



# Metode bag fremskrivning af en ungdoms- årgangs uddannelsesniveau

## Indledning

En ungdomsårgangs kommende uddannelsesniveau er ikke enkelt at forudsige. De unges endelige uddannelsesniveau og det fulde tidsforbrug kendes først, når uddannelsen rent faktisk er opnået. For nogle gælder dette, når de er 25 år, mens det for andre gælder, når de er en del ældre. Det er således nødvendigt at følge en ungdomsårgang over lang tid førend det endelige uddannelsesniveau kan opgøres – og til den tid vil analysen ikke længere give et billede af det aktuelle uddannelsesmønster.

For at imødekomme behovet for at kende konsekvenserne af struktur- og adfærdændringer i uddannelsessystemet foretager Styrelsen for It og Læring med Profilmodellen en fremskrivning af en ungdomsårgangs kommende uddannelsesniveau på baggrund af det uddannelsesmønster, der var da årgangen gik i 9. klasse (og 8. klasse – se uddybning af dette på side 3).

Fremskrivningen tager, som skrevet, udgangspunkt i det uddannelsesmønster, der var da årgangen gik i 8.-9. klasse, dvs. fremskrivningen er fx baseret på, hvorledes årgange af studenter fra samme periode/år bevæger sig videre i uddannelsessystemet.

I Danmarks Statistiks uddannelsesregister registreres løbende alle danskeres status i uddannelsessystemet. Det er således muligt at kortlægge (tegne en profil af) de aktuelle bevægelser i uddannelsessystemet. Det er dette datagrundlag, der benyttes i Profilmodellen.

Den aktuelle version af Profilmodellen tager udgangspunkt i den ungdomsårgang, der gik i 9. klasse i skoleåret 2017/18, og betegnes Profilmodellen 2018. Profilmodellen 2018 viser således, hvordan en ungdomsårgang, som gik i 9. klasse i skoleåret 2017/18, forventes at uddanne sig, hvis uddannelsessystemet og de unges uddannelsesmønster i fremskrivningsperioden er som i dette skoleår (og skoleåret før).

Modellen er en simulation af individuelle uddannelsesforløb for en hel ungdomsårgang, dvs. Profilmodellen 2018 simulerer uddannelsesmønsteret for ca. 66.000 unge. Hvert forløb er afgrænset til at vare i 25 år efter afsluttet 9. klasse<sup>1</sup>, dvs. for hver person simuleres på baggrund af senest kendte uddannelsesmønster i det for fremskrivningen valgte start år, personens uddannelsesforløb 25 år frem i tiden. Herefter optælles personernes endelige uddannelsesniveau. Der optælles både efter 5, 6, 7, 8, 9, 10, 15, 20 og 25 år. På den

---

<sup>1</sup> De unge behøver blot at have påbegyndt, ikke fuldført, 9. klasse. Modelmæssigt bliver det udregnet som 26 år efter påbegyndt 9. klasse.

måde er det muligt at sige noget om 9. klasse-årgangens forventede uddannelsesniveau både på kort, lidt længere og lang sigt.

I Profilmodellen er der endvidere taget højde for 9. klasse-sammensætningen i de enkelte årgange. Den enkelte årgang er fordelt efter befolkningens fordeling på køn og herkomst det pågældende år. Endvidere tages der også højde for fordeling på region.

Metoden i Profilmodel 2018 er ikke ændret i forhold til Profilmodel 2016.

## Datagrundlag

Styrelsen for It og Læring beregner en ungdomsårgangs uddannelsesprofil på baggrund af data fra Danmarks Statistik. Datagrundlaget omfatter følgende registre:

- Elevregisteret (elevers ordinære uddannelser)
- Kvalifikationsregisteret (elevers afsluttede kompetencegivende eksamener via voksenuddannelsesområderne)
- Eksamensdatabasen fra Styrelsen for It og Læring (oplysning om EUX-elever har afsluttet den gymnasiale uddannelsesdel)
- Data fra CPR-registeret (oplysninger om køn og bopæl)
- Vandringsregisteret (oplysninger om død og migration).

For uddybning og dokumentation heraf se: <http://www.dst.dk/Statistik/dokumentation/times.aspx>

Datagrundlaget omfatter alle uddannelser, der har bekendtgørelser, samt uddannelser, der ikke har bekendtgørelser, men er SU-berettigede.

## Generelle antagelser og forudsætninger

I dette afsnit er beskrevet nogle generelle antagelser og forudsætninger for Profilmodellen.

Profilmodellens fremskrivninger baserer sig alene på registreret uddannelsesadfærd hos personer, der har gået i 9. klasse på en dansk grundskoleinstitution. For især indvandrer-/efterkommerpopulationen er der her tale om en væsentlig afgrænsning: Gæstestuderende og andre indvandrere, der er kommet til landet efter grundskolen, medtages således ikke i modellens datagrundlag.

