

# Mortalitet og standardisering

## Mortalitet

SKRIV HER Anna-Vera og Laust

## Standardiserede rater

Vi indfører nu begreberne *aldersspecifikke rater* og *standardiserede rater*. Disse er forskelligt fra de summariske rater, der blev indført i 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. For at motivere aldersspecifikke og standardiserede rater starter vi med at forklare begrænsningen med de summariske rater.

## Sammenligning af summariske rater

Som udgangspunkt har det begrænset interesse at sammenligne forskellige befolkningers summariske rater. Det er især når befolkningerne, som man ønsker at sammenligne, har forskellige aldersfordelinger. Afhængig af formål med undersøgelsen kan det alligevel godt være at man vil sammenligne summariske rater, men det er vigtigt at man ved at de afhænger af aldersfordeling. Problemet er ret nemt at indse ved følgende eksempel.

## Eksempel

Vi beregner de summariske mortalitetsrater i året 2011 i den kvindelige danske befolkning og også i den mandlige 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
# ændre 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="KØN")
# summariske mortalitetsrater per 1000 personaar
dat <- mutate(dat,"Summariske mortalitetsrate"=1000*Doed/RisikoTid)
dat
```

KØN	RisikoTid	Doed	Summariske mortalitetsrate
Women	2806716	26577	9.469073
Men	2760140	25939	9.397712

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. 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 i 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. Dermed kan man fortolke mortalitetsraten som hastigheden befolkningen dør. Denne hastighed 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.

På første blik strider dette resultat imod den gængse viden at dansk kvinder lever længere end danske mænd. Det er problemet, som dette eksempel illustrerer: Fordi den kvindelig befolkning i Danmark er ældre end den mændlige, er den summariske mortalitetsrate højere for kvinder end for mænd. Den summariske mortalitetsrate afspejler 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, da dødeligheden vokser med alderen.

## Aldersfordeling

### 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. Man kan tydeligt se forskellen mellem mænd og kvinder i toppen af pyramiden. En mere sofistikeret og dynamisk version af

den danske alderspyramide findes her <https://extranet.dst.dk/pyramide/pyramide.htm>.

```
library(ggplot2)
library(ggthemes)
## begge køn
dt <- 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
dt <- mutate(dt, ALDER=as.numeric(gsub(" year[s]?", "", ALDER)))
# fjern tomme aldre
dt <- subset(dt, ALDER<106)
# separere køn
dt_m <- subset(dt, KEN=="Men") %>% mutate(INDHOLD=-INDHOLD)
dt_k <- subset(dt, KEN=="Women")
# plot
g <- ggplot(dt, aes(x = ALDER, y = INDHOLD, fill = KEN)) +
  geom_bar(data=dt_m, stat = "identity") +
  geom_bar(data=dt_k, stat = "identity") +
  coord_flip() +
  theme_solarized_2()+ylab("Folketal N(t)") + xlab("Alder (år)") +
  theme(legend.title=element_blank())
g <- ggtitle("Alderspyramide Danmark 1 juli 2023")
g
```

## Folketal i aldersgrupper

Aldersfordelingen af folketal angiver hvor mange personer i en befolkning har en bestemt alder, for alle aldre. Det kan den enten gøre i absolut 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 være lige stor. Da alle personers aldre ændrer sig hele tiden, skal man angive det dato som aldersfordeling 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.

Her er aldersfordeling af hele den danske befolkning den 1 januar 2023 inddelt i 4 intervaller.

