

Stephen P. Jenkins  
(ISER, University of Essex)

Handout for NASUG, July 2006

### Stata programs for analysis of distributions

| Task  | Program   | Source |
|---|---|--------|
| <b>1. Describing distributions</b>  |   |        |
| Simple summary statistics, e.g. mean, median, percentiles, $p_{90}/p_{10}$ , CV   | <code>summarize</code> , <code>count</code>   |        |
| Quantile groups   | <code>xtile</code> , <code>_pctile</code>   |        |
| Quantile group shares, Lorenz and generalized Lorenz ordinates, quantile group membership                                       | <code>sumdist</code>  | ssc    |
| <b>2. Picturing distributions</b>   |   |        |
| Density estimation  | <code>histogram</code> , <code>kdensity</code>  |        |
| Cumulative distribution (Pen's Parade)  | <code>cumul</code>  |        |
| Lorenz and generalized Lorenz curves, TIP curves, etc.  | <code>glcurve</code>  | sj-4-4 |
| <b>3. Inequality and poverty indices</b>  |   |        |
| Inequality indices: $GE(\alpha)$ , $A(\epsilon)$ , Gini, with optional decompositions by population subgroup                    | <code>ineqdeco</code> , <code>ineqdec0</code>   | ssc    |
| Poverty indices: $FGT(\alpha)$ for $\alpha = 0, 1, 2$ , with optional decompositions by population subgroup                     | <code>povdeco</code>  | ssc    |
| <b>4. Variance estimation and statistical inference</b>   |   |        |
| Telling Stata about sample design   | <code>svyset</code> ,<br>version 8: <code>svyset</code>   |        |
| Variance estimation: $GE(\alpha)$ , $A(\epsilon)$   | <code>svygei</code> , <code>svyatk</code>   | ssc    |
| Variance estimation: Gini, income shares, Lorenz ordinates  | <code>svylorenz</code>  | ssc    |
| Variance estimation: total, mean, proportion (can be used for variance estimation of poverty indices with a fixed poverty line) | <code>svy: total</code> , <code>svy: proportion</code> , <code>svy: mean</code><br>(In Stata 8: <code>svytotal</code> , <code>svyprop</code> , <code>svymean</code> ) |        |
| Bootstrapping   | <code>bootstrap</code>  |        |

Programs without an entry in the 'Source' column are built-in to Stata.

The programs marked 'ssc' were written by SPJ (`ssc describe program_name`), with the exception of `glcurve` (SPJ and Philippe Van Kerm; latest version downloadable as Software Update in SJ-4-4), and `svygei` and `svyatk` by SPJ and Martin Biewen (`ssc describe svygei_avyatk`, `replace`).

This is not an exhaustive list of programs! Also see, *inter alia*, `inequal7` by Philippe Van Kerm for calculation of a different portfolio of inequality indices, and `kdens` by Ben Jann for kernel density estimation (both available on SSC).