



System for persondata

Dokumentasjon fra og med årgang 2008

TALL

SOM FORTELLER

NOTATER / DOCUMENTS

2020 / 28

Nadiya Fedoryshyn og Pål Nordby

I serien Notater publiseres dokumentasjon, metodebeskrivelser, modellbeskrivelser og standarder.

© Statistisk sentralbyrå
Ved bruk av materiale fra denne publikasjonen
skal Statistisk sentralbyrå oppgis som kilde.

Publisert 3. september 2020

ISBN 978-82-587-1169-5 (elektronisk)
ISSN 2535-7271 (elektronisk)

Standardtegn i tabeller	Symbol
Tall kan ikke forekomme	.
Oppgave mangler	..
Oppgave mangler foreløpig	...
Tall kan ikke offentligjøres	:
Null	-
Mindre enn 0,5 av den brukte enheten	0
Mindre enn 0,05 av den brukte enheten	0,0
Foreløpig tall	*
Brudd i den loddrette serien	—
Brudd i den vannrette serien	
Desimaltegn	,

Forord

Notatet gir en dokumentasjon av det registerbaserte System for persondata (SFP), som i sin siste versjon (SFP9) foreligger for årgangene 2008 – 2019. Notatet er en oppdatering av notat «System for persondata - versjon 9. Dokumentasjon av årgangene 2008-2009».

Dokumentasjonen følger inndelingen i Statistisk sentralbyrås virksomhetsmodell for arbeidsprosesser i statistikkproduksjonen.

I SFP er registerbasert sysselsettingsstatistikk utvidet til å omfatte også de ikke-sysselsatte, og for begge grupper gis det opplysninger om hvorvidt de er under utdanning eller mottar ulike former for offentlige ytelser. Dette individdatasystemet ble etablert for å beskrive grupperes forhold til arbeidsmarkedet, utdanning og offentlige ytelser.

Fra og med 2015 er tallene for sysselsatte basert på ny datakilde – a-ordningen. Denne dokumentasjonen beskriver endringene i SFP som følge av overgang til a-ordningen.

Publikasjonen er utarbeidet av Nadiya Fedoryshyn og Pål Nordby fra Seksjon for arbeidsmarkeds- og lønnsstatistikk.

Statistisk sentralbyrå, 21. august 2020.

Ann-Kristin Brændvang

Sammendrag

I dette notatet dokumenteres siste versjon av System for persondata (SFP) som foreligger for årgangene 2008 – 2019. SFP består av registre som inneholder data for områder som arbeidsmarked, utdanningsaktivitet og velferdsordninger som er ment å kompensere for manglende arbeidsinntekt.

I SFP er registerbasert sysselsettingsstatistikk utvidet til å omfatte også de ikke-sysselsatte, og for begge grupper gis det opplysninger om hvorvidt de er under utdanning eller mottar ytelser. Systemet ble etablert for å beskrive ulike gruppers forhold til arbeidsmarkedet, utdanning og ytelser. Populasjonen i SFP er personer i alderen 15 år og over, både bosatte og noen grupper ikke-bosatte.

Ved å fokusere ikke bare på sysselsetting og ledighet, men også på andre aktiviteter og ytelser, kan man få en bedre forståelse av arbeidsmarkedet. Analyser av én årgang kan f.eks. beskrive hvilke aktiviteter og ytelser personer utenfor arbeidsstyrken har. Det kan si noe om hvor langt de er fra å bli sysselsatt, f.eks. om de er under utdanning eller uføre. Det er også mulig å se på kombinasjonsstatuser, noe som særlig er interessant blant deltidssysselsatte for å belyse hvorfor de ikke jobber på heltid. I SFP kan man også følge individer over tid, og på den måten beskrive hvor innstrømningen til sysselsetting kommer fra (f.eks. fra utdanning eller ledighet), hvor utstrømningen fra sysselsetting går til (f.eks. til ledighet eller pensjon), og evaluere tiltak (f.eks. se hvordan det gikk 1-2 år etter tiltaket).

SFP inneholder tre hovedtyper av variabler som kan være nyttige for analyser. For det første beregnes én status for den enkelte persons forhold til arbeidsmarkedet, utdanning og ytelser. Personer som er aktive i flere statuser samtidig, blir da plassert på én status ut i fra en bestemt prioritering. Denne variabelen er velegnet til å gi en enkel og fokusert beskrivelse av befolkningsgruppers situasjon. For det andre inneholder SFP et sett av fulltellingsvariabler for den enkelte aktivitet/ytelse (f.eks. "Alderspensjon eller ei", "Under utdanning eller ei", osv.). Disse variablene er velegnet til å beskrive den totale forekomsten av hver enkelt aktivitet/ytelse, og til å gjøre uttrekk av analysepopulasjoner (f.eks. se nærmere på alle personer som mottar alderspensjon). For det tredje inneholder SFP en rekke demografiske og sosioøkonomiske kjennetegn ved personene, og kjennetegn ved jobben/virksomheten for sysselsatte, som man kan bruke bl.a. til å avgrense og beskrive analysepopulasjoner nærmere ved.

SFP brukes av Seksjon for arbeidsmarkeds- og lønnsstatistikk som datagrunnlag for analyser av arbeidsmarkedet, både i betalte oppdrag og i egne prosjekter som formidles gjennom rapporter, artikler og tabellleveranser. Andre seksjoner i SSB kan søke om tilgang til SFP, og det utleveres også mikrodata til forskere. Systemet er siden 2018 brukt til å produsere statistikken «Tilknytning til arbeid, utdanning og velferdsordninger» (<https://www.ssb.no/arbeid-og-lonn/statistikker/arbstatus>).

Innhold

Forord	3
Sammendrag	4
1. Bakgrunn og formål	6
1.1. Formål og historie	6
1.2. Brukere og anvendelsesområder	7
2. Avklare behov - populasjon	7
3. Planlegging og utforming – datakilder	8
3.1. Kilder om forholdet til arbeidsmarkedet, utdanning og ytelser	9
3.2. Kilder om personkjennetegn	13
4. Klargjøre – statusvariabler	14
4.1. Oppbygging av prioritert arbeidsstyrkestatus	14
4.2. Begrunnelse for prioriteringsrekkefølgen på én status	16
5. Datakvalitet	19
5.1. Kontroll og revisjon.....	19
5.2. Frafall av personer og opplysninger	19
5.3. Andre feilkilder og usikkerhet	21
6. Sammenliknbarhet og sammenheng	23
6.1. Sammenheng over tid og sted	23
6.2. Sammenheng med annen statistikk	24
Vedlegg A: Variabler og kodelister	27
Vedlegg B: SFP3 - eldre versjon av SFP	28

1. Bakgrunn og formål

1.1. Formål og historie

Dette notatet er en dokumentasjon av den siste versjonen av System for persondata (SFP9) som foreligger fra og med årgang 2008.

Formålet med SFP er først og fremst å bryte ned den store og heterogene gruppen av personer som er utenfor arbeidsstyrken, ved å fordele dem på noen hovedgrupper av statuser. Dette kan si noe om hvor langt unna sysselsetting de er og hvilke offentlige ytelser de mottar. I SFP kan man også se kombinasjoner av ulike statuser, for eksempel hvordan de sysselsatte kombinerer arbeid med utdanning eller mottak av offentlige ytelser.

Dette individdatasystemet ble etablert for å beskrive bedre grupperes forhold til arbeidsmarkedet, utdanning og ytelser.

Ved å fokusere ikke bare på sysselsetting og ledighet, men også på andre aktiviteter og ytelser, kan man få en bedre forståelse av arbeidsmarkedet. Analyser av én årgang kan f.eks. beskrive hvilke aktiviteter og ytelser personer utenfor arbeidsstyrken har. Det kan si noe om hvor langt de er fra å bli sysselsatt, f.eks. om de er under utdanning eller uføre. Det er også mulig å se på kombinasjonsstatuser, noe som særlig er interessant blant deltidssysselsatte for å belyse hvorfor de ikke jobber på heltid. I SFP kan man også følge individer over tid, og på den måten beskrive hvor innstrømningen til sysselsetting kommer fra (f.eks. fra utdanning eller ledighet), hvor utstrømningen fra sysselsetting går til (f.eks. til ledighet eller pensjon), og evaluere tiltak (f.eks. se hvordan det gikk 1-2 år etter tiltaket).

SFP inneholder tre hovedtyper av variabler som kan være nyttige for analyser. For det første beregnes *én status* for den enkelte persons forhold til arbeidsmarkedet, utdanning og ytelser. Personer som er aktive i flere statuser samtidig, blir da plassert på én status ut i fra en bestemt prioritering. Denne variabelen er velegnet til å gi en enkel og fokusert beskrivelse av befolkningsgrupperes situasjon. For det andre inneholder SFP et sett av *fulltellingsvariabler for den enkelte aktivitet/ytelse* (f.eks. "Alderspensjon eller ei", "Under utdanning eller ei", osv.). Disse variablene er velegnet til å beskrive den totale forekomsten av hver enkelt aktivitet/ytelse, og til å gjøre uttrekk av analysepopulasjoner (f.eks. se nærmere på alle personer som mottar alderspensjon). For det tredje inneholder SFP en rekke demografiske og sosioøkonomiske *kjennetegn ved personene, samt kjennetegn ved jobben/virksomheten* for sysselsatte, som man kan bruke bl.a. til å avgrense og beskrive analysepopulasjoner nærmere ved.

Mange av opplysningene i SFP9 er hentet fra den totale registerbaserte sysselsettingsstatistikken (reg_total), som er den grunnfilen alle de løpende registerbaserte sysselsettingsstatistikken baserer seg på. Til SFP9 brukes en oppdatert versjon av denne grunnfilen, med selvstendig næringsdrivende fra statistikkåret. Opplysninger som er hentet fra reg_total, blir i liten grad dokumentert i dette notatet. For dokumentasjon av reg_total (gyldig til og med 2014), se http://www.ssb.no/emner/06/90/notat_201008/notat_201008.pdf.

For dokumentasjon av eldre versjon av SFP – SFP3 - se vedlegg B.

1.2. Brukere og anvendelsesområder

Løpende statistikk

SFP er grunnlag for statistikken «Tilknytning til arbeid, utdanning og velferdsordninger». Statistikken ble for første gang publisert i desember 2018 og omfatter alle bosatte i alderen 15 år og over per 4. kvartal (tredje uke i november) i referanseåret.

Statistikken har flere formål: For det første å bryte ned den store og heterogene gruppen av personer som er utenfor arbeidsstyrken, ved å fordele de på noen hovedgrupper av statuser. Dette kan si noe om hvor langt unna sysselsetting de er. Statistikken beskriver også om de deltidssysselsatte har andre aktiviteter/ytelser ved siden av eller om deltidsarbeid er eneste registrerte aktivitet. Dette for å belyse hvorfor de ikke jobber på heltid. For det andre er det økt interesse for statistikk over strømmer og bruttoendringer. For eksempel viser statistikken tall over "innstrømmingen" til og "utstrømmingen" fra sysselsetting. Dette kan belyse hvor store strømmene er og hvilke statuser personene beveger seg fra/til.

Les mer om statistikken på statistikkens nettside: <https://www.ssb.no/arbstatus>.

Analyseprosjekter og oppdrag

Seksjon for arbeidsmarkedsstatistikk bruker også SFP som datagrunnlag for analyser, både i betalte oppdrag og egne prosjekter som formidles gjennom rapporter, notater, artikler og tabellleveranser.

Utlevering av mikrodata

Andre seksjoner i SSB kan søke om tilgang til årsfiler, og det utleveres mikrodata til forskere.

