# Using GIS to Examine the Head Start Service Area in Winona County, Minnesota

Aaron M. Thielen[1,2]
[1]*Department of Resource Analysis, Saint Mary's University of Minnesota, Winona, MN 55987;* [2]*Semcac Head Start, Rushford, MN 55971*

*Keywords:* GIS, Head Start, census data, accessibility

## Abstract

Distance from an origin to a destination has long been considered a major factor in determining degrees of accessibility. For this reason, distance to Head Start sites for families in need of their services was examined. Semcac, the Head Start grantee for seven counties in southeastern Minnesota, provided the addresses of Head Start applicants and participants from the past three years in Winona County. These addresses were geocoded for analysis. Distance to the nearest Head Start facility was determined using ESRI Network Analyst for the student origins. Distance for block group centroids and student origin means were determined using the near distance function in ESRI ArcToolbox. The summary statistics for these distance values were compared. A kernel density layer was created using the student origins point features and was used to determine the location of a hypothetical Head Start facility. The summary statistics describing distance were then compared for the two Head Start site and three site datasets to determine the validity of the hypothetical site location. Census data and the student origins were used to create a linear regression model that could predict variability in participant and applicant distribution. Block groups fitting this model were then examined.

## Introduction

Head Start is a federal program that has provided comprehensive early childhood development services to low-income children since 1965. The Office of Head Start (OHS), the Administration for Children and Families (ACF), and the Department of Health and Human Services (DHHS) (ACF, 2007) administer the program.

DHHS regulations require that at least 90% of children enrolled in Head Start must come from families with income at or below the federal poverty guideline or from families that are receiving welfare assistance. 10% of their slots must be reserved for children with disabilities and 10% of children may be from families above the poverty line (Butler and Gish, 2003). The 2008 poverty guideline for a family of 4 is $21,200 (Federal Register, 2008).

Semcac is the Head Start grantee for Winona, Dodge, Fillmore, Freeborn, Steele, Houston, and Mower Counties in southeastern Minnesota. Licensed Head Start facilities are currently located in Austin, Blooming Prairie, Caledonia, Hayfield, Hokah, Kasson, Ostrander, Owatonna, Preston, St. Charles, and Winona, MN. The St. Charles and Winona Head Start facilities are located within Winona County. These two sites and recent participants and applicants are the focus of this study (Figure 1).

Low-income families with children aged 3-5 years are the

Thielen, Aaron. 2008. Using GIS to Examine the Head Start Service Area in Winona County Minnesota. Volume 10, Papers in Resource Analysis. 12pp. Saint Mary's University of Minnesota Central Services Press. Winona, MN. Retrieved (date) from http://www.gis.smumn.edu.

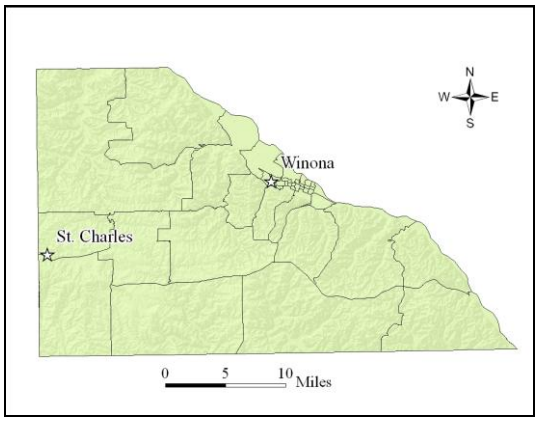


Figure 1. Winona County, Minnesota separated into block groups. The current Head Start sites are labeled by city.

socio-economic group intended to use Head Start facilities. This group of the population is also the most likely to lack the necessary means of transportation for getting to and from these facilities (2003 Community Assessment).

The goal of this study is to visualize the utility and equity of current site locations and to provide information that could assist in the determination of future site locations and the potential for improvement of access to future Head Start applicants. Addresses of past Head Start applicants and year 2000 decennial census data was used for this purpose.

This study did not take into account the concepts of supply constraints or time as factors of access. Instead, only distance was examined. Further analysis of the Head Start service area in Winona County could take these factors into account for a more robust analysis.

## Methods

### *Data Acquisition*

Two Microsoft Excel spreadsheet files containing the addresses of applicants and participants of Head Start for the past three years were obtained from Semcac. There were a total of 384 addresses within the two files.

Winona County road data was acquired from the Winona County Planning Department. The shapefile contained most of the necessary data elements required to geocode the Head Start applicant addresses. Minor editing was required to maximize the address match rate during the geo-coding process.

Using ESRI ArcMap software, a total of 367 addresses were geocoded and saved as point feature classes in a personal geodatabase. The two point feature classes were then merged to provide a picture of the overall distribution of Head Start applicants for the past three years. The majority of the unmatched addresses were rural routes, making the rural applicants slightly underrepresented in this analysis. Other unmatched addresses were P.O. boxes.

