Mortalitet og standardisering

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

1 Mortalitet

SKRIV HER Anna-Vera og Laust

1.1 Den demografiske transition

Udviklingen fra et traditionelt samfund med et højt niveau for dødelighed og fertilitet til et samfund præget af et lavt niveau på begge disse områder betegnes som den demografiske transition. Nedgangen af dødeligheden er det element af transitionen, der kommer først, fertilitetsnedgangen følger bagefter.

2 Standardiserede rater

Vi indfører nu begreberne *aldersspecifikke rater* og *standardiserede rater*. Disse er forskelligt fra de summariske rater fra Kapitel 1. I det følgende betegner vi derfor alle de rater, som Kapitel 1 har omtalt uden prædikat, med prædikatet *summarisk*. *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 begrænsingen med de summariske rater når det kommer til sammenligning af forskelige befolkninger.

2.1 Sammenligning af summariske rater

Som udgangspunkt har det begrænset interesse at sammenligne forskelige befolkningers summariske rater. Det er især problematisk når befolkningerne, som man ønsker at sammenligne, har forskelige 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 de afhænger aldersfordelingen. 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 kvindelig danske befolkning og også i den mændlige danske befolkning.

```
library(danstat)
library(tidyverse)
# risikotid i 2011 baseret på middelfolketal metode 1
# middelfolketal fra K3 bliver ganget med 1 år
x <- get_data("FOLK1a",
                   variables=list(list(code="tid",values="2011K3"),
                                  list(code="køn",values=c(2,1))))
# fjern TID fordi den er konstant
x$TID <- NULL
# aendre variable navn fra INDHOLD til RisikoTid
x <- rename(x,"RisikoTid"="INDHOLD")
# number of doedsfald i 2011
d <- get_data("DOD",variables=list(list(code="tid",values="2011"),
                                   list(code="køn",values=c("K","M"))
    ))
# fjern TID fordi den er konstant
d$TID <- NULL
# navngivning af variable
d <- rename(d,"Doed"="INDHOLD")
# join
dat <- left_join(x,d,by="KŒN")
# summariske mortalitetsrater per 1000 personaar
dat <- mutate(dat,"Summariske mortalitetsrate"=1000*Doed/RisikoTid)
dat
```

```
# A tibble: 2 × 4
  KØN   RisikoTid  Doed `Summariske mortalitetsrate`
  <chr>     <dbl> <dbl>                        <dbl>
1 Women   2806716 26577                         9.47
2 Men     2760140 25939                         9.40
```

Vi ser at den summariske mortalitetsrate i året 2011 var 9,47 døde per 1000 personår for danske kvinder og 9,39 døde per 1000 personår for danske mænd. På første blik strider dette resultat imod den gængse viden at danske kvinder lever længere end danske mænd. Det er problemet som eksemplet illustrerer: Fordi dødeligheden stiger med alder og fordi der er flere kvinder med høj alder end mænd med 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. Da kvinder lever længere end mænd, er der flere ældre kvinder end ældre mænd og det forøger kvindernes summariske mortalitetsrate. Resultatet er dog helt korrekt, kvinderne havde en højre summariske mortalitetsrate end mænd i 2011. Det skyldes bar 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, den kan for eksempel være 20 km per time. Mortalitetsraten er hastigheden befolkningen dør med, den kan for eksempel være 9,39 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

3.1 Alderspyramide

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

```
library(ggplot2)
library(ggthemes)
## begge køn
folk <- get_data("FOLK1a",variables=list(
                          list(code="alder",values=0:125),
                          list(code="køn",values=1:2),
                          list(code="tid",values="2023K3")))
# formatere ALDER til numerisk
folk <- mutate(folk,ALDER=as.numeric(gsub(" year[s]?","",ALDER)))
# fjern tomme aldre
folk <- subset(folk,ALDER<106)
# separere køn
folk_m <- subset(folk,KŒN=="Men") %>% mutate(INDHOLD=-INDHOLD)
folk_k <- subset(folk,KŒN=="Women")
# plot
g <- ggplot(folk, aes(x = ALDER, y = INDHOLD, fill = KŒN)) +
  geom_bar(data=folk_m, stat = "identity")+
  geom_bar(data=folk_k, stat = "identity")+
  coord_flip() +
  theme_solarized_2()+ylab("Folketal N(t)")+xlab("Alder (år)")+
  theme(legend.title=element_blank())
g <- g+ggtitle("Alderspyramide Danmark 1 juli 2023")
g
```

