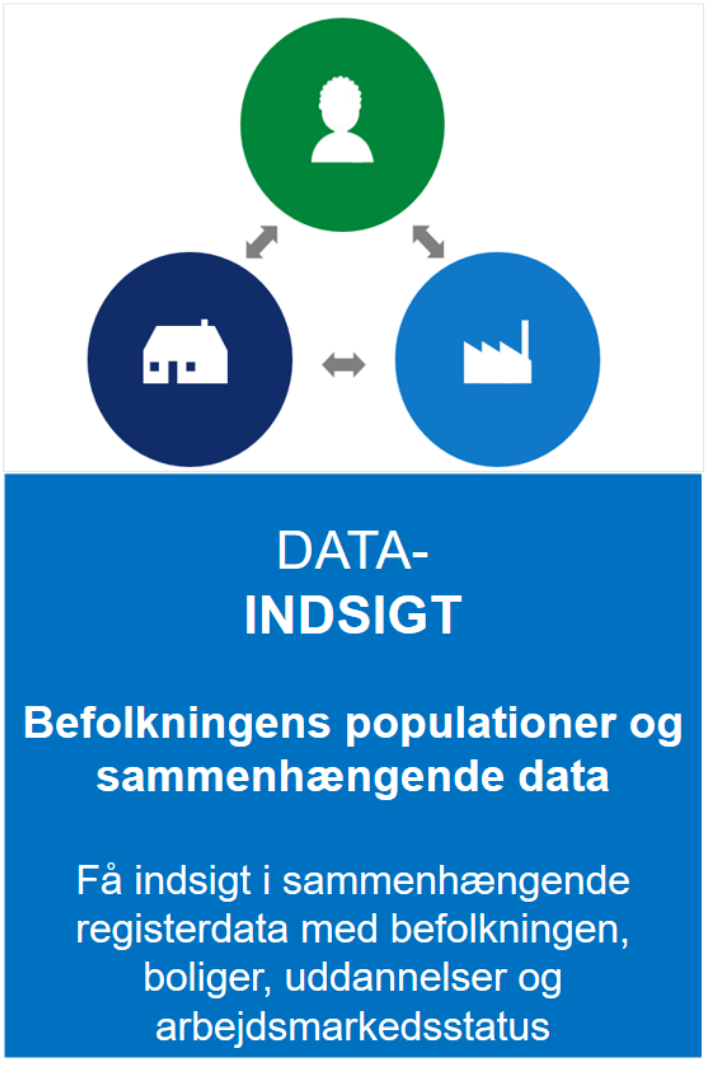
DANMARKS STATISTIK

Danmarks Datavindue



1. Oktober 2026

Forord

Denne publikation er en praktisk og metodisk vejledning til Danmarks Statistiks personstatistikregistre til dig, der skal tage stilling til, hvordan dine populationer skal afgrænses og sammensættes. Formålet er todelt: Dels skabes et klart overblik over, hvilke specifikke datakilder der er til rådighed, og dels præsenteres "best practice" for, hvordan disse kobles mest hensigtsmæssigt. Målet er, at registergrundlaget sammensættes, så det passer bedst muligt til de konkrete analysebehov.

Statistikregistrene i Danmarks Statistik er op- og udbygget siden begyndelsen af 1980'erne. Data er af høj kvalitet og omfatter hele befolkningen. Dette giver helt unikke analysemuligheder for brugerne af data - det være sig på et bestemt tidspunkt og over tid. I Danmarks Datavindue (DDV) gives der adgang til over 500 registre, og via Datasafari kan du få et overblik over de data, der stilles til rådighed. I arbejdet med mikrodata kræves ikke blot kendskab til de enkelte registre og variable, men også en forståelse for de bagvedliggende registerstrukturer. Når vi kobler data på tværs af registrene, opstår der nye muligheder for indsigt, men også metodiske faldgruber.

Vi har lagt vægt på:

- **Datavalg:** En guide til at vælge de mest præcise registre til afgrænsning af befolkningens populationer.
- **Sammenhængskraft:** En gennemgang af nøglevariabler og koblingsmekanismer, der sikrer sammenhængende data.
- **Konsistens:** Hvordan man håndterer ændringer i statistikdesignet.

Vejledningen henvender sig til brugere under forskningsordningerne, der har brug for en sikker hånd i navigationen af de komplekse registerstrukturer. Der kan også være hjælp at hente i den R- og SAS-kode, som vejledningen stiller til rådighed.

Vi arbejder i Danmark Statistiks kontor Consulting, hvor registerdata er fundamentet for forskningen i vores Data Science Lab og for den skræddersyede statistik, vi udarbejder for vores kunder. Vejledningen afspejler vores erfaringer med at sammensætte data til disse formål. Vi er optaget af, at der er sammenhæng i data, og i forbindelse med skræddersyet statistik skal vi så vidt muligt kunne genskabe den publicerede statistik.

Det er vores håb, at vejledningen vil bidrage til en mere effektiv udnyttelse af vores fælles dataressourcer og sikre, at brugernes analyser hviler på det bedst mulige metodiske grundlag.

Louise Rasmussen, Emilie Rune Hegelund og Ole Schnor
Danmarks Statistik, Consulting
Oktober 2026

Indhold

Forord **2**

1. Valg af registre **4**

2. Befolkningens populationer **6**
2.1 Statusregistrene BEF og FAIN 6
2.2 Forløbsregistret Lifelines 9

3. Sammenhængende populationer **11**
3.1 Kvartalsvise populationer baseret på BEF 12
3.2 Månedsvise populationer baseret på Lifelines 16

4. Statusregistrene **19**
4.1 Højest fuldførte uddannelse (UDDA) 19
4.2 Registerbaseret arbejdsstyrke (RAS) 21
4.3 Boligopgørelsen (BOL og BOLIG) 22

5. Forløbsregistrene **24**
5.1 Lifelines 24
5.2 Højest fuldførte uddannelse (UDDF) 28
5.3 Elevregistret (KOTRE) 29
5.4 Arbejdsmarkedsregnskabet (AMRUN) 30

6. Referencetidspunkter, datasætnavne og statustidspunkter **32**

7. Statistiksystemets enheder og nøgler **33**

8. R kode **36**
8.1 Årlige populationer 1981-2007 36
8.2 Befolkningsstatistikkens kvartalsvise populationer for 2023 39
8.3 Lifelines månedsvise populationer for 2023 45

9. SAS kode **50**
9.1 Årlige populationer 1981-2007 50
9.2 Befolkningsstatistikkens kvartalsvise populationer for 2023 53
9.3 Lifelines månedsvise populationer for 2023 56

1. Valg af registre

I vejledningen ser vi på de registre, der indeholder oplysninger om befolkningen, dens boligforhold, dens uddannelsesmæssige baggrund og dens tilknytning til arbejdsmarkedet. Det er et langt stykke af vejen også de registre, der er fundamentet for relationerne mellem enhederne i Danmarks Statistiks registerbaserede statistiksystem:[1]



Vejledningen giver dig en forståelse for registrene og deres relationer, og det kan være nyttigt, når du skal afveje, hvilke krav du har til:

- **Sammenhæng:** Skal der være sammenhæng i populationerne på et givet statustidspunkt, og/eller skal personerne følges over tid?
- **Den publicerede statistik:** Skal resultaterne være sammenlignelige med den publicerede statistik?
- **Historik:** Skal der benyttes historiske data? (1980-2007)

I vejledningen skelnes mellem **status- og forløbsregistre**. For statusregistrene er populationen opgjort på et på forhånd fastlagt statustidspunkt. Med et forløbsregister kan populationen dannes på et vilkårligt tidspunkt, og det giver større fleksibilitet, når registergrundlaget skal sammensættes.

Hvis registergrundlaget skal være sammenligneligt med den publicerede statistik, så skal du som hovedregel benytte statusregistrene. Til forskningsformål kan det være underordnet, om de publicerede tal kan genskabes, men til gengæld kan det være helt afgørende, at populationerne matcher hinanden bedst muligt, og at personerne samtidigt kan følges over tid. I det tilfælde kan forløbsregistrene med fordel benyttes. Historisk (1980-2007) var hovedparten af personstatistikkerne udelukkende baseret på statusregistre, og datidens statistikdesign var med til at sikre sammenhæng i data på tværs af registrene. Fra og med 2008 ses en ændring i statistikdesignet, så sammenhængen på tværs af registrene mistes. Til gengæld er man godt dækket ind mht. forløbsregistre fra og med 2008.

[1] Statistiksystemets enheder og nøgler er nærmere omtalt i afsnit 7.

I tabel 1.1 ses de registre, der indgår i vejledningen med angivelse af datapakke og registertype.

Tabel 1.1. Registre der indgår i vejledningen.

DDV Datapakke	Register	Registertype
Befolkning	**BEF** - Befolkningen	Status
	FAIN - Husstande og familier	Status
	LIFELINES – Befolkning med vandrings- og dødsdatoer	Forløb
Boligforhold	**BOL** - Boligtællingen	Status
	BOLIG - Boligtællingen	Status
	PERBOLIG - Personer og boliger	Status
	BBRENHED - BBR enhed	Status
Befolkningens uddannelsesstatus	**UDDA** – Højest fuldførte uddannelse (status)	Status
	UDDF – Højest fuldførte uddannelse (forløb)	Forløb
Fuldtidsuddannelser & Voksen- og Efteruddannelse	**KOTRE** - Elevregistret	Forløb
Arbejde	**RAS** - Registerbaseret arbejdsstyrkestatistik	Status
	AMRUN - Arbejdsmarkedsregnskab	Forløb

Ved valg af registergrundlag peger vi på nedenstående opmærksomhedspunkter. Bemærk, at det ikke er et spørgsmål om et enten/eller mht. status- og forløbsregistre, idet du sagtens kan kombinere statusregistrene med forløbsregistrene jf. afsnit 3, hvor vi kommer med nogle bud på sammenhængende populationer. De enkelte statusregistre er nærmere beskrevet i afsnit 4, mens forløbsregistrene er nærmere beskrevet i afsnit 5.

Statusregistrene:

- Populationerne opgøres på et på forhånd fastlagt statustidspunkt, og registrene revideres sjældent.
- De fleste publicerede statistikker er baseret på statusregistre, og de publicerede tal i Statistikbanken kan som regel genskabes med udgangspunkt i statusregistrene.
- Vær opmærksom på sammenhængen mellem statusregistrenes referencetidspunkter, datasætnavnene og statustidspunkterne i Statistikbanken. Der henvises til afsnit 6 for en oversigt over disse sammenhænge.
- De historiske data 1980-2007 findes som regel kun som statusregistre.

Forløbsregistrene:

- Populationerne kan udsøges på et vilkårligt statustidspunkt.
- Personerne kan følges over tid uafhængigt af statustidspunkt.
- Flere af forløbsregistrene opdateres årligt og revideres samtidigt bagud i tid. De vil derfor ofte være mere retvisende end statusregistrene.

For alle personregistre i DDV er nøglen CPR-nummeret (PNR), hvilket kan betyde en manglede sammenhæng i data for de personer, der skifter PNR. Det påvirker samtidigt mulighederne for at genskabe de publicerede tal i Statistikbanken. Se afsnit 7 for en beskrivelse af statistiksystemets nøgler.

I afsnit 8 og 9 ses den hhv. R- og SAS-kode, som vi har benyttet til at danne de sammenhængende populationer.

2. Befolkningens populationer

Befolkningsstatistikregistrene er de mest benyttede personregistre i DDV, og initialt skal man som regel forholde sig til spørgsmålet; hvordan skal min population afgrænses i forhold til befolkningen?

Den primære datakilde til Befolkningsstatistikregistrene er Det Centrale Personregister (CPR). Når Danmarks Statistik opgør befolkningen, så er det underforstået, at der er tale om personer, som er bosiddende i Danmark, og som har ret og pligt til at lade sig registrere i CPR-systemet. Ikke alle personer, som opholder sig i Danmark, har ret og pligt til at registrere sig i CPR-systemet fx asylansøgere, som dermed ikke indgår i befolkningen. En nærmere beskrivelse af befolkningen fremgår af Befolkningsstatistikkens Statistikdokumentation.

I dette afsnit ser vi nærmere på statusregistrene **BEF** og **FAIN** samt forløbsregistret **Lifelines**.

2.1 Statusregistrene BEF og FAIN

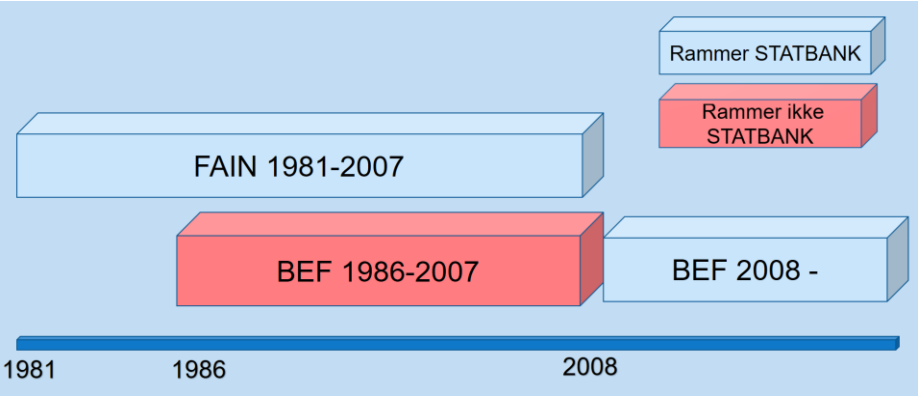
I 2008 implementerede Danmarks Statistik den Personstatistiske Database (PSD), og det fik betydning for den publicerede statistik og for de registre, der i dag stilles til rådighed i DDV. Det betød at:

- Der indføres et nyt familiebegreb, E-familier. Befolkningsstatistikregistret og E-familierne tilbageføres til 1986. På den baggrund dannes en ny statistik i Statistikbanken med familier og husstande fra 1986 og frem.
- Den personrelaterede statistik i Statistikbanken blev ikke opdateret tilbage til 1986. Først fra og med 2008 bliver PSD benyttet som grundlag for den personrelaterede statistik.
- Før 2008 er der tale om en statusopgørelse pr. 1. januar i året. Fra og med 2008 er der tale om en kvartalsvis statusopgørelse pr. 1. januar, 1. april, 1. juli og 1. oktober.

Det er baggrunden for, at der i dag findes to udgaver af Befolkningsstatistikregistret i DDV's Grunddata:

- **FAIN** indeholder befolkningen for perioden 1980-2007. Nøglen er de oprindelige CPR-numre, og populationen svarer meget nøjagtigt til det antal personer, som er offentliggjort i Statistikbanken.
- **BEF** indeholder den PSD-baserede befolkning fra 1986 og frem. Nøglen er CPR-nummer, som er konverteret fra PSD's personløbenummer PERSON_ID. Populationerne i **BEF** er sammenlignelige med Statistikbanken fra 2008 og frem, men ikke før 2008 jf. figur 2.1.

Figur 2.1. DDV's befolkningsregistre og Statistikbanken

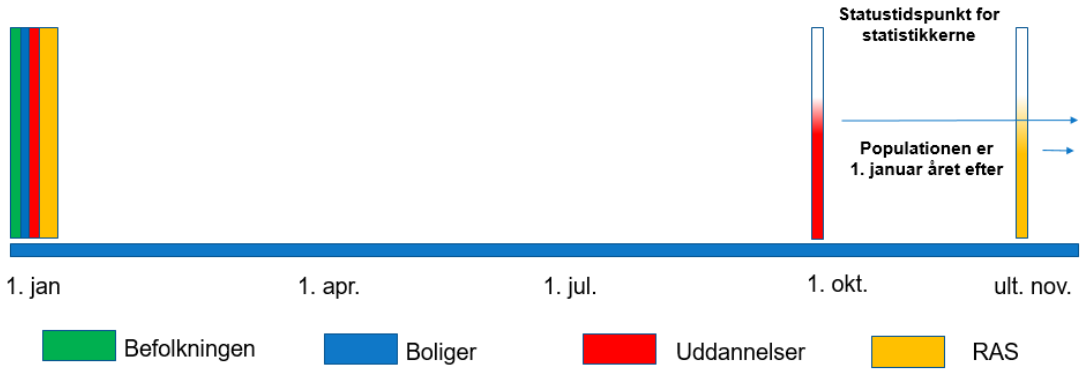


Da **BEF** 1986-2007 er baseret på et nyere udtræk fra CPR-systemet end **FAIN** i samme periode, vil **BEF** alt andet lige indeholde en mere retvisende population. Til gengæld er **FAIN**'s population identisk med populationerne i flere af de andre centrale personstatistikker i perioden 1981-2007, herunder:

- Uddannelser, Højest fuldførte uddannelse, **UDDA**
- Registerbaseret Arbejdsstyrkestatistik, **RAS**
- Boligopgørelsen, **BOL**

I perioden 1981-2007 kan de 4 registre kobles sammen, og resultaterne er næsten altid direkte sammenlignelige med den offentliggjorte statistik i Statistikbanken. Det skal bemærkes, at statustidspunktet for selve statistikken er ultimo november året inden for **RAS** og 1. oktober året inden for uddannelserne i **UDDA**.[2]

Figur 2.2. Populationerne i statistikkerne 1981-2007

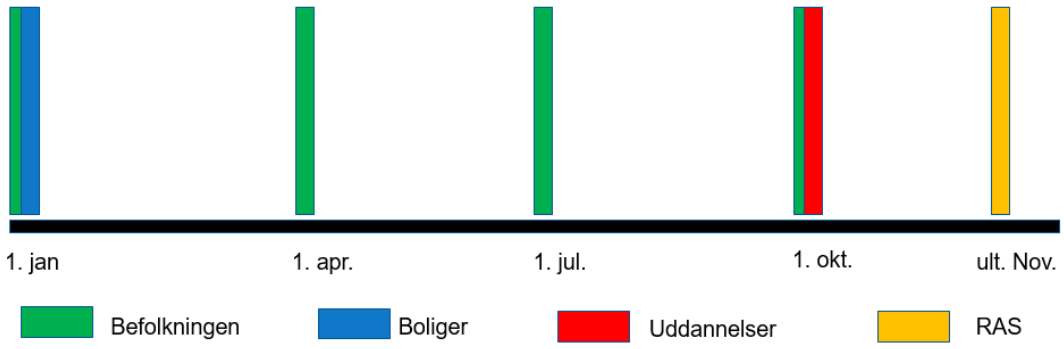


Fra og med 2008 bliver Befolkningsstatistikken **BEF** offentliggjort kvartalsvis, mens der fortsat er tale om årlige statistikker for de andre registre. For **RAS** og **UDDA** gælder, at statustidspunktet for populationerne skifter i 2008, så populationerne nu flugter med statustidspunkterne for selve beskæftigelsesoplysningerne (ultimo november) og selve uddannelsesoplysningerne (1. oktober).