2. Avklare behov - populasjon

Sysselsetting kan sees fra to ulike synsvinkler. Definisjonen av sysselsetting er lik i de to perspektivene, men de har noe ulike anbefalinger om hva som skal være populasjonen. Det ene perspektivet er å beskrive folks tilpasning til arbeidsmarkedet. Da er det vanlig å avgrense populasjonen til personer 15-74 år som er bosatte i Norge (*den hjemmehørende folkemengden i yrkesaktiv alder*). Det andre perspektivet er foretakenes bruk av arbeidskraft i produksjonen av varer og tjenester. Da er det vanlig å inkludere alle *sysselsatte som har sitt arbeidssted i Norge*, dvs. også sysselsatte på korttidsopphold i Norge (ikke-bosatte), og sysselsatte som er 75 år eller eldre. Her kan fokus enten være på *individene og deres hovedarbeidsforhold, eller alle jobber personene har*.

Når det gjelder mottakere av ulike former for offentlige ytelser, kan dette dreie seg både om personer som er bosatte og personer som er ikke-bosatte. Eksempelvis er det mange som utvandrer og mottar alderspensjon fra Norge i utlandet. Noen ytelser kan også mottas av personer som er over 75 år.

Populasjonen i SFP

I SFP er det behov for å kunne beskrive både den totale forekomsten av sysselsatte med arbeidssted i Norge, den totale forekomsten av registrert helt arbeidsledige i Norge og den totale forekomsten av personer som mottar ulike former for offentlige ytelser fra Norge, eller som er under utdanning. Personer som utvandrer og ikke er registrert på noen ytelse eller aktivitet, er mindre interessante for SFP, og disse ville det også være vanskelig å inkludere og følge videre på en god måte i SFP.

Når det gjelder ikke-bosatte, må tre relevante grupper holdes utenfor populasjonen i SFP på grunn av dårlig datakvalitet, nemlig ikke-bosatte lønnsstakere 75 år eller eldre, ikke-bosatte sjøfolk som er bosatt utenfor EØS-området, og ikke-bosatte som

er selvstendig næringsdrivende.¹ Noen av disse personene kan imidlertid komme med i SFP fra kilderegistrene om ytelser eller igangværende utdanning.

Derfor består populasjonen i SFP av følgende grupper:

- **Bosatte:** alle bosatte 15 år eller eldre.
- **Ikke-bosatte:** følgende grupper inngår:
 - o Ikke-bosatte 15 – 74 år som er lønnstakere eller registrert helt arbeidsledige.
 - o Ikke-bosatte 15 år eller eldre som mottar en ytelse eller er under utdanning.

Hva menes med bosatte og ikke-bosatte?

Personer som kommer fra land utenfor Norden og som forventes å oppholde seg i minst seks måneder i landet, skal registreres som bosatt i Det sentrale folkeregisteret. Personer som oppholder seg i Norge i en periode på under seks måneder, eller som ikke bor i Norge, men reiser daglig inn i landet for å arbeide, skal registreres som ikke-bosatt. Sistnevnte gruppe omfatter bl.a personer som pendler over grensen hver dag eller ukentlig for å arbeide i Norge. Det gjelder f.eks. nordmenn og svensker som bor i Sverige og jobber i Norge. En person som er ikke-bosatte kan endre status til bosatt året etter.

SSB publiserer statistikk om sysselsetting blant bosatte 15 – 74 år, samt tabeller for sysselsetting og ledighet blant ikke-bosatte 15 – 74 år. I tillegg til personene som inngår i disse to statistikkene, inkluderer SFP9 også alle bosatte 75 år eller eldre, samt ikke-bosatte 15 år eller eldre som mottar ytelse eller er i utdanning.

Populasjonen i SFP9 er identisk med populasjonen i grunnfilen som kalles den totale registerbaserte sysselsettingsstatistikken (reg_total), men i tillegg inkluderer SFP9 ikke-bosatte som kommer med fra registre om ytelser og utdanningsaktivitet.

Kun hovedarbeidsforhold i SFP

SFP er en individfil med én rad per person. Det betyr at kjennetegn ved arbeidsforholdet/virksomheten for sysselsatte bare gjelder deres hovedarbeidsforhold, i motsetning til grunnfilen for de registerbaserte sysselsettingsstatistikkene (reg_total) der alle hoved- og bijobber har egne rader. I statistikken «Antall arbeidsforhold og lønn» publiseres det tall for antall jobber. I variabelisten i SFP9 finnes det imidlertid en aggregatvariabel for avtalt arbeidstid som viser sum avtalt arbeidstid i alle jobber personen har. Alle øvrige kjennetegn ved jobb eller virksomhet i SFP9, gjelder hovedjobben.

3. Planlegging og utforming – datakilder

I dette kapitlet gis en gjennomgang av kildene som benyttes i produksjonen av SFP9. De brukes til to hovedformål. For det første til å bestemme populasjonen. Populasjonsavgrensningen tar utgangspunkt i en oppdatert versjon av grunnfilen reg_total med selvstendige fra statistikkåret som dekker alle bosatte 15 år eller eldre, og ikke-bosatte 15-74 år som er lønnstakere eller registrert helt arbeidsledige. I tillegg til disse personene fra reg_total, legges det i SFP9 til flere ikke-bosatte 15 år eller eldre fra kildene om ytelser og igangværende utdanning. Populasjonen redegjøres det nærmere for i kapittel 2.

For det andre brukes kildene til å bygge opp variabelisten i SFP9 som består av fem hovedtyper:

- Fulltellingsvariabler for den enkelte aktivitet/ytelse ("AFP eller ei", "Sosialhjelp eller ei", osv.).

¹ Når det gjelder ikke-bosatte selvstendige, har SSB per i dag et for dårlig datagrunnlag for å avgjøre om virksomhetene fortsatt var aktive eller ikke i referanseuken, og derfor er ikke-bosatte selvstendige holdt utenfor i SFP9.

- Én status for den enkelte persons forhold til arbeidsmarkedet, utdanning og ytelser. Personer som er aktive i flere aktiviteter/ytelser samtidig, er her plassert på én status ut i fra en prioritering.
- Personkjennetegn (demografiske, sosioøkonomiske o.l.).
- Kjennetegn ved jobben/virksomheten for sysselsatte.
- Kjennetegn ved ytelsen for personer som mottar dette (uføregrad, pensjoneringsgrad, o.a.).

Den klart viktigste kilden for SFP9 er grunnfilen for de registerbaserte sysselsettingsstatistikkene kalt `reg_total` (oppdatert versjon med selvstendige fra statistikkåret), siden den danner hovedgrunnlag både for avgrensningen av populasjonen (supplert med flere ikke-bosatte som mottar en ytelse eller er under utdanning) og for variabelsettet i SFP9. Alle opplysninger om sysselsetting og ledighet samt kjennetegn ved jobben/virksomheten for sysselsatte, hentes fra `reg_total`. I tillegg hentes de fleste personkjennetegn fra `reg_total` (kjønn, alder, osv.). Det forsøkes å hente flest mulig opplysninger fra `reg_total`, både for å forenkle produksjonen av SFP9 og for å sikre konsistens med `reg_total`. Siden SFP9 har en litt større populasjon enn `reg_total`, må tilsvarende opplysninger for ikke-bosatte som ikke inngår i `reg_total`, hentes fra de samme grunnregistrene som `reg_total` bruker (Det sentrale folkeregisteret, Utdanningsregisteret, o.a.). I SFP9 er det en variabel kalt "fra_regtot" som angir om personen kommer fra `reg_total` eller ikke. Ved hjelp av dette populasjonsmerket kan man også finne ut av om andre opplysninger (f.eks. kjønn og alder) kommer fra `reg_total` eller direkte fra grunnregistrene (for tilleggspopulasjonen som ikke er i `reg_total`). Opplysninger i SFP9 som kommer fra `reg_total`, er dokumentert her:

http://www.ssb.no/emner/06/90/notat_201008/notat_201008.pdf

Noen av kildene som omtales nedenfor, er ikke egne administrative registre formelt sett, men kildefiler som er laget på grunnlag av uttrekk fra administrative registre eller sammenstilling av slike. Gjennomgangen av kildene er her sortert etter hva de opplyser om, ikke etter navnet på kilden.

3.1. Kilder om forholdet til arbeidsmarkedet, utdanning og ytelser

Sysselsetting og arbeidsledighet

Opplysningene om hvem som er sysselsatte eller registrert helt arbeidsledige, dvs. i arbeidsstyrken, er hentet fra grunnfilen for de registerbaserte sysselsettingsstatistikkene (`reg_total`). Det samme gjelder opplysningene om arbeidstid, lønn, samt kjennetegn ved jobben/virksomheten for sysselsatte. Datagrunnlaget som brukes i produksjon av `reg_total` er basert på ulike datakilder i årene før og etter 2015. Hovedkilden fram til og med 2014 var NAVs Arbeidstakerregister (Aa-registeret). I 2015 ble rapporteringen til dette registeret samordnet med rapportering av lønns- og personelldata til Skatteetaten og SSB. Det nye felles rapporteringssystemet kalles a-ordningen. A-ordningen gir generelt et bedre datagrunnlag ved at det er mer korrekt på individnivå, samt at det dekker flere lønnstakerforhold enn Aa-registeret. Blant annet kommer oppdragstakere og frilansere med i rapporteringen fra og med 2015. Denne gruppen registreres med status «0832 Oppdragstaker, frilanser» og `akt_oppdrag='1'` i SFP.

De viktigste endringene i tallene i forbindelse med overgangen til a-ordningen er omtalt i [denne artikkelen](#). Overgang til ny datakilde innebærer at sysselsettingstallene for årgangene fra og med 2015 ikke er sammenlignbare med tall for tidligere årganger.

Til SFP9 brukes en oppdatert versjon av `reg_total` med selvstendige fra statistikkåret, i motsetning til de ordinære registerbaserte sysselsettingsstatistikkene

som av hensyn til produksjonstid må basere seg på gamle data om hvem som er selvstendige. Slik blir kvaliteten bedre på individnivå og ved detaljerte grupper i SFP9, men med samme nivå-tall som ordinær reg_total. Referanseperioden er tredje uke i november.

Foreløpige tall for selvstendige i statistikken «Tilknytning til arbeid, utdanning og velferdsordninger»

For å forbedre aktualiteten i statistikken «Tilknytning til arbeid, utdanning og velferdsordninger» publiseres det foreløpige tall for denne statistikken fra og med statistikkår 2019. De foreløpige tallene vil publiseres i begynnelsen av september hvert år, med første gang september 2020. Opplysningene om selvstendige i statistikkåret er ikke tilgjengelige på dette tidspunkt. Til den foreløpige publiseringen lages derfor en foreløpig SFP-fil med opplysninger om fjorårets selvstendige, på lik linje som ved produksjon av registerbasert sysselsettingsstatistikk.

Når siste årets opplysninger om selvstendige er tilgjengelige, oppdateres SFP-filen og Statistikkbank-tabellene med de nyeste tallene. Kvaliteten i den foreløpige SFP-filen vurderes som god til statistikkformål, men endelig versjon av SFP skal benyttes til analyser og forskning på individnivå.

For mer informasjon om A-ordningen og selvstendig næringsdrivende i SFP, se kapittel 5 om datakvaliteten.

Lønn Fra og med årgang 2016 kobles informasjon om månedslønn og avtalt lønn per november fra A-ordningen på SFP-filene.

Månedslønn er hovedbegrepet i Statistisk sentralbyrås lønnsstatistikk. Månedslønn omfatter avtalt lønn, uregelmessige tillegg og bonuser. Overtidsgodtgjørelser er ikke medregnet i månedslønn.

Avtalt lønn omfatter den faste lønnen (omregnet til månedslønn) som utbetales enten den er definert som time-, måneds-, 14 dagers- eller ukelønn. Avtalt lønn er den faktisk utbetalte lønnen ved tellingstidspunktet, og den blir ofte betegnet som regulativlønn eller fast grunnlønn. Kvalifikasjonstillegg/kompetansetillegg og andre faste personlige tillegg skal inkluderes i denne lønnsarten.