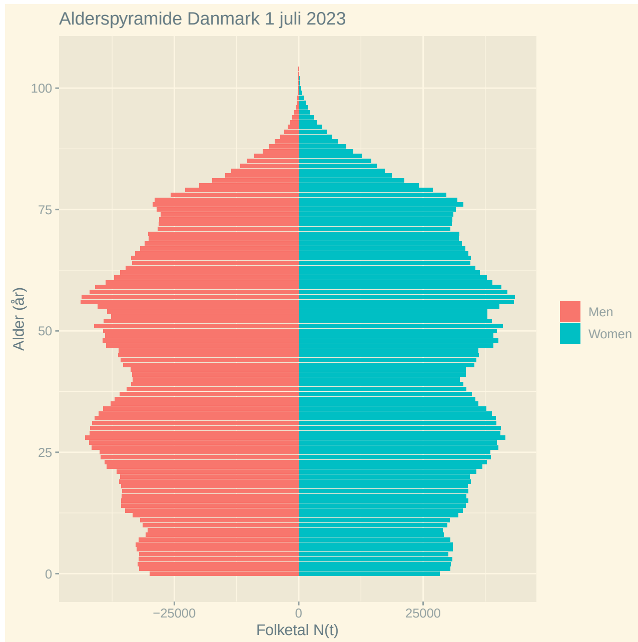

Figure 1: Data fra statistikbankens FOLK1a

3.2 Folketal i aldersgrupper

Aldersfordelingen af folketal angiver hvor mange personer i en befolkning har en bestemt alder, for alle aldre. Det kan den enten gør i absolutte antal, eller som procent i forhold til antal personer i hele befolkingen. 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 være lige stor. Da alle personers aldre ændrer sig hele tiden, skal man angive det dato, som aldersfordelingen referer til. For eksempel kan vi tale om aldersfordeling af kvinder i Danmark den 8 marts 1910 og om aldersfordeling af Fynens population den 1 juli 1989.

3.2.1 Eksempel

Vi finder aldersfordeling af folketal for hele den danske befolkning den 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 rundt 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 fir procenttal er nettop aldersfordelingen med hensyn til de fir intervaller.

```
library(danstat)
library(tidyverse)
## meta <- get_table_metadata("FOLK1a")
## meta$variables[3,]$values[[1]][-1,"id"]
folk <- get_data("FOLK1a",variables=list(
                            list(code="alder",values=0:125),
                            list(code="tid",values="2023K3")))
# formatere ALDER til numerisk
folk <- mutate(folk,ALDER=as.numeric(gsub(" year[s]?","",ALDER)))
# 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
    ))
# procent
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 i formler

En hver definition af aldersintervaller opdeler en befolkning i aldersgrupper. For $x = 1, \ldots, m$ aldersgrupper betegner vi med $N_x(t)$ folketal i aldersgruppe x til kalendertid t. Vi betegner fortsæt med $N(t)$ folketal i hele befolkningen til kalendertid t og udtrykker det som sum af folketal 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 folketal som sum af de aldersspecifikke folketal:

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

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

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

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

$$\text{Aldersfordeling} = \left(\frac{N_1(t)}{N(t)}, \ldots, \frac{N_m(t)}{N(t)}\right).$$

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 gør det enkelt bruger vi inddelingen af befolkningen i de 4 aldersgrupper fra afsnit 3.2.1. Vi henter folketal data fra statistikbankens register BY2 hvor man kan specificere bystørrelse.

```
library(tidyverse)
library(danstat)
## meta <- get_table_metadata("BY2")
b2 <- get_data("BY2",variables=list(
                      list(code="alder",values=0:125),
                      list(code="BYST",values=c("HOVEDS","LAND")),
                      list(code="tid",values="2023")))
# formatere ALDER til numerisk
b2 <- mutate(b2,ALDER=as.numeric(gsub(" year[s]?","",ALDER)))
# 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 højst i hovedstadsområdet og andelen af mennesker mellem 50 og 75 er højst i landdistrikterne.

3.3 Risikotid i aldersgrupper

Med hensyn til mortalitetsrater, har vi brug for aldersfordeling af risikotid 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 befolkingen 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 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 risikotid i aldersgruppen med risikotid i hele befolkningen i perioden $[t_1, t_2]$ og betegner den med V_x:

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

Risikotid beregnes ofte ved at gange middelfolketal med periodens længde. I den særlige situation, hvor perioden er 1 år lang, altså når $t_2 - t_1 = 1$ år, har middelfolketal (antal) og risikotid (personår) den samme værdi, men forskelige endheder. Vi skal bruge V_x som vægte i definitionen af aldersstandardiserede rater (afsnit 5).

