København, den 26. juni 2018

**Den Danske Værdiundersøgelse 2017: Oversigt over data og dokumentation**

Dette notat redegør kort for datafiler og dokumentation i forbindelse med Den Danske Værdiundersøgelse 2017. Det gøres med henblik på at skabe et overblik hos brugeren over hvordan data er indsamlet, bygget op og hvor yderligere information om data forefindes.

**Population og stikprøver**

I 2017 blev der indsamlet data med de samme spørgsmål fra to stikprøver fra den samme population, den danske befolkning fyldt 18 år. Forskellen bestod i indsamlingsmetode: 1) Den ene dataindsamling var intervieweadministreret og 2) den anden selvadministreret.

Den intervieweradministrerede CAPI (dør-til-dør) indsamling resulterede i 1696 valide besvarelser, som udgjorde en svarprocent på 50. Den selvadministrerede indsamling som var en kombination af primært CAWI (web) og sekundær opfølgning med PAPI (postal) resulterede i 1666 valide svar, som udgjorde en svarprocent på 42.

Nærværende datasæt indeholder besvarelserne fra begge stikprøver, som kan identificeres med variablen 'mode'.

**Versioner af data**

Data indsamlingen fra Den Danske Værdiundersøgelse 2017 udkommer i to versioner: 1) ddv1981\_2017 (og ddv2017 med kun 2017 data), der indeholder de danske data integreret med de tidligere runder tilbage til og med 1981 opbygget af det danske forskerhold der har forestået dataindsamlingen og 2) European Value Studies version, efterbehandlet og standardiseret af det centrale europæiske forskerhold, der integrerer med de øvrige Europæiske landes data og i tid tilbage til 1981.

Altovervejende vil de to versioner være identiske og variabelnavne og kodningen af udfaldsrummet vil kun undtagelsesvis afvige. Hvad de substantielle variable angår vil der være flere i den danske version i form af en række ekstra afledte variable. Det følgende fokuserer på det danske datasæts egenskaber og opbygning.

**Datasættets struktur**

Datasættet består indledningsvis af tekniske variable der angår året for dataindsamlingsrunden, interviewpersonernes identifikationsnumre i de forskellige runder, modus for dataindsamlingsmetoden og endelig vægte for 1981 og 2017 data.

Herefter følger variablene fra spørgeskemaet fra ’v1’ til ’v333b’. Disse variabelnavne vil kunne jævnføres med EVS’ Master Questionnaire (se bilag) med undtagelse af variable som er specifikke for Danmark angivet med postfixet ’\_dk’ (for at jævnføre disse variabel med spørgeskemaet skal spørgsmålsnummeret i label anvendes sammen med det danske spørgeskema (se næste afsnit)).

Sidst følger en række variable som ikke er standard i EVS, men som er resultatet af om- og samkodninger af variable for de forskellige år for enten at gøre disse sammenlignelige eller at udlede nye variable. De er navngivet med substantielle navne som f.eks. ’beskæftigelse’ der angiver respondentens beskæftigelsesmæssige status. Disse er beskrevet i bilaget ’Afledte variable’.

**Variabel labels og noter**

Variablenes labels er gennemgående opbygget efter følgende format: ”[Spørgsmåls # i spørgeskema]. [Spørgsmål/indhold]. [Runder uden data hvis aktuelt]. [Angivelse af om der er noter tilknyttet]”. F.eks. er variablen v275b’s label ”105. Respondentens landsdel (NUTS-3). Ej 1981, 1990 & 1999. noter” som fortæller at spørgsmålet er nummer 105 i spørgeskemaet (identisk for det danske spørgeskema og Master Questionnaire, se bilag), indholdet er respondentens landsdel kodet efter NUTS-3 standarden. Der er ikke data før 2008, og til sidst fremgår det at der er tilknyttet noter med yderligere oplysninger. Variabel og værdi labels ligger som bilag separat i do-filerne ’varlab.do’ og ’vallab.do’.

I en række tilfælde er der tilknyttet noter via STATA-kommandoen ’note’. For at fremkalde noterne skrives kommandoen note efterfulgt af det aktuelle variabelnavne og eksekveres. F.eks. ’note v275b’ der vil fremkalde noten ”I 2008 genereret ud fra postnr som i enkelte tilfælde ikke er entydige ift landsdel. Dvs. at enkelte postnumre overlapper kommunegrænser og i endnu færre tilfælde også landsdelsgrænser. I sådanne tilfælde er det tilfældigt hvilken landsdel postnummeret er tildelt.” Noten indeholder i dette tilfælde detaljer og forbehold vedr. hvordan postnummer er omkodet til landsdele i 2008 for at gøre data i de to runder sammenlignelige. Samtidigt eksemplificerer det hvordan det danske datasæt indeholder data som ikke vil være at finde i EVS versionen, da omkodningen af post nummer til landsdel i 2008 ikke er standard men en ekstra information genereret i det danske datasæt.

Da notefunktionen er specifik for STATA og f.eks. ikke findes i SPSS er en oversigt over labels og noter vedlagt som bilag.

**Kodning af manglende svar**

Manglende svar er kodet ud fra EVS standarden hvor værdier startende med 7 (7, 77, 777, osv) angiver at spørgsmålet er irrelevant for respondenten (f.eks. spørgsmål om partner hvor respondenten ikke har en partner), værdier startende med 8 angiver ’Ved ikke’ svar og værdier startende med 9 angiver ’Uoplyst/nægter at svare’. Dertil kommer koden 444 som angiver at spørgsmålet ikke har været stillet i en eller flere runder, f.eks. 1981.

Værdierne er i udgangspunktet ikke sat missing, men som bilag er vedlagt en STATA do-file der automatisk sætter manglende svar startende med 7 = .a, 8=.b, 9=.c og 444=.d.

**Inklusion af tidligere runder**

Datasættets udgangspunkt er spørgsmål og variable fra 2017. De er i så vidt omfang som muligt suppleret med identiske og sammenlignelige variable fra de tidligere runder, som oftest 2008 runder og ofte også længere tilbage. I få tilfælde er spørgsmål der var udgået i 2008 blevet genintroduceret således at tidsserien brydes i 2008. Disse variable er naturligvis også medtaget.

**Stikprøve data**

Med henblik på at give mulighed for at foretage egne bortfaldsanalyser og repræsentativitetsanalyser findes i tilgift til hoveddatasættet et datasæt med stikprøver og metodiske variable.

Det indeholder de to 2017 stikprøver for hhv. dataindsamlingen via intervieweradministrerede dør-til-dør interviews og selvadministrerede web/postal spørgeskemaer beskrevet på baggrundsvariablene køn, alder, landsdel, uddannelse, indkomst samt variabel for deres svar status. Baggrundsvariablene går igen i hoveddatasættet. Individnøglen der gør kobling med svardatasættet muligt er resp\_id1 med label ”Respondentnummer”.

*Jonas Toubøl*

Videnskabelig Assistent, ph.d.

Institut for Sociologi og Socialt Arbejde

Aalborg Universitet, København

**Bilagsoversigt**

EVS2017\_Master\_Questionnaire.pdf

DDV2017\_Spørgeskema.pdf

DDV2017\_Forevisningskort.pdf

DDV2017\_Labels\_noter.xlsx

DDV2017\_Afledte\_variable.pdf

ddv1981\_2017.dta

ddv2017.dta

ddv2017\_stikprøver.dta

ddv1981\_2017missing.do

ddv2017\_varlab.do

ddv2017\_vallab.do