

## Metode bag fremskrivning af en ungdomsårgangs uddannelsesniveau samt deres tidsforbrug

### Indledning

En ungdomsårgangs kommende uddannelsesniveau og deres tidsforbrug frem til opnåelse af en erhvervskompetencegivende uddannelse er ikke enkelt at forudsige. De unges endelige uddannelsesniveau og det fulde tidsforbrug kendes jo først, når uddannelsen rent faktisk er opnået. For nogle gælder dette, når de er 25 år, mens det for andre gælder, når de er 45 år. Det er således nødvendigt at følge en ungdomsårgang over lang tid førend det endelige uddannelsesniveau og det anvendte tidsforbrug 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 UNI•C Statistik & Analyse med den såkaldte Profilmodel en fremskrivning af en ungdomsårgangs kommende uddannelsesniveau på baggrund af 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 2010/11, og betegnes Profilmodellen 2011. Profilmodellen 2011 viser således, hvordan en ungdomsårgang, som gik i 9. klasse i skoleåret 2010/11, 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 2011 simulerer uddannelsesmønsteret for ca. 68.500 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 startår, personens uddannelsesforløb 25 år frem i tiden. Herefter optælles personernes endelige uddannelsesniveau og det anvendte tidsforbrug. Der optælles både efter 5, 6, 7, 8, 9, 10, 15, 20 og 25 år. På den måde er det muligt at sige noget om 9. klasse-årgangens forventede uddannelsesniveau både på kort, lidt længere og lang sigt.

---

<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.

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 etnicitet det pågældende år. Endvidere tages der også højde for fordeling på region. Det bemærkes, at det i Undervisningsministeriets databank også er muligt at vælge en metode, hvor den demografiske fordeling i 2011 fastholdes.

Metoden i Profilmodel 2011 er den samme som i Profilmodel 2009 og 2010.

## **Datagrundlag**

UNI•C Statistik & Analyse beregner en ungdomsårgangs uddannelsesprofil på baggrund af data fra Danmarks Statistik. Datagrundlaget omfatter data fra Det Integrerede Elevregister (elevers ordinære uddannelser), data fra Kvalifikationsregisteret (elevers afsluttede kompetencegivende eksamener via voksenuddannelsesområderne), data fra CPR-registeret (oplysninger om etnisk herkomst og køn) og data fra Vandringsregisteret (oplysninger om død og migration). For uddybning 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.

Ved opgørelse af det anvendte tidsforbrug til opnåelse af første erhvervskompetencegivende uddannelse benyttes beregnede minimumsstudietider for opnåelse af en given uddannelse. Disse er baseret på variabelen PRIA opgjort af Danmarks Statistik i Det Integrerede Elevregister. Minimumsstudietiden er den samlede tid brugt fra afsluttet 9. klasse til fuldførelse af den givne uddannelse. Det bemærkes, at variabelen ændres løbende. Variabelen eksisterer kun tilbage til 2000, hvorfor opgørelsen af tidsforbruget før år 2000 ikke beregnes.

## **Generelle antagelser og forudsætninger**

I dette afsnit er beskrevet nogle generelle antagelser og forudsætninger for både Profilmodellen og den afledte opgørelse af tidsforbruget for opnåelse af en erhvervskompetencegivende uddannelse.

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 2011 dækker over den toårige periode 1. oktober 2009 - 30. september 2011. 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 etnicitet, mens det på kommunalt plan kun er muligt at fordele på køn. Årsagen hertil er et for tyndt datamateriale på kommuneniveau.

Uddannelsessystemet består af over 2000 forskellige uddannelser. Profilmodellen er inddelt i ca. 60 uddannelsesgrupper svarende til UNI•C Statistik & Analyser mellemgrupper.

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 bestandsstavlerne 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. Månedaktiviteten samles efterfølgende i tællingsår. Ved para-

meterfri 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 uddannelsesstilstand 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.

## Beregning af mertidsforbrug

På baggrund af Profilmodellens fremskrivning af en elevs uddannelsesforløb er det muligt at beregne elevens tidsforbrug frem til opnåelse af *første* erhvervskompetencegivende uddannelse. En erhvervskompetencegivende uddannelse er en erhvervsfaglig eller videregående uddannelse.

Det forventede tidsforbrug sammenlignes herefter med den af Danmarks Statistik beregnede normerede studietid for opnåelse af den givne uddannelse. Koblingen heraf er dog ikke entydig, da hver uddannelsesgruppe i Profilmodellen består af flere af uddannelser, som kan have forskellig teoretisk normeret studietid. Ved opgørelse af det anvendte tidsforbrug benyttes således en gennemsnitlig normeret studietid for de uddannelser, der indgår i uddannelsesgruppen.