3.3.1 Eksempel

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

```
library(danstat)
library(tidyverse)
folk23 <- get_data("FOLK1a",variables=list(
                              list(code="alder",values=0:125),
                              list(code="tid",values=c("2022K1","2023
    K1"))))
```

```r
# formatere ALDER som numerisk variable
folk23 <- mutate(folk23,ALDER=as.numeric(gsub(" year[s]?","",ALDER)))
# Risikotid= 1* Middelfolketal metode 2
folk23 <- folk23 %>% group_by(ALDER) %>% summarise(Risikotid=1*mean(
    INDHOLD))
# Aldersintervaller
folk23 <- mutate(folk23,Aldersinterval=cut(ALDER,
                                          breaks=c(0,25,50,75,125),
                                          include.lowest = TRUE))
# antal personår i de 4 aldersintervaller
af23 <- folk23 %>% group_by(Aldersinterval) %>% summarise(Personår=
    sum(Risikotid))
# aldersfordeling i procent
af23 <- af23 %>% mutate(Procent=100*Personår/sum(Personår))
af23
```

```
# 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 Lexis diagram

Et Lexis diagram visualiserer sammenhæng mellem kalendertid (vertikal) og alder (horisontal). Hver person er repræsenteret af sin livslinje (Figur 2). I en *lukket befolkning* (hvor ind- og udvandring ikke forkommer) starter alle livslinjer i fødselsdagen hvor personen er 0 år gamle og ender i dødsdatoen den alder personen har livet til. I en åben befolkning, starter livslinjer for immigranter den dag de immigrerer og slutter for emigranter den dag de emigrerer.

Figur 2 viser 5 personers livslinjer fra en åben befolkning. Den mørkeblå linje representerer en person som bliver født i foråret 2015 og forbliver i befolkingen indtil foråret 2020 hvor lexis diagrammet slutter. Lexis diagrammet kan også bruges til at forklare forskelen mellem kohorteprincippet (man følger en fødselskohorte i en relativt lang periode) og kalenderårsprincippet (man studerer en befolkning i en kort periode). Figur 3 viser et lexis diagram 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.

4 Aldersspecifikke mortalitetsrater

Vi ser på en befolkning i en kalenderperiode $[t_1, t_2]$ og inddeler den i $\{x = 1, \ldots, m\}$ aldersgrupper. Vi betegner med $D_x[t_1, t_2]$ antal dødsfald i perio-

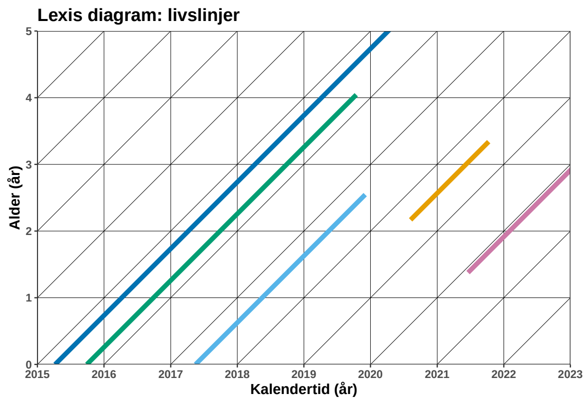

Figure 2: Figuren viser 5 personers livslinjer i (den nederste del af) et Lexis diagram. Livslinjer der ikke starter i alder ’0’ representerer immigranter og livslinjer som stopper representerer enten dødsfald eller emigranter.

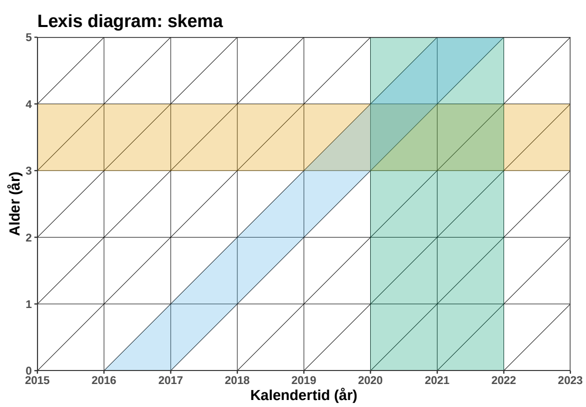

Figure 3: I et Lexis diagram kan man følge en aldersgruppe igennem kalendertid (gul) eller en fødselskohorte igennem både alder og kalendertid (blå). Det grønne område viser en kort kalenderperiode.

den 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 ratio mellem antal dødsfald og risikotid.

$$\text{Aldersspecifikke mortalitetsrate:} \qquad M_x = \frac{D_x}{R_x}, \qquad x = 1, ..., m.$$

Bemærk at den aldersspecifikke mortalitetsrate M_x afhænger også kalenderperioden og den langform notation er $M_x[t_1, t_2]$.

4.0.1 Eksempel

Vi finder antal dødsfald for hele den danske befolkning i perioden mellem den 1 januar 2022 og den 1 januar 2023 og beregner det summariske antal 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 years and over".

