

Metode: Validering af uddannelsesdata fra USF

Når Grønlands Statistik (GS) offentliggør statistik over den grønlandske uddannelsessektor, sker det hovedsageligt på grundlag af udtræk af data fra Uddannelsesstøtteforvaltningens (USF) register over studerendes uddannelsesforløb – det såkaldte elevregister.

Dette dokument beskriver de kontroller og bearbejdningsprocesser som GS foretager på dataudtræk fra USF's elevregister.

Elevregistreringen i USF startede i 1987 med data fra begyndelsen af 1980'erne, men dækker også enkelte uddannelser fra tidligere år. Det betyder at der ikke altid er fyldestgørende oplysninger om kompetencegivende uddannelser for personer født tidligere end begyndelsen af 1970-erne.

Elevregistret fra USF indeholder oplysninger om personer der har været, eller er under uddannelse og som har modtaget eller modtager grønlandsk uddannelsesstøtte.

Registeret indeholder oplysninger om:

- Personnummer
- Uddannelseskode
- Uddannelsessted
- Startdato
- Slutdato (eventuelt forventet slutdato)
- Status¹ den 31. december i tællingsåret

Elevregistret indeholder ofte flere uddannelsesoplysninger for en enkelt person såfremt denne har været registreret på flere uddannelser i tidens løb.

Personer, der ikke har modtaget grønlandsk uddannelsesstøtte indgår ofte heller ikke i de USF-baserede opgørelser. Det drejer sig typisk om personer, der er født uden for Grønland og har uddannet sig i Danmark eller andre lande.

Personer der har fået deres uddannelse gennem ansættelse i uddannelsesstillinger i virksomheder, der udbetaler løn under uddannelsen er heller ikke inkluderet med mindre de har gennemgået uddannelse der på anden måde er registreret i USF eller det danske uddannelsesregister i Danmarks Statistik. Eksempler herpå er Grønlandsbanken, Royal Arctic Line, Air Greenland, KNR, TELE, PISIFFIK, Brugsen, visse dele af Center for Sundhedsuddannelser, Sparbank, Royal Greenland samt voksenlærlinge.

Uddannelsesforløb

I uddannelsesregistret kan et uddannelsesforløb fordele sig over én eller flere uddannelser. Skema A illustrerer et tænkt eksempel på et uddannelsesforløb for en fiktiv person. Eksemplet tager udgangspunkt i en person, der er registreret med i alt seks forløb fordelt på tre uddannelser. Personen har afsluttet to uddannelser med en eksamen, og er den 1. november 2003

¹ Status angiver om personen har afbrudt, afsluttet, fuldført, evt. er på orlov eller endnu er aktiv med sin uddannelse. Hvis en person har status "afs", "akt", "orl" eller "for" ved årets udgang regnes vedkommende som aktiv på uddannelsen.

registreret som aktiv på den tredje uddannelse. Derudover har personen afbrudt en uddannelse, aflagt en trinafsluttende eksamen (trinlut) samt været på orlov².

Skema A - Eksempel på uddannelsesforløb for en fiktiv person

Obs.	Uddannelseskode	Uddannelse	Startdato	Slutdato	Status
1	30.00.11	Folkeskolelærer	01-08-86	30-11-86	Afbrudt
2	30.00.11	Folkeskolelærer	01-08-89	30-06-92	Eksamen
3	80.40.45	Samfundsfag, bach.	01-09-98	30-06-00	Trinlut
4	80.40.45	Samfundsfag, bach.	01-09-00	30-06-01	Eksamen
5	80.40.47	Samfundsfag, kand.	01-09-01	30-06-02	Orlov
6	80.40.47	Samfundsfag, kand.	01-09-02	01-11-03	Aktiv

I den endelige uddannelsesstatistik foretages analysen på baggrund af de studerendes uddannelsesforløb. Hvorvidt den enkelte person er registreret med ét eller flere forløb har ingen betydning, idet den enkelte person udelukkende kan være aktiv på én uddannelse af gangen. Inden elevregistret er klar til brug, er det imidlertid nødvendigt at validere materialet.

Formålet med valideringen er at gennemgå datamaterialet med henblik på at vurdere de enkelte uddannelsesforløb, således at flere uddannelsesforløb på samme uddannelse så vidt muligt fremstår som ét samlet uddannelsesforløb, idet der ønskes en status som *Afbrudt*, *Eksamen* eller *Aktiv*.

For den fiktive person i skema A vil valideringen derfor blandt andet rette sig mod observation nummer tre, hvor personen har aflagt en trinluttende eksamen den 30. juni 2000.

Efter sommerferien genoptager personen den samme uddannelse.

I valideringen opfattes den trinafsluttende eksamen den 30. juni 2000 derfor blot som en mellemliggende registrering i et ellers sammenhængende uddannelsesforløb.

Ved validering fjernes observation tre, idet den oprindelige startdato for uddannelsen dog føres ned til observation fire. En nærmere beskrivelse af valideringen findes i metodekapitlet.

