Training with STATA

1. initial operations: open and explore datasets and variables

2. the importance of using .do files

3. identifying id keys

4. descriptive statistics (TAB, TABLE, SUM) and graphs

    (a) (use sect1.dta file) tabulate (TAB) individual religion and individual religion by zone

    (b) histogram of individual religion with lables / histogram of individual religion by zone with labels (HISTOGRAM)

    (c) proportion of females by religion and zone (be careful with gender coding) (GEN, RECODE, TABLE)

    (d) generate new variables: generate a variable religion, where you recode "others" as missing (GEN)

    (e) variable labels and value labels for religion (LABEL VAR, LABEL DEFINE, LABEL VALUES)

    (f) generate new variables: generate age squared (GEN)

    (g) compute average age in household (SORT, BY, EGEN)

    (h) scatter household heads' age against their spouse age (SCATTER)

    (i) summarize age and spouse age (SUM)

    (j) (use sect11a1.dta) to each household, compute the total amount of land owned and the share of land that each plot represents. Check for outliers (BY, EGEN GEN)

5. merge files (MERGE)

    (a) merge sec1.dta file with sec2.dta (which contains education data)

6. collapse data (COLLAPSE)

    (a) get only one observation per household reporting hh indentifier and the proportion of household members who attended some school

7. regress the total amount of land owned by a family to the number of household members by gender and the proportion of household members with at least some schooling by gender. Control also for zone fixed effects (MERGE, COLLAPSE, EGEN, TAB+GEN, REG)