Census 2000 TIGER/Line block group data for Winona County was acquired online through the ESRI arcdata website (http://arcdata.esri.com/data/tiger2000/tiger_download.cfm). There are a total of 41 block groups in the county. Their populations range from 6 to 1841 people. The approximate total population of the county is 46,285.

Farm Service Administration orthophotos for Winona County were downloaded from the Minnesota DNR Data Deli website (http://deli.dnr.state.mn.us/data_search.html).

Demographic information by block group for Winona County was obtained through the U.S. Census Bureau American Factfinder website (http://factfinder.census.gov/). Demographic data from Summary File 3 was selected based upon a criterion that

pertains to families that would potentially apply for Head Start. A total of 14 variables were chosen for this analysis (Appendix A).

## Analysis

### *Generalized Census Analysis*

The creation of a dataset representing a generalized distribution of poverty levels was performed to gain an idea of poverty concentrations within Winona County by block group.

Where necessary, the Summary File 3 criterion data was consolidated to represent five categories of poverty status:

1) Households with income below poverty level

2) Individuals with income below poverty level

3) Children 5 and under below poverty level

4) Families with income below poverty level

5) Families with children under 5 with income below poverty level

Five raster datasets were created from the block group shapefile and reclassified based upon their corresponding consolidated criterion values. The block group raster datasets were added together using the raster calculator to create a composite poverty raster in order to provide a generalized representation of poverty distribution in Winona County (Figure 2).

The total values for each block group, such as total population and total households, were used to create rasters depicting percentage of poverty per block group for the five categories previously mentioned. Five raster datasets were created from the block group shapefile using values representing their respective poverty

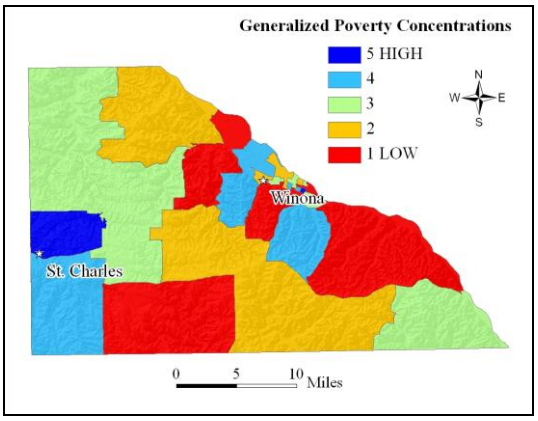


Figure 2. The generalized poverty layer created by combining rasters depicting 5 census variables.

percentages. The block group raster percentage datasets were added together. The resulting composite percent poverty raster shows the generalized distribution of poverty as a percent of the total population in Winona County (Figure 3).

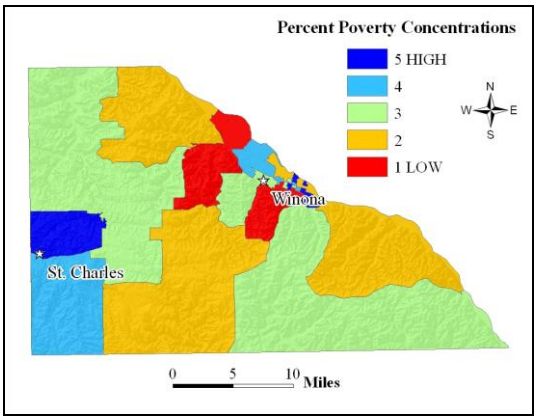


Figure 3. The percent poverty layer created by combining rasters depicting percentages of 5 census variables.

All raster datasets were reclassified on a scale of 1 to 5 based upon natural breaks before and after the raster calculations.

### *Determining Distance*

Close proximity to facilities can be considered to contribute to the health and wellbeing of individuals, and frees up household resources (Pearce et al., 2006). That is why it is important to evaluate the distances that separate served populations and the Head Start facilities they utilize.

The centroid or mean center for each block group was determined. These points were used to calculate the mean straight-line distance for each block group to the nearest Head Start facility using the "Near" proximity tool found in ArcToolbox (Figure 4).

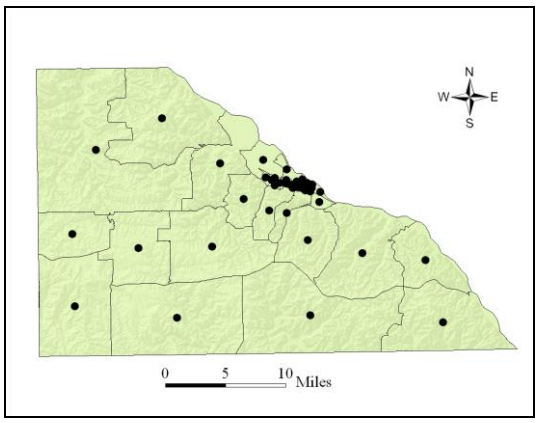


Figure 4. The block group centroids of Winona County. The centroid represents the center of the block group.

