

# Grundbegreber

immediate

November 6, 2023

## Demografi

Demografi er studiet af menneskers befolkningsegenskaber og deres forandringer over tid og sted. Det inkluderer undersøgelser af befolkningens størrelse, aldersfordeling, kønsfordeling, etnicitet, sprog, religion, uddannelse, erhverv og andre relevante faktorer. Demografer anvender forskellige metoder, såsom dataindsamling og statistisk analyse, til at undersøge og beskrive befolkningstrends og -mønstre. Demografi spiller en vigtig rolle i samfundets planlægning og 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, planlægning af byer og samfund, arbejdsmarked og socialpolitik.

## 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 første 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 den 1 januar 2023, en befolkningsgruppe.

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 ofte til at forudsige og planlægge behovet for offentlige tjenester og politikker, såsom sundhedsvæsenet, uddannelse og beskæftigelse.

## 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 [www.statistikbanken.dk](http://www.statistikbanken.dk) var der 5.806.081 personer i Danmarks befolkning den 1 januar 2019:

Befolkning = Hele Danmark

$t_1 = 2019-01-01$

$N(t_1) = 5.806.081$

På et senere tidspunkt vil befolkningens størrelse have ændret sig som følge af fødsler, 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. Her er R-koder som kan bruges til at beregne antal dage mellem to datoer og ændring af folketal i perioden.

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

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

Folketallet ændrer sig jo hver dag, men det er svært at registrere og tælle nøjagtigt for enkelte dage. Heldigvis betyder små unøjagtigheder i tallene typisk ikke meget for demografiske konklusioner. Statistikbanken <sup>1</sup> angiver folketal i starten af hvert kvartal siden 2008 (FOLK1A), den første januar per år siden 1971 (BEFOLK1), den første januar per år siden 1901 (HISB3). Desuden er der summariske tal fra folketællinger hele vejen fra 1769 (FT). Disse tal bliver vist i Figure 1.

## Middelfolketal

Vi indfører nu begrebet middelfolketal, som bruges til at approksimere et gennemsnitlige folketal i en given tidsperiode. Middelfolketallet er en vigtig demografisk indikator, der indgår f.eks., i fertilitetsrater og dødelighedstavler.

---

<sup>1</sup>[www.statistikbanken.dk](http://www.statistikbanken.dk)

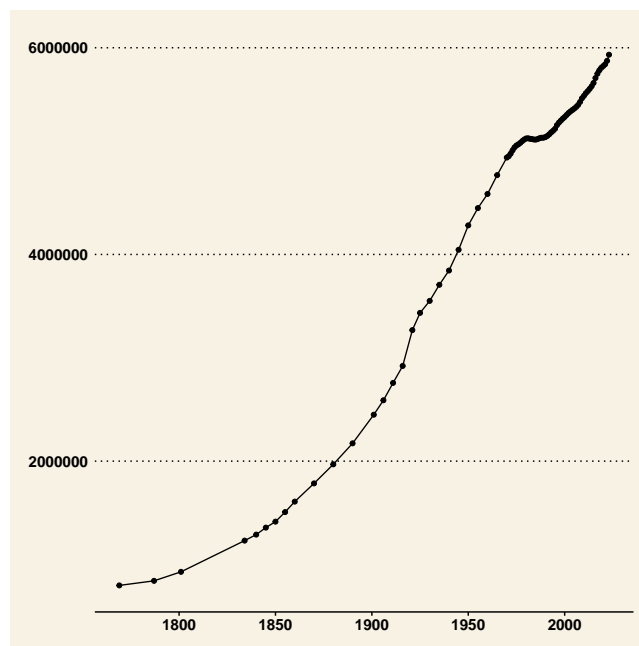


Figure 1: Udvikling af den Danske befolkning siden 1769. Summariske tal fra folketællinger, statistikbankens register FT.

Middelfolketal er defineret som det gennemsnitlige folketal i en given tidsperiode. Vi betegner middelfolketal med  $\tilde{N}$ .

For at beregne middelfolketal helt korrekt, ville man for alle personer være nødt til at tælle hvor mange dage de har levet i perioden i befolkningen. Så kunne man beregne middelfolketal 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.

Praktisk kender man desværre ikke de præcise tal, altså hvor mange dage alle personer fra en befolkning har levet i en given tidsperiode. Derfor er flere metoder, der kan bruges til at beregne middelfolketal approksimativt baseret på enkelte folketal i en given tidsperiode.