```
library(danstat)
library(tidyverse)
## meta <- get_table_metadata("FOLK1a")
## meta$variables[3,]$values[[1]][-1,"id"]
dt <- get_data("FOLK1a", variables=list(
  list(code="alder", values=0:125),
  list(code="tid", values="2023K3")))
# formatere ALDER til numerisk
```

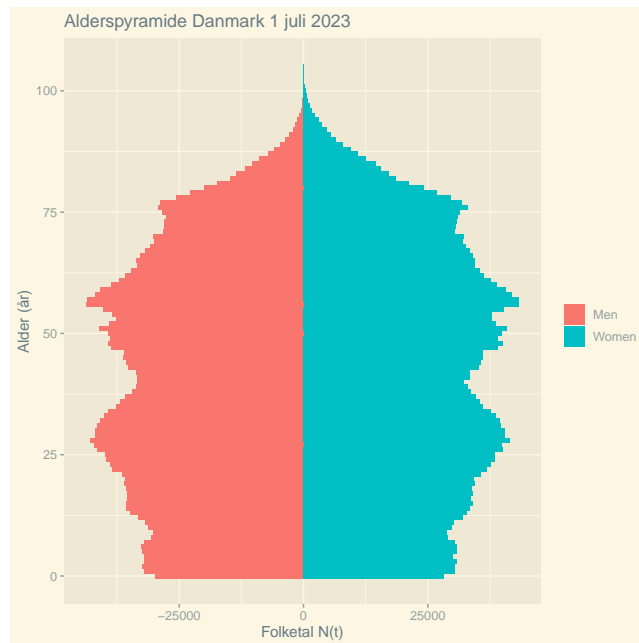


Figure 1: Data fra statistikbankens FOLK1a

```
dt <- mutate(dt, ALDER=as.numeric(gsub(" year[s]?", "", ALDER)))
# Aldersintervaller
dt <- mutate(dt, Aldersinterval=cut(ALDER,
                                   breaks=c(0,25,50,75,125),
                                   include.lowest = TRUE))
# antal person i de 4 aldersintervaller
af <- dt %>% group_by(Aldersinterval) %>% summarise(Antal=sum(INDHOLD))
# procent
af <- af %>% mutate(Procent=100*Antal/sum(Antal))
af
```

Aldersinterval	Antal	Procent
[0,25]	1742979	29.3
(25,50]	1882860	31.7
(50,75]	1778084	29.9
(75,125]	540222	9.09

Vi kan bruge denne inddeling af aldersspektrum i 4 grupper og sammenligne aldersfordeling mellem hovedstadsområdet og landdistrikter i Danmark i 2023. For at gøre det henter vi folketal data fra statistikbankens register BY2.

```
## 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

```

BYST	Aldersinterval	Antal	Procent
Greater Copenhagen Region	[0,25]	424524	31.1
Greater Copenhagen Region	(25,50]	520217	38.2
Greater Copenhagen Region	(50,75]	329994	24.2
Greater Copenhagen Region	(75,125]	88561	6.50
Rural areas	[0,25]	184556	26.8
Rural areas	(25,50]	198151	28.8
Rural areas	(50,75]	258161	37.5
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 unge mennesker er højst i hovedstadsområdet og andelen af mennesker mellem 50 og 75 er højst i landdistrikterne.

## Risikotid i aldersgrupper

En hver definition af aldersintervaller opdeler en befolkning i aldersgrupper. For  $x = 1, \dots, m$  aldersgrupper betegner vi med  $N_x(t)$  folketal i aldersgruppe  $x$  til tid  $t$ . Vi betegner fortsat med  $N(t)$  folketal i hele befolkningen til tid  $t$  og kan udtrykke den som sum af folketal i alle aldersgrupper:

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

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

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

Med hensyn til mortalitetsrater, har vi brug for aldersfordeling af risikotid i en bestemt kalendersperiode. Vi betegner med  $R_x[t_1, t_2]$  den samlede gennemlevede tid i perioden  $[t_1, t_2]$  af alle personer i aldersgruppe  $x$ . En person som har levet i befolkningen i perioden kan bidrage med risikotid til et eller flere aldersintervaller. Det sker for personer som har fødselsdag i perioden og den dag skifter fra aldersgruppe  $x$  til aldersgruppe  $x + 1$ .

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  $V_x$ :

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

Vi skal bruge  $V_x$  som vægte i definitionen af aldersstandardiserede rater.

## Alderstandardisering