NCES 2019-112 U.S. DEPARTMENT OF EDUCATION

Education Demographic and Geographic Estimates (EDGE) Program

School Neighborhood Poverty Estimates, 2016-2017

NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

Education Demographic and Geographic Estimates (EDGE) Program

School Neighborhood Poverty Estimates, 2016-2017

Doug Geverdt
National Center for Education Statistics

NCES 2019-112
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

May 2019

This publication is only available online at the NCES Home Page. To download, view, and print the report as a PDF file, go to the NCES Publications and Products page.

Suggested Citation
Geverdt, D. (2019). *Education Demographic and Geographic Estimates Program (EDGE): School Neighborhood Poverty Estimates, 2016-2017* (NCES 2019-112). 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

Contents

1.0 Purpose

The National Center for Education Statistics (NCES) Education Demographic and Geographic Estimates (EDGE) program developed school neighborhood poverty estimates to provide an indicator of the economic conditions in neighborhoods where schools are located. These spatially interpolated demographic and economic (SIDE) estimates apply spatial statistical methods to existing sources of income and poverty data developed by the U.S. Census Bureau to produce new indicators with additional flexibility to support educational research. The economic conditions of neighborhoods around schools may or may not reflect the neighborhood conditions of students who attend the schools. However, supplemental information about school neighborhoods may be useful to combine with student-level or school-level information to provide a clearer picture of the overall educational environment.

2.0 Data and Methods

2.1 Poverty indicator

School neighborhood poverty estimates are based on income data from families with children ages 5 to 18 who were surveyed over a five-year period as part of the U.S. Census Bureau's American Community Survey (ACS). 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 threshold set for the family's size and structure. The IPR indicator ranges from 0 to 999.[1] Low IPR values indicate a high degree 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. Noncash benefits (such as food stamps and housing subsidies) are excluded, as are capital gains and losses. Unmarried partners and unrelated children are not included in families, and their income is not added to family income. Poverty thresholds for each year are available on the U.S. Census Bureau website.

2.2 School universe

The file includes IPR estimates and standard errors for approximately 100,000+ public school locations in the 50 states and Washington DC. Public school point locations are based on the EDGE Public School Geocode Files. These locations were derived from school addresses reported in the Common Core of Data (CCD) that were provided by states as part of the EDFacts collection. The list of schools and their locations are specific to the school year of the estimates. 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

Traditional neighborhood estimates are based on responses from households located within the boundaries of specific geographies, such as census block groups, census tracts, or school districts. These estimates reflect average conditions across the geographic area, even though different conditions may exist within different parts of the area. Traditional neighborhood estimates also ignore sample cases that may be just outside the geographic boundary.

SIDE IPR estimates are not constrained by predefined census boundaries. Instead, they use the sample cases nearest to a given point (nearest neighbors) to develop spatially optimized estimates. This approach ensures that an estimate for any specific location is informed by conditions from the nearest measured locations, rather than from sample cases contained within the same predefined geographic area that may

[1] Families with incomes above 999% of the poverty threshold were given a value of 999 for privacy reasons.

be located much further away. For instance, the predicted value for a point located near the edge of a census tract boundary would be informed by responses in the neighboring tracts, rather than being limited by the borders of the tract.

SIDE neighborhood IPR estimates represent the IPR 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 the responses of the 25 qualifying ACS sample households that are nearest to each school location. Distances to neighbors were determined by linking ACS responses to the center point (longitude/latitude) of TIGER tabulation blocks. The points were then imported into a Geographic Information System (GIS) and projected.[2] The resulting projected point locations with an IPR value were used as the input for estimation.

2.4 Approach

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.[3]

---

[2] Responses in the continental US were projected into a US contiguous equidistant conic projection. Alaska was projected into the North American Datum (NAD) 1983 Alaska Albers Equal Area Conic, and Hawaii was projected into Hawaii Albers Equal Area Conic.

[3] Weights from the semivariogram do not account for differences in block sampling rates in the ACS sample design.

Kriging also assumes that the underlying data structure is stationary; 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 files are formatted as CSV files and provide the NCES school ID to link with other NCES school-level files. The ACS SIDE dataset vintage and the school year for the school points are identified in the User Notes on the website and in the file name. The fields and summary statistics include:

Name	Description	Type	Length
NCESSCH	NCES school ID	Character	12
NAME	School Name	Character	60
IPR_EST	Income-to-poverty ratio, estimate	Numeric	3
IPR_SE	Income-to-poverty ratio, standard error	Numeric	3

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 Geosciences (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.