

# Grundbegreber

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

## 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, familiedannelse, 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 eksempelvis 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 FSU 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 statistikbanken <sup>1</sup> var der 5.806.081 personer i Danmarks befolkning den 1. januar 2019:

$$\begin{aligned}\text{Befolkning} &= \text{Hele Danmark} \\ t_1 &= 2019-01-01 \\ N(t_1) &= 5.806.081\end{aligned}$$

På et senere tidspunkt vil befolkningens størrelse have ændret sig som følge af fødsler, dødsfald 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
```

---

<sup>1</sup><https://statistikbanken.dk/>

```
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))
```

	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 folketællinger helt tilbage fra 1769 (register FT).

### 3 Instruks til brugen af R koder

For at hente data fra statistikbanken bruger vi en R pakke, som hedder **danstat**. Desuden bruger vi en række andre R pakker og også 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 køre den følgende R-kode:

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

Det vil dog kun virke med adgang til Internet. For at bruge funktionerne også uden adgang til Internet kan I hente filen her

<https://github.com/tagteam/demogRafi/blob/main/R-scripts/demofunk.R>

og gemme den i projekt mappen **demografi** i undermappen **R-scripts**. Hvis det er sket, kan funktionerne køres ved følgende kommando

```
source("R-scripts/demofunk.R").
```

En vigtig funktion hedder **hent\_data**. Den henter data fra statistikbankens register. For at se hvilke registrer 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_ws()+ scale_colour_ws("colors6")
```

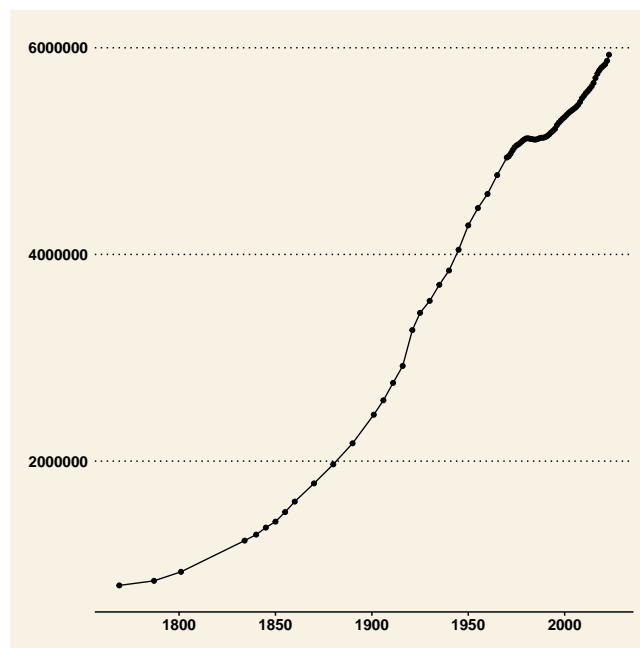
Alle R-koder som indgår i kapitel 1 kan også hentes som R-markdown fil her:  
[https://github.com/tagteam/demogRafi/blob/main/R\\_funk/demofunk.R](https://github.com/tagteam/demogRafi/blob/main/R_funk/demofunk.R)

## 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 Naatsorsueqqis-saartarfik (Grønlands Statistik) og Hagstova Føroya (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



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

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 Statistisk 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 datoerne 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 statistik 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 formelen:

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

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 folketal fra den 1. juli som årets middelfolketal i deres årlige rapporter om befolkningens udvikling <sup>2</sup>.

### 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 folketal:

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

---

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

### 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 folketal 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:

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

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 folketal i starten, midten og slutningen af perioden:

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

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

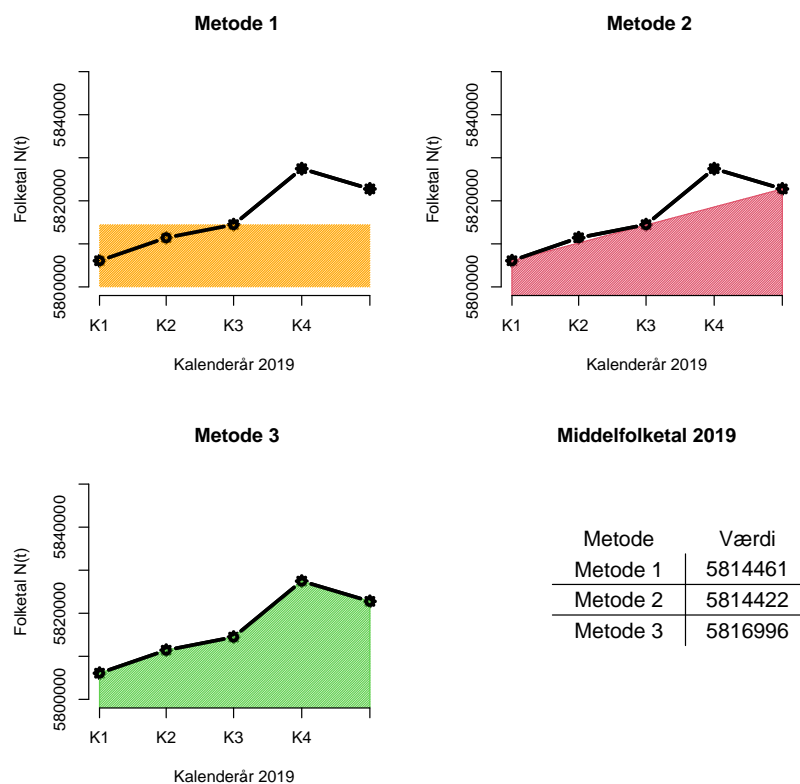
Det er nemmest at anvende metode 1. Med metode 1 er det estimerede middelfolketal  $\tilde{N}[1 \text{ januar } 2016] = 5.707.251$  personer. For at benytte metode 2 bruger vi R som lommerregner:

```
# 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:





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.