[2] I afsnit 8.1 og 9.1 ses den hhv. R- og SAS-kode, der kobler de 4 registre korrekt sammen.

Det betyder, at der fra og med 2008 ikke længere er identiske populationer på tværs af de 4 registre. **BEF** og **UDDA** er identiske pr. 1. oktober, **BEF** og **BOL/BOLIGER** er identiske pr. 1. januar, mens populationen i **RAS** ikke længere har et sammenfald med de andre registre.

Figur 2.3. Populationer i statistikkerne fra 2008



Dermed vil der være personer, der ikke matcher på tværs af registrene, og resultaterne kan være svære at sammenligne med den offentliggjorte statistik.

Et eksempel herpå er statistikbanktabellen HFUDD16, jf. tabel 2.1, hvor befolkningen (15-69 år) pr. 1. oktober 2019 (fra **BEF**) koblet med højest fuldførte uddannelse pr. 1. oktober 2019 (fra **UDDA**) og socioøkonomisk status (fra **RAS)** pr. ultimo november 2019. Kategorien "4. Uden for arbejdsstyrken" indeholder både personer fra RAS med den socioøkonomiske status "4. udenfor arbejdsstyrken" og indeholder derudover også 9.491 personer, der er i befolkningen pr. 1. oktober, men ikke er i RAS ultimo november.

Tabel 2.1. Befolkningens højest fuldførte uddannelse og socioøkonomiske status 2019 fra Statistikbanken

Befolkningens højest fuldførte uddannelse (15-69 år) efter bopælsområde, tid, højest fuldførte uddannelse og socioøkonomisk status

	Under uddannelse	Beskæftigede	Arbejdsløs	Udenfor arbejdsstyrken
Hele landet				
2019				
H10 Grundskole	330 108	336 167	19 656	316 393
H20 Gymnasiale uddannelser	164 434	199 438	5 788	67 247
H30 Erhvervsfaglige uddannelser	37 121	878 167	26 477	233 454
H35 Adgangsgivende uddannelsesforløb	1 116	1 433	63	969
H90 Uoplyst mv.	1 258	21 316	1 216	24 218
Subtotal	534 037	1 436 521	53 200	642 281

Eksemplet viser, at man skal være påpasselig ved match mellem populationer, hvor statustidstidspunkterne er forskellige. Der henvises til afsnit 4 for en detaljeret gennemgang af de omtalte statusregistre.

2.2 Forløbsregistret Lifelines

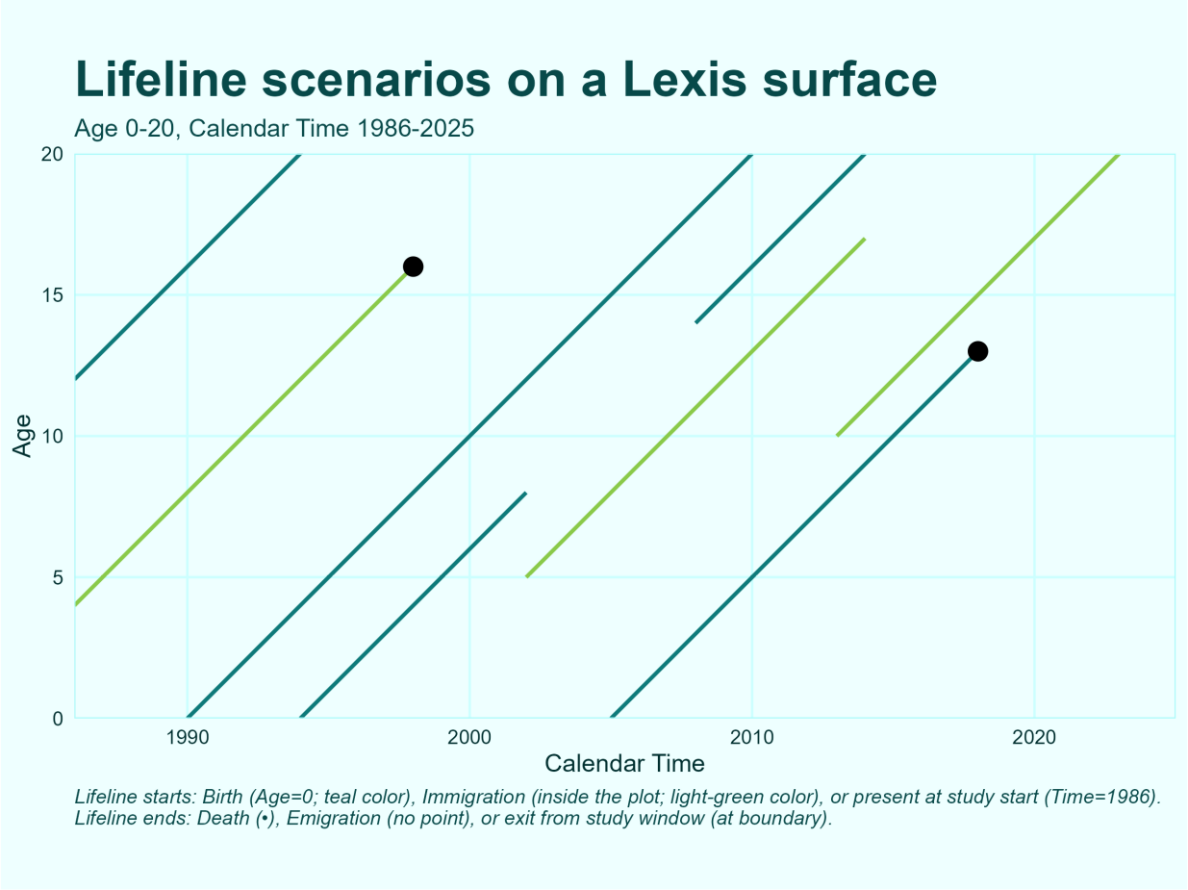
Lifelines indeholder livsforløb for alle individer bosat i Danmark fra 1986[3] og frem. Hver observation repræsenterer en opholdsperiode for et individ i den danske befolkning – startende med enten fødsel, indvandring eller registrering og sluttende med enten død, udvandring eller forsvinding. Individer, der som følge af genindvandring har flere opholdsperioder, er registreret med en separat observation for hver opholdsperiode. Dette gør det muligt at udvælge individer ud fra primære demografiske karakteristika, såsom alder, periode, kohorte, observationslængde m.m. Registret indeholder oplysninger tilbage til 1968, men er først komplet for hele befolkningen fra 1986 og frem.

Lifelines adskiller sig fra **BEF/FAIN**:

- Der er tale om et forløbsregister med angivelse af en start/slutdato for hvert forløb.
- Registret er udviklet til brug for forskere og benyttes ikke til publicering, så der er ikke et sammenligningsgrundlag i Statistikbanken.
- Registret revideres hvert år bagud i tid, og der er ikke nogen tidsmæssig bagkant i forhold til revisionerne. Den seneste udgave af registret vil som udgangspunkt altid være mere retvisende end de tidligere udgaver. Der er typisk kun tale om mindre revisioner.
- Der vil typisk være færre personer i **Lifelines** end i **BEF**, når man sammenligner de to populationer på et bestemt statustidspunkt. Det kan forklares ved de rettelser, som sker i CPR-systemet efter, at Danmarks Statistik har trukket **BEF** populationen. Der er helt overvejende tale om de udvandringer, som først registreres på et senere tidspunkt i CPR-systemet. Se mere om dette i afsnit 5.1.
- Hele registret revideres hvert år, og ved DDV's konverteringen fra PERSON_ID til PNR benyttes den aktuelle nøgle for samtlige observationer i registret. Personer med et skift i CPR-nummer vil derfor optræde med deres aktuelle CPR-numre i **Lifelines**. Forløbene i selve **Lifelines** er hermed intakte, men når man matcher **Lifelines** mod statusregistrene, så vil der optræde personer, der ikke finder match. Se mere om dette i afsnit 7.

[3] Der findes befolkningsregister fra 1980 og frem, men dækningen af fødsler, død, ind- og udvandringer er først fra 1986 vurderet tilstrækkelig til dannelse af Lifelines.

Figur 2.4. Eksempel på kohorte fra Lifelines, 0-20-årige i perioden 1986-2025.



I kohortestudier hvor populationen skal udvælges på bestemte statustidspunkter og følges over tid, kan man med fordel benytte **Lifelines** jf. Lexis diagrammet. I sit design skal der tages stilling til, hvordan populationen skal 'skæres til'. Se mere om dette i afsnit 5.1, hvor vi bl.a. omtaler 2 studier udført i regi af Danmarks Statistiks Data Science Lab.

Lifelines kan også benyttes, hvis du skal trække din population på andre statustidspunkter end de kvartalsvise **BEF** statustidspunkter. Se mere om dette i afsnit 3.2, hvor vi danner månedsvise statuspopulationer med udgangspunkt i **Lifelines**.

3. Sammenhængende populationer

I dette afsnit ser vi på muligheder for at danne sammenhængende populationer med udgangspunkt i hhv. statusregistret **BEF** og forløbsregistret **Lifelines**.

- Hvis dine analyser skal foretages på årlige eller kvartalsvise statustidspunkter, så kan du med fordel tage udgangspunkt i **BEF**. **BEF** er direkte sammenlignelig med den publicerede Befolkningsstatistik (fra 2008), og registret indeholder centrale oplysninger om bopæl, civilstand, husstands- og familierelationer samt herkomst, oprindelsesland og seneste indvandring.

- **Lifelines** kan med fordel benyttes, hvis du skal følge personer over tid og/eller skal trække din population på andre statustidspunkter end de kvartalsvise **BEF** statustidspunkter. I dette afsnit har vi valgt at danne månedsvise populationer med sammenhængende data for månederne i 2023. I praksis kan du afgrænse dine populationer på et utal af måder, så eksemplet i afsnit 3.2 er blot til inspiration.

I vores arbejde med at danne sammenhængende populationer inddrager vi disse forløbsregistre:

- Højest fuldførte uddannelse med forløb – **UDDF**. Indeholder historikken tilbage fra 1970'erne.
- Elevregistret med den igangværende uddannelse – **KOTRE**. Indeholder historikken tilbage fra 1970'erne.
- Arbejdsmarkedsregnskabet – **AMRUN.** Findes fra og med 2008. Ud over arbejdsmarkedsoplysninger indeholder registret også oplysninger om bopælskommune, herkomst og oprindelsesland.

Udvælgelsen af observationer fra et forløbsregister (på et givet statustidspunkt) kan med fordel ske, inden der matches videre mod andre registre. Herved kan man sikre sig, at man ikke får dannet et udtræk med mere end én observation pr. person på statustidspunktet. Erfaringen viser, at forløbsregistrene desværre i nogle tilfælde kan indeholde dubletter på et givet statustidspunkt, også selvom det ikke burde være tilfældet. Selvom dubletterne som regel udgør en ganske ubetydelig andel, så kan forekomsten af blot nogle få dubletter give anledning til dannelse af fejlagtige analysedata. Og hvis der ikke er sikret mod uønskede dubletter fra start, så kan det være endog rigtigt svært at gennemskue fejlen, når først analyserne er igangsat.

I afsnit 5 beskrives forløbsregistrene hver for sig med henblik på at kunne foretage en korrekt udvælgelse af observationer inden den videre match.

3.1 Kvartalsvise populationer baseret på BEF

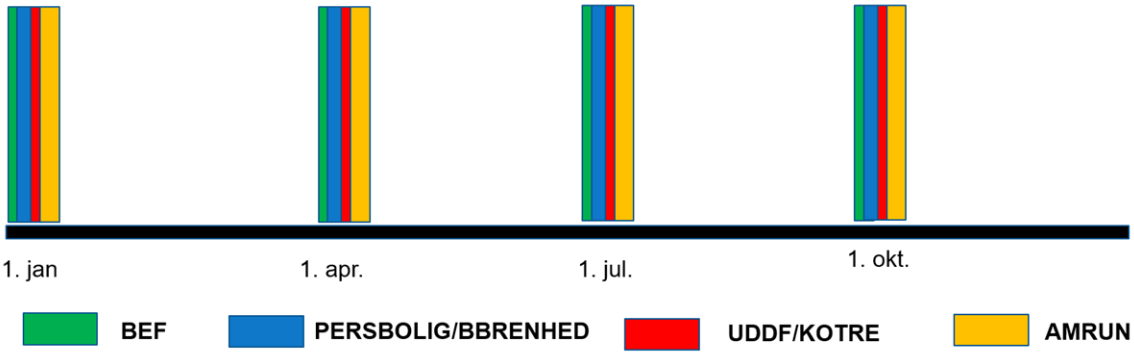
Hvis man trækker sin population i **BEF**, så er den sammenlignelig med den publicerede Befolkningsstatistik fra 2008 og frem. Befolkningsstatistikken publiceres kun ca. 6 uger efter statustidspunktet, så **BEF** kan også være relevant, hvis man ønsker en meget aktuel population.

I eksemplet nedenfor har vi dannet data for de 4 kvartaler i 2023 med udgangspunkt i disse registre:

- **BEF** - Befolkningsstatistikregistret med status hhv. d. 1. januar, 1. april, 1. juli og 1. oktober. Svarende til de publicerede tal for befolkningen.
- **UDDF** – Forløbsregister med højest fuldførte uddannelse med registerstatus 30/09 2025. Med nedslag på statustidspunkterne.
- **KOTRE** – Elevregistret med registerstatus 30/09 2025. Med nedslag på statustidspunkterne.
- **AMRUN** – Forløbsregister for året 2023. Med nedslag på statustidspunkterne.
- **PERSBOL/BBRENHED** – Boligoplysninger med statustidspunkt d. 1. januar, 1. april, 1. juli og 1. oktober. Kun 1. januar 2023 findes som publicerede tal.

Som skitseret i figur 3.1:

Figur 3.1. Kvartalsvise sammenhængende populationer



Højest fuldførte uddannelse er opgjort på hovedområderne i tabel 3.1, hvor der også sammenlignes med de publicerede tal i HFUDD16 pr. 1. oktober 2023. **UDDF** er et levende register, og der er derfor afvigelser i forhold til de publicerede tal, som er baseret på statusregistret **UDDA**. Som bruger af mikrodata skal man notere sig, at **UDDF** data er mere korrekte end de data, der er publiceret. Ved match mod **UDDF** får vi flere uoplyste end i Statistikbanken, og denne afvigelse kan tilskrives problematikken omkring konvertering til CPR-numre i DDV jf. afsnit 7.

Tabel 3.1. Befolkningen 15-69 år, fordelt på højest fuldførte uddannelse i 2023.

	Statustidspunkt								Statistikbanken	
	01/01/2023		01/04/2023		01/07/2023		01/10/2023		HFUDD16 2023	
	Antal	Pct.	Antal	Pct.	Antal	Pct.	Antal	Pct.	Antal	Pct.
I alt	4.097.219	100,0	4.104.040	100,0	4.103.272	100,0	4.116.473	100,0	4.116.473	100,0
Højste fuldførte uddannelse										
Uoplyst	49.806	1,2	53.616	1,3	50.310	1,2	46.710	1,1	46.301	1,1
10 Grundskole	968.600	23,6	976.111	23,8	937.229	22,8	947.441	23,0	948.017	23,0
20 Gymnasiale	454.802	11,1	441.660	10,8	466.115	11,4	462.292	11,2	459.639	11,2
30 Erhvervsfaglige	1.144.260	27,9	1.140.489	27,8	1.134.795	27,7	1.135.581	27,6	1.132.496	27,5
35 Adgangsgivende	2.133	0,1	2.056	0,1	2.100	0,1	2.091	0,1	2.126	0,1
40 Korte videregående	213.129	5,2	213.241	5,2	218.033	5,3	218.547	5,3	218.235	5,3
50 Mellemlange videregående	640.739	15,6	647.988	15,8	651.109	15,9	647.402	15,7	643.478	15,6
60 Bachelor	107.516	2,6	108.570	2,6	110.327	2,7	120.906	2,9	131.064	3,2
70 Lange videregående	470.513	11,5	473.831	11,5	486.060	11,8	487.583	11,8	488.711	11,9
80 Ph.d. og forsker	45.721	1,1	46.478	1,1	47.194	1,2	47.920	1,2	46.406	1,1

I tabel 3.2 ses antal elever med en igangværende uddannelse. I Statistikbankens opgørelse pr. 1/10 2023 (UDDAKT10) indgår også elever, der ikke har bopæl i Danmark, så tallene er ikke direkte sammenlignelige. De procentvise fordelinger er dog tæt på at være identiske. Af tabellen kan ses, at der er markant færre elever i grundskole og i gymnasiale uddannelser pr. 1. juli. Dette skyldes, at elever udskrives fra et klassetrin inden sommerferien og først indskrives på næste klassetrin efter sommerferien.

Tabel 3.2. Elever i befolkningen, fordelt på igangværende uddannelse i 2023.

	Statustidspunkt								Statistikbanken	
	01/01/2023		01/04/2023		01/07/2023		01/10/2023		UDDAKT10 2023	
	Antal	Pct.	Antal	Pct.	Antal	Pct.	Antal	Pct.	Antal	Pct.
I alt	1.185.415	100,0	1.170.151	100,0	292.802	100,0	1.193.032	100,0	1.200.895	100,0
Igangværende uddannelse										
10 Grundskole	679.710	57,3	679.496	58,1	9	0,0	675.999	56,7	677.256	56,4
15 Forberedende	13.307	1,1	13.626	1,2	13.144	4,5	14.177	1,2	14.845	1,2
20 Gymnasiale	138.862	11,7	136.878	11,7	1.416	0,5	139.910	11,7	140.860	11,7
29 Erhvervsfaglige grundforløb	10.331	0,9	393	0,0	119	0,0	11.773	1,0	11.811	1,0
30 Erhvervsfaglige	93.460	7,9	103.352	8,8	86.041	29,4	97.399	8,2	97.003	8,1
35 Adgangsgivende	412	0,0	407	0,0	114	0,0	498	0,0	502	0,0
40 Korte videregående	22.647	1,9	21.456	1,8	14.722	5,0	22.709	1,9	22.968	1,9
50 Mellemlange videregående	89.489	7,5	82.856	7,1	72.179	24,7	89.071	7,5	90.121	7,5
60 Bacheloruddannelser	70.006	5,9	66.369	5,7	54.092	18,5	71.052	6,0	72.216	6,0
70 Lange videregående	57.951	4,9	56.067	4,8	41.878	14,3	60.955	5,1	63.417	5,3
80 Ph.d. og forsker	9.240	0,8	9.251	0,8	9.088	3,1	9.489	0,8	9.896	0,8

I tabel 3.3 ses resultatet af match mod Arbejdsmarkedsregnskabet **AMRUN**. Her optræder der 10.000-12.000 personer, som ikke finder match i **AMRUN**. Det skyldes, at **AMRUN**s populationer er baseret på et senere CPR-udtræk end Befolkningsstatistikkens udtræk. De har samtidigt ikke nogen aktiv tilstand i **AMRUN** på statustidspunktet, så de er med stor sandsynlighed udvandret. Man kan derfor overveje, om de evt. helt skal udgå af populationen.