Både månedslønn og avtalt månedslønn brukes til publisering av statistikkbanktabellene i statistikken «Tilknytning til arbeid, utdanning og velferdsordninger».

Heltidsekvivalent

Statistikken viser månedslønn og avtalt månedslønn i heltidsekvivalenter. I statistikken blir ansatte med stillingsandel på mindre enn 100 prosent definert som deltidsansatte. For å kunne sammenligne lønn mellom hel- og deltidsansatte omregnes lønna til de deltidsansatte til hva den ville vært hvis de jobbet heltid. Dette gjøres ved å benytte stillingsprosenten til hver enkelt deltidsansatt som omregningsfaktor. Månedslønn per heltidsekvivalent for de deltidsansatte kan da slås sammen med månedslønn med de heltidsansatte slik at man kan beregne en gjennomsnittlig månedslønn for alle ansatte.

Tiltak og ytelser for arbeidssøkende

NAVs ARENA-register over registrerte arbeidssøkende gir opplysninger om hvilke aktiviteter og ytelser de arbeidssøkende går på. Referanseperioden vi henter data for er tredje uke i november. Det er tre dimensjoner vi er interesserte i: 1) nedsatt arbeidsevne eller ikke, 2) type aktivitet og 3) type ytelse. Variabelen tiltak fra grunnfilen reg_total beskriver de to første dimensjonene for de som er helt ledige eller på tiltak, mens ytelsesstatusen fra ARENA beskriver ytelse for alle i ARENA. I praksis lages det i SFP9 fulltellingsvariabler (0 – 1) basert på utvalgte aktivitetsstatuser og ytelsesstatuser, som så settes sammen i en bestemt prioritering

ved beregningen av én status der tiltak (aktivitet) prioriteres foran ytelse fra ARENA. De arbeidssøkende som ikke har aktivitet i ARENA, kan da få synliggjort ytelsen fra ARENA, mens de som har begge deler kan få ytelsen som kombinasjonsstatus, avhengig hvordan prioriteringsrekkefølgen slår ut for den enkelte siden kun de høyest prioriterte statusene synliggjøres.

Aktivitet beskrives av variabelen tiltak fra grunnfilen reg_total. Det er en aggregert versjon SSB lager av aktivitetsopplysninger fra ARENA for personer som er helt ledige eller på tiltak i ARENA:

- 0 Helt ledige
- 1 Ordinære tiltaksdeltakere (inkl. kvalifiseringstiltak)
- 6 Nedsatt arbeidsevne, ikke på tiltak
- 7 Nedsatt arbeidsevne, på tiltak
- Uoppgitt

I SFP9 lages det fulltellingsvariabler basert på hver av disse aktivitetskodene, med to unntak. Verdien 0 (Helt ledig ifølge ARENA) er det ikke hensiktsmessig å bruke ved beregning av én status. Hvem som er registrert helt arbeidsledig, hentes isteden fra variabelen arb_arbmark_status i reg_total, som er basert på verdien 0 på tiltak fra ARENA, men hvor noen personer er omklassifisert til lønnstakere fordi SSB tror mer på dateringene i Aa-registeret (a-ordningen fra og med 2015). Hvis vi også hadde laget en fulltellingsvariabel basert på verdien 0 fra variabelen tiltak, ville det gitt intetsigende opplysninger på én status. De fleste av disse ville da fått statusen "Registrert helt arbeidsledig og helt ledig i ARENA", og noen "Lønnstaker og helt ledig i ARENA". Informasjon om hvem ARENA anser som helt ledige, kan brukere finne på variabelen pers_tiltak fra reg_total, som vil inngå i variabellisten i SFP9. Uoppgitt-verdier på variabelen tiltak i reg_total (arbeidssøkende som ikke er helt ledige eller på tiltak), lages det heller ikke fulltellingsvariabel av, fordi det er en sammensatt gruppe som heller bør beskrives ved eventuelle ytelser de får (delvis ledige, stønadsmottakere som ikke er helt ledig/på tiltak, og personer som burde vært registrert som inaktive).

Denne behandlingen av personer med verdi 0 i ARENA fører til at det ikke er helt samsvar mellom antall ledige i SFP sammenliknet med NAV sine tall. Se mer om dette i kapittel 6.

I tillegg til utnyttelsen av aktivitetsopplysninger fra ARENA, lages det også egne fulltellingsvariabler i SFP9 for noen av ytelsene i ARENA. Ytelsesstatusen i ARENA er i prinsippet ikke en opplysning om utbetalinger/mottakere, men betyr at personen har rett på ytelsen. På den måten kan vi belyse massen i ARENA som ikke er i noen aktivitet slik vi definerer det, men som mottar en relevant ytelse. De som har både aktivitet og ytelse i ARENA kan da få synliggjort ytelsen som kombinasjonsstatus, avhengig av hvordan prioriteringsrekkefølgen slår ut for den enkelte person ved plasseringen på én status.

Kodeliste for de ulike ytelsene i ARENA-filene er tilgjengelig her: <https://www.ssb.no/klasse/klassifikasjoner/394>.

I ARENA er det en del personer som verken har verdi på variabelen tiltak i reg_total eller på noen av rettighetsstatusene i ARENA. Den massen/opplysningen brukes ikke til noe i SFP9. De er i prinsippet ikke interessante for SFP, fordi de verken er i aktivitet slik vi definerer det, eller på noen relevant ytelse fra ARENA. I denne gruppen kan det også være en del manglende registreringer av opphør.

<i>Introduksjonsstønad</i>	<p>Introduksjonsstønad utbetales <u>etter introduksjonsloven</u> til nyankomne innvandrere som deltar på introduksjonsprogram. Stønaden er ment som inntektssikring til deltakerne og utbetales av kommunen.</p> <p>Opplysningene om introduksjonsstønad i SFP er basert på Kostra-skjema.</p>
<i>Ordinær utdanning</i>	<p>Datagrunnlaget for befolkningens utdanningsnivå trekkes ut fra Nasjonal utdanningsdatabase (NUDB). Siden SSB ikke har opplysninger om igangværende utdanning på ungdomsskolenivå, gjøres følgende bearbeidinger av grunndataene:</p> <ul style="list-style-type: none">- 15- og 16-åringer som mangler igangværende utdanning og som ikke har fullført ungdomsskole, settes som i gang med utdanning. Dette vil medføre at noen 15-16 åringer feilaktig klassifiseres som i gang med utdanning (hvis de har droppet ut av ungdomsskolen), men i de fleste tilfeller vil dette være riktig en riktig bearbeiding, slik at dette gir færre feil totalt sett blant 15-16 åringerne.- 15- og 16-åringer som mangler igangværende utdanning og som har fullført ungdomsskole, settes som ikke i gang med utdanning.
<i>Sykepenger</i>	<p><u>Sykepenger</u> erstatter arbeidsinntekt når du ikke kan jobbe på grunn av sykdom eller skade. SSB benytter data om mottakere i referanseuken (tredje uke i november) fra NAV. Personer som mottar både foreldre- og sykepenger, er i våre data ikke klassifisert som sykepengemottakere. Siden vi får våre data om referanseuken i februar året etter, vil det mangle en del etterregistreringer av krav fremsatt fra arbeidsgivere inntil tre måneder etter fraværet.</p>
<i>Arbeidsavklaringspenger</i>	<p><u>Arbeidsavklaringspenger (AAP)</u> skal sikre inntekt i perioder man på grunn av sykdom eller skade har behov for hjelp fra NAV for å komme i arbeid. Denne ytelsen erstattet rehabiliteringspenger, attføringspenger og tidsbegrensa uførestønad fom årgang 2010.</p> <p>SSB får data om mottakere i november fra NAV.</p>
<i>Uførepensjon/uføretrygd</i>	<p><u>Uføretrygd</u> kan være aktuelt for personer som har varig nedsatt inntektsevne på grunn av sykdom eller skade. I årene 2008-2009 inkluderer uføretrygd både foreløpig og varig uførepensjon. Etter endringer i uføreordningen fra 1. januar 2015 ble uførepensjon erstattet med uføretrygd.</p> <p>Data om mottakere i november er basert på opplysninger fra NAV til og med årgang 2018. Fra og med årgang 2019 hentes informasjon om mottakere fra a-ordningen. Tallene i a-ordningen er basert på rapportering fra NAV.</p>
<i>Etterlattepensjon inkl. barnepensjon</i>	<p><u>Etterlattepensjon</u> og <u>barnepensjon</u> er ytelser som skal bidra til å sikre ektefelle/samboer og barn inntekt til livsopphold etter dødsfall.</p> <p>Data om mottakere i november er basert på opplysninger fra NAV til og med årgang 2018. Fra og med årgang 2019 hentes informasjon om mottakere fra a-ordningen. Tallene i a-ordningen er basert på rapportering fra NAV.</p>
<i>AFP-pensjon</i>	<p><u>Avtalefestet pensjon (AFP)</u> er tidligpensjon for personer som har fylt 62 år, og som arbeider innenfor områder der det er tariffavtaler og hvor AFP inngår i tariffavtalen.</p> <p>Til og med årgang 2018 ble data om mottakere i november hentet fra NAV. Det er mangelfull informasjon om to grupper av nye AFP-mottakere f.o.m. 2009, nemlig nye statlige AFP-mottakere 65 – 67 år og nye kommunale AFP-mottakere. Hvis disse personene ikke er registrert i noen andre ytelser eller aktiviteter i SFP9, vil de havne i ukjent status i SFP9.</p> <p>Fra og med årgang 2019 hentes data om mottakere fra a-ordningen, siden a-ordningen gir bedre dekningsgrad av AFP-mottakere i offentlig sektor.</p>

<i>Supplerende stønad</i>	<u>Supplerende stønad</u> kan gis til personer som har fylt 67 år og som har liten eller ingen alderspensjon grunnet kort botid i Norge. Stønad ble innført fra 1. januar 2006. SSB får data om mottakere i november fra NAV.
<i>Alderspensjon</i>	<u>Alderspensjon</u> fra folketrygden er en ordning som sikrer inntekt når man blir pensjonist. Data om mottakere i november er basert på opplysninger fra NAV til og med årgang 2018. Fra og med årgang 2019 hentes informasjon om mottakere fra a-ordningen. Tallene i a-ordningen er basert på rapportering fra NAV.
<i>Pensjoner fra andre enn NAV/Folketrygden</i>	Datakilden er her en samlekategori av utvalgte poster i Skatteetatens lønns- og trekkoppgaveregisteret (LTO-registeret). Kilden ble benyttet til og med 2014. Fra og med 2015 ble LTO-registeret erstattet av a-ordningen. Personer som har et beløp på en eller flere av følgende poster i LTO-registeret regnes som aktive på fulltellingsvariabelen for denne typen pensjoner: <ul style="list-style-type: none"> - Itokode_211: Pensjon i og utenfor arbeidsforhold og livrenter i arbeidsforhold m.v. (fast rubrikk). - Itokode_212: Pensjon og livrenter i arbeidsforhold m.v. Medfører arbeidsgiveravgift. - Itokode_214: Skattepliktig etterlønn og etterpensjon etter dødsfall. - Itokode_216: Føderåd. - Itokode_238: Nye livrenter i arbeidsforhold (kollektive livrenter). - Itokode_247: Uførepensjon fra andre enn folketrygden. <p>Referanseperioden for LTO-registeret er hele året, men siden referanseperioden for øvrige opplysninger i SFP9 er i november og mottak av pensjoner som oftest representerer en varig tilstand, er ikke mangelen på datering av dette registeret noe stort problem. Det er altså særlig personer som begynner å motta slike pensjoner etter referanseuken for SFP9, dvs. i løpet av årets 5 – 6 siste uker, som vil bli feilaktig klassifisert som aktive mottakere av disse pensjonene i referanseuken i SFP9.</p>
<i>Kontantstøtte</i>	<u>Kontantstøtte</u> er penger man kan få hvis man har barn mellom 1 og 2 år som ikke har fått tildelt fulltidsplass i barnehage. Kontantstøtte gis til husholdningen, og i parforhold kan det dermed være mer eller mindre tilfeldig hvem av ektefellene/samboerne som står registrert som mottaker. Data om mottakere i november hentes fra NAV.
<i>Enslig forsørgerstønad</i>	Enslig forsørgerstønad kan inkludere <u>flere ytelser</u> som kan være aktuelle for personer som er alene med barn, som overgangsstønad, stønad til barnetilsyn, stønad til skolepenger osv. Data om mottakere i november hentes fra NAV.
<i>Sosialhjelp</i>	<u>Økonomisk sosialhjelp</u> fra NAV skal sikre personer midlertidige inntekter til et forsvarlig livsopphold. Sosialhjelp gis til husholdningen, og i parforhold kan det dermed være mer eller mindre tilfeldig hvem av ektefellene/samboerne som står registrert som mottaker. Data om mottakere i november hentes fra NAV.

