Assignment 1: HIV prevalence

Sample

The sample is from the GapMinder Data, the largest observational study for understanding of statistics about social, economic, and environmental development at local, national, and global levels.
The sample includes annual estimates for prevalence of HIV infection (hivrate quantitative response variable) among individual persons 15-49 years old and total estimated numbers of infected, all ages per country. The sample includes estimates of total urban population (urbanrate quantitative explanatory variable) for each country compiled by national statistical offices. Number of countries (current sample size equals to N=214) is considered where the number of inhabitants is more than 100,000.
My hypothesis is that there is a strong correlation between the HIV rate and urban rate for 214 countries considered in my sample.

Data Collection Procedure

Data for HIV rate were collected by data reporting procedure in countries with different income. In high-income countries, national case reporting is combined with variable extent of HIV testing in risk groups. In most medium and low-income countries, HIV prevalence is monitored using antenatal care surveys and demographic health surveys. Data for most countries (N =214) are estimates obtained from UNAIDS/WHO in 2009. Data for urban rate were calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects in 2008.

Measures

The measure of HIV prevalence (hivrate – quantitative response variable) was drawn from country level surveillance data compiled by the UNAIDS organization in their database (http://www.unaids.org/en/dataanalysis), and made available for download through the Gapminder web site (www.gapminder.org). It measures the estimated HIV prevalence in percentage (scale is in percent 0-100%) for number of people living with HIV per 100 population of age group 15-49 for each country in 2009. For the current study, the quantitative response variable will be used as it is (to study linear and multiple regressions) or converted to the categorical response variable (to study the logistic regression).

The measure of urban rate (urbanrate – quantitative explanatory variable) was drawn from country level surveillance data compiled by national statistical offices (calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects), and made available for download through the Gapminder web site (www.gapminder.org). It measures the percentage of total urban population (scale is in percent 0-100%) for each country in 2008. Since my explanatory variable is quantitative, I centered it to its mean value.