### Metode 1

Her skal man kun kende et enkelt folketal, nemlig folketal cirka i midten af perioden. Hvis perioden starter i tidspunkt  $t_1$  og slutter i tidspunkt  $t_2$ , så er tidspunkt i midten givet som  $(t_1 + t_2)/2$  og folketal i midten er givet ved  $\tilde{N}[t_1, t_2] = N((t_1 + t_2)/2)$ .

Dette tal bruges som et estimat for middelfolketal. Denne metode er simpel og tilstrækkeligt for mange formål, især når folketal ikke ændrer sig særlig meget i den given tidsperiode. For eksempel bruger Danmark Statistik folketal fra den 1. juli som årets middelfolketal i de årlige rapporter om befolkningsudvikling <sup>2</sup>.

### Metode 2

For at bruge denne metode skal man kende folketal i starten og i slutningen af perioden. Det estimerede middelfolketal er gennemsnit af de to folketal:  $\tilde{N}[t_1, t_2] = (N(t_1) + N(t_2))/2$ .

### Metode 3

Her skal man kende folketal 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:

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

Alle tre metoder kan bruges ligeværdig hvis folketal er relativt stabil i perioden. Ændrer folketal sig meget i perioden, er metode 2 en bedre tilnærmelse

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

til det sande ukendte middelfolketal end metode 1, og ligeligt er metode 3 er bedre end metode 2. Figure 2 visualiserer forskellen mellem de 3 metoder.

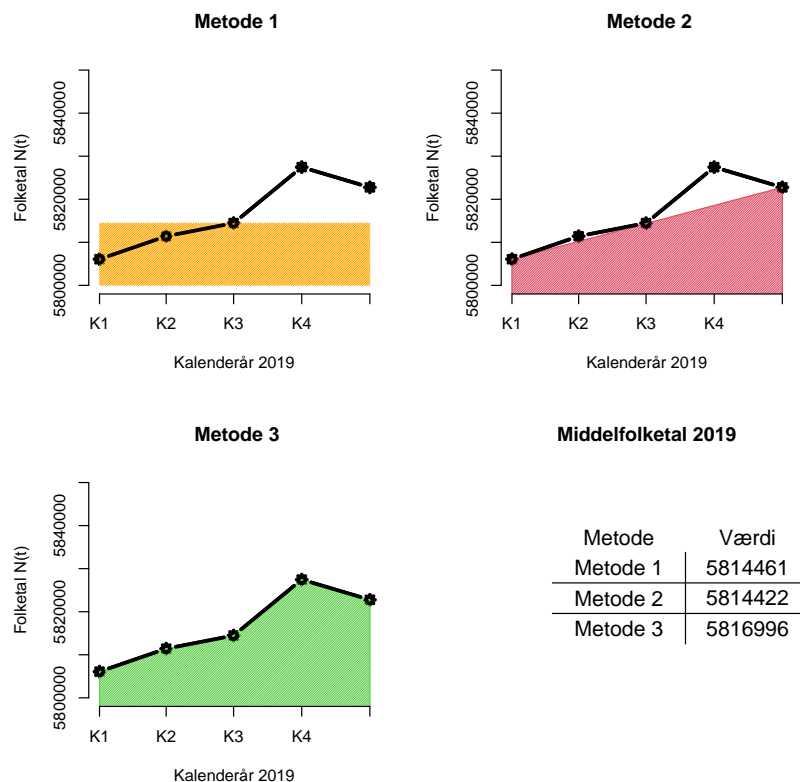


Figure 2: Figuren viser de 3 metoder for at beregne middelfolketal baseret på 5 folketal: 1 januar 2019, 1 april 2019, 1 juli 2019, 1 oktober 2019, 1 januar 2023.

## Eksempel

Vi beregner middelfolketal for den Danske befolkning i en periode, som starter den 1 januar 2016 og slutter den 1 januar 2023. Fra statistikbankens BEFOLK1 henter vi folketal i starten, midten og slutningen af perioden:

```
library(danstat)
library(tidyverse)
library(stringi)
dt=get_data("BEFOLK1",variables=list(list(code="tid",values=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  $\tilde{N}[1 \text{ jan } 2016, 1 \text{ jan } 2023] = 5.707.251$  personer. For at beregne metode 2 bruger vi R som lommerregner:

```
# Metode 2
(5511451 + 5707251)/2
```

```
[1] 5609351
```

Med metode 2 estimerer vi middelfolketallet i perioden 2009–2023 dermed til 5.609.351 personer. For metode 3 er beregningen i R den følgende:

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