A shapefile containing roads within Winona County was used to create a network dataset in ESRI ArcCatalog. This dataset contained a system of interconnected junctions and linear features capable of being used for distance analysis. Using the Network Analyst extension and the Closest Facility function in ArcMap, a set of distances were calculated from the 367 geocoded addresses to the nearest Head Start facility using the Winona County road network dataset.

In order to get a better idea of the areas with higher population densities, the locations of the student origins were averaged per block group, providing 41 mean centers or averages for the student origins. It is assumed the student origins are geographically located in areas with the highest populations of potential need families. Where a block group contained no student origins, the block group centroid was used for this distance analysis.

These points were used to calculate the straight-line distance for the high population density areas for each block group to the nearest Head Start facility using the same process as the block group centroids (Figure 5).

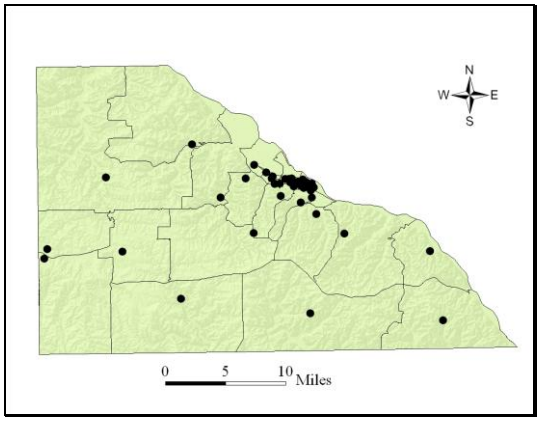


Figure 5. The student origin means are meant to represent areas of higher population density.

## Results

### *Nearest Facility*

Of the 41 block group centroids, 5 were found to be nearer to the St. Charles site while 36 were found to be closer to the Winona site (Figure 6). The student

origin means showed the same results. The respective populations for these sets of block groups are 8,107 for the St. Charles site and 38,178 for the Winona site.

Of the 41 block groups, 12 or 29.3% were within 2 miles of the nearest Head Start. 24 or 58.5% of the block groups were between 2-10 miles away. 5 or 12.2% of the block groups were greater than 10 miles away, up to 18.5 miles away.

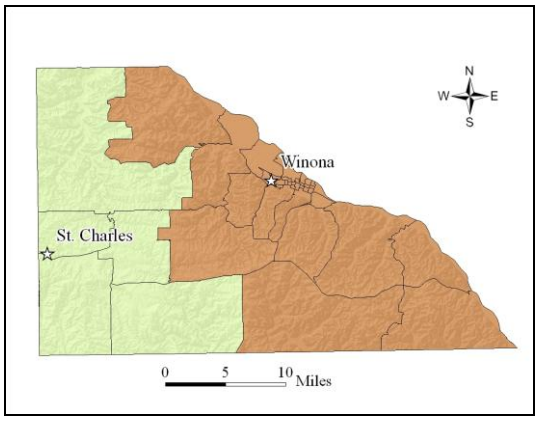


Figure 6. The orange block groups have centroids and student origin means nearer to the Winona facility. The olive block groups are nearer to the St. Charles site.

The greatest distance value determined by the network analyst to separate a student origin and a Head Start facility was 15.68 miles. This could be considered the maximum distance an applicant or participant is willing or able to travel for the applicants examined.

Out of all 367 geocoded addresses, 52.6% of them were within 2 miles of the nearest Head Start facility. 43.6% of student origins were between 2-10 miles of the nearest facility. This leaves the other 3.8% of the student origins being greater than 10 miles away, up to 15.68 miles away. 128 or 34.9% of the student origins were nearer to the St. Charles site and 239 or 65.1% of student origins were closer to the Winona site. It can be assumed that resource needs for the Winona site are greater than the St. Charles site, given its greater number of applicants and/or participants.

***Comparison of Summary Statistics***

The overall summary statistics describing distance for the centroids, student origins, and student origin means differed in a variety of ways (Table 1).

Table 1. The summary statistics for the student origins, student origin means, and the block group centroids. All values refer to miles.

| Student Origins | |
|---|---|
| N= | 367.00 |
| Mean | 2.626 |
| Median | 1.834 |
| Std. Dev. | 2.768 |
| Minimum | 0.04 |
| Maximum | 15.68 |

| Student Origin Means | |
|---|---|
| N= | 41.00 |
| Mean | 3.890 |
| Median | 2.650 |
| Std. Dev. | 3.995 |
| Minimum | 0.053 |
| Maximum | 18.50 |

| Block Group Centroids | |
|---|---|
| N= | 41.00 |
| Mean | 4.399 |
| Median | 2.947 |
| Std. Dev. | 4.169 |
| Minimum | 0.10 |
| Maximum | 18.50 |

A comparison between the centroids and the geocoded student origins showed that actual applicants are, on average, 40% closer. The student

origin median distance is 37% less than the block group centroid median distance. The minimum distance was 0.06 miles less and the maximum distance was 15% less. The standard deviation was 34% closer to the mean.

