

Nepal, much like hundreds of countries around the world, is working to address poverty throughout its seventy-five districts. Using data from the GeoDa data center website, this study compares the model results of multilinear regression and geographically weighted regression analyses on the relationships of seven poverty indicator variables (per capita income, percentage of malnourished children under age five, percentage of the population without safe water, overall population, number of children enrolled in grades 1-5, adult illiteracy rate, and deprivation in economic provisioning) with the poverty index. The data used in this project were downloaded from the GeoDa data center, which compiled information from AidData and Open Nepal. The data include a vector polygon shapefile of all seventy-five districts in Nepal and twenty-six attribute variables.

Using the statistical software package RStudio, a pairwise scatterplot of the seven variables indicated correlations between several population variables, which prompted tests of linear regression for each variable. Adult illiteracy produced the most correlated value at 0.9, which indicated that there is a risk of multicollinearity in this study. A multilinear regression model (MLR) produced r-squared values of 0.9923 and 0.9915, indicating high levels of model fitness that may be somewhat skewed by the collinearity of some variables. To try to improve model fit, two geographically-weighted regression models were run, one with a Gaussian kernel and fixed bandwidth, the other with a Gaussian kernel and adaptive bandwidth. Both GWR models improved on the MLR model's performance. The goodness-of-fit value calculated for the adaptive bandwidth model was 1.0, or perfect fit. To investigate the spatial variation of the seven variables, a bootstrap resampling test was run for each dependent-independent variable set and mapped to the Nepalese districts.

As expected for such a complicated topic, poverty is spatially varying throughout Nepal, and the regression coefficients of the GWR do seem to be related to the distribution of urban and rural populations: per capita income, malnourished children, no safe water, and economic provisioning all returned high-value clusters in more peripheral districts to the northwest of the country. In comparison, population, children in grades 1-5, and adult illiteracy produced higher-value clusters in the southeastern districts. All seven variables comparatively lower values in the center of the country, where the capital Kathmandu is located. Further analysis into the multicollinearity of the variables may produce a clearer exploration of the variables' relationship with the poverty index. It is clear that poverty is a spatially-influenced issue, with possible connections to urban and rural characteristics. A smaller-scale study may be more successful in identifying the specifics of such relationships.