3.2. Kilder om personkjennetegn

I SFP9 hentes flest mulig opplysninger fra den oppdaterte versjonen av reg_total (selvstendige fra statistikkåret), både for å forenkle produksjonen av SFP9 og for å sikre konsistens i nivåtall med reg_total. Det gjelder også personkjennetegn som f.eks. kjønn og alder. Siden SFP9 har en litt videre populasjon, må tilsvarende opplysninger for ikke-bosatte som ikke inngår i reg_total, men som kommer med

fra kilderegistrene om ytelser eller igangværende utdanning, hentes fra de samme grunnlagsregistrene som reg_total bruker (Det sentrale folkeregister, utdanningsregisteret, o.a.). Variabellisten i SFP9 inkluderer også noen variabler som ikke finnes i reg_total, og som derfor for hele populasjonen i SFP9 må hentes direkte fra grunnlagsregistre i SSB over slike personkjennetegn.

Familietype og barn Informasjon om familietype, antall barn og barnas alder er basert på opplysninger fra Det sentrale folkeregister.

Utdanningsnivå Opplysninger om høyeste fullførte utdanningsnivå er levert fra Seksjon for utdannings- og kulturstatistikk i SSB. SSB mangler opplysninger om utdanningsnivå for en del innvandrere. For disse har man beregnet (imputert) utdanningsnivå for å kunne gi bedre statistikk på aggregert nivå over utdanningsnivå for innvandrere som står med uoppgitt i registeret. Imputerte tall brukes i publiseringen av statistikken «Tilknytning til arbeid, utdanning og velferdsordninger» fra og med 2008. SSB har kun imputert tallene for utdanningsnivå, ikke fagfelt. De beregnede tallene brukes til statistikk og analyse på aggregert nivå, og er ikke egnet til analyse på individnivå. For nærmere beskrivelse av imputeringsmetoden, se Jentoft (2014) og statistikken om befolkningens utdanningsnivå.

4. Klargjøre – statusvariabler

I SFP9 er det to hovedtyper statusvariabler om forholdet til arbeidsmarkedet, utdanning og ytelser:

- Et sett av fulltellingsvariabler som angir aktivitet eller ei i hver enkelt tilstand/ytelse/aktivitet.
- Beregning av én status hvor hver person plasseres på én status ut i fra en bestemt prioritering.

Fulltellingsvariabler Fulltellingsvariablene (variablene med navn som begynner med akt_ i SFP) har verdien 1 for de som var aktive i den aktuelle ytelsen/tilstanden i referanseuken, og verdien 0 for de som ikke var aktive. F.eks. er det en egen fulltellingsvariabel for alderspensjon, en for sosialhjelp, en for kontantstøtte, osv. Hvis analyseformålet er å beskrive alle som mottok alderspensjon i SFP, er det fulltellingsvariabelen for alderspensjon (=1) man må ta utgangspunkt i.

Prioritert arbeidsstyrkestatus Ved beregning av én status blir personer som er aktive i flere aktiviteter/ytelser samtidig, plassert på én status ut i fra en bestemt prioritering. Dvs. at alle fulltellingsvariablene sees i forhold til hverandre i en bestemt prioriteringsrekkefølge og så synliggjøres den høyest prioriterte tilstanden/ytelsen personen er aktiv i.

SSB publiserer flere statistikker som viser befolkningens tilknytning til velferdsordninger. Tallene i disse statistikkene er ikke direkte sammenliknbare med antall personer på ulike stater i SFP. Les nærmere om dette i kapittel 6 om sammenheng med andre statistikker.

4.1. Oppbygging av prioritert arbeidsstyrkestatus

Variabelen *status* har ulike aggregeringsnivåer ut i fra en åttesifret tallstruktur. Det betyr at man kan velge å ta ut ulik grad av detaljering på denne variabelen, som på det mest detaljerte nivået (åtte siffer) synliggjør inntil to aktiviteter (de to høyest prioriterte) for hver person. På åttesiffernivå vil altså personer som er aktive i flere stater samtidig, få synliggjort de to høyest prioriterte aktivitetene som en kombinasjonsstatus, f.eks. statusen ”08164008 Lønnstaker deltid og ordinær

utdanning” hvor de fire første sifrene angir den høyest prioriterte aktiviteten/ytelsen personen er aktiv i (her 0816 Lønnstaker deltid), og de fire siste sifrene angir den nest høyest prioriterte aktiviteten/ytelsen personen er aktiv i (her 4008 Ordinær utdanning). På tosiffernivå synliggjøres én aktivitet/ytelse per person på det groveste aggregeringsnivået, f.eks. ”08 Lønnstaker”. På firesiffernivå synliggjøres fortsatt bare én aktivitet/ytelse per person, men mer detaljert beskrevet, f.eks. ”0816 Lønnstaker deltid”.

Tabell 4.1 viser kodelista for variabelen én status på firesiffernivå, hvor prioriteringsrekkefølgen er ovenfra og nedover. Begrunnelsen for prioriteringsrekkefølgen kommer vi tilbake til i neste avsnitt.

Eksempel: Hvis en person er 50 prosent lønnstaker og 50 prosent varig ufør, vil han eller hun få følgende koder på de ulike siffernivåene av status gitt prioriteringsrekkefølgen for plassering på status:

- 08 Lønnstaker
- 0816 Lønnstaker deltid
- 081648 Lønnstaker deltid og helserelatert ytelse
- 08164864 Lønnstaker deltid og varig ufør
- 08 48 Lønnstaker og helserelatert ytelse

Tabell 4.1. Kodeliste for variabelen én status på firesiffernivå. Prioriteringsrekkefølgen ved plassering av personer på én status er ovenfra og nedover

Kodeliste
Arbeidsstyrken
08 Lønnstaker
0808 Lønnstaker heltid
0816 Lønnstaker deltid
0824 Lønnstaker basert på kun LTO (t.o.m. årgang 2014)
0832 Oppdragstaker, frilanser (f.o.m. årgang 2015)
16 Selvstendig næringsdrivende
1608 Selvstendig næringsdrivende
24 Registrert helt arbeidsledig
2408 Registrert helt arbeidsledig
Utenfor arbeidsstyrken
32 Andre arbeidssøkende
3208 Ordinære tiltaksdeltakere
3216 Individstønnad
3224 Dagpenger
3232 Ventestønnad (t.o.m. årgang 2007)
3240 Introduksjonsordning
40 Utdanning
4008 Ordinær utdanning
48 Helserelaterte ytelser
4808 Sykepenger
4816 Nedsatt arbeidsevne, på tiltak
4848 Arbeidsavklaringspenger (erstattet 4824 Attføringspenger, 4832 Rehabiliteringspenger og 4840 Tidsbegrensa uførestønnad f.o.m 2010)
4856 Foreløpig uførepensjon (t.o.m. årgang 2013)
4864 Varig uføretrygd (varig uførepensjon t.o.m. årgang 2014)
56 Pensjoner
5608 Etterlattepensjon inkl. barnepensjon
5616 AFP-pensjon
5624 Supplerende stønnad
5632 Alderspensjon
5640 Pensjoner fra andre enn NAV/Folketrygden
64 Andre grupper
6408 Kontantstøtte
6416 Enslig forsørgerstønnad
6424 Sosialhjelp
6432 Nedsatt arbeidsevne, ikke på tiltak
72 Ukjent status
7208 Ukjent status

4.2. Begrunnelse for prioriteringsrekkefølgen på én status

Hvilke hensyn ligger til grunn for prioriteringen av aktivitetene/ytelsene når personer som er aktive i flere aktiviteter/ytelser samtidig, skal plasseres på én status? Begrunnelsene faller i fire grupper:

- Et ønske om konsistens med nivåfall i andre registerbaserte sysselsettingsstatistikker i SSB.
- Blant personer utenfor arbeidsstyrken skal grad av nærhet til arbeidsstyrken synliggjøres, og midlertidige aktiviteter/ytelser skal synliggjøres foran mer varige aktiviteter/ytelser.
- Et ønske om og lett kunne samle årsaker til manglende sysselsetting i noen hovedgrupper.
- Plasseringen på én status skal bygge på god datakvalitet.

Konsistens med andre registerbaserte sysselsettingsstatistikker

En hovedbegrunnelse for prioriteringsrekkefølgen er ønsket om **konsistens med andre registerbaserte sysselsettingsstatistikker** i SSB. For å oppnå samme metodikk og nivåfall for arbeidsstyrken i SFP9, hentes opplysninger om hvem som er lønnstakere, selvstendig næringsdrivende eller registrert helt arbeidsledige, direkte fra grunnfilen for den totale registerbaserte sysselsettingsstatistikken (reg_total). Arbeidsstyrken blir på den måten prioritert høyest ved beregning av én status i SFP9. Dette er i tråd med ILOs anbefalinger for arbeidsmarkedsstatistikk, og vil være relevant bl.a. for statistikk i SSB om bruttostrømmer i arbeidsmarkedet. Til endelig versjon av SFP9 brukes en oppdatert reg_total basert på data fra statistikkåret om hvem som er selvstendige.

Ytelser som er nær arbeidsstyrken skal synliggjøres foran andre

Blant personene utenfor arbeidsstyrken, er en hovedbegrunnelse for prioriteringsrekkefølgen at **tilstander/ytelser som er nær arbeidsstyrken skal synliggjøres foran andre**. Det betyr at aktiviteter som innebærer at personene gjør noe for å komme i arbeid (f.eks. tiltak, introduksjonsordning eller utdanning), prioriteres foran statuser som ikke indikerer at personen gjør noe for å komme i arbeid (f.eks. alderspensjon eller varig uførhet). På samme måte skal statuser som betyr at personen i utgangspunktet bare er midlertidig utenfor arbeidsstyrken (f.eks. utdanning eller tidsbegrensa uførhet), synliggjøres foran statuser som betyr at personen er mer varig utenfor (f.eks. varig uførepensjon).

Grupper årsaker til manglende sysselsetting

En tredje hovedbegrunnelse for prioriteringsrekkefølgen er ønsket om og lett kunne **samle årsaker til manglende sysselsetting i noen hovedgrupper** med tilhørende undergrupper. Dette betyr f.eks. at ulike helserelaterte ytelser prioriteres rett etter hverandre, slik at man lett kan aggregere de opp til én hovedgruppe kalt helserelaterte ytelser. Det samme resonnetet gjelder for ulike typer pensjoner.