```
library(danstat)
library(tidyverse)
library(tibble)
meta <- get_table_metadata("dod",variables_only=TRUE)
agevals <- filter(meta,id=="ALDER")[["values"]][[1]][["id"]][-1]
dd23 <- get_data("dod",variables=list(
                      list(code="alder",values=agevals),
                      list(code="tid",values=c("2022"))))
# formatere ALDER som numerisk variable
dd23 <- mutate(dd23,ALDER=as.numeric(gsub(" year[s]?| years and over"
    ,"",ALDER)))
# Aldersintervaller
dd23 <- mutate(dd23,Aldersinterval=cut(ALDER,
                                    breaks=c(0,25,50,75,125),
                                    include.lowest = TRUE))
# antal døde i de 4 aldersintervaller
group_dd23 <-dd23 %>% group_by(Aldersinterval) %>% summarise(antal_dø
    de=sum(INDHOLD))
group_dd23
```

```
# 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. Det gør vi med et left-join:

```
x <- left_join(af23,group_dd23,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
```

4.1 Sammenligning af aldersspecifikke mortalitetsrater

For at sammenligne mortalitet i to befolkninger (vi kalder dem studiebefolkning A versus befolkning B) kan man sammenligne de aldersspecifikke mortalitetsrater mellem de to befolkninger (M_x^A versus M_x^B). Det giver lige så mange resultater som der er aldersintervaller, altså et resultat for hver aldersgruppe (Figur 4). 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 forskelen i en figur som viser de aldersspecifikke mortalitetsrater af de to befolkninger ved siden af hinanden.

4.1.1 Eksempel

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

5 Aldersstandardisering

Formålet med alderstandardisering er at sammenligne mortalitetsrater (og andre rater) mellem to eller flere befolkninger, som har forskelige aldersfordelinger. Den overordnede ide er at udskifte den rigtige aldersfordeling med en anden aldersfordeling og at beregne mortalitetsraten som den ville havde 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 typisk ikke så vigtigt hvilken aldersfordeling man vælger. For eksempel, kan vi spørge hvor meget højere er mortalitetsraten blandt danske mænd sammenlignet med danske kvinder hvis aldersfordeling havde været den samme blandt mænd og kvinder. Vi mangler kun at specificere den aldersfordeling som de standardiserede rater skal have i fælles. Her er

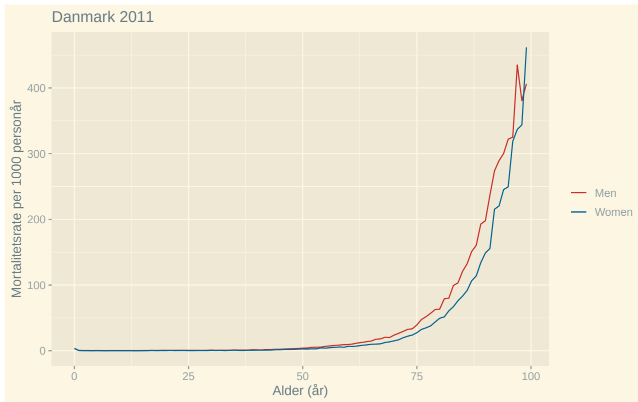

Figure 4: Figuren viser aldersspecifikke mortalitetsrater fra hele den danske befolkning i 2011. Vi ser at dødeligheden var højre for mænd for alle aldrer undtagen aldersgruppe 99+

der umiddelbart flere forskelige muligheder: aldersfordeling bland mænd, aldersfordeling bland kvinder, aldersfordeling bland alle dansker uanset køn, og en helt tredje 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 metoderne direkte versus indirekte standardisering.

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

$$V_x = \frac{R_x}{R}$$

hvor R betegner befolkningens total risikotid i perioden. Vi omskriver formlen for den aldersspecifikke mortalitetsrate sådan at antal dødsfald i aldersgruppen står isoleret:

$$D_x = M_x R_x.$$

Vi betegner fortsæt med M befolkningens summariske mortalitetsrate og med D antal dødsfald i perioden. Det følgende regnestykke viser at M er et vægtet gennemsnit af M_x hvor vægtene er aldersfordelingen af risikotid.

