens register DOD, BEFOLK2.

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 overlevelsestavler, som vi beskriver 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.

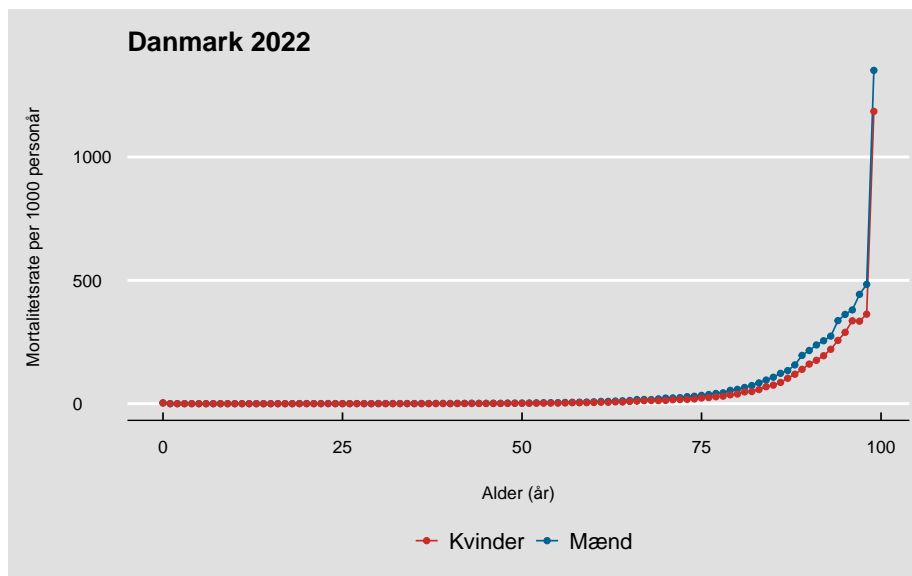


Figure 3: Aldersspecifikke mortalitetsrater fra hele den danske befolkning i 2022. Vi ser at dødeligheden blandt de ældre var højere for mænd end for kvinder.