Validering

Datagrundlaget for meget af Grønlands Statistiks uddannelsesstatistik hidrør hovedsageligt fra USF – Uddannelses-støtteforvaltningen. Da USF-data er administrative data og ikke naturlige statistik-data valideres alle USF-data inden de kan anvendes til statistik. Valideringen af data er en maskinel rutine, der foregår ved afvikling af et specielt tilrettet statistikprogram.

Selve valideringen forløber i fire trin. I det første trin foretages et generelt kvalitetstest af data. I testen fejlsøges datamaterialet for ugyldige cpr-numre, datoer, uddannelseskoder, statuskoder osv.

I tilfælde af fejl korrigeres materialet i henhold til USFs anbefalinger.

I valideringens efterfølgende trin vurderes de registrerede uddannelsesforløb for hver enkelt person separat. Valideringen sætter trin for trin hver enkelt

² På mellemlange og længerevarende uddannelser kan der søges orlov dvs. en studiepause. Et orlov har som regel en varighed af et eller to semestre.

person under lup, vurderer oplysningerne og foretager validering i henhold til et fastlagt sæt regler.

I trin to udvælges samtlige personer, der udelukkende er registreret én gang i registret. Disse personer har et entydigt uddannelsesforløb, der kun består af én observation. Disse unikke uddannelsesforløb betragtes som valide.

I trin tre udvælges samtlige personer, der er registreret med flere – dog entydige uddannelseskoder. Disse personer har ligeledes et sæt af entydige uddannelsesforløb, hvor hver uddannelseskode kun forekommer én gang. Disse unikke uddannelsesforløb betragtes ligeledes som valide.

Skema B. Eksempel på gruppering i uddannelsesforløb for en fiktiv person

Obs.	Uddannelseskode	Uddannelse	Startdato	Slutdato	Status	Gruppe
1	30.00.11	Folkeskolelærer	01-08-86	30-11-86	Afbrudt	1
2	30.00.11	Folkeskolelærer	01-08-89	30-06-92	Eksamen	2
3	80.40.45	Samfundsfag, bach.	01-09-98	30-06-00	Trinslut	3
4	80.40.45	Samfundsfag, bach.	01-09-00	30-06-01	Eksamen	3
5	80.40.47	Samfundsfag, kand.	01-09-01	30-06-02	Orlov	4
6	80.40.47	Samfundsfag, kand.	01-09-02	01-11-03	Aktiv	4

Etablering af uddannelsesforløb

I valideringens fjerde trin behandles den resterende mængde af data. Denne restgruppe består af personer, der er registreret flere gange på mindst én uddannelse. Valideringen udføres for én person af gangen i henhold til følgende valideringsregler:

1. Materialet sorteres kronologisk efter variablene *Startdato*.
2. Der oprettes grupper af data afhængig af uddannelseskoden og statuskoden. Hver gang uddannelseskoden skifter, opstår en ny gruppe. Endvidere opstår en ny gruppe, hvis statuskoden er *afbrudt* og næste observations startdato ligger mere end én måned efter den foregående slutdato, eller hvis statuskoden er *trinslut* og næste observations startdato ligger mere end tre måneder efter den foregående slutdato, jf. skema B. Valideringsmæssigt tillades det således at en afbrudt uddannelse kan genstartes på samme uddannelsesforløb indenfor én måned, samt at en trinafsluttende eksamen kan videreføres i et uddannelsesforløb, hvis næste startregistrering på samme uddannelseskode foretages indenfor tre måneder fx efter en sommerferie.
3. Grupper bestående af én observation betragtes som *gode*.
4. I grupper med flere data udvælges startdatoen fra den første observation samt slutdatoen fra den sidste observation, jf. skema C på næste side.

Skema C. Eksempel på behandling af grupperede uddannelsesforløb for en fiktiv person

Obs.	Uddannelseskode	Uddannelse	Startdato	Slutdato	Status	Gruppe
------	-----------------	------------	-----------	----------	--------	--------

1	30.00.11	Folkeskolelærer	01-08-86	30-11-86	Afbrudt	1
2	30.00.11	Folkeskolelærer	01-08-89	30-06-92	Eksamen	2
4	80.40.45	Samfundsfag, bach.	01-09-98	30-06-01	Eksamen	3
6	80.40.47	Samfundsfag, kand.	01-09-01	01-11-03	Aktiv	4

Valideringen udføres maskinelt for samtlige personer med mere end én registrering på samme uddannelseskode.

Valideret datasæt

Efter valideringen samles de validerede data i et datasæt i SAS-format. Herefter kan tabeller til Statistikbank, publikationer og særlige analyseopgaver udarbejdes på grundlag af dette datasæt.

Selvkontrol på skolerne

Det skal bemærkes at Grønlands Statistik tidligere foretog særlige kontroller af uddannelsesdata på individniveau ved at sende USF-oplysningerne til verifikation ude på de enkelte skoler.

Denne praksis, der var langsommelig og som gav tvivlsom gevinst mht. præcision i data, har GS fra og med 2010 publikationen opgivet efter samråd med IIN.

I dag står de enkelte skoler selv for at holde statusoplysninger opdaterede på deres elever.