Tabel 3.3. Befolkningen fordelt på socioøkonomisk status i 2023.

	Statustidspunkt								Statistikbanken	
	01/01/2023		01/04/2023		01/07/2023		01/10/2023		RAS201 2023	
	Antal	Pct.	Antal	Pct.	Antal	Pct.	Antal	Pct.	Antal	Pct.
I alt	5.932.654	100,0	5.941.388	100,0	5.944.145	100,0	5.959.464	100,0	5.955.637	100,0
Socioøkonomisk status										
01 Beskæftigede	3.043.698	51,3	3.072.352	51,7	3.091.858	52,0	3.101.422	52,0	3.112.462	52,3
02 Arbejdsløse	60.944	1,0	78.790	1,3	78.366	1,3	74.414	1,2	75.450	1,3
03 Uden for arbejdsstyrken	2.817.107	47,5	2.780.129	46,8	2.762.536	46,5	2.772.517	46,5	2.767.725	46,5
99 Ikke i AMR	10.905	0,2	10.117	0,2	11.385	0,2	11.111	0,2	0	0,0

Ved tildeling af socioøkonomisk status i **AMRUN** følges de internationale retningslinjer fra ILO. Her bliver beskæftigelse prioriteret, også for de studerende der samtidigt har et lønmodtagerjob.[4] I tabel 3.4 er eleverne pr. 1. april 2023 fordelt på deres socioøkonomiske indplacering i **AMRUN**. Bortset fra grundskoleeleverne er store dele af eleverne indplaceret som en del af arbejdsstyrken. I skræddersyet statistik fra Consulting korrigerer vi ofte den socioøkonomiske status, så elever med en igangværende uddannelse får status som studerende. Det fremgår også af tabel 3.4, at der praktisk talt er 100 pct. match mellem eleverne i befolkningen og **AMRUN**, idet kun 5 elever ikke finder match i **AMRUN**.

Tabel 3.4. Elever pr 1. april 2023 fordelt på socioøkonomisk status.

	Socioøkonomisk status			I alt	Socioøkonomisk status			I alt
	I arbejdsstyrken	Uden for arbejdsstyrken	Ikke i AMRUN		I arbejdsstyrken	Uden for arbejdsstyrken	Ikke i AMRUN	
	Antal	Antal	Antal	Antal	Pct.	Pct.	Pct.	Pct.
I alt	404.433	765.713	5	1.170.151	34,6	65,4	0,0	100,0
Uddannelseskode								
10 Grundskole	72.965	606.531	0	679.496	10,7	89,3	.	100,0
15 Forberedende	2.333	11.293	0	13.626	17,1	82,9	.	100,0
20 Gymnasiale	86.699	50.177	2	136.878	63,3	36,7	0,0	100,0
29 Erhvervsfaglige grundforløb	230	163	0	393	58,5	41,5	.	100,0
30 Erhvervsfaglige	88.715	14.635	2	103.352	85,8	14,2	0,0	100,0
35 Adgangsgivende	166	241	0	407	40,8	59,2	.	100,0
40 Korte videregående	13.679	7.777	0	21.456	63,8	36,2	.	100,0
50 Mellemlange videregående	51.216	31.640	0	82.856	61,8	38,2	.	100,0
60 Bacheloruddannelser	41.633	24.735	1	66.369	62,7	37,3	0,0	100,0
70 Lange videregående	38.115	17.952	0	56.067	68,0	32,0	.	100,0
80 Ph.d. og forsker	8.682	569	0	9.251	93,8	6,2	.	100,0

[4] Se mere om definitionerne i Statistikdokumentation for Arbejdsmarkedsregnskab.

Boligopgørelsen (fx BOL201) publiceres kun pr. 1. januar i året, men nøglen mellem personer og boliger **PERBOLIG** findes kvartalsvis, så man er på sikker grund mht. personernes bopæl på statustidspunktet. Herefter kan man matche mod **BBRENHED,** som ligeledes findes kvartalsvis. I tabel 3.5 er der sammenlignet med Statistikbanken BOL201 pr. 1/1 2023. I vores match får vi lidt flere uoplyste end i Statistikbanken, og det kan tilskrives problematikken omkring konvertering til CPR-numre i DDV jf. afsnit 7.

Tabel 3.5. Befolkningen fordelt på boligernes anvendelse i 2023.

	Statustidspunkt								Statistikbank BOL201	
	01/01/2023		01/04/2023		01/07/2023		01/10/2023		01/01/2023	
	Antal	Pct.	Antal	Pct.	Antal	Pct.	Antal	Pct.	Antal	Pct.
I alt	5.932.654	100,0	5.941.388	100,0	5.944.145	100,0	5.959.464	100,0	5.932.654	100,0
Enhedsanvendelse										
Uoplyst	33.480	0,6	31.512	0,5	31.174	0,5	29.882	0,5	33.252	0,6
01 Parcel/Stuehuse	3.031.348	51,1	3.030.030	51,0	3.031.938	51,0	3.027.469	50,8	3.031.404	51,1
02 Række-, kæde- og dobbelthuse	798.841	13,5	800.864	13,5	804.011	13,5	807.739	13,6	798.866	13,5
03 Etageboligbebyggelse	1.940.105	32,7	1.947.798	32,8	1.946.717	32,8	1.959.322	32,9	1.940.227	32,7
04 Kollegier	49.657	0,8	49.892	0,8	47.697	0,8	51.128	0,9	49.673	0,8
05 Døgninstitutioner	16.786	0,3	17.315	0,3	17.687	0,3	18.176	0,3	16.789	0,3
06 Fritidshuse	39.876	0,7	40.775	0,7	41.599	0,7	41.510	0,7	39.876	0,7
07 Andet	22.561	0,4	23.202	0,4	23.322	0,4	24.238	0,4	22.567	0,4

3.2 Månedsvise populationer baseret på Lifelines

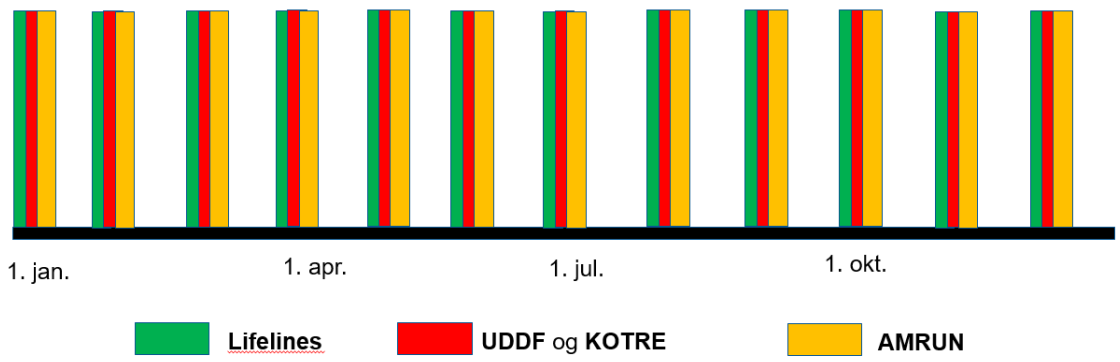
Hvis du skal bruge andre statustidspunkter end de kvartalsvise fra **BEF** registret, så kan populationer baseret på **Lifelines** være et alternativ. Fra **AMRUN** kan man ud over oplysninger om arbejdsmarkedstilknytning også hente månedsvise oplysninger om fx alder, køn, herkomst, oprindelsesland og bopæl/husstand (opgjort ultimo i måneden). Bemærk, at de familiemæssige relationer kun findes i det kvartalsvise **BEF** register.

Følgende registre indgår i eksemplet:

- **Lifelines** – Forløbsregister med alle personer som er bosiddende i Danmark med status pr. 31/12 2025, og med et nedslag på de månedsvise statustidspunkter.
- **AMRUN** – Forløbsregister for året 2023. Med nedslag på de månedlige statustidspunkter og med oplysninger opgjort ultimo i måneden.
- **UDDF** – Forløbsregister med højest fuldførte uddannelse med status 30/09 2025. Med et nedslag på de månedlige statustidspunkter.
- **KOTRE** – Elevregistret med status 30/09 2025. Med et nedslag på de månedlige statustidspunkter.

Som skitseret i figur 3.2:

Figur 3.2. Månedsvise sammenhængende populationer



I tabel 3.6 og tabel 3.7 ses de månedlige **Lifelines** populationer fordelt på socioøkonomisk status fra **AMRUN**, hhv. antal og pct. Kun ganske få personer fra **Lifelines** har ikke fundet match med **AMRUN**.

Tabel 3.6. Lifelines befolkning i 2023 fordelt på socioøkonomisk status. Antal.

	Statustidspunkt											
	01/01/2023	01/02/2023	01/03/2023	01/04/2023	01/05/2023	01/06/2023	01/07/2023	01/08/2023	01/09/2023	01/10/2023	01/11/2023	01/12/2023
	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal
I alt	5.913.865	5.918.271	5.921.346	5.923.796	5.924.736	5.925.534	5.925.714	5.926.604	5.935.873	5.942.329	5.945.696	5.947.543
Socioøkonomisk status												
01 Beskæftigede	3.043.411	3.048.575	3.062.304	3.072.049	3.093.439	3.107.941	3.091.455	3.097.516	3.107.450	3.100.964	3.107.680	3.104.720
02 Arbejdsløse	60.915	84.810	81.809	78.750	75.254	72.833	78.306	79.787	75.136	74.355	75.980	75.875
03 Uden for arbejdsstyrken	2.808.571	2.783.898	2.776.235	2.772.038	2.755.113	2.743.858	2.755.070	2.748.418	2.752.401	2.766.150	2.761.175	2.766.089
99 Ikke i AMR	968	988	998	959	930	902	883	883	886	860	861	859

Tabel 3.7. Lifelines befolkning i 2023 fordelt på socioøkonomisk status. Pct.

	Statustidspunkt											
	01/01/2023	01/02/2023	01/03/2023	01/04/2023	01/05/2023	01/06/2023	01/07/2023	01/08/2023	01/09/2023	01/10/2023	01/11/2023	01/12/2023
	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.
I alt	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Socioøkonomisk status												
01 Beskæftigede	51,5	51,5	51,7	51,9	52,2	52,4	52,2	52,3	52,4	52,2	52,3	52,2
02 Arbejdsløse	1,0	1,4	1,4	1,3	1,3	1,2	1,3	1,3	1,3	1,3	1,3	1,3
03 Uden for arbejdsstyrken	47,5	47,0	46,9	46,8	46,5	46,3	46,5	46,4	46,4	46,5	46,4	46,5
99 Ikke i AMR	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

I tabel 3.8 er **Lifelines** populationerne sammenlignet med de kvartalsvise **BEF** populationer og fordelt på socioøkonomisk status fra **AMRUN**.

Tabel 3.8. BEF og Lifelines populationerne fordelt på socioøkonomisk status 2023.

	Statustidspunkt							
	01/01/2023		01/04/2023		01/07/2023		01/10/2023	
	BEF	Lifelines	BEF	Lifelines	BEF	Lifelines	BEF	Lifelines
I alt	5.932.654	5.913.865	5.941.388	5.923.796	5.944.145	5.925.714	5.959.464	5.942.329
Socioøkonomisk status								
01 Beskæftigede	3.043.698	3.043.411	3.072.352	3.072.049	3.091.858	3.091.455	3.101.422	3.100.964
02 Arbejdsløse	60.944	60.915	78.790	78.750	78.366	78.306	74.414	74.355
03 Uden for arbejdsstyrken	2.817.107	2.808.571	2.780.129	2.772.038	2.762.536	2.755.070	2.772.517	2.766.150
99 Ikke i AMR	10.905	968	10.117	959	11.385	883	11.111	860

Resultaterne i tabel 3.8 illustrerer, at vi sammenligner tre registre, som hver især har trukket sine populationer på forskellige tidspunkter i CPR-systemet:

- **BEF**: Befolkningen trækkes i CPR-systemet ca. 1 måned efter referencetidspunktet og revideres aldrig.
- **AMRUN**: Udtræk fra CPR-systemet sker med samme tidsmæssige spænd (10 måneder) mellem reference- og registertidspunkter, måned for måned. Fx er befolkningen for november 2023 trukket i september 2024 og december 2023 er trukket i oktober 2024.
- **Lifelines:** Revideres hvert år med nye udtræk fra CPR-systemet, og der er ikke nogen tidsmæssig begrænsning i forhold til rettelserne bagud i tid.

Både **BEF** og **AMRUN** indeholder derfor personer som ifølge det seneste **Lifelines** CPR-udtræk ikke var i befolkningen på statustidspunktet. Der er helt overvejende tale om personer, der er udvandret. Læs mere om dette i afsnit 5.1.

I tabel 3.9 og tabel 3.10 ses de månedlige **Lifelines** populationer fordelt på højest fuldførte uddannelse i 2023, hhv. antal og pct.

Tabel 3.9. Lifelines befolkning 15-69 år, fordelt på højest fuldførte uddannelse. Antal.

	Statustidspunkt											
	01/01/2023	01/02/2023	01/03/2023	01/04/2023	01/05/2023	01/06/2023	01/07/2023	01/08/2023	01/09/2023	01/10/2023	01/11/2023	01/12/2023
	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal
I alt	4.080.025	4.084.210	4.086.559	4.087.616	4.087.542	4.087.167	4.086.058	4.086.339	4.094.889	4.100.866	4.103.764	4.105.824
Højste fuldførte uddannelse												
Uoplyst	47.834	50.480	51.369	51.412	51.227	49.564	47.814	46.560	46.220	46.732	47.448	47.892
10 Grundskole	967.445	969.476	972.617	974.936	978.321	981.934	935.656	939.360	943.701	945.829	949.868	953.536
20 Gymnasiale	452.192	441.728	440.241	439.280	438.742	437.877	463.678	460.535	459.107	459.720	459.455	458.745
30 Erhvervsfaglige	1.141.915	1.139.567	1.138.579	1.138.343	1.136.489	1.135.006	1.132.643	1.132.496	1.132.559	1.133.300	1.131.817	1.130.704
35 Adgangsgivende	2.128	2.069	2.061	2.052	2.044	2.043	2.097	2.081	2.091	2.086	2.079	2.073
40 Korte videregående	211.782	211.755	211.898	212.016	211.986	211.919	216.723	216.631	216.941	217.113	217.130	217.190
50 Mellemlange videregående	637.686	646.853	646.067	645.041	643.862	643.042	648.245	646.845	645.835	644.539	643.522	642.532
60 Bachelor	106.354	106.450	107.069	107.516	107.767	107.725	109.173	111.305	116.783	119.513	120.077	120.510
70 Lange videregående	467.675	470.542	471.121	471.248	471.130	471.809	483.542	483.830	484.704	484.957	485.065	485.049
80 Ph.d. og forsker	45.014	45.290	45.537	45.772	45.974	46.248	46.487	46.696	46.948	47.077	47.303	47.593

Tabel 3.10. Lifelines befolkning 15-69 år, fordelt på højest fuldførte uddannelse. Pct.

	Statustidspunkt											
	01/01/2023	01/02/2023	01/03/2023	01/04/2023	01/05/2023	01/06/2023	01/07/2023	01/08/2023	01/09/2023	01/10/2023	01/11/2023	01/12/2023
	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.
I alt	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Højste fuldførte uddannelse												
Uoplyst	1,2	1,2	1,3	1,3	1,3	1,2	1,2	1,1	1,1	1,1	1,2	1,2
10 Grundskole	23,7	23,7	23,8	23,9	23,9	24,0	22,9	23,0	23,0	23,1	23,1	23,2
20 Gymnasiale	11,1	10,8	10,8	10,7	10,7	10,7	11,3	11,3	11,2	11,2	11,2	11,2
30 Erhvervsfaglige	28,0	27,9	27,9	27,8	27,8	27,8	27,7	27,7	27,7	27,6	27,6	27,5
35 Adgangsgivende	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1
40 Korte videregående	5,2	5,2	5,2	5,2	5,2	5,2	5,3	5,3	5,3	5,3	5,3	5,3
50 Mellemlange videregående	15,6	15,8	15,8	15,8	15,8	15,7	15,9	15,8	15,8	15,7	15,7	15,6
60 Bachelor	2,6	2,6	2,6	2,6	2,6	2,6	2,7	2,7	2,9	2,9	2,9	2,9
70 Lange videregående	11,5	11,5	11,5	11,5	11,5	11,5	11,8	11,8	11,8	11,8	11,8	11,8
80 Ph.d. og forsker	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,2	1,2

I tabel 3.11 og tabel 3.12 ses de månedlige **Lifelines** populationer fordelt på elever med en igangværende uddannelse i 2023, hhv. antal og pct.

Tabel 3.11. Lifelines befolkning, elever fordelt på igangværende uddannelse. Antal.