## 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 aldere, der har folketal 0
folk <- subset(folk, alder<106)
# Mænd skal vises på venstresiden, derfor bliver INDHOLD negativt
folk_m <- subset(folk, KEN=="Mænd") %>% mutate(INDHOLD=-INDHOLD)
# For kvinder på højre siden er INDHOLD positivt
folk_k <- subset(folk, KEN=="Kvinder")
# Grafik
g <- ggplot(folk, aes(x = alder, y = INDHOLD, fill = KEN)) +
  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
```

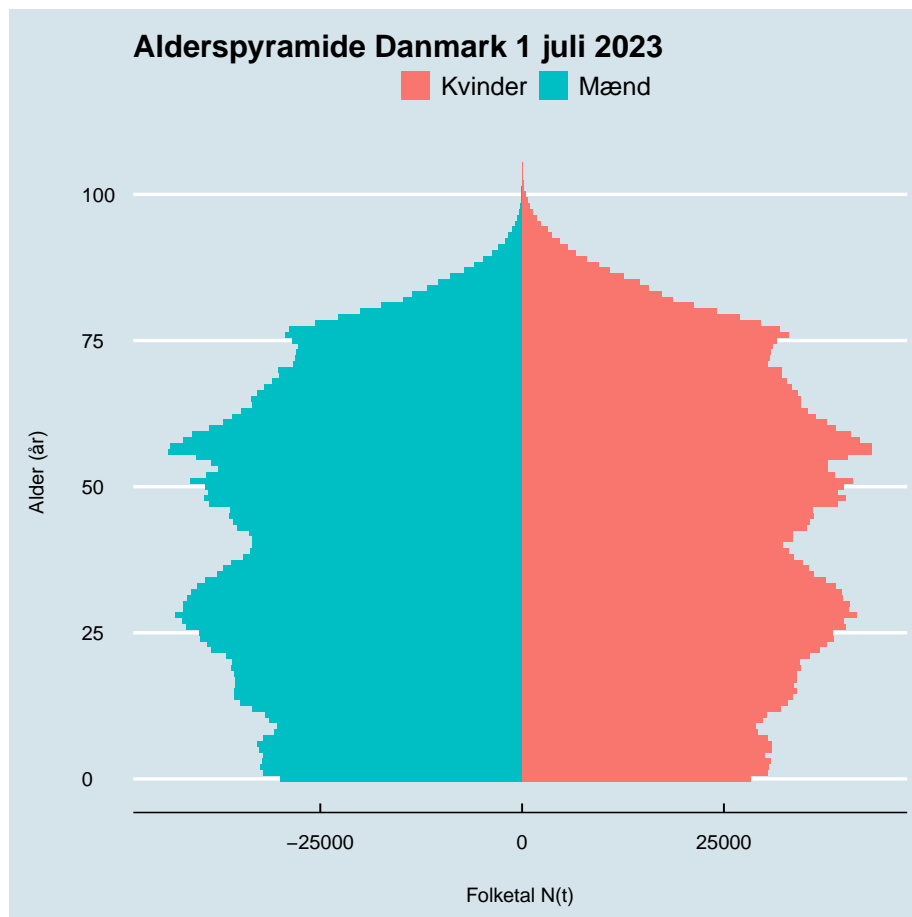


Figure 4: Data fra statistikbankens FOLK1a

### 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 parenteser er rund og inkluderet hvis parenteser er firkantede. 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:

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

I eksemplet fra afsnit 3.2.1 er der  $m = 4$  aldersgrupper og når vi indsætter tal i formelen 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 fortsat 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
<fct>           <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 befolkningen 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 aldersgruppe igennem kalendertid, og en fødselskohorte igennem både kalendertid og alder.

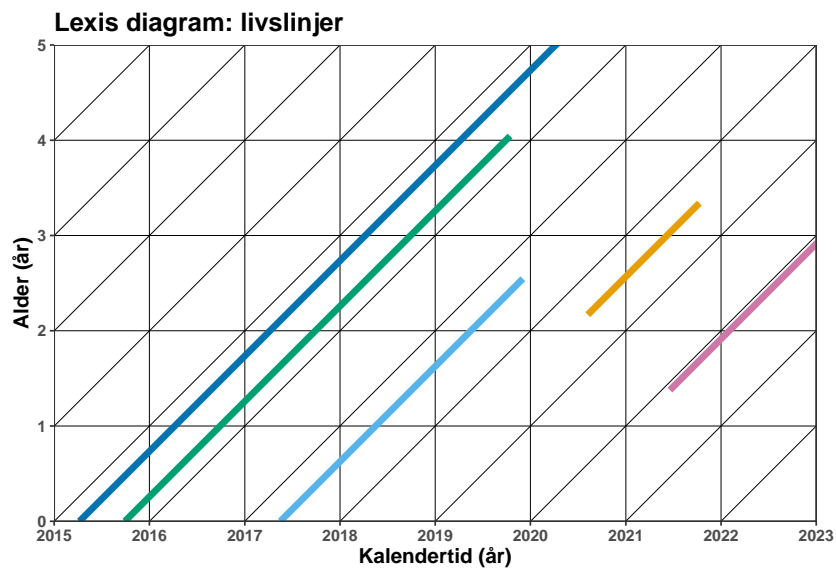


Figure 5: Figuren viser 5 personers livslinjer i (den nederste del af) et Lexis diagram. Livslinjer der ikke starter i alder '0' repræsenterer immigranter og livslinjer som stopper repræsenterer enten dødsfald eller emigranter.

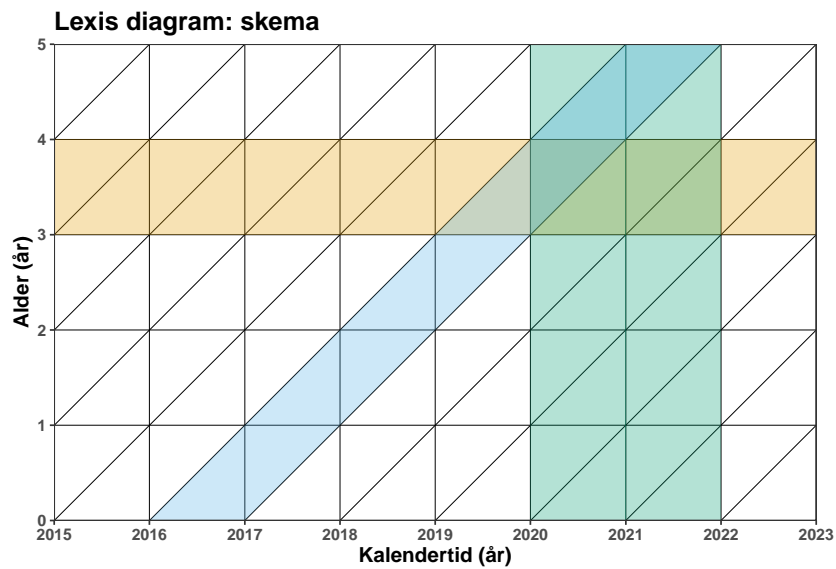


Figure 6: I et Lexis diagram kan man følge en aldersgruppe igennem kalendertid (gul) eller en fødselskohorte igennem både alder og kalendertid (blå). Det grønne område viser en kort kalenderperiode.

## 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 persons 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 eksemlet 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 (afnit 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 overskuelig 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","KEN"))
```

```
# 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()
```

