

# 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)

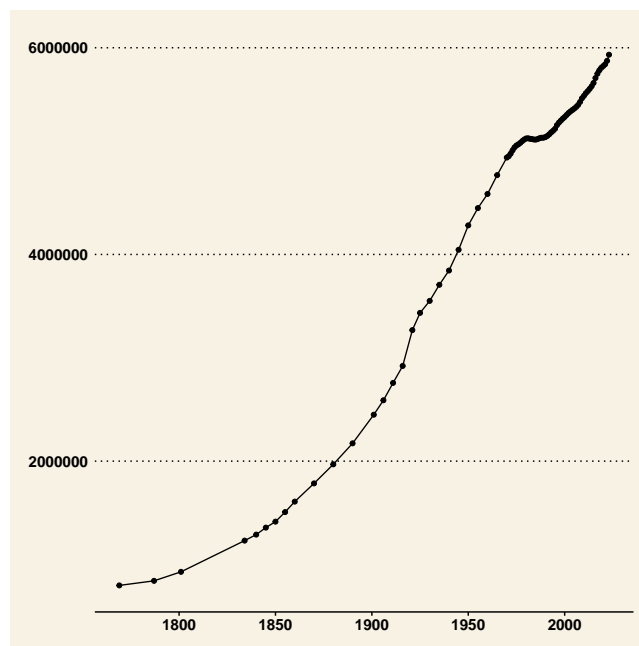


Figure 1: Udvikling af den Danske befolkning siden 1769. Summariske tal fra folketællinger, statistikbankens register FT.

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.

### Standard 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 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 folketallet ikke ændrer sig særlig meget i den given tidsperiode. For eksempel bruger Danmark Statistiks den 1. juli som årets middelfolketal i de årlige rapporter om befolknings udvikling <sup>2</sup>.

### Standard 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:  $\tilde{N}[t_1, t_2] = (N(t_1) + N(t_2))/2$ .

### Standard 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 standard metoder kan bruges ligeværdig hvis folketallet er relativt stabil i perioden. Ændrer folketallet sig meget i perioden, er standard metode 2 en bedre tilnærmelse til det sande ukendte middelfolketal end standard metode 1, og ligeligt er standard metode 3 er bedre end standard metode 2.

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

## 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 standard metode 1. Med standard metode 1 er det estimerede middelfolketal  $\tilde{N}[1 \text{ jan } 2016, 1 \text{ jan } 2023] = 5.707.251$  personer. For at beregne standard metode 2 bruger vi R som lommerregner:

```
# standard metode 2
(5511451 + 5707251)/2
```

```
[1] 5609351
```

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

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

```
[1] 5714652
```

Med standard metode 3 estimerer vi middelfolketal 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 standard metode 3 og standard 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 middelfolketal med standard metode 3. Følgende R-koder henter alle folketal fra statistikbankens register FOLK1a og anvender metode 3.

```

# Standard metode 3
# konstruere vektor 2009K1, 2009K2, ..., 2022K4
kvartal_years <- paste0(rep(2009:2022,rep(4,14)), "K", 1:4)
# tilføej 2023K1
kvartal_years <- c(kvartal_years, "2023K1")
# hent data fra FOLK1a
dt <- get_data("FOLK1a", variables=list(list(code="tid", values=kvartal_years)))
# transform rstal + kvartal til dato
patt <- pattern=c("Q1", "Q2", "Q3", "Q4")
repl <- c("-01-01", "-01-04", "-01-07", "-01-10")
dt <- mutate(dt,
              TID=as.Date(stri_replace_all_regex(TID,
                                                  pattern=pat,
                                                  replace=repl,
                                                  vectorize_all=FALSE))
              )
# middelfolketal_metode3
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 matematisk formel, der bruges i demografisk analyse til at beskrive forholdet mellem antallet af fødsler, dødsfald og migration i en befolkning. Ligningen er som følger:

\$\$

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

\$\$

Hvor:

- $N(t_1)$  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.