Når det gjelder prioritering av aktiviteter/ytelser innenfor hver enkelt hovedgruppe, er det igjen nærhet til arbeidsstyrken som prioriteres. Innenfor hovedgruppen ”48 Helserelaterte ytelser” betyr dette at sykepenges prioriteres foran tidsbegrensa uførestønad, som igjen prioriteres foran varig uførhet. De fleste sykepengemottakerne blir klassifisert som sysselsatte ved plassering på én status i SFP, men det er også noen registrert helt arbeidsledige som mottar sykepenges, og også noen personer utenfor arbeidsstyrken som mottar det. Den lille andelen av sykepengemottakerne som blir klassifisert utenfor arbeidsstyrken i SFP, kan dreie seg om spesielle grupper av ikke-sysselsatte som allikevel har rett på sykepenges,² eller personer som egentlig skulle vært klassifisert som sysselsatte i SFP (målefeil).

Innenfor hovedgruppen ”32 Andre arbeidssøkende” prioriteres tiltak (aktivitet) foran ytelser fra ARENA. Individstønad prioriteres foran dagpenges, fordi 98 prosent av de som mottar individstønad er på tiltak (i aktivitet), mens bare 11

² Et eksempel er personer med opptjent rett på sykepenges som midlertidig har vært ute av inntektsgivende arbeid i mindre enn 1 måned (Folketrygdloven kapittel 8).

prosent av dagpengemottakerne er på tiltak. Introduksjonsstønad prioriteres lavest i gruppen ”32 Andre arbeidssøkende”, siden dette er en noe mindre arbeidsrettet støtte enn de andre tiltakene/ytelsene fra ARENA. De fleste dagpengemottakerne vil falle inn under arbeidsstyrken i SFP9 som registrert helt ledige eller sysselsatte. Som nevnt i kapittel 3.1 om bruk av ARENA-registeret, omklassifiseres noen helt ledige til sysselsatte i SFP9 fordi vi tror mer på dateringene i NAVs Aa-register/ordningen. De relativt få dagpengemottakerne som plasseres utenfor arbeidsstyrken på SFP9-statusen, er delvis ledige/arbeidssøkende på dagpenger i ARENA som ikke regnes som helt ledige eller sysselsatte. I lys av NAV's definisjon av delvis ledige kan det tenkes tilfeller av delvis ledige som ikke er sysselsatte i referanseuken, og også Folketrygdlovens kapittel om dagpenger tyder på at noen grupper kan ha rett på dagpenger selv om de er utenfor arbeidsstyrken.³ Det kan også være et lite innslag av feilklassifiseringer i datagrunnlaget (f.eks. delvis ledige som skulle vært klassifisert som sysselsatte, eller andre typer feil/forsinkelser ved registreringen i kildene).

Innenfor den lavest rangerte hovedgruppa, ”64 Andre grupper”, tar prioriteringen hensyn til ytelsenes potensielle varighet. Kontantstøtte har i utgangspunktet kortest varighet og prioriteres derfor først, foran enslig forsørgerstønad som kan mottas over noe lengre tid, og sosialhjelp som potensielt kan mottas over lang tid. Begrunnelsen for å rangere ”6432 Nedsatt arbeidsevne, ikke på tiltak” nederst, kommer vi tilbake til i et senere avsnitt om kvalitetshensyn.

Generelt er det slik at jo høyere en ytelse/tilstand prioriteres, jo mer vil den overskygge de andre, siden det bare skal synliggjøres inntil to aktiviteter/ytelser per person ved plassering på én status (8 siffer). Det er årsaken til at vi prioriterer sysselsetting og ledighet høyest, siden vi ønsker å synliggjøre om personene er i arbeidsstyrken. Det samme gjelder prioriteringen av nærhet til arbeidsstyrken, som betyr at utdanning prioriteres foran helserelaterte ytelser, helserelaterte ytelser foran pensjoner, og tidsbegrensede ytelser foran mer varige. I den grad ulike ytelser rent juridisk eller i praksis ikke kan kombineres med hverandre, vil rangeringen dem i mellom ikke bety noe for hva som blir synliggjort.

Hensyn til datakvalitet

Det ligger også **kvalitetshensyn** bak prioriteringsrekkefølgen. En type kvalitetsproblem kan være manglende datering til referanseuken. Kilden om hvem som er i arbeidsstyrken fra oppdatert reg_total, er generelt sett godt datert til referanseuken, med to unntak. Det gjelder for det første en liten og spesiell gruppe lønnstakere som hadde LTO-registeret som eneste kilde (årsfil) i årene før 2015, og som derfor er prioritert lavest innenfor hovedgruppa ”08 Lønnstakere” i disse årene. Det er også noen andre små grupper av lønnstakere fra LTO-registeret hvor man finner kobling i andre lønnsregistre som er datert til 1. oktober, men for disse kan man anta at de fleste fortsatt var aktive i referanseuken. Før overgang til ordningen i 2015 ble bl.a. størrelse på årslønn lagt til grunn for om de ble

³ Ifølge NAV omfatter delvis ledige alle som søker inntektsgivende arbeid ved NAV og har hatt inntektsgivende arbeid i løpet av de to siste ukene, dog mindre enn normal arbeidstid og som søker arbeid med lengre arbeidstid og er tilgjengelig for det arbeid som søkes (<http://www.nav.no/Om+NAV/Tall+og+analyse/Arbeidsmarked/Delvis+ledige>). I lys av denne definisjonen, kan det forekomme noen delvis ledige i ARENA som ikke regnes som sysselsatte i SFP9. I paragraf 4-5 i Folketrygdloven heter det at ”... Dersom alder eller tungtveiende sosiale hensyn knyttet til helse, omsorgsansvar for små barn eller pleietrengende i nær familie tilsier det, kan medlemmet regnes som reell arbeidssøker selv om vedkommende søker deltidsarbeid eller bare søker arbeid innenfor et begrenset geografisk område...” (https://lovdata.no/dokument/NL/lov/1997-02-28-19/KAPITTEL_4-1#KAPITTEL_4-1). I paragraf 4-9 heter det videre at ”... Det kan ytes dagpenger når medlemmet har vært arbeidsløs og stått tilmeldt Arbeids- og velferdsetaten som reell arbeidssøker (se § 4-5) i minst tre av de siste femten dager. Lørdag og søndag regnes ikke med...”. Personer under utdanning skal normalt ikke kunne få dagpenger, men ifølge paragraf 4-6 kan det gjøres spesielle unntak. Dette kan være forklaringer på hvorfor noen dagpengemottakere er utenfor arbeidsstyrken i SFP9. Permitterte på dagpenger, som omtales i paragraf 4-7, kan også i noen tilfeller være utenfor arbeidsstyrken i SFP9.

klassifisert som sysselsatte i referanseuken. Nivået på antall lønnstakere ble tilpasset nivået i AKU. Fra og med 2015 brukes ikke denne kategorien lenger, siden informasjon om lønnstakere hentes fra a-ordningen.

For det andre er selvstendige hentet fra en årsfil basert på skattemeldinger i statistikkåret. Størrelsen for denne gruppen estimeres basert på blant annet næringsinntekt og tilpasses nivået i AKU til og med årgang 2014.

For lønnstakere basert på kun LTO og selvstendige, gir kodelista for én status ikke angivelse av avtalt arbeidstid, fordi SSB mangler opplysninger om arbeidstid for disse personene.

En annen ytelse som er basert på en årsfil, er samlekategorien som vi kaller ”5640 Pensjoner fra andre enn NAV/Folketrygden”. Denne ytelsen er rangert lavest blant pensjoner dels på grunn av manglende datering, men også fordi dette i mange tilfeller kan dreie seg om pensjoner som man mottar på toppen av andre pensjoner, og derfor kan prioriteres lavt. Siden mottak av slike pensjoner vanligvis innebærer en varig tilstand for de fleste personer, vil problemet for datakvaliteten særlig gjelde de personene som begynner å motta slike pensjoner først etter referanseuken, dvs. i løpet av årets siste 5-6 uker.

En tredje ytelse det er dateringsproblemer ved, er opplysningene om igangværende utdanning som gjelder pr. 1. oktober, men de fleste studier varer over hele høstsemesteret. Alle de øvrige ytelsene i SFP9 kan dateres til referanseuken eller til november hvis data bare kan dateres til måneden.

Når det gjelder aktiviteten ”6432 Nedsatt arbeidsevne, ikke på tiltak” blant arbeidssøkere fra ARENA, vil de fleste av disse i praksis være mottakere av en ytelse. For at ikke denne vage aktivitetsangivelsen skal overskygge ytelsen(e) de mottar, har vi valgt å prioritere den helt til slutt i SFP9, etter sosialhjelp, slik at vi heller får synliggjort hvilke ytelser disse personene mottar (f.eks. arbeidsavklaringspenger).

Utover disse kvalitetshensynene, gjøres det ikke flere kvalitetsvurderinger ved beregning av én status i SFP9. SFP9 drar imidlertid nytte av de mange konsistensbehandlingene som gjøres i produksjonen av grunnfilen `reg_total` ved klassifisering av personer som sysselsatte eller registrert helt arbeidsledige.

Vi har valgt å ikke ta hensyn til beløpsstørrelser på ytelsene ved beregning av én status i SFP9. Det ville vanskeliggjøre formidlingen om man la inn denne dimensjonen i beregningen i tillegg til den grunnleggende arbeidsmarkedsfaglige prioriteringsrekkefølgen mellom ytelsene. Det er også slik at det varierer mellom kilderegistrene hvorvidt det finnes opplysninger om beløpsstørrelse eller ikke, og de beløpsvariablene som finnes er til dels nokså forskjellige typer variabler som det ville være krevende rent metodisk å sammenstille (ulike målestørrelser og telleperioder). I inntektsstatistikken til SSB kan man finne informasjon om hvor stor andel av personens samlede inntekt kommer fra yrkesinntekt, kapitalinntekt eller ulike overføringer (for eksempel, pensjon eller kontantstøtte).

Vi legger heller ikke opp til å vaske dataene for ulovlige eller ulogiske kombinasjoner utover den konsistensbehandling som skjer i `reg_total` ved plassering av personer som lønnstakere, selvstendige, registrert helt arbeidsledige eller utenfor arbeidsstyrken. Vi ser ikke grunn til å skjule eventuelle tilfeller av juks eller saksbehandlingsfeil ved utbetaling av ytelser, og det ville også være vanskelig å vurdere hvorvidt en kombinasjon er ulovlig. De fleste

aktiviteter/ytelser kan kombineres delvis (f.eks. 50 prosent ufør og 50 prosent jobb). Ulogiske kombinasjoner kan også skyldes feil eller forsinkelser i registrene.

5. Datakvalitet

Datakvaliteten i grunnfilen for de registerbaserte sysselsettingsstatistikkene (reg_total) omtales i et eget notat, men vi vil trekke frem noen punkter derfra som er spesielt relevante for SFP. Vi vil også redegjøre for kvalitetsproblemer ved kildene om ytelser og igangværende utdanning, samt ved metoden for beregning av status i SFP9.

Generelt regnes datakvaliteten på kildene som benyttes i SFP som god.

5.1. Kontroll og revisjon

Opplysningene om hvem som er i arbeidsstyrken, dvs. sysselsatte (lønnstakere og selvstendige) eller registrert helt arbeidsledige, kontrolleres rutinemessig gjennom produksjonen av grunnfilen for de registerbaserte sysselsettingsstatistikkene (reg_total). Der gjøres det også kontroll av resten av variabellisten og populasjonsavgrensningen i reg_total. SFP9 drar nytte av disse kontrollene når det hentes opplysninger fra den oppdaterte versjonen av reg_total med selvstendige fra statistikkåret. Fra og med 2015 er reg_total basert på ny datakilde. Hovedkilden fram til og med 2014 var NAVs Arbeidstakerregister (Aa-registeret) og Skatteetatens Lønns- og trekkoppgaverregister (LTO-registeret). I 2015 ble rapporteringen til dette registeret samordnet med rapportering av lønns- og personelldata til Skatteetaten og SSB. Det nye felles rapporteringssystemet kalles a-ordningen. A-ordningen gir generelt et bedre datagrunnlag ved at det er mer korrekt på individnivå, samt at det dekker flere lønnstakerforhold enn Aa-registeret. Samtidig innebærer denne overgangen at tallene for sysselsatte er ikke sammenliknbare i årene før og etter 2015. I tillegg ble noen av variabelnavnene i SFP endret for å samsvare med reg_total etter overgang til a-ordningen.

