

## Help on Copy Mean method

Copy Mean imputes missing values in Longitudinal (a panel of cross section and time series) data thus, it takes into account other series too (multivariate). It fulfills the missing values (NA) by a combination of linear interpolation (see UniLinear entry in the Univariate Imputation Menu) and imputation using the population's mean trajectory, as follows.

Let  $Y_{it}$  be the missing data in a trajectory where  $i$  refers to a series in the dataset ( $i^{th}$  column) and  $t$  reflect the date (time). Additionally,  $Y_{ia}$  and  $Y_{ib}$  are the closest preceding and following non-missing values, respectively ( $a < t < b$ ).

Then, using linear interpolation (as in UniLinear), we get the value obtained by applying a linear interpolation between  $a$  and  $b$  on the mean trajectory:

$$\overline{Y_{it}^{LI}} = \overline{Y_{ia}} + \frac{t - a}{b - a}(\overline{Y_{ib}} - \overline{Y_{ia}})$$

Also let  $\overline{Y} = (\overline{Y_{i1}}, \dots, \overline{Y_{it}})$  denote the mean trajectory of the population. Then the average variation  $AV_{it}$  at time  $t$  is the difference between the mean and the linear imputed value i.e.,

$$AV_{it} = \overline{Y_{it}} - \overline{Y_{it}^{LI}}$$

Finally, the Copy Mean imputes  $Y_{it}$  by adding the average variation  $AV_{it}$  to the result of the linear interpolation,  $\overline{Y_{it}^{LI}}$  such that:

$$\overline{Y_{it}^{CM}} = \overline{Y_{it}^{LI}} + AV_{it}$$

Screen example is depicted below. In all imputation figures, the non-NA observations are the colored lines while the imputed values ( $Y_{it}^*$ ) are the colored dots.