```
[1] 5714652
```

Med metode 3 estimerer vi middelfolketallet i perioden 2009–2023 til 5.714.652 personer baseret på de 3 folketal fra perioden. Vi ser i det her eksempel at der er mere end 100.000 personer forskel mellem metode 3 og metode 2. Om denne forskel er vigtigt eller ligegyldigt, ville afhænge formålet med den konkrete demografiske undersøgelse. Er den vigtigt, ville man hente så mange folketal som muligt fra perioden, og beregne middelfolketallet med metode 3. Følgende R-koder henter alle folketal fra statistikbankens register FOLK1a og anvender metode 3.

```
library(tidyverse)
library(stringi)
# Metode 3 baseret paa 57 folketal mellem 2009 og 2023
# konstruere vektor 2009K1, 2009K2, ..., 2022K4
kvartal_years <- paste0(rep(2009:2022,rep(4,14)), "K", 1:4)
# tilfoej 2023K1
kvartal_years <- c(kvartal_years, "2023K1")
# hent data fra FOLK1a
dt <- get_data("FOLK1a", variables=list(list(code="tid",
                                             values=kvartal_years)))
# transform aarstal + kvartal til dato
patt <- pattern=c("Q1", "Q2", "Q3", "Q4")
repl <- c("-01-01", "-01-04", "-01-07", "-01-10")
dt <- mutate(dt,
```

```

TID=stri_replace_all_regex(TID,
                           pattern=pat,
                           replace=repl,
                           vectorize_all=FALSE))

# transform fra character til dato
dt <- mutate(dt, TID=as.Date(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
})

```

5717974

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

## 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 periode i en befolkning. Jo flere dødsfald sker og jo flere personer emigrerer jo mindre er folketal i slutningen af perioden samlignet med starten af perioden. Ligeledes er folketal i slutningen jo højere jo flere personer bliver født og indvandrer til befolkningen. Den demografiske ligevægtsligning for en periode  $[t_1, t_2]$  er:

$$N(t_2) = N(t_1) + (F[t_1, t_2] - D[t_1, t_2]) + (I[t_1, t_2] - U[t_1, t_2]).$$

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 forskel mellem fødsler og dødsfald ( $F[t_1, t_2] - D[t_1, t_2]$ ) for *naturlig vækst* og forskel mellem ind- og udvandring ( $I[t_1, t_2] - U[t_1, t_2]$ ) for *netto migration*.

$U[t_1, t_2]$ ) for *nettovandring*. Figure 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 folketal i denne periode hvorimod fødsler og dødsfald holder et rimelig konstant niveau. Man kan også se at udvandring er stigende helt op til 2019 men knækker i 2020 på grund af corona krisen.

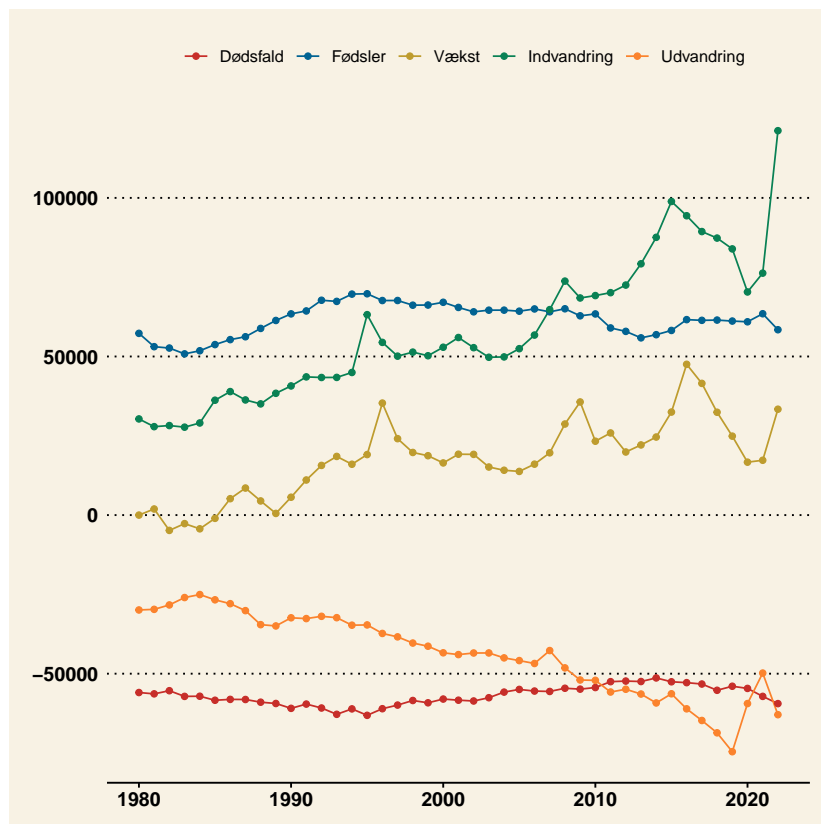


Figure 3: Siden 1980 ændring i folketal (vækst), antal fødsler, dødsfald, ind- og udvandring.

### Eksempel

Vi henter tal for den Danske befolkning i år 2022 fra statistikbankens register FOLK1a, DOD, FOD, INDVAN og UDVAN.