	Statustidspunkt											
	01/01/2023	01/02/2023	01/03/2023	01/04/2023	01/05/2023	01/06/2023	01/07/2023	01/08/2023	01/09/2023	01/10/2023	01/11/2023	01/12/2023
	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal	Antal
I alt	1.185.392	1.181.639	1.178.385	1.170.132	1.167.647	1.162.224	292.726	1.045.046	1.195.988	1.193.329	1.190.189	1.185.009
Igangværende uddannelse												
10 Grundskole	679.605	680.045	679.799	679.404	679.094	678.957	9	675.432	675.576	675.906	676.121	675.978
15 Forberedende	13.360	13.484	13.604	13.676	13.834	13.661	13.189	13.662	14.215	14.207	14.427	14.455
20 Gymnasiale	138.992	138.208	137.659	136.979	136.573	136.121	1.421	79.193	140.573	139.991	139.406	138.690
29 Erhvervsfaglige grundforløb	10.336	352	379	394	378	403	119	158	12.027	11.778	11.546	11.194
30 Erhvervsfaglige	93.497	108.554	107.781	103.392	103.466	101.462	86.066	88.405	100.758	97.417	97.022	95.495
35 Adgangsgivende	414	419	421	409	406	398	114	231	507	497	486	469
40 Korte videregående	22.599	22.227	21.753	21.417	21.202	20.866	14.709	14.704	22.423	22.711	22.489	22.151
50 Mellemlange videregående	89.465	83.698	84.011	82.833	82.136	81.387	72.162	71.877	88.333	89.055	88.207	87.470
60 Bacheloruddannelser	70.065	68.264	67.077	66.395	65.932	65.393	54.106	51.554	71.699	71.193	70.546	69.901
70 Lange videregående	57.841	57.110	56.618	55.990	55.398	54.390	41.756	40.723	60.482	61.089	60.439	59.787
80 Ph.d. og forsker	9.218	9.278	9.283	9.243	9.228	9.186	9.075	9.107	9.395	9.485	9.500	9.419

Tabel 3.12. Lifelines befolkning, elever fordelt på igangværende uddannelse. Pct.

	Statustidspunkt											
	01/01/2023	01/02/2023	01/03/2023	01/04/2023	01/05/2023	01/06/2023	01/07/2023	01/08/2023	01/09/2023	01/10/2023	01/11/2023	01/12/2023
	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.	Pct.
I alt	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Igangværende uddannelse												
10 Grundskole	57,3	57,6	57,7	58,1	58,2	58,4	0,0	64,6	56,5	56,6	56,8	57,0
15 Forberedende	1,1	1,1	1,2	1,2	1,2	1,2	4,5	1,3	1,2	1,2	1,2	1,2
20 Gymnasiale	11,7	11,7	11,7	11,7	11,7	11,7	0,5	7,6	11,8	11,7	11,7	11,7
29 Erhvervsfaglige grundforløb	0,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	1,0	1,0	0,9
30 Erhvervsfaglige	7,9	9,2	9,1	8,8	8,9	8,7	29,4	8,5	8,4	8,2	8,2	8,1
35 Adgangsgivende	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
40 Korte videregående	1,9	1,9	1,8	1,8	1,8	1,8	5,0	1,4	1,9	1,9	1,9	1,9
50 Mellemlange videregående	7,5	7,1	7,1	7,1	7,0	7,0	24,7	6,9	7,4	7,5	7,4	7,4
60 Bacheloruddannelser	5,9	5,8	5,7	5,7	5,6	5,6	18,5	4,9	6,0	6,0	5,9	5,9
70 Lange videregående	4,9	4,8	4,8	4,8	4,7	4,7	14,3	3,9	5,1	5,1	5,1	5,0
80 Ph.d. og forsker	0,8	0,8	0,8	0,8	0,8	0,8	3,1	0,9	0,8	0,8	0,8	0,8

4. Statusregistrene

4.1 Højest fuldførte uddannelse (UDDA)

Grundpopulationen i registret med befolkningens højest fuldførte uddannelse (**UDDA**) lægger sig helt op ad Befolkningsstatistikkens populationer jf. afsnit 2.1.

Uddannelseskoden er den centrale variabel i registret. Der er tale om en 4-cifret kode, der benyttes i forbindelse med indberetningerne til Danmarks Statistik. Når statistikken offentliggøres, omsættes den 4-cifrede kode til DISCED-15 koder, som er Danmarks Statistiks hierarkiske klassifikationssystem for uddannelser. Nedenfor ses et eksempel, hvor den 4-cifrede kode '5082 Multimediedesigner' er indplaceret i forhold til DISCED-15 *Hovedområder*. På det mest overordnede niveau er uddannelsen defineret som værende en kort videregående uddannelse.



Se nærmere i Klassifikationsdatabasen: https://www.dst.dk/da/Statistik/dokumentation/nomenklaturer/disced15-audd

Det betyder, at der skal tages stilling til, om og hvordan uddannelserne skal grupperes, når uddannelseskoden benyttes. Ud over *Hovedområde* er det også muligt at opdele på *Uddannelsestyper*, *Uddannelsesniveau* og *Fagområde*. Hertil findes de internationale inddelinger *ISCED Level* og *ISCED Field*.[5] I det omfang der sker ændringer i klassifikationerne, kan det påvirke mulighederne for at sammenligne med de offentliggjorte tal i Statistikbanken.

- 1981-1990: Her findes ikke nogen offentliggjort statistik.
- 1991-2007: Når der sammenlignes med HFU1, er det ikke muligt genskabe de publicerede tal. Dels er det ikke muligt at genskabe den klassifikation af uddannelseskoden (*Forspalten*), der blev benyttet ved publiceringen. Dels er der ikke et sammenfald mellem de versioner af registret, der findes i **UDDA**, og dem der er benyttet i forbindelse med de offentliggjorte tal.
- 2008 - : Populationerne opgøres pr. 30 september. Det senest publicerede år er altid at betragte som foreløbigt, og den endelige version af registret vil først foreligge, når det efterfølgende år publiceres. Der sker løbende ændringer i DISCED-15 klassifikationen, og den aktuelle klassifikation benyttes kun ved publiceringen af de to seneste år i Statistikbanken, dvs. for de foreløbige tal for år x og de endelige tal for år x-1. Hovedreglen er derfor, at Statistikbankens tal kun kan genskabes for de to seneste år. Hvis man som forsker skal sammenligne data hen over årerne, så er det mest korrekt at bruge den seneste version af

[5] Her kan man med stor fordel benytte de SAS formater og kodelister, som stilles til rådighed på forskermaskinerne.

klassifikationen, upåagtet at resultatet dermed ikke er i overensstemmelse med Statistikbanken.

For at sammenligne med Statistikbanken skal befolkningen aldersafgrænses til de 15-69-årige via match med **BEF**. Bemærk, at brugere i DDV ikke opnår 100 pct. match mellem registrene, selvom de har samme population og samme statustidspunkt. Afvigelsen kan tilskrives problematikken omkring konvertering til CPR-numre i DDV jf. afsnit 7.

Oplysninger om højest fuldførte uddannelse bliver dannet på baggrund af elevregisteret og kvalifikationsregisteret, hvor der bl.a. foretages imputering af indvandreres medbragte uddannelse. Oplysningerne bliver prioriteret på baggrund af kildens troværdighed, og ved brug af variablen HF_KILDE kan der skelnes mellem, hvor oplysningen om højest fuldførte uddannelse stammer fra, hvilket i nogle henseender kan være væsentligt i forhold til brugen af data.

UDDA er et statusregister og er baseret på forløbsregistret **UDDF**. **UDDF** revideres hvert år bagud i tid, idet der ikke er nogen tidsmæssig bagkant. Dette i modsætning til **UDDA**, som kun revideres et år bagud i tid. Hvis forløb i personernes højest fuldførte uddannelse er af interesse, så anbefales det at benytte **UDDF** og ikke **UDDA**. Det skal bemærkes, at indvandrernes medbragte uddannelse imputeres, når det ikke har været muligt at få en uddannelsesoplysning på anden måde. Imputerede værdier er oplysninger, som erstatter uoplyste værdier, baseret på besvarelser fra tilsvarende personer. De imputerede data er anvendelige i overordnede statistiske opgørelser, men kan ikke betragtes som valide oplysninger om enkeltpersoners uddannelsesniveau. Når **UDDA** benyttes, kan en imputeret uddannelse på personniveau skifte fra det ene statusår til det næste. Det er endnu en årsag til, at **UDDF** bør benyttes, når personer skal følges over tid.

UDDA indeholder oplysninger om en evt. igangværende uddannelse, men den findes ikke i **UDDF**. Her kan man i stedet benyttes Elevregistret **KOTRE** jf. afsnit 3. .

4.2 Registerbaseret arbejdsstyrke (RAS)

I **RAS**' grundpopulation indgår alle, som er i befolkningen på statustidspunktet jf. afsnit 2.1. Derudover er der nogle variationer i grundpopulationerne, når der ses på perioden fra og med 1980:

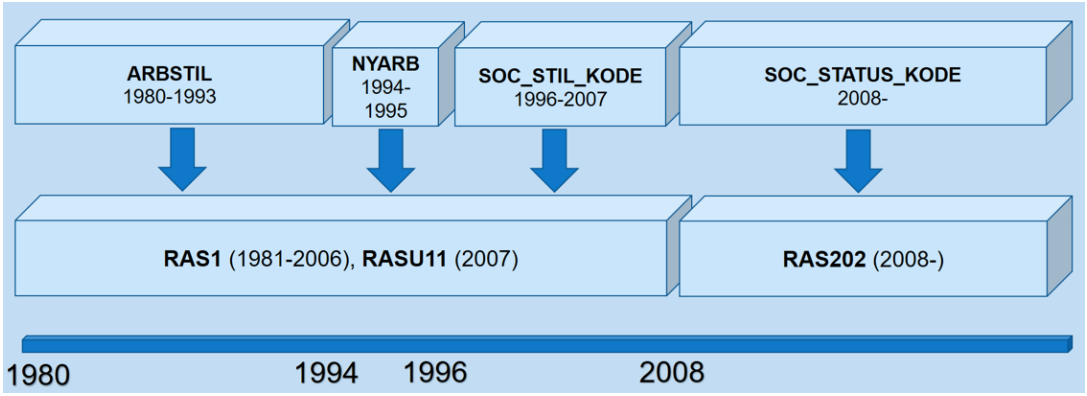
- 1980-1989: Her er populationen identisk med den population, der indgår i befolkningen (**FAIN**). Indeholder kun befolkningens primære beskæftigelse.
- 1990-2007: Indeholder ikke kun den primære beskæftigelse men også sekundær og tertiær beskæftigelse samt ikke november beskæftigelse. For selektion af den primære beskæftigelse svarende til den offentliggjorte statistik bruges variablen NOVPRIO=1.
- 2008-: For at afgrænse til den offentliggjorte statistik benyttes de to variable PRIMAER_STATUS_KODE=1 og I_BEFOLKNINGEN_KODE=1.

Ved match med **FAIN** skal bemærkes, at årsangivelsen er pr. 1. januar året efter, dvs. at **RAS1980** fx skal matches med **FAIN1981**. Bemærk endvidere, at **RAS2007** med status pr. november 2007 skal matches med **BEF200712** og ikke **FAIN**. Fra og med **RAS2008** (status pr. november 2008) er der ikke længere sammenfald mellem **RAS** populationen og Befolkningsstatistikken jf. tidligere afsnit.

RAS er fra og med 2008 baseret på Arbejdsmarkedsregnskabet (**AMRUN**), som dækker hele kalenderåret. For at få en så retvisende befolkning som mulig, så vælges et CPR-udtræk senere end det, der benyttes i Befolkningsstatistikken. Således er befolkningen for december 2021 trukket i oktober 2022, befolkningen i november 2021 er trukket i september 2022 osv. Derved er der samme tidsmæssige spænd mellem reference- og registertidspunkter måned for måned, og det sikrer, at beskæftigelses- og erhvervsfrekvenserne bliver sammenlignelige hen over året.

For **RAS** gælder, at de offentliggjorte tal i Statistikbanken i udgangspunktet kan rammes. Der ses dog nogle marginale afvigelser i enkelte år før 2008. En af de helt centrale variable i **RAS** angiver befolkningens tilknytning til arbejdsmarkedet. Variablen har i perioden været genstand for flere databrud og har haft 4 forskellige navne ARBSTIL, NYARB, SOCSTIL_KODE samt SOC_STATUS_KODE. Hvis selektionen af **RAS**' grundregister (som angivet ovenfor) følges, ses der stor overensstemmelse i forhold til de offentliggjorte tal i Statistikbanken jf. figur 4.1.

Figur 4.1. Socioøkonomisk status og Statistikbanken



4.3 Boligopgørelsen (BOL og BOLIG)

Boligopgørelsen opgør antal boliger/personer med oplysninger om boligerne og deres beboere. Der tages udgangspunkt i den samme population, som findes i Befolkningsstatistikregistret pr. 1. januar i året. Boligopgørelsens fundament er et match på adresserne mellem CPR- og BBR-registret. Der er tilfælde, hvor adresserne er registreret forskelligt i de to registre, hvilket betyder, at der findes personer/boliger uden match. Der vil dermed aldrig være en fuldstændig overensstemmelse mellem de to registre. Dette forhold er beskrevet i Statistikdokumentationen.

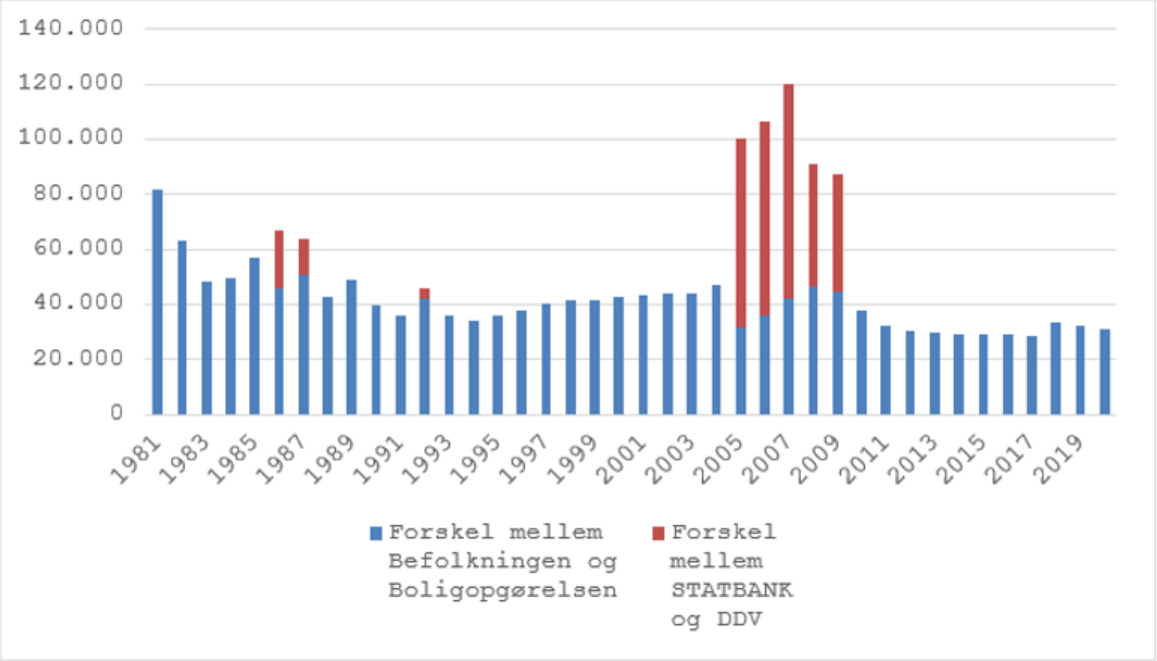
Boligopgørelsen blev omlagt i 2010, og når forskningsdataserierne benyttes, er registret opdelt i to dataserier; **BOL** (1981-2009) og **BOLIG** (2010 og frem).

BOL: Årerne 1981 til 2009 (statustidspunktet for dataserien er 31/12 året inden). Dataserien indeholder en lang række bolig- og bygningsoplysninger fra BBR. I praksis skal **BOL** matches med befolkningen i **FAIN** dataserien i perioden 1981-2007. Match mod **FAIN** sker via adressekoden KOM og BOPIKOM. Da **BOL** indeholder både beboede og ubeboede boliger og adresser i CPR uden match, skal disse koder benyttes for en korrekt udvælgelse i årerne 1981-2004: CPR_BBR_MATCHKODE_GL in('1',2') og BOLIGART_GL ^='88'.

For en sammenligning med de publicerede tal i Statistikbanken gælder:

- Det er ikke muligt at genskabe de offentliggjorte tal i Statistikbanken BOL6 for årene 1986-1987. Årsagen er, at metoden for match mellem BBR- og CPR-registret i 1988 blev omlagt med tilbagevirkende kraft for årene 1986 og 1987.
- I 1992 ses en mindre afvigelse i forhold til BOL6.
- Det er ikke muligt at genskabe de offentliggjorte tal i Statistikbanken for årene 2005-2009, da der mangler et stort antal boligenheder i **BOL**. Et sted mellem 70.000-80.000 personer pr. år finder ikke match, når der sammenlignes med de publicerede tal i Statistikbanken jf. figur 4.2.

Figur 4.2. Antal personer, der ikke finder match mellem CPR og BBR samt antal personer, der ikke finder match mellem Statistikbanken og BOL/BOLIG



For 2008 og 2009 skal **BOL** matches mod **BEF** dataserien. Nøglen i **BOL** er fortsat KOM og BOPIKOM, mens **BEF** indeholder variablen ADRESSE_ID. Sammenhængen mellem de to nøgler findes i dataserien **BEFADR**. Bemærk samtidigt, at **BOL** ikke er retvisende i 2008 og 2009, idet der mangler et stort antal boligenheder.

BOLIG: Fra 2010 og frem (statustidspunktet for dataserien er 31/12 året inden). I forbindelse med omlægningen af Boligopgørelsen i 2010 blev der dannet en ny nøgle mellem personerne fra Befolkningsstatistikken og boligenhederne, og denne ligger i dataserien **PERBOLIG**. Nøglen mellem **BOLIG** og **PERBOLIG** er variablen DST_BOLIGERHVERVSENHED_ID, og i **PERBOLIG** kan variablen DST_BOLIGMATCH med fordel benyttes for at udvælge de personer, der har fundet match med en bolig. De personer, der ikke matcher en bolig, optræder med DST_BOLIGMATCH=999. **BOLIG** indeholder kun et begrænset antal variable, og BBR-variablene skal hentes i dataserien **BBRENHED**. Her er nøglen også DST_BOLIGERHVERVSENHED_ID.

Med udgangspunkt i dataserierne **BOLIG**, **PERBOLIG** og **BBRENHED** er det muligt at genskabe de publicerede tal i Statistikbanken både mht. antal boliger og antal personer.

5. Forløbsregistrene

5.1 Lifelines

Lifelines opgør personers opholdsperioder i befolkningen ud fra start- og slutdatoer, samt hvilke begivenheder som ligger til grund for hhv. start- og slutdatoer (fx fødsel, indvandring, udvandring, osv.). Ligeledes er individets fødselsdato registreret, så alder nemt kan beregnes på et givent tidspunkt. Registeret benyttes i udgangspunktet ikke til publiceret statistik, men kan være fordelagtigt at bruge i tilfælde, hvor en population med statustidspunkter udenfor kvartalsskift skal matches med befolkningen.