Profilmodellen på regionalt og kommunalt niveau er beregnet på basis af adfærd for de elever, som havde bopæl i regionen/kommunen ved afslutning af 9. klasse, uanset om de senere er flyttet. De elever, som afslutter deres grundskole på en efterskole, er optalt i den region/kommune, hvor de havde bopæl inden efterskoleopholdet. De få elever, som ikke har en oplyst bopælsregion/bopælskommune, er optalt i den region/kommune, hvor de har gået i grundskole.

Personer, der forlader landet for mere end 10 år eller dør, fjernes fra beregningerne på henholdsvis udrejse- eller dødstidspunktet. Hvis det ikke er muligt at følge en person 10 år frem i registeret på grund af afslutning af registeret, skønnes hvorvidt denne person på baggrund af empiriske betingede sandsynligheder vil vende tilbage til Danmark inden for 10 år, dvs. de observerede personers adfærd benyttes til at imputere det mest sandsynlige udfald med den nuværende viden om personens adfærd.

Profilmodellens beregninger af uddannelsesprofiler for delbefolkning, dvs. f.eks. unge af anden herkomst end dansk eller unge i de enkelte kommuner, er generelt mere usikre end resultater på et mere aggregeret niveau. Årsagen hertil er, at det empiriske datamateriale kan være meget begrænset eller helt mangle. For at imødekomme dette er det valgt, at et år i Profilmodellen dækker over et tværsnit af to år, dvs. år 2018 dækker over den toårige periode 1. oktober 2016 - 30. september 2018. Det bemærkes endvidere, at Profilmodellen og den afledte opgørelse af tidsforbruget bygger på såkaldte tællingsår (1. oktober – 30. september).

Der kan imidlertid fortsat være overgange mellem uddannelser, hvor datamaterialet er meget tyndt. Ved simuleringen er det således valgt at supplere med data fra et mere aggregeret niveau, hvis der er færre end 10 personer i datamaterialet, dvs. hvis der er færre end 10 personer, der i en given kommune, forventes at opnå en given uddannelse, benyttes data fra det regionale niveau, mens beregningerne på regionalt niveau suppleres med data på landsplan. Herved gøres modellen mere robust.

På lands- og regionalt plan er det muligt at fordele resultaterne på køn og herkomst, mens det på kommunalt plan kun er muligt at fordele på køn. Årsagen hertil er et for tyndt datamateriale på kommuneniveau.

Uddannelsessystemet består af ca. 2000 forskellige uddannelser. Profilmodellen er inddelt i ca. 80 uddannelsesgrupper.

Endelig bemærkes, at det er valgt at sammensætte en elevs uddannelsesforløb til ét forløb, hvis eleven har flere uddannelser inden for samme uddannelsesgruppe med pauser mellem uddannelserne på mindre end 16 måneder.

## **Modelbeskrivelse**

Profilmodellen er en simulationsmodel, hvor elever fra en ungdomsårgang efter afsluttet 9. klasse individuelt tildeles et uddannelsesforløb (afbrud, fuldførelser og overgange mellem uddannelser) på baggrund af beregnede sandsynligheder for senest kendte adfærd i uddannelsessystemet.

Input til Profilmodellen er betingede sandsynligheder for overlevelse i og overgange mellem tilstande i uddannelsessystemet. Disse sandsynligheder opregnes i såkaldte overlevelsestavler.

Overlevelsestavlerne findes som henholdsvis bestands- og overgangstavler. I bestandstavlerne følges elever, som er i gang med en uddannelse, indtil afbrud eller fuldførelse, mens elever i overgangstavlerne følges fra de har fuldført henholdsvis afbrudt en uddannelse indtil eventuel påbegyndelse af en ny uddannelse. Både bestands- og overgangstavlerne opregner måned for måned, hvor mange der fra en startpopulation fortsat er tilbage eller er afgået til én af flere konkurrerende afgangsmuligheder. Månedssaktiviteten samles efterfølgende i tællings år. Ved parameterfri metode (Kaplan-Meier) beregnes overlevelsesfunktion samt kumulerede og betingede overgangssandsynligheder.

Profilmodellen simulerer en ungdomsårgangs fremtidige uddannelse ved anvendelse af de estimerede betingede sandsynligheder i bestands- og overgangstavlerne. Den stokastiske proces i Profilmodellen opfattes som en markovkæde, dvs. sandsynligheden for overgang fra én uddannelsestilstand til en anden bestemmes alene af den aktuelle tilstand og er uafhængig af elevens forudgående uddannelseshistorik.

For at minimere usikkerheden i Profilmodellen simuleres hver elev fra en ungdomsårgang fem gange. Alle de simulerede uddannelsesforløb indgår i den samlede beregning.