Ved produksjon av nye årganger av SFP9 gjøres det rimelighetskontroller av nivåtallet over mottakere av de enkelte stønadene i SFP9 sammenliknet med NAVs egne statistikker over antall mottakere. Selv når datakilden og populasjonen er den samme, kan det allikevel være visse avvik mellom de to statistikkgrunnlagene som følge av avvikende referanseperioder eller uttakstidspunkter fra registrene. Nivåttall fra NAVs egen statistikk er bl.a. publisert på følgende nettsider hos NAV:

- NAV-statistikk om mottakere av stønader fra NAV
- NAV-statistikk om mottakere av arbeidsavklaringspenger i november (f.o.m. 2010)
- NAV-statistikk om personer med nedsatt arbeidsevne og mottakere av uføretrygd
- NAV-statistikk om mottakere av pensjoner

En annen type kontroll som gjøres ved produksjon av ny årgang av SFP9, er at nivåtallene for alle fulltellingsvariablene og variabelen én status, sammenliknes med forrige årgang av SFP9. Ved å se nærmere på *endringstallene*, kan eventuelle mangler i kilderegistrene eller andre typer feil avdekkes.

5.2. Frafall av personer og opplysninger

Frafall av personer

Det oppnås totaltelling av bosatte personer i referanseuken gjennom kontroller mot Det sentrale folkeregisteret i produksjonen av grunnfilen reg_total. Blant ikke-bosatte derimot, vil SFP9 mangle tre relevante grupper, nemlig ikke-bosatte lønnstakere som er 75 år eller eldre, ikke-bosatte som er selvstendig nærings-

drivende, og ikke bosatte sjøfolk som er bosatt utenfor EØS. Dette handler egentlig ikke om frafall, siden disse gruppene er holdt utenfor populasjonsrammen for SFP9, som redegjort for i kapittel 2 om populasjonsavgrensningen. Noen av disse ikke-bosatte personene kan allikevel komme med i SFP9 fra kilderegistrene om offentlige ytelser eller igangværende utdanning.

Frafall av opplysninger I registerdata kan det forekomme frafall av enkeltopplysninger. Personkjennetegn vil sjelden mangle siden dette er opplysninger alle skal ha i grunnregistre som f.eks. Det sentrale folkeregisteret.

Når det gjelder statusopplysninger, kan det forekomme manglende (eller forsinkede) inn-/utmeldinger av lønnstakerforhold fra arbeidsgivere til NAVs Aa-register, som var hovedkilden til og med 2014. Omfanget av dette problemet reduseres i grunnfilen reg_total ved at man der venter til ca. 4 måneder etter referanseuken med å hente ut data om referanseuken fra Aa-registeret. Likedan kan det være manglende eller forsinkede inn-/avregistreringer av stønadsmottakere i registrene over ytelser, men siden dette handler om utbetalinger av penger, er det grunn til å tro at kvaliteten på registre over stønader generelt er god.

Selv om kvalitet generelt anses som god i a-ordningen er det likevel noen utfordringer. Det er dels knyttet til at det særlig rapporteres for mange arbeidsforhold som SSB ikke regner som aktive og dels knyttet til manglende rapportering av arbeidstid. Den første utfordringen omtales nedenfor, mens den andre utfordringen omtales i kapittel 5.3.

Det rapporteres inn mange arbeidsforhold i a-ordningen som etter SSBs vurdering ikke er «aktive» i statistikkperioden. Dette vil kunne gjelde sesongarbeidere som ikke har utført arbeid i referansemåneden (og derfor ikke har fått lønn), der det (feilaktig) ikke er meldt sluttdato på opphørte arbeidsforhold, og tilfeller der personer feilaktig har blitt rapportert med aktivt arbeidsforhold. Sistnevnte gjelder f.eks. personer på vikarlistor som ikke har utført arbeid i perioden. Hovedkriteriet for å bestemme om et arbeidsforhold skal regnes som aktivt eller ikke er at man har et arbeidsforhold i referanseuka. I tillegg benyttes opplysninger om lønn.

Klassifiseringen skjer på følgende måte:

1. Startdato og eventuell sluttdato for arbeidsforholdet tilsier at man har et arbeidsforhold i referanseuka (som er uka med den 16. i)
2. Informasjon om lønn og type lønnsutbetaling avgjør om arbeidsforholdet skal regnes som sysselsetting og om det er aktivt i statistikkperioden. Størrelsen på lønnen har ingen betydning.
3. Personer som ikke har mottatt lønn, men som er midlertidig fraværende pga. sykdom, ferie, lønnet permisjon e.l. skal også regnes som sysselsatte (jmf. ILO-definisjon beskrevet lenger ned). For å identifisere dette benyttes informasjon om NAV-ytelser som erstatter lønn (sykepenger, svangerskapsenger eller foreldrepenger) og permisjoner/mitteringer. Permisjoner/mitteringer uten lønn på under 90 dager regnes som sysselsetting (permisjon/mittering med lønn regnes som sysselsetting uansett lengde).
4. I tillegg sjekkes arbeidsforhold uten lønns- og fraværsopplysning for ulike tidsforsinkelser, f.eks. kan det være en tidsforsinkelse i utbetaling av lønn ved nyoppstartede arbeidsforhold, eller manglende utbetaling av lønn for timelønte ved avvikling av ferie. Derfor inngår også arbeidsforhold med fersk startdato og arbeidsforhold med utbetalt lønn i måneden før og etter statistikkperioden i populasjonen.

Om lag 98 prosent av de innrapporterte arbeidsforholdene blir regnet som aktive på bakgrunn av punkt 1 og 2 over.

Permisjoner og permitteringer over 14 dager skal rapporteres i a-ordningen. I henhold til internasjonale anbefalinger fra Internasjonal Labour Organization (ILO) skal permisjoner/permitteringer under 90 dager regnes som sysselsetting. Permisjoner med lønn regnes som sysselsetting uansett lengde, som nevnt over.

To av ytelsene i SFP9, sosialhjelp og kontantstøtte, gis til husholdningen. Det kan i parforhold derfor være mer eller mindre tilfeldig hvem av ektefellene/samboerne som står registrert som mottaker.

For ytelsen AFP-pensjon, er det mangelfull informasjon om to grupper av nye mottakere for årgangene 2009-2018, nemlig nye statlige AFP-mottakere 65 – 67 år og nye offentlige AFP-mottakere. Hvis disse personene ikke er registrert i andre ytelser eller aktiviteter i SFP9, vil de havne i ukjent status i SFP9. Fra og med 2019 benyttes a-ordningen til å hente inn disse dataene til SFP. Ved å benytte A-ordningen, fanger vi opp flere mottakere av offentlig AFP.

For ytelsen sykepenger, kan arbeidsgivere fremsette krav inntil 3 måneder i etterkant og det kan også ta tid før mottakerne blir registrert i NAVs register. I SFP9 baserer vi oss på data om sykepengemottakere i referanseuken som SSB får levert i februar året etter. Det betyr at vi ikke får med alle etterregistreringer i datagrunnlaget vårt, og dermed blir nivået i SFP9 litt for lavt. Personer som mottar både foreldre- og sykepenger, er i våre data ikke klassifisert som sykepengemottakere.

Når det gjelder virksomhetskjennetegn, kan lønnstakere som kommer fra andre kilder enn Aa-registeret, mangle en del virksomhetskjennetegn. F.eks. mangler personer i førstegangstjeneste og sivilarbeidere opplysninger om både virksomhets- og foretakstilknytning i data fra Vernepliktsverket. Lønnstakerforhold som kom kun fra LTO-registeret (ikke i Aa-registeret) i årene t.o.m.2014 manglet i utgangspunktet opplysninger om hvilken virksomhet de arbeidet i (men hadde foretaksopplysninger). Men for en del av disse jobbene ble organisasjonsnummer virksomhet imputert ved å knytte de til den virksomheten som flest andre ansatte i foretaket med samme bostedskommune var knyttet til. Dersom det ikke fantes andre ansatte i foretaket som har organisasjonsnummer virksomhet og bor i samme kommune, klarer vi ikke å imputere organisasjonsnummer virksomhet, men de blir tildelt en næringskode.

Når det gjelder høyeste fullførte utdanningsnivå, mangler det opplysninger for visse grupper, bl.a. for en del innvandrere som er bosatt i Norge etter år 2000 og ikke har fullført noen utdanning i Norge. Det er påkoblet informasjon om beregnet (imputert) utdanningsnivå på SFP-filene tilbake til 2008.

Bosatte personer i Norge som arbeider i et annet land (f.eks. bor i Østfold og arbeider i Sverige), vil inngå i populasjonen i SFP, men de vil ikke være registrert som sysselsatte i SFP. Kun arbeidsforhold med arbeidssted i Norge, fanges opp som sysselsetting i registerbasert sysselsettingsstatistikk. I intervjuundersøkelsen AKU derimot, vil bosatte som arbeider i et annet land, bli klassifisert som sysselsatte. Sammenhengen mellom SFP og AKU drøftes nærmere i kapittel 6.

5.3. Andre feilkilder og usikkerhet

Datering

Som nevnt i kapittel 4.1 om kvalitetshensyn ved prioriteringsrekkefølgen for plassering på én status, er det *dateringsproblemer* ved noen av kildene om status. Data om hvem som er selvstendig næringsdrivende, er hentet fra en årsfil basert på skattemeldinger til Skatteetaten - som ikke kan dateres til referanseuken for SFP9. Med andre ord har vi kun opplysninger om næringsinntekt i løpet av året, og ikke i

referanseuken. For å velge ut hvilke som var aktive i referanseuken tas det hensyn til flere forhold:

- Personer som er registrerte arbeidsledige kan ikke inngå blant selvstendig næringsdrivende, selv om de har næringsinntekt;
- Hvis det viktigste lønnstakerforholdet er et verneplikts- eller siviltjenesteforhold velges dette som viktigste arbeidsforhold;
- Hvis avtalt stillingsprosent tilsvarer en heltidsstilling (avtalt stillingsprosent ≥ 100) velges lønnstakerforholdet som viktigste arbeidsforhold;
- Hvis avtalt stillingsprosent er lavere enn det som tilsvarer heltid (avtalt stillingsprosent < 100), men sum kontantlønn $>$ sum næringsinntekt velges lønnstakerforholdet som viktigste arbeidsforhold;
- For andre kombinasjoner av lønnstaker- og selvstendigforhold, velges selvstendigforholdet som viktigste arbeidsforhold.

Det kan altså forekomme feil på individnivå og i små grupper når det gjelder hvem som er klassifisert som selvstendige blant bosatte. Som nevnt i forrige avsnitt, mangler SFP9 opplysninger om hvem som er selvstendige blant ikke-bosatte.

For mer informasjon om valg av selvstendig næringsdrivende i årene før 2015, se dokumentasjonsnotat for registerbasert sysselsettingsstatistikk.

