

Data Craft Exercises

John Fox

FIOCRUZ November 2009

The data file <http://socserv.socsci.mcmaster.ca/jfox/Books/Applied-Regression-2E/datasets/UnitedNations.txt> contains data on social indicators for the countries of the world published in 1998 by the United Nations. See the file <http://socserv.socsci.mcmaster.ca/jfox/Books/Applied-Regression-2E/datasets/UnitedNations.pdf> for a description of the variables in the data set.

1. Examine the distributions of total fertility rate, GDP per capita, female illiteracy, and contraception. Characterize the distribution of each variable in terms of symmetry or skewness; non-normality or apparent normality; number of modes; and the presence or absence of unusual values.
2. If a distribution is not symmetric, attempt to make it symmetric by an appropriate transformation on the ladder of powers and roots. If a variable is a proportion, consider using a logit transformation.
3. Examine and characterize the relationship between the total fertility rate and each of GDP per capita, female illiteracy, contraception, and region.
4. Consider the relationships between the total fertility rate and each of GDP per capita, female illiteracy, and contraception. In each case, if the relationship is nonlinear, attempt to make the relationship linear by transforming one or both variables, or explain why a transformation to linearity is not an appropriate strategy.
5. Consider the relationship between the total fertility rate and region. If the spread of total fertility is not constant, attempt to make it constant by a transformation or explain why a transformation is inappropriate.