Quality/ Confidentiality for Tabular Data--Larry Cox

Local data quality: characteristics of one or a few cells
- cell values (income; number of employees)
- time trends of cell values
- computed quantities (income per employee)

Preserving local quality: preserve local characteristics
subject to confidentiality demands, or, a balance

Global data quality: characteristics of data set or subsets
- distributional parameters such as mean, variance
- statistics such as quantiles
- distributional shape

Preserving global quality: preserve global characteristics
subject to confidentiality demands, or, a balance

Local and global quality **overlap**
- measurement error
- correlation
- rank order statistics
- time trends
- analyzability

Univariate setting is familiar
Multivariate setting
-preserve univariate quality for each variable
-preserve relationships between variables
- covariance
- regressions

Quality characteristics of familiar SDL methods

Complementary cell suppression
- preserves local quality for unsuppressed cells
- not for suppressed cells
- inhibits analyzability
- pokes holes in the distribution
- multivariate: problems magnified significantly

Random rounding and perturbation
- can preserve local and global quality well for
    univariate and multivariate data

Controlled tabular adjustment
- can preserve local and global quality well for
    univariate and multivariate data
- QP-CTA and MDI-CTA: nearly opposite quality
    strengths and limitations

Perturbing underlying microdata
- insufficient information to judge