NCES 2018–027 U.S. DEPARTMENT OF EDUCATION

Education Demographic and Geographic Estimates (EDGE) Program

School Neighborhood Poverty Estimates, 2015-2016

NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

Education Demographic and Geographic Estimates (EDGE) Program

School Neighborhood Poverty Estimates, 2015-2016

NOVEMBER 2018

Doug Geverdt
National Center for Education Statistics

NCES 2018-027
U.S. DEPARTMENT OF EDUCATION

NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

U.S. Department of Education
Betsy DeVos
Secretary

Institute of Education Sciences
Mark Schneider
Director

National Center for Education Statistics
James Lynn Woodworth
Commissioner

Administrative Data Division
Ross Santy
Associate Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

NCES, IES, U.S. Department of Education
550 12th Street SW
Washington, DC 20202

November 2018

The NCES Home Page address is http://nces.ed.gov and the Publications and Products address is http://nces.ed.gov/pubsearch. This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES Publications and Products address shown above.

Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Suggested Citation
Geverdt, D. (2018). *School Neighborhood Poverty Estimates, 2015-2016* (NCES 2018-027). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from http://nces.ed.gov/pubsearch/.

Content Contact
Doug Geverdt
(202) 245-8230
Douglas.Geverdt@ed.gov

Acknowledgement

The NCES EDGE program is grateful to Laura Nixon of the U.S. Census Bureau's Education Demographic, Geographic, and Economic Statistics Branch for her long-term collaboration on this project. Her encouragement, patience, persistence, and insight have been essential.

Contents

1.0 Purpose

The National Center for Education Statistics (NCES) Education Demographic and Geographic Estimates (EDGE) program developed spatially interpolated demographic and economic (SIDE) estimates as a way to extract new value from existing sources of poverty data to provide estimates of neighborhood poverty around specific geographic locations (Geverdt and Nixon, 2018). The estimates provided in the School Neighborhood Poverty Estimates file reflect economic conditions of neighborhoods where schools are physically located. They are not designed to reflect the economic conditions of students who are enrolled in schools. Economic conditions of students inside a school may vary considerably from neighborhood conditions outside a school. In cases where small elementary schools limit enrollment to a local neighborhood, the economic conditions of the school neighborhood may be relatively consistent with that of students enrolled in the school. In cases where schools draw students from multiple neighborhoods – as often occurs with larger schools, schools that serve secondary grade levels, and schools that operate open enrollment plans – the economic conditions of the school neighborhood may be quite different from the conditions of neighborhoods where students live.

The primary role of schools is to educate children, but schools are also important neighborhood institutions whose physical facilities are used for cultural events, emergency shelters, polling stations, and other types of public engagement. The investment or lack of investment in school facilities, or the potential closing of local schools, can have significant implications for surrounding neighborhoods, and neighborhood residents may provide considerable support and advocacy for the care of local school facilities that also serve as important neighborhood amenities. Schools are geographic places as well as institutions for learning, and the School Neighborhood Poverty Estimates provide an indicator of the economic landscape around these locations. The NCES EDGE program is working to extend the SIDE framework so it can be used by local school officials to produce similar estimates for student neighborhoods, rather than school neighborhoods. If successful, these estimates could then be aggregated to schools to provide supplemental indicators of school poverty.

2.0 Data and Methods

2.1 Poverty indicator

The 2015-2016 SNP estimates are based on income data from families with children ages 5 to 18 in the U.S. Census Bureau's 2012-2016 American Community Survey (ACS) 5-year collection. The ACS is a continuous household survey that collects social, demographic, economic, and housing information from the population in the United States each month. The estimates reflect the income-to-poverty ratio (IPR) which is the percentage of family income that is above or below the federal poverty level determined for the corresponding family size and structure. The IPR indicator ranges from 0 to a top-coded value of 999. Lower IPR values indicate a higher level of poverty. A family with income at the poverty threshold has an IPR value of 100. The Census Bureau calculates the IPR based on money income reported for families relative to the poverty thresholds. Noncash benefits (such as food stamps and housing subsidies) are excluded, as are capital gains and losses.

2.2 School universe

The file includes IPR estimates and standard errors for 100,551 public school locations in the 50 states and Washington DC. Public school point locations were based on the 2015-2016 EDGE Public School Geocode File. These locations were derived from school addresses reported in the 2015-2016 Common Core of Data (CCD) that were provided by states as part of the 2015-2016 EDFacts collection. The file includes public schools of all levels (elementary, middle, and high school) as well as charter and vocational schools.