For personer som er definert som sysselsatte og lønnstakere kun på grunnlag av opplysninger fra lønns- og trekkoppgaveregisteret i årene før 2015 (noen få prosent av lønnstakerne), er arbeidsforholdet ikke datert fordi LTO-registeret er en årsfil. Kontantlønnens størrelse legges til grunn for om personen anses som sysselsatt. Det er dermed knyttet en viss usikkerhet om disse faktisk var sysselsatt i referanseuken. Slike arbeidsforhold er derfor skilt ut med egen kode ved plassering på status i SFP9, slik at man i analyser kan ha mulighet for å se om de skiller seg vesentlig fra andre og hvor dette kan skyldes dårlig kvalitet på datering. Andelen slike arbeidsforhold av alle lønnstakere lå på 7,6 prosent for bosatte og 5,3 prosent for ikke-bosatte i 2009.

Som nevnt i kapittel 3.1 produseres det fra og med årgang 2019 en foreløpig SFP fil for å forbedre aktualiteten i statistikken «Tilknytning til arbeid, utdanning og velferdsordninger». Denne filen er basert på fjorårets informasjon om selvstendig næringsdrivende. Som følge av dette vil noen personer bli feilklassifisert som sysselsatte selv om de er utenfor arbeidsstyrken, og motsatt. En sammenlikning for årgang 2017 viser at bruk av fjorårets selvstendige slår mest ut på gruppene «5632 Alderspensjon» og «7208 Ukjent status». Man får om lag tre tusen færre alderspensjonister og om lag fem tusen flere personer med ukjent status ved bruk av 2016-opplysninger om selvstendige.

Virksomhetstilknytning

En annen type feilkilde gjelder *opplysningene om virksomhetstilknytning og tilhørende virksomhetskjennetegn*. For noen store foretak med mange virksomheter under seg, kan fordelingen av ansatte på virksomhet i arbeidstakerregisteret til tider være mangelfull. Det kan gi merkbare utslag på virksomhetsnivå, arbeidsstedskommunenivå og andre berørte aggregeringsnivåer når slike feil oppstår og når de rettes.

For personer som var definert som lønnstakere kun på grunnlag av lønns- og trekkoppgaveregisteret i årene før 2015, kunne jobben bare knyttes til foretaksnivået. Her er det utarbeidet en rutine for på best mulig måte å tildele organisasjonsnummer virksomhet der det er mulig (se kapittel 5.2).

Avtalt arbeidstid En tredje type feilkilde gjelder *opplysningene om avtalt arbeidstid*. Til og med 2014 ble arbeidstid for de fleste lønnstakerne hentet fra Aa-registeret og av god kvalitet. For to grupper av sysselsatte er arbeidstiden estimert, og det kan innebære visse kvalitetsproblemer på individnivå og i små grupper. Det gjelder selvstendig næringsdrivende og lønnstakere basert på kun LTO-registeret i årene før 2015, hvor arbeidstiden er estimert basert på bl.a. størrelsen på henholdsvis kontantlønnen eller næringsinntekten.

Fra og med 2015 er opplysninger om arbeidstid hentet fra a-ordningen. Avtalt arbeidstid er erstattet med avtalt stillingsprosent, antall timer som utgjør fulltid per uke og betalte timer for timelønne. Avtalt arbeidstid og heltid/deltid utledes fra disse variablene. Vi har bedre informasjon om arbeidstid i a-ordningen. I Aa-registeret var det ingen informasjon om hva som tilsvarte heltid. Normal fulltid i Norge er 37,5 time per uke, men store grupper har et lavere timeantall for fulltid, f.eks. skift- og turnusarbeidere. Vi kunne dermed ikke vite om en oppgitt arbeidstid på f.eks. 33 timer var deltid eller heltid. Den største forbedringen ved overgangen til a-ordningen var at man fikk data om stillingsprosent, som gir mulighet til å skille mellom heltid og deltid. En annen viktig forbedring var at a-ordningen har data om arbeidstid for alle arbeidsforhold, mens vi ikke hadde opplysning om arbeidsforhold for små og kortvarige arbeidsforhold i registrene før 2015.

A-ordningen gir kun informasjon om lønnstakere. For selvstendig næringsdrivende mangler vi informasjon om arbeidstid. Seksjon for arbeidsmarkeds- og lønnsstatistikk har startet arbeidet med å utvikle nytt opplegg for å beregne arbeidstid for selvstendig næringsdrivende.

Selv om informasjonen om arbeidstid i a-ordningen regnes som god er det likevel noen utfordringer. Det går dels på at det ikke rapporteres stillingsprosent for mange timelønne og dels på at det for noen timelønne rapporteres standardverdier på 0 eller 100. Sistnevnte gjør det vanskelig å skille en reell stillingsprosent på 100 fra en feilaktig verdi på 100. SSB har derfor utviklet en metode som gir bedre data-grunnlag for arbeidstid. Metoden er nærmere beskrevet i artikkelen [Metode for bedring av informasjon om arbeidstid i a-ordningen](#).

6. Sammenliknbarhet og sammenheng

6.1. Sammenheng over tid og sted

Versjon 9 av SFP foreligger for årgangene 2008 – 2019 og oppdatert årlig. Siden det i SFP9 er implementert en helt ny metodikk for beregning av én status, vil statusvariabelen i eldre versjon av SFP (SFP3 2004 - 2009) ikke være direkte sammenliknbar med SFP9. SFP9 omfatter også flere typer ytelser og har en videre populasjon enn eldre versjon (ikke-bosatte, 15-åringer og personer over 75 år). For mer informasjon om eldre versjon av SFP, se vedlegg B.

*Brudd i
sysselsettingstallene i 2015*

Tallene om sysselsetting er ikke direkte sammenliknbare i årene før og etter 2015. Dette skyldes overgangen til ny datakilde – a-ordningen. Endringer må tolkes med forsiktighet. Denne endringen er nærmere beskrevet her: <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/ending-i-datagrunnlaget-for-registerbasert-sysselsettingsstatistikk>.

I a-ordningen samles det inn andre opplysninger om arbeidstid enn i Aa-registeret, som omtalt i kapittel 5. Det betyr at arbeidstid fra og med 2015 *ikke* er sammenliknbar med arbeidstid i tidligere årganger.

Arbeidsavklaringspenger og varig uføretrygd

1. mars 2010 ble tre ytelser fra NAV slått sammen til én ytelse kalt Arbeidsavklaringspenger. Dette gjaldt ytelsene atføringspenger, rehabiliteringspenger og tidsbegrensa uførestønad. I kodelista for variabelen én status i SFP9 er denne overgangen håndtert som vist i tabell 4.1.

Etter endringer i uføreordningen fra 1. januar 2015 ble varig uførepensjon erstattet med varig uføretrygd. Begge ytelsene har samme kode for variabelen prioritert arbeidsstyrkestatus i SFP, 4864.

6.2. Sammenheng med annen statistikk

Den totale registerbaserte sysselsettingsstatistikken

Opplysningene om hvem som er sysselsatte (lønnstakere og selvstendig næringsdrivende) eller registrert helt arbeidsledig, er i SFP9 hentet fra den totale registerbaserte sysselsettingsstatistikken (reg_total) som er grunnfilen for de forskjellige registerbaserte sysselsettingsstatistikkenes. Mens de ordinære registerbaserte sysselsettingsstatistikkenes av hensyn til produksjonstiden må bruke gamle data om hvem som er selvstendig næringsdrivende (skattemelding fra året før statistikkåret), baserer endelig versjon av SFP9 seg på en oppdatert reg_total med selvstendige fra statistikkåret. Selv om nivåtallene vil være like i den ordinære og oppdaterte versjonen av reg_total, kan personer bytte status. Foreløpig versjon av SFP9 er i likhet med annen registerbasert statistikk basert på informasjon om selvstendige fra året før statistikkåret.

Arbeidskraftundersøkelsen (AKU)

Nivåtallene på landsnivå når det gjelder sysselsetting samt fordelingen på lønnstakere og selvstendige, vil være likt med AKU for bosatte 15 – 74 i årene før 2015. Dette er som følge av at produksjonen av registerbasert sysselsettingsstatistikk tilpasset seg nivået i AKU før overgang til a-ordningen i 2015. Brutt ned på ulike kjennetegn som kjønn, alder, næring, arbeidssted, og arbeidstid vil det forekomme nivåforskjeller grunnet ulike kilder til dataene. Fra og med 2015 blir ikke nivåtallet for lønnstakere eller selvstendige bestemt av AKU. Man slipper derfor utvalgsusikkerheten som fulgte med AKU-tallene. På den annen side blir nivåtallet på lønnstakere litt for lavt. Fra og med 26. april 2018 ble derimot AKU-tall tilbake til 2006 publisert etter ny estimeringsmetode, noe som gjør at tallene ikke lenger er like på landsnivå. Mer om ny metode i artikkelen Ny metode gir bedre AKU-tall.

SFP9 har en videre populasjon enn AKU (personer 75 år og over, og visse grupper av ikke-bosatte).

AKU har noe flere selvstendige blant bosatte 15 – 74 år enn nivået i reg_total. Det skyldes bl.a. at en del respondenter i AKU svarer at de er selvstendige, selv om de egentlig er ansatt (lønnet av eget AS), og disse blir derfor regnet som lønnstaker og ikke selvstendige i de registerbaserte sysselsettingsstatistikkenes og SFP.

I AKU er det ingen eksplisitte spørsmål om hvorvidt personen mottar ytelser. De som i AKU klassifiseres som arbeidsledige eller utenfor arbeidsstyrken spørres om hva de oppfatter som sin hovedsakelige virksomhet, hvor svaralternativene er bl.a. ”student/skoleelev”, ”hjemmearbeidende”, ”alderspensjonist”, ”førtidspensjonist” og ”arbeidsufør”. De sysselsatte får blant annet status etter om de jobber heltid, lang deltid eller kort deltid.

Det er gjort analyser av samsvaret mellom status og eventuelt hovedsakelig virksomhet i AKU (4. kvartal) og én status i SFP9 (november) blant personer som har svart i AKU. På forhånd kunne en forvente visse avvik, hovedsakelig grunnet ulike referanseperioder (alle ukene i kvartalet i AKU) og ulike definisjoner av arbeidsledige. I AKU blir en person klassifisert som arbeidsledig ut fra kriterier om arbeidssøking og tilgjengelighet, mens tilsvarende status i SFP (registrert helt

arbeidsledig) fastsettes på bakgrunn av registrert helt arbeidsledige hos NAV. En annen grunn til visse avvik mellom AKU og SFP er at registerstatistikken bare vil fange opp sysselsetting blant personer med arbeidssted i Norge. Hvis en person f.eks. bor i Østfold og arbeider i Sverige, så vil han eller hun bli fanget opp som sysselsatt i AKU, mens samme person i registerdata vil fremstå som ikke-sysselsatt. Forskjeller mellom arbeidsledige i register (NAV) og AKU er beskrevet i artikkelen Hvorfor ulike arbeidsledighetstall?

Analysene viser at det store flertallet av AKU-respondentene hadde sammenfallende status i AKU og SFP9 når det gjaldt å være enten i eller utenfor arbeidsstyrken. Blant de som var i arbeidsstyrken både i AKU og SFP var det stor grad av sammenfallende statuser når det gjaldt de sysselsatte, men noen avvik når det gjaldt arbeidsledighet. Dette henger sammen med de ulike definisjonene av arbeidsledige. Blant de som var utenfor arbeidsstyrken både i AKU og SFP9, var det også høy grad av samsvar mellom status i SFP og hva respondentene oppga som hovedsakelig virksomhet i AKU.

Blant den lille gruppen av AKU-respondenter som hadde ukjent status i SFP9, ble om lag halvparten klassifisert som sysselsatte i AKU. Forklaringer på dette avviket kan bl.a. være forskjellen i referanseperiodene eller at disse arbeidsforholdene har vært så små at de ikke har blitt fanget opp i SFP9 via de ulike registrene. Av AKU-respondentene med ukjent status i SFP9 og som var utenfor arbeidsstyrken i AKU, oppga over halvparten hjemmearbeid som hovedsakelig virksomhet i AKU. Personer som bare er hjemmearbeidende, vil naturlig nok få ukjent status i SFP9.