Her ses nogle tænkte eksempler omfattende 3 personer:

PNR	FOED_DATO	EVENT_START_DATE	EVENT_FINAL_DATE	EVENT_START_CAUSE	EVENT_FINAL_CAUSE
555555555555	15/05/1916	01/01/1968	14/11/1987	4. Født før 1968	1. Død
777777777777	19/06/1956	20/07/1972	17/10/1974	6. Indvandret	5. Udvandret
777777777777	19/06/1956	09/12/1974	28/03/2012	6. Indvandret	5. Udvandret
777777777777	19/06/1956	29/10/2017	17/07/2024	6. Indvandret	5. Udvandret
999999999999	07/03/1988	07/03/1988	31/12/2025	3. Født	3. I befolkningen

Personer, der var i befolkningen ved CPR-systemets oprettelse i 1968, vil altid have EVENT_START_DATE=”01-01-1968”. Personer, der fortsat er i befolkningen, når registret opdateres, vil altid have registrets referencedato som slutdato. Referencedatoen for det aktuelle register er 31/12/2025. Det gælder fx for person ’999999999999’ der fortsat er i befolkningen og har fået EVENT_FINAL_DATE=”31-12-2025”.

Da der ikke er overlap mellem perioderne for den enkelte person, kan man udsøge den danske befolkning på et vilkårligt statustidspunkt fx d. 14. marts 2018:

EVENT_START_DATE<="14mar2018"d<=EVENT_FINAL_DATE

Med udgangspunkt i de tænkte eksempler fås dette resultat:

PNR	FOED_DATO	EVENT_START_DATE	EVENT_FINAL_DATE	EVENT_START_CAUSE	EVENT_FINAL_CAUSE
777777777777	19/06/1956	29/10/2017	17/07/2024	6. Indvandret	5. Udvandret
999999999999	07/03/1988	07/03/1988	31/12/2025	3. Født	3. I befolkningen

Hvis man skal trække en kohorte, kan det dermed ske ved brug af **Lifelines**. Man kan udvælge personer, som har været i landet i en afgrænset tidsperiode vha. EVENT_START_DATE og EVENT_FINAL_DATE. En afgrænsning på alder kan ske ved at bruge fødselsdagen til at udregne alderen på de relevante nedslagstidspunkter. Eksempelvis kan vi udsøge personer på 18 år eller derover, som var i befolkningen på et vilkårligt tidspunkt i perioden 2010-2019 med dette resultat:

PNR	FOED_DATO	EVENT_START_DATE	EVENT_FINAL_DATE	EVENT_START_CAUSE	EVENT_FINAL_CAUSE
777777777777	19/06/1956	09/12/1974	28/03/2012	6. Indvandret	5. Udvandret
777777777777	19/06/1956	29/10/2017	17/07/2024	6. Indvandret	5. Udvandret
999999999999	07/03/1988	07/03/1988	31/12/2025	3. Født	3. I befolkningen

I eksemplet ses en person med to forløb i løbet af tidsperioden, idet vedkommende er ud- og indvandret igen. Skal begge forløb indgå i analyserne eller evt. kun det første forløb? Det er den type spørgsmål, man skal tage stilling til, og det gør det mere kompliceret at afgrænse kohortepopulationerne sammenlignet med statuspopulationerne.

Til inspiration peger vi her på to studier, hvor Data Science Lab har trukket populationerne i **Lifelines**:

Faldet i incidensen af demens

I et studie ser man på faldet i incidensen af demens, og hvilken betydning uddannelse, formue og sundhed har herpå. Fra **Lifelines** udvælges alle personer på 65 år eller derover, som boede i Danmark på et tidspunkt i perioden 2005-2018. De skal samtidigt have haft bopæl i Danmark i mindst et år før deres indtræden i kohorten. Populationen er på i alt 1.757.168 personer, som er født i perioden 1896-1953. De øvrige registeroplysninger hentes i forløbsregistrene **LPR** (Landspatientregistret), **LMDB** (Lægemidler) og **UDDF** (højest fuldførte uddannelse), statusregistret **BEF** (Befolkningen) og årsregistret **FAIK** (Familieindkomster). Idet man vælger at inddrage **FAIK**, så skal man også matche mod **BEF**, hvor man finder sammenhængen mellem personer og familier. Samlet set er registergrundlaget dermed sammensat af både forløbs-, status- og årsregistre.

Nedenfor ses populationen i studiet for hhv. 2005 og 2018.

	2005		2018	
	Men (*N* = 350,557)	Women (*N* = 466,812)	Men (*N* = 516,540)	Women (*N* = 607,760)
Age				
65-69 years	117,236 (33.4)	125,217 (26.8)	155,369 (30.1)	162,925 (26.8)
70-74 years	89,358 (25.5)	104,111 (22.3)	157,813 (30.6)	169,262 (27.9)
75-79 years	68,349 (19.5)	89,969 (19.3)	100,965 (19.5)	116,704 (19.2)
80-84 years	45,347 (12.9)	74,791 (16.0)	60,593 (11.7)	79,100 (13.0)
85+ years	30,267 (8.6)	72,724 (15.6)	41,800 (8.1)	79,769 (13.1)

Kilde: Hegelund et al. (2022); The plasticity of late-onset dementia[6]

Effekter ved langtidspåvirkning af luftforurening

I et andet studie ser man på de sundhedsmæssige effekter ved langtidspåvirkning af luftforurening.[7] Her består populationen af personer på 30 år og derover med bopæl i Danmark pr. 1. januar 2000,

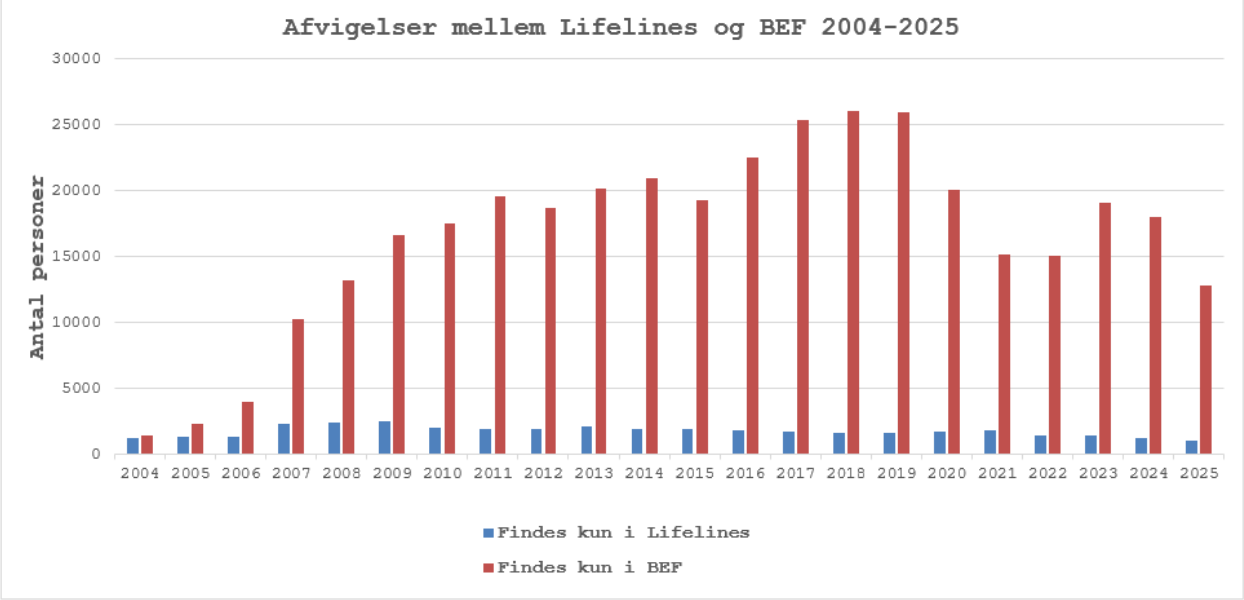
[6] Hegelund, Emilie R., et al. "The Plasticity of Late-Onset Dementia: A Nationwide Cohort Study in Denmark." *Alzheimer's & Dementia*, vol. 18, no. 7, 2022, pp. 1287–95, doi:10.1002/alz.12469.

[7] Hegelund, Emilie Rune, et al. "Air Pollution and Human Health: A Phenome-Wide Association Study." *BMJ Open*, vol. 14, no. 2, 2024, p. e081351, doi:10.1136/bmjopen-2023-081351.

som havde boet i landet mindst ét år, hvilket er 3.323.612. Personer uden adresser med geolokationer ekskluderes, hvilket giver en endelig population på 3.111.988 personer. Personerne følges frem til første forekomst af udfaldet af interesse, udvandring, død eller 31. december 2017, hvad end der forekommer først. De øvrige registeroplysninger hentes i forløbsregistrene **LPR** (Landspatientregistret), **DODSAARS/DODSAASG** (Dødsårsagsregistret) og **UDDF** (højest fuldførte uddannelse), statusregistret **BEF** (Befolkningen) og årsregistrene **AKM** (Arbejdsklassifikationsmodulet) og **FAIK** (Familieindkomster). Desuden kobles med eksterne data, fx geolokationer (x, y-koordinater) for luftforurening.

I begge studier hentes oplysninger fra statusregistret **BEF**. I denne match skal man være opmærksom på, at der er færre personer i **Lifelines** end i **BEF**, når man sammenligner de to populationer på et bestemt statustidspunkt jf figur 5.1. Det kan forklares ved de rettelser, som sker i CPR-systemet efter, at Danmarks Statistik har trukket **BEF** populationen.

Figur 5.1. Antal personer uden match mellem Lifelines og BEF pr. 1. januar i året



Der er helt overvejende tale om ud- og indvandringer. som registreres på et senere tidspunkt i CPR-systemet. Hvis man fx sammenligner de to populationer pr. 1. januar 2023. så er der 19.112 personer, der kun findes i **BEF,** og 1.419 personer, der kun findes i **Lifelines**. Ud af de 19.112 personer så kan man genfinde 18.164 af disse i **Lifelines** på andre statustidspunkter, hvilket fremgår af tabel 5.1. Et stort antal på 15.635 personer er, ifølge **Lifelines,** udvandret før 1. januar 2023, og 2.401 er indvandret efter 1. januar 2023.

Tabel 5.1. Personer fra BEF uden match pr. 1. januar 2023, som kan genfindes på andre tidspunkter i Lifelines.

		EVENT_FINAL_CAUSE					I alt
		Død, forsvundet, korrigeret		I befolkningen	Udvandret		
		År for hændelse		År for hændelse	År for hændelse		
		1968-2022	2023-2025	2023-2025	1968-2022	2023-2025	
EVENT_START_CAUSE	År for hændelse						
Født, Fundet	1968-2022	15	0	0	653	0	668
	2023-2025	0	1	75	0	3	79
Indvandret	1968-2022	34	0	0	14.982	0	15.016
	2023-2025	0	12	1.978	0	411	2.401
I alt		49	13	2.053	15.635	414	18.164

Ud af de 1.419 personer der kun findes i **Lifelines**, så er 1.204 personer indvandret før 1. januar 2023 jf. tabel 5.2.

Tabel 5.2. Personer fra Lifelines uden match pr. 1. januar 2023.

		EVENT_FINAL_CAUSE			I alt
		Død, forsvundet, korrigeret	I befolkningen	Udvandret	
		År for hændelse	År for hændelse	År for hændelse	
		2023-2025	2023-2025	2023-2025	
EVENT_START_CAUSE	År for hændelse				
Født, Fundet	1968-2022	62	118	35	215
Indvandret	1968-2022	125	634	445	1.204
I alt		187	752	480	1.419

5.2 Højest fuldførte uddannelse (UDDF)

UDDF er et forløbsregister med en start/slutdato for personernes højest fuldførte uddannelse. Registret opdateres hvert år i forbindelse med publiceringen pr. 1. oktober. Det gælder at:

- Registret indeholder hele historikken. Der er således én observation pr. person pr. tilstand af højest fuldførte uddannelse. Ingen personer udgår af registret, også selvom de er døde eller udvandrede.
- Registret revideres hvert år bagud i tid, og der er ikke er nogen tidsmæssig bagkant i forhold til revisionerne. Bemærk, at dette er i modsætning til statusregistret **UDDA**, som kun revideres et år bagud i tid.

Her ses et tænkt eksempel, hvor en person indgår i registret med 6 observationer.

Obs	PNR	HFAUDD	HF_VFRA	HF_VTIL
1	9999999999	1106 6. klasse	16/02/2012	14/02/2013
2	9999999999	1107 7. klasse	15/02/2013	14/02/2014
3	9999999999	1108 8. klasse	15/02/2014	15/02/2015
4	9999999999	1109 9. klasse	16/02/2015	12/02/2018
5	9999999999	5098 3. år, hhx	13/02/2018	09/07/2022
6	9999999999	4802 HA almen erhvervsøkonomi, bach.	10/07/2022	31/12/9999

Datoerne i eksemplet er vilkårligt sat, men det gælder, at der ikke er overlap mellem perioderne, og der er ikke nogen huller mellem perioderne, dvs. at en afsluttet periode altid afløses af en ny periode, der starter dagen efter. Den senest opnåede fuldførte uddannelse vil altid have tidsangivelsen HF_VTIL='31/12/9999'. Den højest fuldførte uddannelse kan helt entydigt udsøges på et vilkårligt tidspunkt fx d. 1. marts 2022:

```
HF_VFRA<="01mar2022"d<=HF_VTIL
```

Og så fås dette resultat i forhold til det tænkte eksempel:

Obs	PNR	HFAUDD	HF_VFRA	HF_VTIL
5	9999999999	5098 3. år, hhx	13/02/2018	09/07/2022

5.3 Elevregistret (KOTRE)

De igangværende uddannelser kan hentes i Elevregistret **KOTRE**. Registret blev etableret tilbage i starten af 1970'erne og er et forløbsregister, hvor de enkelte studerende kan følges gennem deres uddannelseskarriere helt fra 0. klasse til evt. forskeruddannelse. Registeret omfatter studerende ved de ordinære uddannelser, der er offentligt regulerede.

Registret opdateres hvert år i forbindelse med publiceringen pr. 1. oktober. Det gælder at:

- Registret indeholder hele historikken. Der er én observation pr. person pr. påbegyndt uddannelse/uddannelsesdel. De afbrudte uddannelser indgår, og der vil være en ny observation, hvis eleven har skiftet institution eller har holdt en pause i studieforløbet.
- Registret revideres hvert år bagud i tid, idet der ikke er nogen tidsmæssig bagkant i forhold til revisionerne.
- Elevregistret er et grundregister, der benyttes til publicering af oplysninger om fuldtidsuddannelser fx UDDAKT10. I den forbindelse skal det bemærkes, at Statistikbanken kun revideres èt år bagud i tid.

Her ses et tænkt eksempel, hvor en person indgår i Elevregistret med 14 observationer.

Obs	PNR	UDD	UDEL	ELEV3_VFRA	ELEV3_VTIL
1	8888888888	1114 Grundskole til og med 6. kl	1. klassetrin	08/05/2006	07/04/2007
2	8888888888	1114 Grundskole til og med 6. kl	2. klassetrin	09/05/2007	06/04/2008
3	8888888888	1114 Grundskole til og med 6. kl	3. klassetrin	08/05/2008	06/04/2009
4	8888888888	1114 Grundskole til og med 6. kl	4. klassetrin	08/05/2009	06/04/2010
5	8888888888	1114 Grundskole til og med 6. kl	5. klassetrin	08/05/2010	07/04/2011
6	8888888888	1114 Grundskole til og med 6. kl	6. klassetrin	09/05/2011	06/04/2012
7	8888888888	1116 Grundskole 7.-9. kl.	7. klassetrin	08/05/2012	06/04/2013
8	8888888888	1116 Grundskole 7.-9. kl.	8. klassetrin	08/05/2013	17/11/2013
9	8888888888	1116 Grundskole 7.-9. kl.	8. klassetrin	18/11/2013	06/04/2014
10	8888888888	1116 Grundskole 7.-9. kl.	9. klassetrin	08/05/2014	07/04/2015
11	8888888888	5090 Hhx	1. klassetrin	21/05/2015	06/04/2016
12	8888888888	5090 Hhx	2. klassetrin	16/05/2016	05/04/2017
13	8888888888	5090 Hhx	3. klassetrin	08/05/2017	06/04/2018
14	8888888888	5127 Socialrådgiver, prof.bach.	Udelt uddannelse	08/06/2018	24/10/2021

Datoerne i eksemplet er vilkårligt sat, men det gælder, at der ikke er overlap mellem perioderne. Det giver den fordel, at en evt. igangværende uddannelse entydigt kan udsøges på et vilkårligt tidspunkt, fx d. 1. marts 2021:

ELEV3_VFRA<="01mar2021"d<=ELEV3_VTIL

Og så fås dette resultat i forhold til det tænkte eksempel:

Obs	PNR	UDD	UDEL	ELEV3_VFRA	ELEV3_VTIL
14	8888888888	5127 Socialrådgiver, prof.bach.	Udelt uddannelse	08/06/2018	24/10/2021

5.4 Arbejdsmarkedsregnskabet (AMRUN)

Arbejdsmarkedsregnskabet er et årligt forløbsregister med befolkningens tilknytning til arbejdsmarkedet. Hvert forløb indeholder oplysninger om timer i forløbet, og for de beskæftigede findes oplysninger om deres arbejdssteder. Den registerbaserede arbejdsstyrkestatistik **RAS** er et afledt statusregister, hvor statustidspunktet er sat til ultimo november i året. For **AMRUN** gælder følgende:

- Samme person kan have flere forskellige arbejdsmarkedsforløb på samme tidspunkt, men det vil som regel kun være interessant at se på personens primære tilknytning til arbejdsmarkedet. Der skal således selekteres på PRIMAER_STATUS_KODE=1
- Registret indeholder tilstande for personer, som ikke nødvendigvis er i befolkningen fx grænsependlere fra Sverige og Tyskland. Omvendt indeholder det også personer, som er i befolkningen på et givet tidspunkt, selvom de ikke er registreret med en tilstand. Hvis populationen skal afgrænses til personer, der er i befolkningen, skal der selekteres på I_BEFOLKNINGEN_KODE=1.

Her ses et tænkt eksempel, hvor en person indgår med 16 observationer i 2023. Personen indgår i befolkningen, og kun primær beskæftigelse er medtaget.