Den gennemsnitlige normerede studietid beregnes forskelligt afhængig af, om den pågældende uddannelse er omfattet af Danmarks Statistisk elevregister. Den gennemsnitlige studietid for de uddannelsesgrupper, hvor alle uddannelser er registreret, beregnes ved, at hver uddannelse i gruppen vægtes med det antal personer, som i det seneste registerår har afsluttet pågældende uddannelse. Derved opnås vægtede gennemsnitlige teoretiske normerede studietider for uddannelsesgrupperne.

Den beregningsteknik kan dog ikke anvendes for de uddannelser, som ikke er registreret i registeret fra Danmarks Statistik. Denne restgruppe omfatter uddannelser, som ikke afsluttes. Det gælder fx åbne uddannelser, enkeltfag m.m.. For disse uddannelsesgrupper er beregnet simple gennemsnitlige teoretiske normerede studietider<sup>2</sup>.

Ved opgørelse af det samlede tidsforbrug opdeles tiden i to kategorier: Normeret studietid og samlet mertidsforbrug. Det samlede mertidsforbrug opdeles herefter i fem underkategorier: Dobbelt ungdomsuddannelse, 10. klasse, studieskift, uddannelsespause og resttid. Ved denne opdeling er der dog en række situationer, hvor det ikke er entydigt klart, hvilken opgørelsesmetode der er mest korrekt at anvende. I disse tilfælde er truffet nogle valg. Nedenstående punkter beskriver disse valg.

---

<sup>2</sup> Det vil eventuelt være muligt at beregne vægtede gennemsnit ved brug af supplerende datakilder.

- a) Tid brugt på *dobbelt ungdomsuddannelse* er tid, som bruges på ungdomsuddannelser efter afslutning af første ungdomsuddannelse. Påbegyndelse af en ekstra ungdomsuddannelse tæller som tid brugt på dobbelt ungdomsuddannelse uanset om den ekstra ungdomsuddannelse fuldføres eller afbrydes.
- b) Tid brugt i *10. klasse* omfatter al tid, der bruges på 10. klasse inkl. de to måneders sommerpause efter afslutning. En normal afsluttet 10. klasse betyder således 12 måneders mertidsforbrug.
- c) Tid brugt på *studieskift* omfatter al tid, der bruges på studier, som afbrydes, bortset fra studieafbrudssituationer, der kan henføres til punkt a eller b.
- d) *Uddannelsespauser* er de perioder undervejs til opnåelse af en erhvervskompetencegivende uddannelse, hvor personen ikke er under uddannelse. Da der ofte i forbindelse med studieskift er "tvungen" uddannelsespause over sommeren på 2 måneder er det valgt kun at medtælle pausetid på mere end 2 måneder pr. uddannelsesforløb. En person, som fx har tre uddannelsespauser på henholdsvis 2, 14 og 10 måneder i sit uddannelsesforløb, får således medregnet  $0 + 12 + 8 = 20$  måneders uddannelsespause.
- e) *Resttid* er den tid, som ikke kan henføres til punkterne a, b, c eller d. Tiden er primært studietidsforsinkelse, dvs. den studietid, som bruges udover den normerede studietid. Studietidsforsinkelsen kan være såvel negativ som positiv. Resttiden er negativ, hvis studiet gennemføres hurtigere end den normerede tid, hvilket hyppigt sker i tilfælde, hvor en person får merit fra anden uddannelse taget i Danmark eller udlandet. Resttiden kan dog også være negativ som følge af fejl eller mangler i datagrundlaget relateret til Danmarks Statistiks opgørelser af normerede studietider for de forskellige uddannelser.

Negativ studietidsforsinkelse modregnes om muligt i eventuel tid brugt på dobbelt ungdomsuddannelse, subsidiært tid brugt på studieskift og tertiært tid brugt på uddannelsespauser. Årsagen hertil er, at alle disse kategorier kan dække over tid brugt på opnåelse af merit og dermed negativ studietidsforsinkelse.

Ud over ovenstående punkter bemærkes, at universitetsbachelorer har en særlig status i beregningen af mertidsforbruget, da universitetsbachelorgraden som udgangspunkt er en mellemstation mod opnåelse af en kandidatuddannelse. Såfremt personen inden for den 25-årige simulationsramme fuldfører en kandidatgrad, antages kandidatuddannelsen for den afsluttende erhvervskompetencegivende uddannelse.