2.3 Neighbors and neighborhoods

Unlike neighborhood estimates that identify conditions for predefined census geographic areas, SIDE estimates attempt to identify conditions for specific point locations. Traditional neighborhood estimates are based on household responses contained within predefined geographies like census block groups, census tracts, or school districts. These estimates reflect average conditions across the full geographic area, even though different conditions may exist in different parts of the area, and these estimates are not informed by sample cases that may be just outside the geographic border. SIDE IPR estimates are not constrained by predefined census boundaries. Instead, they rely on a collection of nearest neighbors to develop optimized estimates for unmeasured locations. This ensures that an estimate for a specific unmeasured location is informed by conditions from the nearest measured locations rather than from sample cases contained within the same predefined geographic area that may be located much further away. SIDE neighborhood IPR estimates represent the income-to-poverty ratio predicted for an eligible household if it were present at the predicted location. The estimates included in this file rely on school locations as anchor points to predict the household IPR value that would be expected if a qualifying household were present at the school location. Neighborhood IPR estimates are based on responses of 25 neighbors (qualifying ACS sample households) that are nearest to a school location. Distances to neighbors were determined by linking ACS responses to the center point (longitude/latitude) of 2015 TIGER tabulation blocks. The points were then imported into a GIS and projected to a US equidistant conic projection (Alaska and Hawaii are projected into North American Datum (NAD) 1983 Alaska Albers and Hawaii Albers Equal Area Conic, respectively). The resulting projected point locations with an IPR value were used as the input for estimation.

2.4 Process

SIDE IPR estimates rely on a spatial estimation process known as kriging, a least squares statistical interpolator that uses the weighted sum of values from measured locations to predict values at non-measured locations (Cressie 1989; Cressie 1993). A kriging model produces a prediction surface for a study region that can be used to predict the IPR for households in non-measured locations. The general interpolation function can be represented as:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

where the predicted value at a non-measured location ($\hat{Z}(s_o)$) is the product of the sum of all individually measured locations ($Z(s_i)$) conditioned by their individual weights (λ_i). Kriging assumes that data are spatially autocorrelated, meaning that the value of sample cases partly depends on their distance to other cases. The closer a sampled location is to a non-sampled location, the more weight it will have on the final predicted value of the non-sampled location. However, kriging weights are not solely a function of individual distance. They incorporate prior information about the covariance structure across the full set of measured points. The process first examines the spatial relationships in the data to quantify spatial dependence and formalize a covariance function, and then applies that function back to the data to enable prediction. Therefore, individual weights are informed by what is known about the relationship between cases with a similar distance elsewhere in the data. This relationship is modeled with a semivariogram. The goal of a semivariogram is to quantify the spatial dependence within a dataset. It accomplishes this by modeling the interaction between the semivariance of all potential pairs of cases (y-axis) against the distance of all potential pairs of cases (x-axis). The resulting empirical semivariogram model provides the weights that are applied to each neighbor in a kriging model based on the known distance between the unmeasured location (the prediction point) to each of the neighbors (i.e., ACS sample households) and based on the known income measurements of those neighbors.

Kriging assumes that the underlying data structure is stationary, i.e., that the relationship between variance and distance can be modeled consistently across all portions of the study area. This is frequently not the case for large study areas or for spatial distributions of social and economic conditions. As a result, spatial statisticians developed extensions to traditional kriging methods to help address this concern, including a technique referred to as empirical Bayesian kriging, or EBK (Krivoruchko 2012; Krivoruchko and Gribov 2014). EBK provides a relatively robust response to the problem of non-stationarity by dividing large study areas into smaller regional subsets and then developing a unique model for *each* individual subset. This computational strategy produces a collection of local models that reflect the unique spatial dependencies of each regional subset. The harmonized effect of these local models allows EBK to provide better predictions and more accurate standard errors than are usually achieved through classical kriging methods that rely on a single model of the study area. Also, unlike traditional kriging, EBK does not assume that the default empirical semivariogram model is the true function. Instead, EBK relies on restricted maximum likelihood and subsequent simulation to develop a distribution of semivariogram models for each region. The process develops an initial model and uses it to simulate new data at each of the originally measured locations. The simulated data are then used to produce a new semivariogram model, and the process is repeated to produce a distribution of empirical semivariograms. This refinement produces a more accurate measure of the standard error.

3.0 File format and variables

The SNP estimates file (EDGE_SIDE1216_PUBSCHS1516.csv) is formatted as a CSV file and provides the NCES school ID to link with other NCES school-level files. The fields and summary statistics include:

Name	Description	Type	Length	Mean	Median	Min	Max
NCESSCH	NCES school ID	Character	12	NA	NA	NA	NA
NAME	School Name	Character	32	NA	NA	NA	NA
IPR_EST	Income-to-poverty ratio, estimate	Numeric	3	294	258	35	986
IPR_SE	Income-to-poverty ratio, standard error	Numeric	3	71	64	2	374

4.0 References

Cressie, N. (1989). Geostatistics. *The American Statistician, 43*(4): 197–202.

Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons.

Geverdt, D. and Nixon, L. (2018). *Sidestepping the box: Designing a supplemental poverty indicator for school neighborhoods (NCES 2017-039)*. U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Krivoruchko, K. (2012). Empirical Bayesian Kriging – Implemented in ArcGIS Geostatistical Analyst. *ArcUser, 15*(4): 6–10.

Krivoruchko, K. and Gribov, A. (2014) Pragmatic Bayesian Kriging for Non-Stationary and Moderately Non-Gaussian Data. In E. Pardo-Igúzquiza, C. Guardiola-Albert, J. Heredia, L. Moreno-Merino, J. Durán, and Vargas-Guzmán J. (Eds.) *Mathematics of Planet Earth: Proceedings of the 15th Annual Conference of the International Association for Mathematical Geociences (Lecture Notes in Earth System Sciences)* (pp. 61–64). Berlin: Springer-Verlag.

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography, 46*(2): 234–240.