Obs	PNR	I_BEFOLKNINGEN_KODE	PRIMAER_STATUS_KODE	SOC_STATUS_KODE	FRA_DATO	TIL_DATO
1	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/01/2023	31/01/2023
2	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/02/2023	06/02/2023
3	7777777777	1	1	200 Arbejdsløse	07/02/2023	26/02/2023
4	7777777777	1	1	132 Lønmodtagere på højeste niveau	27/02/2023	28/02/2023
5	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/03/2023	31/03/2023
6	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/04/2023	30/04/2023
7	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/05/2023	01/05/2023
8	7777777777	1	1	200 Arbejdsløse	02/05/2023	31/05/2023
9	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/06/2023	30/06/2023
10	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/07/2023	31/07/2023
11	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/08/2023	31/08/2023
12	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/09/2023	30/09/2023
13	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/10/2023	31/10/2023
14	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/11/2023	26/11/2023
15	7777777777	1	1	132 Lønmodtagere på højeste niveau	27/11/2023	30/11/2023
16	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/12/2023	31/12/2023

Bemærk, at der er dannet mindst en observation for hver måned i 2023, også selvom personen har været i uafbrudt beskæftigelse i samme job fra 1/6 til 31/12. Ved skift i tilstanden er der dannet en ny observation, fx er personen arbejdsløs fra 7/2 til 26/2. Det gælder, at der ikke er overlap mellem perioderne.

Hvis tilstanden ultimo i en måned ønskes, skal datoen ultimo måneden fx 31. oktober 2023 angives:

TIL_DATO="31oct2023"d.

Og så fås dette resultat:

Obs	PNR	I_BEFOLKNINGEN_KODE	PRIMAER_STATUS_KODE	SOC_STATUS_KODE	FRA_DATO	TIL_DATO
13	7777777777	1	1	132 Lønmodtagere på højeste niveau	01/10/2023	31/10/2023

Ved en vilkårlig dato i måneden fx 15. februar 2023 kan denne angives:

FRA_DATO<="15feb2023"d<=TIL_DATO

Derved fås dette resultat:

Obs	PNR	I_BEFOLKNINGEN_KODE	PRIMAER_STATUS_KODE	SOC_STATUS_KODE	FRA_DATO	TIL_DATO
3	7777777777	1	1	200 Arbejdsløse	07/02/2023	26/02/2023

6. Referencetidspunkter, datasætnavne og statustidspunkter

Når man benytter statusregistrene i DDV, skal man være opmærksom på sammenhængene mellem **datasætnavne**, **referencetidspunkter** og **statustidspunkterne i Statistikbanken.** Når kilden til registret er den Personstatistiske Database, så vil referencetidspunktet for registret som regel være ultimo året/kvartalet/måneden, hvor statustidspunktet i Statistikbanken er primo i efterfølgende år/kvartal/måned. I **BEF** finder man referencetidspunktet som en variabel i registret, mens man for de øvrige statusregistre skal udlede referencetidspunktet via datasætnavnet.

Tabel 6.1. Oversigt over datasætnavne, referencetidspunkter og statustidspunkter i Statistikbanken.

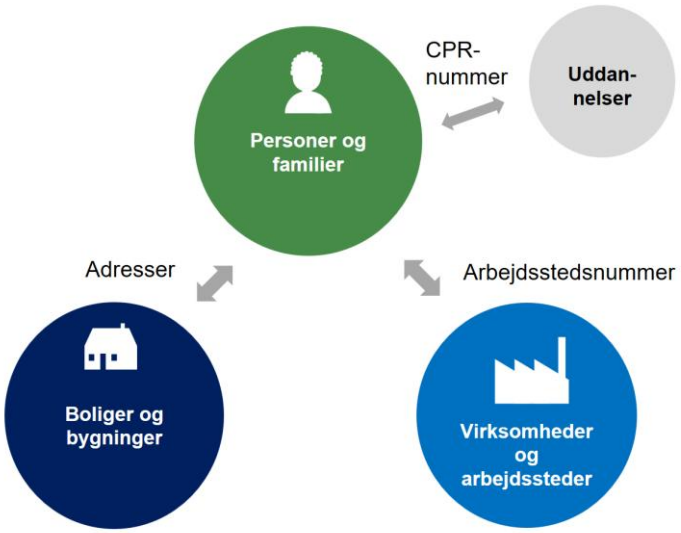
		Udvalgte eksempler		
Register	**Bemærkning**	**Datasætnavn**	**Referencetidspunkt i registret**	**Statustidspunkt i Statistikbanken**
BEF	Referencetidspunktet i registret er dagen inden statustidspunktet i Statistikbanken. I Statistikbanken er statustidspunktet angivet som den 1. i kvartalet.	BEF201803 BEF201806 BEF201809 BEF201812	31/03/2018 30/06/2018 30/09/2018 31/12/2018	2018K2 2018K3 2018K4 2019K1
FAIN	Statustidspunktet er d. 1. januar i året. Registret indeholder ikke et referencetidspunkt, så det er årstallet i datasætnavnet. der angiver statustidspunktet.	FAIN1980 FAIN1981 FAIN1982	- - -	1. januar 1980 1. januar 1981 1. januar 1982
BOL **BOLIG** **PERBOLIG** **BBRENHED**	Datasætnavnet er december året inden statustidspunktet i Statistikbanken som er d. 1. januar i året. Registret indeholder ikke et referencetidspunkt, så det er årstallet i datasætnavnet. der angiver statustidspunktet.	BOL198012 BOL198112 BOLIG202312 BOLIG202412 PERBOLIG202312 PERBOLIG202412 BBRENHED202312 BBRENHED202412	- - - - - - - -	1. januar 1981 1. januar 1982 1. januar 2024 1. januar 2025 1. januar 2024 1. januar 2025 1. januar 2024 1. januar 2025
UDDA 198012 til 200712	Datasætnavnet angiver december året inden statustidspunktet i Statistikbanken som er d. 1. januar i året. Registret indeholder ikke et referencetidspunkt, så det er år/måned i datasætnavnet. der angiver statustidspunktet.	UDDA199712 UDDA199812 UDDA199912	- - -	1. januar 1998 1. januar 1999 1. januar 2000
UDDA 200809 -	Datasætnavnet angiver september. hvor statustidspunktet i Statistikbanken er d. 30. september. Registret indeholder ikke et referencetidspunkt, så det er år/måned i datasætnavnet. der angiver statustidspunktet.	UDDA202309 UDDA202409 UDDA202509	- - -	30. september 2023 30. september 2024 30. september 2025
RAS 1980 til 2007	Datasætnavnet angiver året inden statustidspunktet i Statistikbanken som er d. 1. januar i året. Registret indeholder ikke et referencetidspunkt, så det er året i datasætnavnet, der angiver statustidspunktet.	RAS1994 RAS1995 RAS1996	- - -	1. januar 1995 1. januar 1996 1. januar 1997
RAS 2008 -	Datasætnavnet angiver året, hvor statustidspunktet i Statistikbanken er ultimo november i året. Registret indeholder ikke et referencetidspunkt, så det er året i datasætnavnet, der angiver statustidspunktet.	RAS2015 RAS2016 RAS2017	- - -	Ultimo november 2015 Ultimo november 2016 Ultimo november 2017

7. Statistiksystemets enheder og nøgler

Et af formålene med statistik er at beskrive virkeligheden og udviklingen i denne bedst muligt gennem generelle observationer. I Danmarks Statistik skelnes mellem *administrative registre* og *statistiske registre*, idet formålet med de statistiske registre ikke er at beskrive den administrative virkelighed, og derfor kan det være nødvendigt at afgrænse oplysningerne eller kombinere oplysninger fra flere administrative registre. Danmarks Statistik var først i verden til at opbygge et komplet statistisk registersystem baseret på administrative registre.[8]

Populationen kan variere alt efter, hvad der søges belyst. Derfor kan registre også mere generelt beskrives som indeholdende *enheder*. Enhederne kan være personer, virksomheder, biler, bygninger mv. alt efter registerets formål. Statistikregistre kan kobles med hinanden gennem *entydige identifikationsnøgler,* som deles mellem registrene. I indeværende vejledning indgår registre og deres relationer som skitseret i figur 7.1.

Figur 7.1. Enheder og nøgler i vejledningen



Når mikrodata stilles til rådighed for forskerne i DDV, er nøglerne krypteret i relation til det enkelte forskningsprojekt. Ved opslag i data ser du derfor en række løbenumre, som kan benyttes til at koble enhederne på tværs af registrene, men de kan ikke bruges til at identificere hverken personer eller virksomheder.

[8] For en beskrivelse af historikken omkring det danske personstatistiske registersystem, se *Personstatistik i Danmark, et registerbaseret statistiksystem*, Danmarks Statistik (1994) jf.:
https://www.dst.dk/da/Statistik/nyheder-analyser-publ/Publikationer/VisPub?cid=54567

I tabel 7.1 har vi opgjort registrene ud fra deres enhed og deres nøgler.

Tabel 7.1. Registrenes enheder og nøgler.

Enhed	Register	Nøgle for personer	Nøgler for boliger og bygninger	Nøgler for virksomheder og arbejdssteder	Andre nøgler
Personer	**BEF**	PNR			
	FAIN	PNR	KOM BOPIKOM		
	LIFELINES	PNR			
Boliger og bygninger	**BOL**		KOM BOPIKOM		
	BOLIG		DST_BOLIGERHVERVSENHED_ID		
	PERBOLIG	PNR	DST_BOLIGERHVERVSENHED_ID		
	BBRENHED		DST_BOLIGERHVERVSENHED_ID		
Uddannelser	**UDDA**	PNR			INSTNR
	UDDF	PNR			INSTNR
	KOTRE	PNR			INSTNR
Virksomheder og arbejdssteder	**RAS**	PNR		ARBGNR ARBNR CVRNR ARBNR SENR OK_NR	
	AMRUN	PNR		CVRNR ARBNR SENR OK_NR	

Den centrale personnøgle i DDV er CPR-nummeret, som optræder med variabelnavnet **PNR**.

FAIN indeholder populationer for perioden 1980-2007, og her er **PNR** nøglen baseret på de oprindelige CPR-numre. Jf. også SAS-koden i afsnit 9.1.

BEF indeholder befolkningen fra den Personstatistisk Database (PSD), og **PNR-**nøglen er CPR-numre, som er konverteret fra PSD's personløbenummer **PERSON_ID**. Her benyttes den aktuelle konverteringsnøgle den dag, at data konverteres. Konverteringsnøglen ændrer sig fra dag til dag, og dermed kan personer med et skift i CPR-nummer optræde med forskellige **PNR** registrene imellem.

Det betyder at:

- Det er ikke muligt at opnå fuldstændig match registrene imellem, også selv om der er tale om samme population og samme referencetidspunkt.
- Der optræder blanke **PNR**, og det skyldes, at nogle relativt få **PERSON_ID** er udgået af den aktuelle konverteringsnøgle.
- Det ikke er muligt at genskabe de publicerede tal i Statistikbanken i de tilfælde, hvor det kræver en kobling af flere registre.
- Hvis et projekt får en genleverance af historiske data, kan der også være foretaget en konvertering, som påvirker mulighederne for at genskabe den oprindelige match registrene imellem.
- Man kan ikke altid følge de personer, som skifter **PNR**. For forløbsregistrene **Lifelines**, **UDDF** og **KOTRE** konverteres hele registret med samme nøgle, hver gang der kommer en ny udgave, og man kan derfor godt følge de personer, der har haft skift i **PNR**. For de øvrige

registre i denne vejledning vil man ikke kunne følge de personer, der har haft et skift i CPR-nummer. I perioden 1980-2025 drejer det sig om ca. 15.000 personer, som har haft et skift i PNR i Befolkningsstatistikken (FAIN og BEF).

For at hente oplysninger om boliger og bygninger skal der matches på adressekoden. Boligopgørelsen blev omlagt i 2010, og der er derfor tale om to forskellige sæt nøgler:

- 1981-2009: Her skal man matche på kommune (**KOM**) og bopæl i kommunen (**BOPIKOM**). Jf. også SAS-programmet i afsnit 9.1.
- 2010 og frem: Her skal man matche på variablen **DST_BOLIGERHVERVSENHED_ID**

Se yderligere beskrivelse af dette i afsnit 4.3.

Oplysninger om befolkningens tilknytning til arbejdsmarkedet hentes i **RAS/AMRUN**. Begge registre indeholder oplysninger på personniveau, og nøglerne mellem personer og deres arbejdssteder/virksomheder er indeholdt i selve registret, så selve koblingen mod **RAS/AMRUN** sker derfor på **PNR**. Det er kun, hvis du skal matche videre mod andre erhvervsstatistiske registre, at nøglerne for virksomheder og arbejdssteder kommer i spil. Mulighederne for at matche videre mod andre erhvervsstatistiske registre indgår ikke i vejledningen.

Endeligt skal det nævnes, at registrene med uddannelsesoplysninger indeholder nøglen institutionsnummer (**INSTNR**), som gør det muligt at matche videre mod Institutionsregistret. Mulighederne for at matche videre mod Institutionsregistret indgår ikke i vejledningen.

8. R kode

8.1 Årlige populationer 1981-2007

```
#-------------------------------------------------------------------------------
# Programmet danner årlige sammenhængende populationer med FAIN, BOL, UDDA og RAS
#-------------------------------------------------------------------------------

#-------------------------------------------------------------------------------
#Biblioteker
#-------------------------------------------------------------------------------

#Datamanipulation
library(tidyverse)
library(haven)

#-------------------------------------------------------------------------------
#Parametre
#-------------------------------------------------------------------------------
E
# Sti til grunddata - skal opdateres
grunddata_sti <- "//nas4/FseData/Grunddata/"

#-------------------------------------------------------------------------------
# Danner funktion
#-------------------------------------------------------------------------------

# Danner en funktion til brug for dannelse af årlige funktioner med inputtet y svarende til år
pop_funktion <- function(y)

{
    ##############################
    # FAIN, RAS og UDDA
    ##############################

    fain <- read_sas(paste0(grunddata_sti, "FAIN/fain", y ,".sas7bdat" )) |>
            rename_all(tolower) # Variablene omdøbes til lowercase, da R er case sensitive

    if (y <= 1990) {
    ras <- read_sas(paste0(grunddata_sti, "RAS/ras", y-1 ,".sas7bdat" )) |> # y-1 da RAS og
FAIN ligger forskudt
            rename_all(tolower)
    }

    # Fra 1991 og frem til 2007 skal RAS afgrænses til personens primære
beskæftigelse  (novemberprioritering)
```

```
    if (y >= 1991) {
    ras <- read_sas(paste0(grunddata_sti, "RAS/ras", y-1 ,".sas7bdat" )) |>
            rename_all(tolower) |>
            filter(novprio == 1)
    }

    udda <- read_sas(paste0(grunddata_sti, "UDDA/udda", y-1 ,"12.sas7bdat" )) |> # y-1 da UDDA
og FAIN ligger forskudt
            rename_all(tolower)

    # Merger data på cpr-nummer-niveau
    pop1 <- fain |>
      left_join(ras, by = "pnr") |>
      left_join(udda, by = "pnr")

    # Sletter dubletter og sorterer
    pop1 <- pop1 |>
      arrange(pnr) |>
      distinct(pnr, .keep_all = TRUE) |>
      arrange(kom, bopikom) |>
      mutate(aar =  y)

    ##############################
    # BOL
    ##############################

    # Før 2005 skal der afgrænses til at enheden findes både i BBR og CPR, og boliger uden
enhedsopdeling frasorteres
    if (y <= 2004) {
    bol <- read_sas(paste0(grunddata_sti, "BOL/bol", y-1 ,"12.sas7bdat" )) |>
            rename_all(tolower) |>
            filter(cpr_bbr_matchkode_gl == "1", boligart_gl != "88")
    }

    # Fra og med 2005 skal der afgrænses til boligforhold oplyst
    if  (y >= 2005) {
    bol <- read_sas(paste0(grunddata_sti, "BOL/bol", y-1 ,"12.sas7bdat" )) |>
            rename_all(tolower) |>
            filter(husstand == "1")
    }

    # Fjerner dubletter på KOM + BOPIKOM
    bol <- bol |>
      arrange(kom, bopikom) |>
      distinct(kom, bopikom, .keep_all = TRUE) |>
      mutate(match_bol =1 )
```

```
    ##############################
    # Samler data
    ##############################

    # Joiner populationen fra udda, fain og bef med bol og justerer cpr_bbr_matchkode_gl
    # cpr_bbr_matchkode_gl sættes til 3, hvis match_bol=1 (ellers beholder de gamle værdier)
    pop2 <- pop1 |> left_join(bol, by = c("kom", "bopikom"))|>
            mutate(cpr_bbr_matchkode_gl = if_else(match_bol==1 , "3", cpr_bbr_matchkode_gl)) |>
            select(-match_bol)

    # Sorter og fjern dubletter på PNR
    pop2 <- pop2 |>
    arrange(pnr) |>
    distinct(pnr, .keep_all = TRUE)

   return(pop2)
}

#-------------------------------------------------------------------------------
#  Kalder pop funktionen for årene
#-------------------------------------------------------------------------------

# Definer årstal der skal køres over (her er det eksempelvis 1981-1985)
aar <- 1981:1985

# Kør funktionen for hvert år og gem i en liste
resultater_liste <- set_names(aar, aar) %>%
  map(~ pop_funktion(.x))

# Gem f.eks. 1985 som sin egen data frame
df_1985 <- resultater_liste[["1985"]]
```