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

I afsnit 2.1 har vi diskuteret at forskelen mellem kvinders og mænds summariske mortalitetsrater skyldes ikke kun kønsforskellen af mortalitetsrater men også kønsforskellen af aldersfordelinger. Kitagawas dekomposition viser dette klart og mere generel 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 (1) 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:

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

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, vil man indse hvorfor Kitagawas komposition holder. For hvert aldersinterval x gælder

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

5.2 Direkte standardisering

Formålet med den såkaldte direkte standardisering er at sammenligne mortalitetsrater mellem to befolkninger uanset forskele i aldersfordeling.

Vil man sammenfatte forskellen i kun et tal, kan man bruge direkte standardisering. Vi fortolker den

$$\text{direkte standardiserede mortalitetsrate} = \sum_{x=1}^{m} M_x^A V_x^S,$$

som den mortalitetsrate vi ville havde set i studiebefolkning A, hvis aldersfordeling af risikotid havde været den samme som i referencebefolkning S. Vi kalder denne mortalitetsrate den standardiserede mortalitetsrate for studiebefolkning B med hensyn til studiebefolkning A som standard population. Ideen er at vi nu kan direkte sammenligne den standardiserede mortalitetsrate fra studiebefolkning B med den summariske mortalitetsrate fra befolkning A :

$$\sum_{x=1}^{m} M_x^B V_x^A \text{ med } \sum_{x=1}^{m} M_x^A V_x^A.$$

Her har vi brugt aldersfordeling af risikotid fra studiebefolkning A som reference. Vi kan ligeledes bruge aldersfordeling af risikotid fra studiebefolkning B eller en helt anden befolkning som reference. Hvis vi bruger en helt tredje befolkning som reference, lad os kalde den referencebefolkning S, kan vi direkte sammenligne de to standardiserede mortalitetsrater:

$$\sum_{x=1}^{m} M_x^B V_x^S \text{ med } \sum_{x=1}^{m} M_x^A V_x^S.$$

Den hyppigste form af rapportere denne sammenligning mellem to standardiserede mortalitetsrater er det såkaldte standardiserede rate ratio:

$$\text{SRR} = \frac{\sum_{x=1}^{m} M_x^B V_x^S}{\sum_{x=1}^{m} M_x^A V_x^S}.$$

5.3 Indirekte standardisering

Formålet med den såkaldte indirekte standardisering er også at sammenligne mortalitetsraterne mellem to befolkninger. Man sammenligner det totale antal dødsfald i studiebefolkning A med det forventede antal døde i studiebefolkning A hvis (hypotetisk) de aldersspecifikke mortalitetsrater havde været lige som i en reference befolkning, vi kalder den igen referencebefolkning S. Er de forventede antal dødsfald højere, kan man konkludere, at den samlede dødelighed (det vil sige de aldersspecifikke mortalitetsrater samlet set) var højere i reference befolkningen end i studiebefolkning A.

Beregningen kræver kendskab til de aldersspecifikke mortalitetsrater i referencebefolkning S, de aldersspecifikke risikotider i studiebefolkning A og det totale antal dødsfald i studiebefolkning A. Det totale antal dødsfald i studiebefolkning A er given ved

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

Relativt til den totale risikotid R^A er det forventede antal døde hvis dødeligheden havde været lige som i referencebefolkning S given 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 referencebefolkning S er det såkaldte standardiserede mortalitetsratio:

$$\begin{aligned} \text{SMR} &= \frac{\sum_{x=1}^{m} M_x^A V_x^A}{\sum_{x=1}^{m} M_x^S V_x^A} \\ &= \frac{\sum_{x=1}^{m} M_x^A R_x^A}{\sum_{x=1}^{m} M_x^S R_x^A} \\ &= \frac{\sum_{x=1}^{m} D_x^A}{\sum_{x=1}^{m} M_x^S R_x^A} \\ &= \frac{\text{Observeret antal døde}}{\text{Forventet antal døde}} \end{aligned}$$

Den indirekte standardiserede mortalitetsrate i befolking A er given ved

$$SMR * M^S$$

5.4 Direkte versus indirekte standardisering

Direkte og indirekte standardisering er meget tæt beslægtet. Det ses når vi bruger aldersfordeling fra studiebefolkning A som reference i formlen for SRR:

Direkte standardisering kræver kendskab til aldersfordeling af risikotid i reference befolkningen (vi kalder den referencebefolkning S).

Kender man ikke aldersfordeling fra referencebefolkning S kan man ikke andvende direkte standardisering og dermed ikke beregne SRR. Hvis man tilgengæld kender de aldersspecifikke mortalitetsrater i befolkning S kan man i stedet for beregne SMR. Man kan dog ikke direkte sammenligne SMR for studiebefolkning A med SMR for studiebefolkning B.