```
N <- get_data("FOLK1a",list(list("tid",c("2022K1","2023K1")))[["INDHOLD"]])
D <- get_data("DOD",list(list("tid", 2022)))[["INDHOLD"]]
F <- get_data("FOD",list(list("tid", 2022)))[["INDHOLD"]]
```



```
I <- get_data("INDVAN",list(list("tid", 2022)))[["INDHOLD"]]
U <- get_data("UDVAN",list(list("tid", 2022)))[["INDHOLD"]]
})
data.frame(X=c("Folketal jan 2022",
               "Folketal jan 2023",
               "Foedsler 2022",
               "Doedsfald_2022",
               "Indvandring 2022",
               "Udvandre 2022"),
           Antal=c(N[1],N[2],F,D,I,U))
```

	X	Antal
1	Folketal jan 2022	5873420
2	Folketal jan 2023	5932654
3	Foedsler 2022	58430
4	Doedsfald_2022	59435
5	Indvandring 2022	121183
6	Udvandre 2022	62927

Baseret på disse tal beregner vi at den Danske befolkingens vækst i perioden til  $(5.932.654 - 5.873.420) = 59.234$  personer. Den naturlige vækst i perioden er negativt:  $(58.430 - 59.345) = -915$  personer og nettovandring i perioden positivt:  $(121.183 - 62.927) = 58.256$  personer. Vi ser at ligevægtsligningen ikke går op, der mangler  $(59.234 - 58.256) = 978$  personer. Det vil sige at de forskellige register som statistikbanken bogfører er ikke intern konsistent. Det kan være mange grunde til. En vigtig grund er at det er svært at registrere de præcise datoer hvor ind- og udvandring sker.

## Rater

I demografi bruger vi rater til at beskrive relative ændringer. For eksempel beskriver dødsraten antal døde relativt til befolkingens størrelse. Som enhed for dødsraten bruges ofte *antal døde per personår*. Her dividerer man antal døde i en periode med antal personår som personer fra befolkingen har livet i samme periode. Mere generalt 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*. Der findes også dimensionsløse rater som er kvotient af to forskellige størrelser med samme måleenhed. Disse kan udtrykkes som en procentdel. De fleste demografiske rater bruger risikotid i nævneren.

## Risikotid

Vi betegner med  $R[t_1, t_2]$  det samlede gennemlevede tid i perioden  $[t_1, t_2]$  af alle personer i en befolkning og kalder den også for *risikotid*. Udtrykket *risikotid*

giver egentlig kun mening når man studerer en risikabel hændelse som fx død i en befolkning som er eksponeret for denne hændelse. Men, vi bruger det alligevel også i andre sammenhænge. Enheden for risikotid er antal personår. For at beskrive risikotid i lille befolkninger kan den regnes om til antal personuger eller antal person dage. For stor befolkninger vil man typisk regne om til enheder som *100.000 personår* eller *1.000.000 personår*. Kender man det eksakte antal dage som alle personer fra en befolkning har livet i en given periode beregner man risikotid som sum af alle person dage. Det kræver dog at man kender datoer for fødsler, dødsfald, ind- og udvandring. Man kan også estimere risikotid baseret på registerdata. For at estimere risikotid i en befolkning baseret på registerdata multiplicerer vi middelfolketal med periodens længde.

## Begivenhed per personår

Mange demografiske rater er defineret som antal begivenheder (f.eks., dødsfald eller indvandring) relativt til antal gennemlevede personår i en periode i en befolkning:

$$\text{Rate}_X[t_1, t_2] = \frac{\text{Antal begivenheder X i perioden } [t_1, t_2]}{R[t_1, t_2]}$$