8.2 Befolkningsstatistikkens kvartalsvise populationer for 2023

```
#------------------------------------------------------------------------------
# Programmet danner sammenhængende kvartalsvise populationer med udgangspunkt
# i BEF-data
#------------------------------------------------------------------------------

#------------------------------------------------------------------------------
#Biblioteker
#------------------------------------------------------------------------------

#Datamanipulation
library(tidyverse)
library(haven)

#------------------------------------------------------------------------------
#Parametre
#------------------------------------------------------------------------------

# Sti til grunddata - skal opdateres
grunddata_sti <- "//nas4/FseData/Grunddata/"

# Sti til textfiler med formater - skal opdateres
formater_sti <- "G:/K15/Prodgruppe/Projekter og arbejdsgrupper/Formidling af mikrodata/Danmarks
Datavindue - Data indsigt/Befolkningens populationer/R-programmer/Tekstfiler/"


#------------------------------------------------------------------------------
# Funktion
#------------------------------------------------------------------------------

udtraek <- function(status_dato, bef_ref, amrun_ref, uddf_ref, kotre_ref)

{

#------------------------------------------------------------------------------
# Grunddataudtræk
#------------------------------------------------------------------------------

##############################
# Befolkningen
##############################

# Henter data
bef <- read_sas(paste0(grunddata_sti, "BEF/bef", bef_ref ,".sas7bdat" ),
                col_select=c
                          (PNR
```

```
                          ,KOEN
                          ,ALDER
                          ,FAMILIE_TYPE
                          ,IE_TYPE
                          ,KOM
                          ,REFERENCETID)) |>
        rename_all(tolower) |>
        mutate(status_dto = status_dato)

#############################
# Højeste fuldførte uddannelse
#############################

# Henter data
# Her er det altid seneste år, der skal indlæses, der filtreres i stedet på statusdatoen
uddf <- read_sas(paste0(grunddata_sti, "UDDF/uddf", uddf_ref,".sas7bdat" ),
                 col_select=c
                           (PNR
                           ,HFAUDD
                           ,HF_KILDE
                           ,INSTNR
                           ,HF_VFRA
                           ,HF_VTIL)) |>
        rename_all(tolower) |>
        filter(hf_vfra <= as.Date(status_dato)
              ,hf_vtil >= as.Date(status_dato)
              ,pnr != '') |>  # Sletter rækker med blanke pnr
        distinct(pnr, .keep_all = TRUE) # Beholder kun første række pr. pnr og sletter dermed
evt. dubletter


#############################
# Igangværende uddannelser fra elevregisteret
#############################

# Henter data
# Her er det altid seneste år, der skal indlæses, der filtreres i stedet på statusdatoen
kotre <- read_sas(paste0(grunddata_sti, "KOTRE/kotre", kotre_ref,".sas7bdat" ),
                 col_select=c
                           (PNR
                           ,UDD
                           ,ELEV3_VFRA
                           ,ELEV3_VTIL)) |>
         rename_all(tolower) |>
         filter(elev3_vfra <= as.Date(status_dato)
               ,elev3_vtil >= as.Date(status_dato)
               ,pnr != '') |>  # Sletter rækker med blanke pnr
```

```
         distinct(pnr, .keep_all = TRUE)  # Beholder kun første række pr. pnr og sletter dermed
evt. dubletter

#############################
# Arbejdsmarkedsregnskabet
#############################

amrun <- read_sas(paste0(grunddata_sti, "AMRUN/amrun", amrun_ref,".sas7bdat"),
                  col_select=c
                            (PNR
                            ,I_BEFOLKNINGEN_KODE
                            ,PRIMAER_STATUS_KODE
                            ,SOC_STATUS_KODE
                            ,FRA_DATO
                            ,TIL_DATO
                            ,CPRTJEK)) |>
          rename_all(tolower) |>
          filter(fra_dato == as.Date(status_dato)
                 ,primaer_status_kode == 1 # Udvælger personers primære status
                 ,cprtjek == "0"
                 ,pnr != '') |>  # Sletter rækker med blanke pnr
          distinct(pnr, .keep_all = TRUE) |> # Beholder kun første række pr. pnr og sletter
dermed evt. dubletter
          select(-cprtjek)


#############################
# Boligoplysninger
#############################

# Indlæser data om personers bolig pr. ultimo året (her benyttes bef_ref, da denne er ultimo
året)
perbolig <- read_sas(paste0(grunddata_sti, "PERBOLIG/perbolig", bef_ref,".sas7bdat" ),
                     col_select = c(PNR
                                   ,DST_BOLIGERHVERVSENHED_ID
                                   ,CPRTJEK)) |>
            rename_all(tolower) |>
            filter(,cprtjek == "0"
                   ,pnr != '') |>
            select(-cprtjek)

# Indlæser data om BBR-enhederne
bbrenhed_1 <- read_sas(paste0(grunddata_sti, "BBRENHED/2 temp_behandlet/bbrenhed",
bef_ref,".sas7bdat" ),
                       col_select = c(DST_BOLIGERHVERVSENHED_ID
                                     ,DST_BOLIGMATCH
                                     ,ENH_ANVENDELSEKODE
```

```
                                    ,DST_ENH_EJERFORHOLDSKODE)) |>
              rename_all(tolower)

# Sammenkobler om boliger og bbr-enheder
bbrenhed <- perbolig |>
    inner_join(bbrenhed_1, by = "dst_boligerhvervsenhed_id") |>
    distinct(pnr, .keep_all = TRUE) |> # Beholder kun første række pr. pnr og sletter dermed
evt. dubletter
    mutate(bbrenhed = 1)

#-------------------------------------------------------------------------------
# Data samles
#-------------------------------------------------------------------------------



samlet_bef <- bef |>
  # Left-joiner datasæt på befolkningsdatasættet
  left_join(uddf, by = "pnr") |>
  left_join(kotre, by = "pnr") |>
  left_join(amrun, by = "pnr") |>
  left_join(bbrenhed, by = "pnr")


#-------------------------------------------------------------------------------
# Funktion afsluttes og eksekveres
#-------------------------------------------------------------------------------

}

# Kører funktionen for de fire kvartaler i 2023 - for uddannelsesforløbsregistre benyttes altid
den seneste version af datasættene
samlet_k1_2023 <- udtraek(status_dato = "2023-01-01", bef_ref = "202212", amrun_ref = "2023",
uddf_ref = "202509", kotre_ref = "2025")
samlet_k2_2023 <- udtraek(status_dato = "2023-04-01", bef_ref = "202303", amrun_ref = "2023",
uddf_ref = "202509", kotre_ref = "2025")
samlet_k3_2023 <- udtraek(status_dato = "2023-07-01", bef_ref = "202306", amrun_ref = "2023",
uddf_ref = "202509", kotre_ref = "2025")
samlet_k4_2023 <- udtraek(status_dato = "2023-10-01", bef_ref = "202309", amrun_ref = "2023",
uddf_ref = "202509", kotre_ref = "2025")

#-------------------------------------------------------------------------------
# Samler og tilretter data
#-------------------------------------------------------------------------------

samlet <- rbind(samlet_k1_2023, samlet_k2_2023, samlet_k3_2023, samlet_k4_2023)

# Checker for dubletter
```

```
dubletter <- samlet |>
  add_count(pnr, status_dto) |>
  filter(n > 1, pnr != '')

# Indlæser tekstfiler til brug for formatering af variable
fmt_audd_hoved_l1l5_kt <- read.csv2(paste0(formater_sti, "N_AUDD_HOVED_L1L5_KT.txt"),
fileEncoding="Latin1")
fmt_udd_hoved_l1l5_kt <- read.csv2(paste0(formater_sti, "N_UDD_HOVED_L1L5_KT.txt"),
fileEncoding="Latin1")
fmt_socstatus_gr3_sb <- read.csv2(paste0(formater_sti, "N_SOCSTATUS_GR3_SB.txt"),
fileEncoding="Latin1")
fmt_enh_anvendelseskode_sb <- read.csv2(paste0(formater_sti, "N_ENH_ANVENDELSESKODE_SB.txt"),
fileEncoding="Latin1")
fmt_enh_anvendelsekode_ds <- read.csv2(paste0(formater_sti, "N_ENH_ANVENDELSEKODE_DS.txt"),
fileEncoding="Latin1")


# Tilkobler de formaterede variable og justerer i indholdet
samlet_ny <- samlet |>
  left_join(fmt_audd_hoved_l1l5_kt, by = c("hfaudd"="START")) |>
  left_join(fmt_udd_hoved_l1l5_kt, by = c("udd"="START"))  |>
  left_join(fmt_socstatus_gr3_sb, by = c("soc_status_kode"="START")) |>
  left_join(fmt_enh_anvendelsekode_ds, by = c("enh_anvendelsekode" = "START")) |>
  left_join(fmt_enh_anvendelseskode_sb, by = c("ENH_ANVENDELSEKODE_DS" = "START")) |>
  rename_all(tolower) |>
       mutate(socstatus_gr3_sb = replace_na(socstatus_gr3_sb, "Ikke i AMRUN")
              ,socstil_gr2_sb = case_when(is.na(soc_status_kode) ~ "Ikke i AMRUN"
                                          ,soc_status_kode >= 110 & soc_status_kode <= 200 ~ "I
arbejdsstyrken"
                                          ,soc_status_kode >= 300 & soc_status_kode <= 517 ~
"Uden for arbejdsstyrken"
                                          ,TRUE ~ "Andet" )
              ,socstatus_elev = case_when(soc_status_kode == 29 ~ "04 Personer under
uddannelse"
                                          ,is.na(soc_status_kode) ~ "99 Ikke i AMRUN"
                                          ,TRUE ~ as.character(socstatus_gr3_sb)
                                          )
              ,status = referencetid + 1 # statustidspunktet justeres med en dag, så det
stemmer overens med Statistikbanken
              )

#-------------------------------------------------------------------------
# Danner tabeller
#-------------------------------------------------------------------------

##############################
# Højest fuldførte uddannelse
```

```r
##############################

hfaudd_tabel <- samlet_ny |>
  filter(alder >= 15 & alder <= 69 ) |>
  group_by(status, audd_hoved_l1l5_kt) |>
  summarise(antal = n()) |>
  ungroup()

##############################
# Igangværende uddannelse
##############################

udd_tabel <- samlet_ny |>
  filter(!is.na(udd_hoved_l1l5_kt)) |>
  group_by(status, udd_hoved_l1l5_kt) |>
  summarise(antal = n()) |>
  ungroup()

##############################
# Socioøkonomisk status
##############################

socstatus_tabel <- samlet_ny |>
  group_by(status, socstatus_gr3_sb ) |>
  summarise(antal = n()) |>
  ungroup()

##############################
# Socioøkonomisk status og igangværende uddannelse
##############################

socstatus_udd_tabel <- samlet_ny |>
  filter(udd_hoved_l1l5_kt != "Ukendt"
       ,status == as.Date("2023-04-01")) |>
  group_by(status, udd_hoved_l1l5_kt, socstil_gr2_sb ) |>
  summarise(antal = n()) |>
  ungroup()

##############################
# Anvendelseskode for boliger
##############################

enh_anvendelse_tabel <- samlet_ny |>
  group_by(status, enh_anvendelseskode_sb) |>
  summarise(antal = n()) |>
  ungroup()
```

8.3 Lifelines månedsvise populationer for 2023

```
#------------------------------------------------------------------------------
# Beskrivelse: Programmet danner sammenhængende populationer med udgangspunkt i Lifelines.
#------------------------------------------------------------------------------

#------------------------------------------------------------------------------
# Biblioteker
#------------------------------------------------------------------------------

# Til datamanipulation
library(lubridate)
library(tidyverse)
library(purrr)

# Til at læse SAS filer
library(haven)


#------------------------------------------------------------------------------
# Parametre
#------------------------------------------------------------------------------

# Sti til grunddata - skal opdateres
grunddata_sti <- "//nas4/fsedata/Grunddata/"

# Sti til formater (kodelister) - skal opdateres
formater_sti <- "//nas1/data/K15/Prodgruppe/Projekter og arbejdsgrupper/Formidling af
mikrodata/Danmarks Datavindue - Data indsigt/Befolkningens populationer/R-
programmer/Tekstfiler/"


#------------------------------------------------------------------------------
# Reference
#------------------------------------------------------------------------------

# Referencetidspunkter for uddannelses forløbsregistre.
# Der findes kun én version og altid den seneste version.

# Lifelines -altid seneste
life_ref <- "2025"

# Højest fuldført uddannelse - altid seneste
uddf_ref <- "202509"

# Elevregistret - altid seneste
kotre_ref <- "2025"
```

```r
#------------------------------------------------------------------------------
# Funktion
#------------------------------------------------------------------------------

# Udvælg populationen i Lifelines. Personer der udgør af befolkningen på statusdagen indgår
ikke

# Danner life funktion.
# Life funktionen afvikles pr. måned. I funktionen angives to argumenter:
# 1. Statusdagen (status_dato)
# 2. Referencetid for SAS-datasæt med amr (amr_ref)

life <- function(status_dato, amr_ref) {

  # Henter lifelines, tager ca. 2 min at hente
  pop <- read_sas(paste0(grunddata_sti, "LIFELINES/lifelines", life_ref ,".sas7bdat")) |>
      rename_all(tolower) |>
      filter(
      event_start_date <= status_dato,
      status_dato < event_final_date,
      pnr != "") |>
    arrange(pnr) |>
    distinct(pnr, .keep_all = TRUE)

  # Henter højest fuldførte uddannelse, tager ca. 2 min at hente
  uddf <- read_sas(paste0(grunddata_sti, "UDDF/uddf", uddf_ref ,".sas7bdat"),
                   col_select=c(PNR, HFAUDD, HF_KILDE, INSTNR, HF_VFRA, HF_VTIL)) |>
      rename_all(tolower) |>
      filter(
      hf_vfra <= status_dato,
      status_dato <= hf_vtil,
      pnr != "") |>
    arrange(pnr) |>
    distinct(pnr, .keep_all = TRUE)

  # Henter igangværende uddannelser, tager ca. 3 min
  kotre <- read_sas(paste0(grunddata_sti, "KOTRE/kotre", kotre_ref ,".sas7bdat"),
                   col_select=c(PNR, UDD, ELEV3_VFRA, ELEV3_VTIL)) |>
    rename_all(tolower) |>
    filter(
      elev3_vfra <= status_dato,
      status_dato <= elev3_vtil,
      pnr != "") |>
    arrange(pnr) |>
    distinct(pnr, .keep_all = TRUE)
```

```r
  # Henter arbejdsmarkedsregnskabet, tager ca. 10 min
  amr <- read_sas(paste0(grunddata_sti, "AMRUN/amrun", amr_ref, ".sas7bdat"),
    col_select = c(
      PNR,
      I_BEFOLKNINGEN_KODE,
      PRIMAER_STATUS_KODE,
      SOC_STATUS_KODE,
      FRA_DATO,
      TIL_DATO,
      BOPAEL_KOM_KODE,
      ARB_BEL_KOM_KODE,
      ARB_HOVED_BRA_DB07,
      IE_TYPE,
      CPRTJEK)) |>
    rename_all(tolower) |>
    filter(
      fra_dato == status_dato,
      primaer_status_kode == 1,
      cprtjek == "0",
      pnr != ""
    ) |>
    select(-cprtjek) |>
    arrange(pnr) |>
    distinct(pnr, .keep_all = TRUE)

  # Data samles, der leftjoines på lifelines så alle fra lifelines medtages
  output <- pop |>
    left_join(uddf, by = "pnr") |>
    left_join(amr, by = "pnr") |>
    left_join(kotre, by = "pnr") |>
    mutate(
      # Status_dato
      status_dato = status_dato,
      # Alder på statustidspunktet
      alder = floor(
      time_length(
      interval(foed_dato, status_dato),
      "years"
      )
    )
  )

return(output)
}
```

```r
#------------------------------------------------------------------------------
# Kalder funktion
#------------------------------------------------------------------------------

# Genererer den 1. i hver måned for 2023
datoer <- seq(as.Date("2023-01-01"), as.Date("2023-12-01"), by = "month")

# Danner navne til listen, f.eks. "20230101", "20230201" osv.
dato_navne <- format(datoer, "%Y%m%d")

# Kører funktionen for hver dato og gem i en navngivet liste
life_liste <- set_names(datoer, dato_navne) %>%
  map(~ life(as.string <- as.character(.x), "2023"))

# Saml listen og behold statustidspunkt som første kolonne
life_df <- list_rbind(life_liste, names_to = "statustidspunkt")

#------------------------------------------------------------------------------
# Tilretter data
#------------------------------------------------------------------------------

# Indlæser kodelister
socstatus_gr3_sb <- read.csv2(paste0(formater_sti,
"N_SOCSTATUS_GR3_SB.txt"),  fileEncoding="Latin1")
audd_hoved_l1l5_kt <- read.csv2(paste0(formater_sti,
"N_AUDD_HOVED_L1L5_KT.txt"),  fileEncoding="Latin1")
udd_hoved_l1l5_kt <- read.csv2(paste0(formater_sti,
"N_UDD_HOVED_L1L5_KT.txt"),  fileEncoding="Latin1")


# Tilkobler kodelister på data
samlet <- life_df |>
  left_join(socstatus_gr3_sb, by = c("soc_status_kode"="START")) |>
  left_join(audd_hoved_l1l5_kt, by = c("hfaudd"="START")) |>
  left_join(udd_hoved_l1l5_kt, by = c("udd"="START")) |>
  rename_all(tolower) |>
  mutate(socstatus_gr3_sb = replace_na(socstatus_gr3_sb, "Ikke i AMRUN"))

# Checker for dubletter
dubletter <- samlet |>
  add_count(pnr, status_dato) |>
  filter(n > 1, pnr != '')

#------------------------------------------------------------------------------
# Tabeller
#------------------------------------------------------------------------------
```

```r
##############################
# Højest fuldførte uddannelse
##############################

hfaudd_tabel <- samlet |>
  filter(!is.na(audd_hoved_l1l5_kt) & alder >= 15 & alder <= 69 ) |>
  group_by(status_dato, audd_hoved_l1l5_kt) |>
  summarise(antal = n()) |>
  ungroup()

##############################
# Igangværende uddannelse
##############################

udd_tabel <- samlet |>
  filter(!is.na(udd_hoved_l1l5_kt)) |>
  group_by(status_dato, udd_hoved_l1l5_kt) |>
  summarise(antal = n()) |>
  ungroup()

##############################
# Socioøkonomisk status
##############################

socstatus_tabel <- samlet |>
  group_by(status_dato, socstatus_gr3_sb ) |>
  summarise(antal = n()) |>
  ungroup()
```