```
# 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, FOD, 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
```

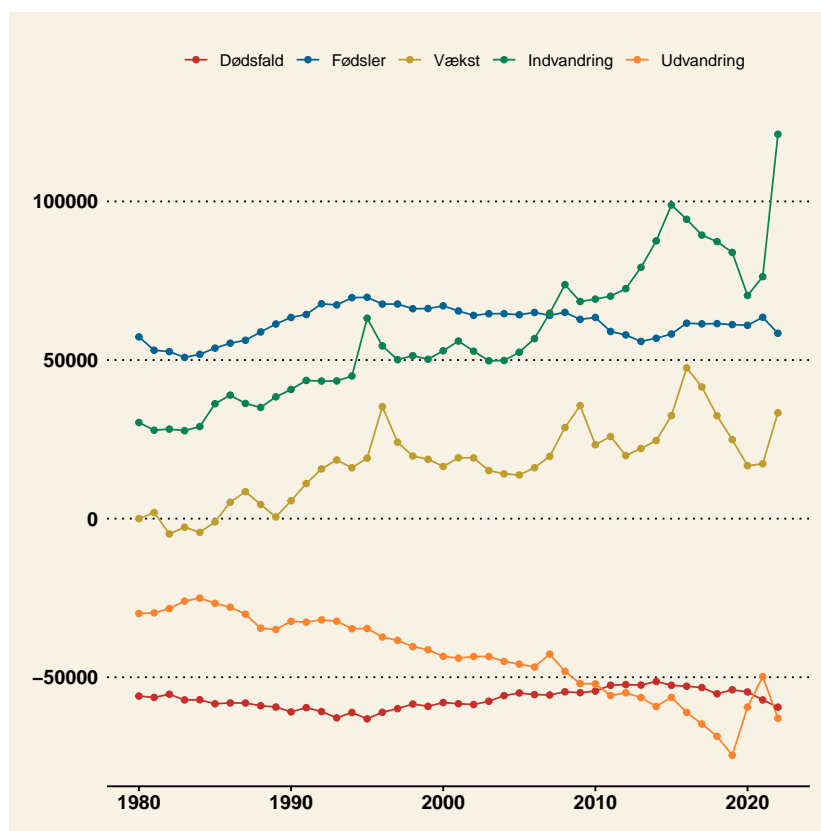
Baseret på disse tal beregner vi væksten i den danske befolkning i perioden til  $(5.932.654 - 5.873.420) = 59.234$  personer. Den naturlige vækst i perioden er negativ:  $(58.430 - 59.345) = -915$  personer og nettovandring i perioden positiv:  $(121.183 - 62.927) = 58.256$  personer. Vi ser, at ligevægtsligningen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

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

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 udvandring 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 dominerende faktor for ændringer af folketallet i denne periode, hvorimod 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=1980:2022)
V = V %>% mutate(INDHOLD = INDHOLD - c(INDHOLD[1],INDHOLD[-length(
  INDHOLD)]))
D = hent_data("dod",tid=1980:2022)
D = D %>% mutate(INDHOLD = -INDHOLD)
F = hent_data("FOD",tid=1980:2022)
I = hent_data("INDVAN",tid=1980:2022)
U = hent_data("UDVAN",tid=1980:2022)
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_ws() + scale_colour_ws("colors6") +
  theme(legend.title=element_blank())
```



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

## 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 størrelse. 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*.

### 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 interessere 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 person dage. 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 person dage. Det kræver dog, at man kender eksakte datoer for alle fødsler, dødsfald samt ind- og udvandring 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 indvandring, relativt til risikotid, altså antal gennemlevede personår i samme periode i en befolkning:

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

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 flytningsrater 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

Risikotiden 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 formelen 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 sider af den demografiske ligevægtsligning (formel (K1.4)) og dividerer på begge sider 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 middelfolketal 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
```

```

# vækstrate
Vrate <- 1000*vækst/risikotid
# naturlige vækstrate
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,Frater,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 nettovandringsraten var 9,9 personer per 1000 personår. Ligesom den demografiske ligevægtsligning ikke går op, gør denne formel heller ikke, og der er en fejlrate:

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