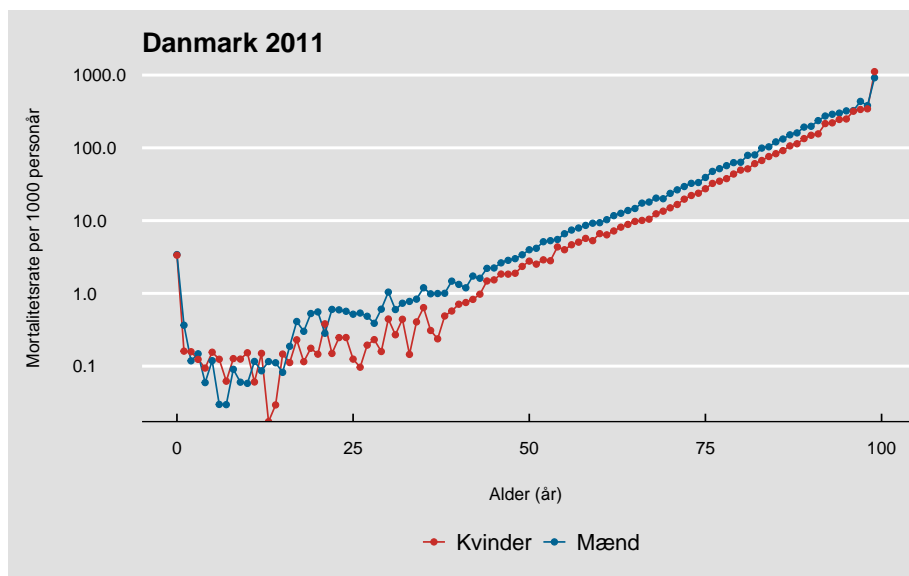


Figure 7: Aldersspecifikke mortalitetsrater for hele den danske befolkning i 2011. Vi ser at dødeligheden var højere for mænd i de fleste aldersgrupper.

## 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 formelen 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 formelen 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 sammen 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  $A$ .

Beregningsen 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 standardbefolkningen, 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 af aldersfordeling af studiebefolkning  $A$  og  $\text{SMR}(B, S)$  afhænger af 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:

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