9. SAS kode

9.1 Årlige populationer 1981-2007

```
*************************************************************************************************
  Program: Danner årlige sammenhængende data 1981-2007 med disse registre:
           - FAIN: Husstande og familier
           - BOL:  Boligtællingen
           - UDDA: Højest fuldførte uddannelse
           - RAS:  Registerbaseret arbejdsstyrke
*************************************************************************************************;
libname FAIN 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname RAS  'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname UDDA 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname BOL  'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx';

libname temp 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx';

options DKROCOND=WARN;

*-----------------------------------------------------*
* 1981-1990                                           *
*-----------------------------------------------------;
%macro pop1(aar);
*-----------------------*
* FAIN, RAS og UDDA.    *
*-----------------------;
data pop1;
retain aar;
merge fain.fain&aar(in=ok1)
      ras.ras%eval(&aar-1) (drop=koen)
      udda.udda%eval(&aar-1)12;
by pnr;
if ok1;
aar=&aar;
run;

* Tjek for dupletter;
proc sort data=pop1 nodupkey;
by pnr;
run;

* Sorter på adressekoden;
proc sort data=pop1;
by kom bopikom;
run;

*-----------------------*
* Hent BOL.             *
* Ikke alle matcher.    *
*-----------------------;
data bol;
set bol.bol%eval(&aar-1)12;
where CPR_BBR_MATCHKODE_GL='1' and BOLIGART_GL ^='88';
run;

proc sort data=bol nodupkey;
by kom bopikom;
run;

data pop2;
merge pop1(in=ok1) bol(in=ok2);
by kom bopikom;
if ok1;
if ok1 and ^ok2 then CPR_BBR_MATCHKODE_GL='3';
run;

* Permanent datasæt;
proc sort data=pop2 out=temp.fain_udda_ras_bol_&aar nodupkey;
by pnr;
run;
```

```
%mend;

%pop1(1981);
%pop1(1982);
%pop1(1983);
%pop1(1984);
%pop1(1985);
%pop1(1986);
%pop1(1987);
%pop1(1988);
%pop1(1989);
%pop1(1990);

*-----------------------------------------------------*
* 1991-2004                                           *
*-----------------------------------------------------;
%macro pop2(aar);
*------------------------*
* FAIN, RAS og UDDA.     *
*------------------------;
data pop1;
retain aar;
merge fain.fain&aar(in=ok1)
      ras.ras%eval(&aar-1) (drop=koen where=(novprio=1))
      udda.udda%eval(&aar-1)12;
by pnr;
if ok1;
aar=&aar;
run;

* Tjek for dupletter;
proc sort data=pop1 nodupkey;
by pnr;
run;

proc sort data=pop1;
by kom bopikom;
run;

*------------------------*
* Hent BOL.              *
* Ikke alle matcher.     *
*------------------------;
data bol;
set bol.bol%eval(&aar-1)12;
where CPR_BBR_MATCHKODE_GL='1' and BOLIGART_GL ^='88';
run;

proc sort data=bol nodupkey;
by kom bopikom;
run;

data pop2;
merge pop1(in=ok1) bol(in=ok2);
by kom bopikom;
if ok1;
if ok1 and ^ok2 then CPR_BBR_MATCHKODE_GL='3';
run;

* Permanent datasæt;
proc sort data=pop2 out=temp.fain_udda_ras_bol_&aar nodupkey;
by pnr;
run;

%mend;

%pop2(1991);
%pop2(1992);
%pop2(1993);
%pop2(1994);
%pop2(1995);
%pop2(1996);
%pop2(1997);
```

```
%pop2(1998);
%pop2(1999);
%pop2(2000);
%pop2(2001);
%pop2(2002);
%pop2(2003);
%pop2(2004);

*-----------------------------------------------------*
* 2005-2007.                                          *
*-----------------------------------------------------;
%macro pop3(aar);
*------------------------*
* FAIN, RAS og UDDA.     *
*------------------------;
data pop1;
retain aar;
merge fain.fain&aar(in=ok1)
      ras.ras%eval(&aar-1) (drop=koen where=(novprio=1))
      udda.udda%eval(&aar-1)12;
by pnr;
if ok1;
aar=&aar;
drop cprtjek cprtype version;
run;

* Tjek for dupletter;
proc sort data=pop1 nodupkey;
by pnr;
run;

proc sort data=pop1;
by kom bopikom;
run;

*------------------------*
* Hent BOL.              *
* Ikke alle matcher.     *
*------------------------;
data bol;
set bol.bol%eval(&aar-1)12;
where Husstand=1;
run;

proc sort data=bol nodupkey;
by kom bopikom;
run;

data pop2;
merge pop1(in=ok1) bol(in=ok2);
by kom bopikom;
if ok1;
if ok1 and ^ok2 then CPR_BBR_MATCHKODE_GL='3';
run;

* Permanent datasæt;
proc sort data=pop2 out=temp.fain_udda_ras_bol_&aar nodupkey;
by pnr;
run;
%mend;

%pop3(2005);
%pop3(2006);
%pop3(2007);
```

9.2 Befolkningsstatistikkens kvartalsvise populationer for 2023

```
*************************************************************************************************
   Program: Danner sammenhængende kvartalsvise
            populationer med udgangspunkt i BEF-data.
*************************************************************************************************;
*-----------------------------------*
    Biblioteker.
*-----------------------------------;
libname BEF      'xxxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname UDDF     'xxxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname AMRUN    'xxxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname KOTRE    'xxxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname PERBOLIG 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname BBRENHED 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxx';

libname temp     'xxxxxxxxxxxxxxxxxxxxxxxxxxxxx';

*-----------------------------------------*
  Referencetidspunkter for forløbsregistre.
  Der findes kun een version.
*-----------------------------------------;
%let kotre_ref=2025;        * Elevregistret - altid seneste;
%let uddf_ref=202509;       * Højest fuldført uddannelse - altid seneste;

*-------------------------------------------------------------*
  Makro.
*-------------------------------------------------------------;
%macro bef_data(status_dato,bef_ref,amr_ref);

*-----------------------------------*
  Befolkningen.
*-----------------------------------;
proc sql;
create table bef as
select referencetid,pnr,koen,alder,familie_type,ie_type,kom
from bef.bef&bef_ref.;
quit;

*-----------------------------------*
  Højest fuldførte uddannelser.
*-----------------------------------;
proc sql;
create table uddf as
select pnr,hfaudd,hf_kilde,instnr,HF_VFRA,HF_VTIL
from uddf.uddf&uddf_ref.(where=(HF_VFRA<="&status_dato."d<=HF_VTIL and PNR ne ''));
quit;

* Fjern nogle evt. dubletter;
proc sort data=uddf nodupkey;
by pnr;
run;

*-----------------------------------*
  Igangværende uddannelser.
*-----------------------------------;
proc sql;
create table kotre as
select  pnr,udd,ELEV3_VFRA,ELEV3_VTIL
from    kotre.kotre&kotre_ref.
         (where=(ELEV3_VFRA<="&status_dato."d<=ELEV3_VTIL and PNR ne ''));
quit;

* Fjern nogle evt. dubletter;
proc sort data=kotre nodupkey;
by pnr;
run;

*-----------------------------------*
  Arbejdsmarkedsregnskabet.
*-----------------------------------;
proc sql;
create table amr as
```

```
select pnr,I_BEFOLKNINGEN_KODE,PRIMAER_STATUS_KODE,SOC_STATUS_KODE,FRA_DATO,TIL_DATO
from amrun.amrun&amr_ref.(where=(FRA_DATO="&status_dato."d
                                    and PRIMAER_STATUS_KODE=1
                                    and cprtjek='0'
                                    and PNR ne ''));
quit;

* Fjern nogle evt. dubletter;
proc sort data=amr nodupkey;
by pnr;
run;

*----------------------------------*
  Boligoplysninger.
*----------------------------------;
proc sql;
create table perbolig as
select pnr,DST_BOLIGERHVERVSENHED_ID
from perbolig.Perbolig&bef_ref.(where=(cprtjek='0' and PNR ne ''));
quit;

proc sql;
create table bbrenhed as
select  a.*,b.dst_boligmatch,b.ENH_ANVENDELSEKODE,b.DST_ENH_EJERFORHOLDSKODE
from    perbolig a
        inner join
        bbrenhed.Bbrenhed&bef_ref. b
        on a.DST_BOLIGERHVERVSENHED_ID=b.DST_BOLIGERHVERVSENHED_ID
order by a.pnr
;
quit;

* Fjern nogle evt. dubletter;
proc sort data=bbrenhed nodupkey;
by pnr;
run;

*----------------------------------*
  Data samles.
  Markering af match.
*----------------------------------;
data temp.bef_&bef_ref.;
merge bef(in=ok1) uddf amr kotre bbrenhed;
by pnr;
if ok1;        * Alle i befolkningen medtages;
run;
%mend;

*------------------------------------------------------------------*
 Makroen afvikles pr. kvartal.
 Der skal angives 3 parametre:
   1. Statusdagen iflg. Statistikbanken
   2. Referencetid for SAS-datasæt med BEF, PERBOLIG og BBRENHED
   3. Referencetid for SAS-datasæt med AMR.
*------------------------------------------------------------------;
%bef_data(01JAN2023,202212,2023);
%bef_data(01APR2023,202303,2023);
%bef_data(01JUL2023,202306,2023);
%bef_data(01OCT2023,202309,2023);

*----------------------------------------------------------*
  Tabeller.
*----------------------------------------------------------;
*----------------------------------*
  Formater.
*----------------------------------;
proc format;
VALUE SOC_STATUS
other=[SOCSTATUS_GR3_SB.]
.='99 Ikke i AMR'
;
VALUE SOC_STATUS_ELEV
other=[SOCSTATUS_GR3_SB.]
29='04 Personer under uddannelse'
```

```
.='99 Ikke i AMR'
;
Proc Format;
value SOCSTIL_GR2_SB
. ='Ikke i AMRUN'
110-200=' I arbejdsstyrken'
300-517=' Uden for arbejdsstyrken';
run;

*-----------------------------------*
   Alle data samles.
*-----------------------------------;
data tab1;
set temp.Bef_202212 temp.Bef_202303 temp.Bef_202306 temp.Bef_202309;
status=referencetid+1;    * Referencetiden i Statistikbanken;
enh_anvendelse=put(ENH_ANVENDELSEKODE,ENH_ANVENDELSEKODE_DS.);
keep status hfaudd alder udd  SOC_STATUS_KODE mk_bbr enh_anvendelse;
run;

*-----------------------------------------*
   Tabel med højest fuldførte uddannelse.
*-----------------------------------------;
proc tabulate data=tab1 missing order=formatted;
where 15 le alder le 69;
class status hfaudd;
format hfaudd AUDD_HOVED_L1L5_KT. status ddmmyy10.;
table all hfaudd, status='Statustidspunkt'*(n*f=commax12. colpctn*f=5.1) ;
keylabel n='Antal' colpctn='Pct.' all='I alt';
run;

*-----------------------------------------*
   Tabel med igangværende uddannelse.
*-----------------------------------------;
proc tabulate data=tab1 missing order=formatted;
where udd ne .;
class status udd;
format udd UDD_HOVED_L1L5_KT. status ddmmyy10.;
table all udd, status='Statustidspunkt'*(n*f=commax12. colpctn*f=5.1) ;
keylabel n='Antal' colpctn='Pct.' all='I alt';
run;

*-----------------------------------------*
   Tabel med soc_status_kode.
*-----------------------------------------;
proc tabulate data=tab1 missing order=formatted;
format SOC_STATUS_KODE SOC_STATUS. status ddmmyy10.;
class status soc_status_kode;
table all soc_status_kode, status='Statustidspunkt'*(n*f=commax12. colpctn*f=5.1) ;
keylabel n='Antal' colpctn='Pct.' all='I alt';
run;

*-----------------------------------------*
   Tabel med soc_status_kode og dem
   med en igangværende uddannelse.
*-----------------------------------------;
proc tabulate data=tab1 missing order=formatted;
where udd ne . and status='01apr2023'd;
format SOC_STATUS_KODE SOCSTIL_GR2_SB. status ddmmyy10. UDD UDD_HOVED_L1L5_KT.;
class status soc_status_kode udd  ;
table status,all udd,((soc_status_kode all)*n*f=commax12.)
                        ((soc_status_kode all)*rowpctn*f=5.1) / misstext='0';
keylabel n='Antal' rowpctn='Pct.' all='I alt';
run;

*-----------------------------------------*
   Tabel med bolig anvendelseskode.
*-----------------------------------------;
proc tabulate data=tab1 missing order=formatted;
format enh_anvendelse $ENH_ANVENDELSESKODE_SB. status ddmmyy10.;
class status enh_anvendelse;
table all enh_anvendelse='Enhedsanvendelse', status='Statustidspunkt'*(n*f=commax12.
colpctn*f=5.1) ;
keylabel n='Antal' colpctn='Pct.' all='I alt';
run;
```

9.3 Lifelines månedsvise populationer for 2023

```
*************************************************************************************************
   Program: Danner månedsvise sammenhængende
            populationer med udgangspunkt i Lifelines.
*************************************************************************************************;
*-----------------------------------*
   Biblioteker.
*-----------------------------------;
libname LIFE  'xxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname UDDF  'xxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname AMRUN 'xxxxxxxxxxxxxxxxxxxxxxxxxxxx';
libname KOTRE 'xxxxxxxxxxxxxxxxxxxxxxxxxxxx';

libname temp  'xxxxxxxxxxxxxxxxxxxxxxxxxxxx';

*---------------------------------------*
  Referencetidspunkter for uddannelses
  forløbsregistre. Der findes kun een
  version og altid den seneste version.
*---------------------------------------;
%let life_ref=2025;           * Lifelines - altid seneste;
%let kotre_ref=2025;          * Elevregistret - altid seneste;
%let uddf_ref=202509;         * Højest fuldførte uddannelse - altid seneste;

*-------------------------------------------------------------*
  Makro.
*-------------------------------------------------------------;
%macro life(status_dato,amr_ref);
*-------------------------------------------------------------*
  Udvælg populationen i Lifelines.
  Personer der udgår af befolkningen på statusdagen indgår
  ikke.
*-------------------------------------------------------------;
proc sql;
create table pop as
select *
from life.lifelines&life_ref.(where=(event_start_date<="&status_dato."d<event_final_date and PNR
ne ''))
;
quit;

* Fjern evt. dubletter;
proc sort data=pop nodupkey;
by pnr;
run;

*-----------------------------------*
  Højest fuldførte uddannelser.
*-----------------------------------;
proc sql;
create table uddf as
select pnr,hfaudd,hf_kilde,instnr,HF_VFRA,HF_VTIL
from   uddf.uddf&uddf_ref.(where=(HF_VFRA<="&status_dato."d<=HF_VTIL and PNR ne ''))
;
quit;

* Fjern evt. dubletter;
proc sort data=uddf nodupkey;
by pnr;
run;

*-----------------------------------*
 Igangværende uddannelser.
*-----------------------------------;
proc sql;
create table kotre as
select  pnr,udd,ELEV3_VFRA,ELEV3_VTIL
from    kotre.kotre&kotre_ref.
         (where=(ELEV3_VFRA<="&status_dato."d<=ELEV3_VTIL and PNR ne ''));
quit;

* Fjern evt. dubletter;
```

```
proc sort data=kotre nodupkey;
by pnr;
run;

*----------------------------------*
  Arbejdsmarkedsregnskabet.
*----------------------------------;
proc sql;
create table amr as
select pnr,I_BEFOLKNINGEN_KODE,PRIMAER_STATUS_KODE,SOC_STATUS_KODE,FRA_DATO,TIL_DATO,
       BOPAEL_KOM_KODE,ARB_BEL_KOM_KODE,ARB_HOVED_BRA_DB07,IE_TYPE
from amrun.amrun&amr_ref.(where=(FRA_DATO="&status_dato."d
                                 and PRIMAER_STATUS_KODE=1
                                 and cprtjek='0'
                                 and PNR ne ''));
quit;

proc sort data=amr nodupkey;
by pnr ;
run;

*----------------------------------*
  Data samles.
  Markering af match.
*----------------------------------;
data temp.life&status_dato.;
retain pnr status_dato;
merge pop(in=ok1) uddf amr kotre;
by pnr;
if ok1;        * Alle fra Lifelines medtages;
status_dato="&status_dato"d;
format status_dato mmyy.;
* alder på statustidspunktet;
alder=FLOOR(YRDIF(foed_dato,status_dato,'AGE'));
run;
%mend;

*-------------------------------------------------------*
 Makroen afvikles pr. måned.
 Der skal angives 4 parametre:
   1. Statusdagen.
   2. Referencetid for SAS-datasæt med AMR.
 -------------------------------------------------------;
options obs=max;
%life(01JAN2023,2023);
%life(01FEB2023,2023);
%life(01MAR2023,2023);
%life(01APR2023,2023);
%life(01MAY2023,2023);
%life(01JUN2023,2023);
%life(01JUL2023,2023);
%life(01AUG2023,2023);
%life(01SEP2023,2023);
%life(01OCT2023,2023);
%life(01NOV2023,2023);
%life(01DEC2023,2023);

*-----------------------------------------------------------*
  Tabeller.
*-----------------------------------------------------------;
*----------------------------------*
  Formater.
*----------------------------------;
proc format;
VALUE SOC_STATUS
other=[SOCSTATUS_GR3_SB.]
.='99 Ikke i AMR'
;
run;

*----------------------------------*
  Alle data samles.
*----------------------------------;
data tab1;
```

```
set temp.life01:;
alder=FLOOR(YRDIF(foed_dato,status_dato,'AGE'));  * Alder på statusdatoen;
keep status_dato I_befolkningen_kode hfaudd udd
     SOC_STATUS_KODE alder;
run;

*-----------------------------------------*
   Tabel med højest fuldførte uddannelse.
*-----------------------------------------;
proc tabulate data=tab1 missing order=formatted;
where 15 le alder le 69;
class status_dato hfaudd;
format hfaudd AUDD_HOVED_L1L5_KT. status_dato ddmmyy10.;
table all hfaudd, status_dato='Statustidspunkt'*(n*f=commax12.);
table all hfaudd, status_dato='Statustidspunkt'*(colpctn*f=5.1) ;
keylabel n='Antal' colpctn='Pct.' all='I alt';
run;

*-----------------------------------------*
    Tabel med igangværende uddannelse.
*-----------------------------------------;
proc tabulate data=tab1 missing order=formatted;
where udd ne .;
class status_dato udd;
format udd UDD_HOVED_L1L5_KT. status_dato ddmmyy10.;
table all udd, status_dato='Statustidspunkt'*(n*f=commax12.) ;
table all udd, status_dato='Statustidspunkt'*(colpctn*f=5.1) ;
keylabel n='Antal' colpctn='Pct.' all='I alt';
run;

*-----------------------------------------*
    Tabel med socioøkonomisk status.
*-----------------------------------------;
proc tabulate data=tab1 missing order=formatted;
format SOC_STATUS_KODE SOC_STATUS. status_dato ddmmyy10.;
class status_dato soc_status_kode;
table all soc_status_kode, status_dato='Statustidspunkt'*n*f=commax12.;
table all soc_status_kode, status_dato='Statustidspunkt'*colpctn*f=5.1 ;
keylabel n='Antal' colpctn='Pct.' all='I alt';
run;
```