Stata 14 og ældre datasæt, do-files mm.

Stata kan som standard læse datasæt, do-files og andre Stata-filer selvom de er oprettet i en tidligere version af Stata. Med Stata 14 er der dog potentielt en udfordring med ældre filer, såfremt de indeholder såkaldte udvidede ASCII-karakterer. Æ, Ø og Å er eksempler på udvidede ASCII-karakterer, men eksemplerne indbefatter også á eller €. Åbnes en ældre fil indeholdende eksempelvis en variable med ø i navnet i Stata 14 vil ø'et blive erstattet med andre karakterer, hvilket naturligvis er uheldigt. Den tekniske forklaring på dette er, at med overgangen fra Stata 13 til 14 har Stata skiftet fra at bruge ASCII-karakterkoder til unicode. Det er en stor fordel, da man nu uden videre kan anvende praktisk talt alle karakterer i Stata, og man eksempelvis have variable der starter med æ, ø eller å. Udfordringer er så bare gamle filer indeholdende udvidede ASCII-karakterer.
Det er heldigvis nemt at rette op på. Nedenstående procedure undersøger først alle Stata-filer (typisk datasæt og do-filer) i en given mappe og rapporterer hvilke filer, der indeholder udvidede ASCII-karakterer. Herefter erstattes de udvidede ASCII-karakterer med unicode-karakterer, og filen kan herefter anvendes uden problemer. Stata gemme tillige en kopi af den oprindelige fil.

clear
cd ***"sti_til_relevant_mappe"***
unicode encoding set ***Windows-1252*** **//eller** ***latin1***
unicode analyze *
unicode translate *

Proceduren fordrer, at man ikke har et åbent datasæt, hvorfor proceduren stater med **clear**, der lukker eventuelle åbne datasæt (uden at gemme). Proceduren fordrer også, at den/de filer man skal have udbedret befinder sig i Statas aktive mappe, og man kan derfor have brug for at ændre Statas aktive mappe. Dette gøres med kommandoen **cd**, der også har en dialogboks. Herefter skal man specificere hvilken ASCII-kodning, der er anvendt i det oprindelige datasæt. For datasæt med latinske karakterer vil det oftest være *Windows-1252* eller *latin1* – se evt. http://www.stata.com/manuals14/dunicodeencoding.pdf. Næstsidste skridt kan i princippet springes over, da det alene undersøger hvilke Stata-filer i den aktive mappe, der indeholder udvidede ASCII-karakterer. Kommandoen resulterer i et output, hvoraf problematiske filer fremgår. Sidste skridt erstatter ASCII-kodningen med unicode i de filer, der indeholder udvidede ASCII-karakterer. Alle berørte filer er gemt når proceduren afsluttes, og man kan nu åbne og arbejde med dem. De oprindelige filer er placeret i en undermappe med navnet **bak.stunicode**.