

Beskrivelse af CALDISS workshopserie: Håndtering af registerdata med SAS og STATA

Titel:

Håndtering af registerdata med SAS og STATA

Placering:

November 2018

Workshop 1 (SAS): 13. november kl. 13-16

Workshop 2 (STATA): 19. november kl. 13-16

Ansvarlig:

Kristian Gade Kjellmann (kgk@adm.aau.dk)

Undervisere:

Kristian Gade Kjellmann (kgk@adm.aau.dk)

Jeevitha Yogachandiran Qvist (jy@dps.aau.dk)

Rolf Lyneborg Lund (rolfl@socsci.aau.dk)

Type og sprog:

Workshopserie

Dansk

Mål:

Workshopserien giver overblik over de forskellige dataformater, som der er på Danmarks Statistiks registre. Formålet er at udvikle softwaretekniske kompetencer til at håndtere typiske, indledende datahåndteringsudfordringer, når man arbejder med registerdata og på tværs af dataformater (sammenkobling/merge af registre, udtræk fra registre, konvertere filformater og typiske variabel-type udfordringer).

Omfang og forventet arbejdsindsats:

Workshop-serien rummer 2 workshops a 3 timer fordelt over 2 dage. Der er ingen forberedelse til workshop-serien.

Deltagerforudsætninger:

Det er en fordel, at man i forvejen er bekendt med at arbejde med registerdata, men det er ikke et krav. Workshop-serien kan *ikke* betragtes som et introduktionskursus i registerdatas indhold, men som en teknisk introduktion af registerdatas formater, samt hvordan man arbejder med disse.

Workshop-serien bruger eksempler fra følgende registre: UDD, DREAM, IDAP, BEF. Kendskab til et eller flere af disse registre er en fordel, men ikke et krav.

Det antages, at deltagere i forvejen er bekendt med en af softwareløsningerne anvendt i workshop-serien (SAS og STATA), og at man kan arbejde enten med editor-filer (SAS) eller do-filer (STATA).

Det antages desuden at deltagere har disse softwareløsninger installeret på egen PC/MAC, da workshop-serien indeholder praktiske øvelser.

Aktiviteter:

De to workshops beskæftiger sig med hver sin softwareløsning. Workshop 1 beskæftiger sig med SAS og workshop 2 beskæftiger sig med STATA.

Selvom de to workshops kan betragtes som individuelt afgrænset, så gennemgår de ikke de samme temaer.

Workshop 1: Registerdataformater og håndtering af registerdata med SAS

v/ Jeevitha Y. Qvist og Kristian G. Kjellmann

Workshoppen gennemgår de forskellige typer af registerdataformater på Danmarks Statistik (forløb, hændelse, årsregistrering) og illustrerer, hvordan de adskiller sig fra hinanden i deres struktur. Hertil gives case-eksempler på typiske udfordringer med at arbejde med de forskellige formater.

Da data på Danmarks Statistik ofte som udgangspunkt er lettest tilgængeligt i SAS-format, fokuserer denne workshop på, hvordan man foretager de indledende udtræk og kørsler i SAS. Workshoppen gennemgår således de typiske overgange i SAS fra rå data til det specifikke datasæt, som man er interesseret.

Med udgangspunkt i en række simulerede datasæt, undervises der i datahåndteringsværktøjer med SAS. Det gennemgås blandt andet, hvordan man sammenkobler/merger data fra forskellige registre ud fra en nøglevariabel (PNR), hvordan man kan sammenkoble data fra forskellige dataformater og kilder (fx surveydata) samt hvordan man laver tilfældige udtræk af data eller udtræk, der møder bestemte kriterier. Workshoppen arbejder med et årsregisterdatasæt i long-format som udgangspunkt.

Viden, færdigheder, kompetencer

- Viden om forskellige typer af registerdataformater på Danmarks Statistik
- Viden om typiske problemstillinger ved at arbejde med specifikke registre eller på tværs af registre
- Viden om forskellige merge-metoder i SAS
- Viden om aggregerede mål ved DST
- Færdigheder inden for håndtering og bearbejdning af registerdata i SAS
- Færdigheder inden for databearbejdning med SAS
- Kompetencer til at indlæse data fra Danmarks Statistik i SAS
- Kompetencer til at merge registerdata i SAS ud fra nøglevariabel
- Kompetencer til at foretage tilfældige udtræk af registerdata med SAS
- Kompetencer til at foretage udtræk af observationer, der møder bestemte kriterier med SQL

Workshop 2: Håndtering af registerdata med STATA

v/ Rolf Lyneborg Lund og Kristian Gade Kjellmann

Workshoppen gennemgår diverse databearbejdningværktøjer i STATA, der finder anvendelse i arbejde med registerdata. Med udgangspunkt i en række simulerede datasæt undervises der i, hvordan man konverterer data fra SAS til STATA med Stat-Transfer, hvad man skal være opmærksom på, når man åbner SAS-data i STATA, hvordan man sammenkobler/merger data fra forskellige registre og kilder ud fra nøglevariabel samt hvordan man løser typiske datamanagementudfordringer, når der arbejdes med datoer og streng-værdier (tekst).

Desuden gennemgår workshoppen, hvordan man kan automatisere processer, der kræver at samle data fra mange forskellige datasæt.

Viden, færdigheder, kompetencer

- Viden om forskellige merge-metoder i STATA
- Færdigheder inden for håndtering og bearbejdning af registerdata med STATA
- Færdigheder inden for håndtering af dato- og strengværdier i STATA
- Kompetencer til at konvertere data fra SAS-format til STATA-format
- Kompetencer til at merge registerdata i STATA ud fra nøglevariabel
- Kompetencer til at konvertere variabelformater i STATA