Sammenheng mellom registrerte arbeidsledige hos NAV og registrerte ledige i SFP

Antall registrerte arbeidsledige i SFP er noe lavere enn i statistikken over registrerte ledige som er basert på registrerte arbeidsledige hos NAV. Noen registrerte ledige vil i SFP (som i registerbasert sysselsettingsstatistikk) være omklassifisert til sysselsatte basert på data om lønnstakere fra a-ordningen. I tillegg brukes ulike måleperioder i datagrunnlaget for registrerte ledige (NAV's ARENA-register) og for registerbasert sysselsettingsstatistikk (a-ordningen). I a-ordningen telles lønnstakere per 3. uke i referansemåneden, mens NAV's tall i hovedsak gjelder siste dag i måneden. Tallet i SFP vil derfor være lavere enn i statistikken over registrerte ledige som baseres på NAV's tall.

Andre statistikker om antall ytelsesmottakere

SSB publiserer flere statistikker som viser befolkningens tilknytning til velferdsordninger:

- Statistikken over alderspensjonister gir oversikt over personer som er alderspensjonister i henhold til regelverket for alderspensjon fra Folketrygden.
- Statistikken over uføretrygdede gir en beskrivelse av personer som har fått et positivt vedtak om uføretrygd fra Folketrygden.
- Statistikken om økonomisk sosialhjelp omfatter alle som mottar sosialhjelp i det aktuelle året.

På grunn av blant annet ulike populasjoner og måletidspunkter er ikke tallene i disse statistikkene direkte sammenliknbare. Formålet med statistikkene om alderspensjonister og uføretrygdede er å gi en totaltelling av alle som benytter disse ordningene. Statistikken om uføretrygdede teller derfor alle personer som er registrert med positivt vedtak om uføretrygd i minst en måned i referanseåret, mens statistikken om alderspensjonister teller alle som har hatt utbetalt alderspensjon i minst en måned.

I SFP og i statistikken om tilknytning til arbeid ønsker man imidlertid å synliggjøre at personer kan være registrert med flere aktiviteter samtidig. I 2016 var for

eksempel om lag 16 prosent av alle alderspensjonistene også sysselsatte, mens 19 prosent utføre kombinerte uføretrygd med arbeid.

I statistikken om tilknytning til arbeid sammenstilles alle aktivitetene som hver person er aktiv i på et gitt tidspunkt (i referanseuken, som er 3. uke i november). For personer som er registrert på flere aktiviteter/ytelser samtidig, prioriteres aktivitetene basert på en bestemt rekkefølge.

Les mer om ulike statistikker som viser tilknytning til velferdsordninger her: <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/hva-viser-de-ulike-statistikkene-om-tilknytning-til-velferdsordninger>.

FD-trygd Til forskjell fra SFP9 som er en individdatabase over opplysninger per referanseuken/november, er FD-Trygd en forløpsdatabase over begivenheter gjennom året i den enkelte persons livsløp (fra og med 1992). Disse registreringene kan settes sammen til individuelle hendelses- eller forløpshistorier (forløpsdata), og aggregeres over en gruppe av individer eller en hel populasjon. SFP bruker ofte de samme kilderegistrene (in-put data) som i FD-trygd, men det kan være forskjeller i begreper og måling. For mer informasjon om FD-trygd, se <https://www.ssb.no/omssb/tjenester-og-verktoy/data-til-forskning/fd-trygd> .

Vedlegg A: Variabler og kodelister

Tabell A1. viser overordnet beskrivelse av variablene i SFP.

Tabell A1. Oversikt over variabler i SFP

Type variabel	Forklaring	Eksempler	
Personkjennetegn	Variabler som gir informasjon om personlige kjennetegn - blant annet alder, kjønn, innvandringsbakgrunn, utdanning og opplysninger om barn i familien.	PERS_ALDER	Alder per 16. i statistikkmåned
		PERS_INVKAT	Innvandringskategori
		PERS_BOTID_FORST	Antall år siden første gang bosatt
		PERS_KOMMNR	Bostedskommune
Informasjon om arbeidsforholdet	Variabler som gir informasjon om hovedarbeidsforholdet og arbeidsgiveren - blant annet arbeidstid, lønn, yrke, næring og sektor	ARB_ARBKOMM	Arbeidsstedskommune
		VIRK_ID_SSB	Virksomhetens orgnr eller snr
		PERS_SUM_STILLINGSPST	Samlet stillingsprosent for personen
		ARB_YRKE	Yrkeskode i henhold til arbeidsoppgaver
Aktivitetsvariabler	Dikotomiske (1/0) variabler som angir om personen er aktiv eller ikke i hver enkel ytelse/aktivitet.	AKT_LODELTID AKT_LOHELTID AKT_ALDERSP AKT_ORDUTD	Lønnstaker deltid Lønnstaker heltid Alderspensjon Ordinær utdanning
Prioritert arbeidsstyrkestatus	Beregnes ved å sammenstille aktivitetsvariablene i en bestemt prioriteringsrekkefølge (nærmere beskrevet i kapittel 4.1).	PERS_STATUS8	SFP-status i forh.til arb.markedet, utdanning og off.ytelser, 8-sifra
Informasjon om velferdsordninger	Variabler som gir mer detaljert informasjon om mottak av ulike ytelser - blant annet grad av pensjon, type AFP (offentlig/privat) og uføregrad.	PERS_AS_F PERS_ANT_D_AS PERS_AFP_GRUPPERING	Arbeidssøkerstatus fingergruppe i ARENA Antall dager i arbeidssøkerstatus i ARENA AFP-ordning

Variabelen pers_status8 som viser en prioritert arbeidsstyrkestatus er beskrevet i kapittel 4.1, og tabell 4.1 viser kodelisten for denne variabelen. De øvrige variablene er dokumentert i en Excel-fil som er publisert sammen med dette dokumentasjonsnotatet.

<https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/system-for-persondata>

Tabellene i Excel-vedlegget inneholder følgende informasjon:

- oversikt over variablene som er tilgjengelige på SFP-filene i årene fra og med 2008;
- kort beskrivelse av variablene (inkludert lengde og type variabel- numerisk/karakter);
- kodelister.

Vedlegg B: SFP3 - eldre versjon av SFP

I tillegg til SFP9, er det for årene 2004-2009 tilgjengelig en eldre versjon av SFP som kalles SFP3. Denne versjonen har noen forskjeller fra SFP9 når det gjelder datakilder, definisjoner/måling og populasjonsrammer. Den nyeste versjonen, SFP9 for årgangene fra og med 2008, er utførlig dokumentert i dette notatet. Nedenfor gis en kort presentasjon av SFP3.

SFP3 har en smalere populasjon enn SFP9 (kun bosatte 16-74 år inngår) og skiller seg også fra SFP9 på grunn av noe ulik beregning av sysselsetting og noe ulike referansetidspunkt. Kodelista for variabelen status i SFP9, hvor personer plasseres på én status i forhold til arbeidsmarkedet, utdanning og ytelser, er ikke direkte sammenlignbar med statusvariabelen i SFP3. Det skyldes både at SFP9 har en ny metode for plassering på status (ny prioriteringsrekkefølge og ny kodeliste) og det at SFP9 inkluderer flere ytelser.

I SFP3 er personene knyttet til én enkelt status basert på følgende prioritering:

- sysselsetting,
- sysselsettingstiltak,
- attføringstiltak,
- arbeidsledighet,
- tidsbegrenset uførhet,
- noen spesielle grupper langtidssykemeldte,
- under utdanning,
- varig uførepensjon,
- annen attføring,
- alderspensjon og
- annet

Bakgrunnen for revisjonen i 2008 var at:

1. Nye kilderegistre var tilgjengelige - introduksjonsstønad, kontantstøtte, supplerende stønad, foreløpig uførepensjon i påvente av vedtak, pensjoner fra andre enn NAV/Folketrygden, og en ny versjon av den registerbaserte sysselsettingsstatistikken med utvidet populasjon (reg_total).
2. Man ønsket å utvide populasjonen - inkludere 15-åringene, personer 75 år eller eldre, og ikke-bosatte som er lønnstakere, helt ledige, i utdanning eller på ytelse.
3. Man ønsket en full tilpasning til Den registerbaserte sysselsettingsstatistikken når det gjelder metodikk og nivåfall for arbeidsstyrken. Opplysninger om hvem som er sysselsatte, registrert helt arbeidsledige og utenfor arbeidsstyrken, hentes i SFP9 fra den totale registerbaserte sysselsettingsstatistikken (reg_total). SFP9 bruker her en spesialversjon med selvstendige fra statistikkåret.

Siden SFP9 bruker forskere data om hvem som er selvstendige enn det den ordinære registerbaserte sysselsettingsstatistikken gjør (skattemelding fra statistikkåret), vil det kunne være avvik mellom de to systemene på individnivå. Det var også noen andre svakheter ved kodelista for variabelen én status i SFP3. Kodelista i SFP3 på det mest detaljerte nivået var lite egnet for aggregering opp til hovedgrupper av statuser, det ble synliggjort bare noen få typer av doble aktiviteter, og prioriteringsrekkefølgen hadde visse svakheter sett i forhold til SFP9. Dette er blitt forbedret i den nye versjonen SFP9.

SFP3 (2004 – 2009)

SFP3 har en såkalt ny statuskode (status) hvor opplysninger om sysselsetting hentes fra REGSYS inklusive LTO-arbeidsforhold, og en ”gammel statuskode” (gml_status) som ikke er koplet med REGSYS og ikke tar hensyn til LTO-forhold (gammel status i SFP3 er lik status i SFP2). Det betyr at kun Aa-registeret legges til grunn for identifisering av lønnstakere i gammel status. Både for ny og gammel status er opplysninger om hvem som er selvstendige fra året før, som i REGSYS.

Det er visse avvik mellom ny status i SFP3 og REGSYS i hvordan sysselsetting og helt arbeidsledige beregnes. Et unntak fra hovedregelen om at opplysninger om sysselsetting i ny status i SFP3 hentes fra REGSYS, er at en del personer som er sysselsatte i REGSYS kodes om til attførings- eller opplæringstiltak i ny status i SFP3. SFP3 har også et høyere antall helt arbeidsledige enn REGSYS. Alle som er helt arbeidsledige i REGSYS får samme status i SFP3 (ny status), men blant de som er utenfor arbeidsstyrken i REGSYS er det noen som er kodet som helt arbeidsledige i SFP3. Dette skyldes først og fremst forskjeller mellom SFP3 og REGSYS i rutinene for konsistensbehandling av Aa-registeret mot ARENA.

Hva er referansetidspunktet for ny status i SFP3? Opplysninger om sysselsetting hentes fra REGSYS, som har uke 47 som referansetidspunkt i årgangene 2006-2007 (uke 44/45 i SFP3 2004 – 2005, og en dag i slutten av november før årgang 2004). Øvrige statusopplysninger i ny status i SFP3 har én dag i slutten av november (21.11.2006) som referansetidspunkt, samt noen opplysninger fra andre tidspunkt i november, pr 4. kvartal eller for hele året.

Referansetidspunktet for aldersopplysningene i SFP3 er per 31. desember, i motsetning til i REGSYS i senere år hvor alder er per referanseuken. Det medfører visse populasjonsmessige avvik mellom SFP3 og Regsys etter at Regsys begynte å basere seg på alder i referanseuken.

Siden utplukking av viktigste arbeidsforhold skjer på ulik måte i variablene ny og gammel status i SFP3, finnes to potensielt ulike variabler for organisasjonsnummer virksomhet i SFP3, et for hver type status.