A comparison between the block group centroids and the student origin means showed that the maximum distance for the origin means was 15% greater than the student origins and the minimum distance was 0.01 miles greater. The average distance for the mean origins was 32% greater and the median was 31% greater. The standard deviation was 31% further from the mean.

The maximum distances turned out to be the same for the student origin means and block group centroids. This occurred because the block groups with these distance values did not contain student origins. The minimum distance for the origin means was 0.05 miles less than the mean and the median distances for the origin means were 10% less than the centroids. The standard deviation was 5% lower.

The results of the comparison between the centroids and the origin means indicate that the shift from the center of the block groups to the center of the student origins within the block groups only slightly decreases the distance to the nearest Head Start facility when determining proximity to potential users.

In all cases, the summary statistics for the student origins were less than the centroids and the origin means. This indicates that the sites are closer to populations utilizing them than they are to the overall population of Winona County. It could also indicate that the populations utilizing the sites are doing so because they live closer to them.

The difference in summary statistics could be the result of the difference in sample sizes among the datasets. It is also possible the differences could be caused by using the Network Analyst to determine distances for the student origins and by using straight line (near) distance for the block group centroids and the student origin means.

**Hypothetical Head Start Site and its Effect on Nearest Facility Distance**

Geographic information systems are an excellent tool for community planning. The potential for determining a new Head Start site became clear after creating a kernel density raster layer using the student origin point features and the Spatial Statistics tool from the ArcToolbox. This raster layer also helped to visualize the effectiveness of the current site locations (Figure 7).

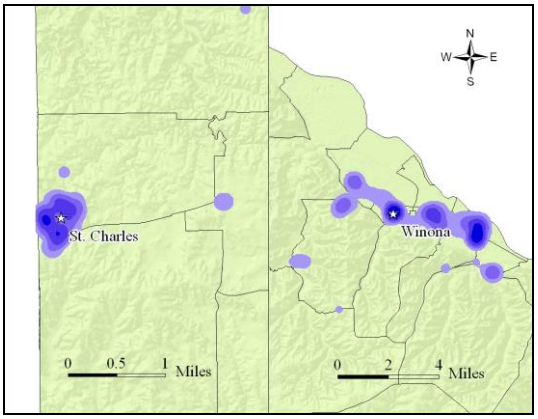


Figure 7. Close-up views of kernel density layer. The darker the color, the higher the density of Head Start applicants and participants.

The kernel density layer indicated a high density of student origins directly around the Winona Head Start facility. The St. Charles site is shown to be surrounded by a series of smaller high-density areas. This would

indicate that this site is equidistant to a series of target populations. Both sites appear to be located in equitable locations beneficial to dense populations of need families.

The location for a new Head Start site was chosen based upon the three-year trend of applicants and participants of Head Start as represented by the kernel density layer. The hypothetical location chosen for this study was close to the center of a high-density area and shifted slightly north based upon the location of a pre-existing school (Figure 8). Washington-Kosciusko Elementary is located on Mankato Avenue. The hypothetical site is located on the same block as the school, at the intersection of High Forest and Sanborn Streets.


Figure 8. The location of the hypothetical Head Start facility as determined by the kernel density raster layer.

The hypothetical school location was chosen visually. By adding this facility to the Network Analyst closest facility and by running the near tool for the centroids and origin means, a quantitative analysis was done to determine if the new site would actually improve proximity to Head Starts.

The summary statistics for almost all of the variables improved from the addition of a third facility (Table 2). Only the minimum values for the student origin means and the block group centroids stayed the same. The median distances for all of the variables were reduced by over a mile.

Table 2. The summary statistics for the student origins, student origin means, and the block group centroids after including the hypothetical site in the analysis. All values refer to miles.

| Student Origins | |
|---|---|
| N= | 367.00 |
| Mean | 1.827 |
| Median | 0.812 |
| Std. Dev. | 2.595 |
| Minimum | 0.01 |
| Maximum | 13.18 |

| Student Origin Means | |
|---|---|
| N= | 41.00 |
| Mean | 2.786 |
| Median | 1.323 |
| Std. Dev. | 3.708 |
| Minimum | 0.053 |
| Maximum | 15.48 |

| Block Group Centroids | |
|---|---|
| N= | 41.00 |
| Mean | 3.368 |
| Median | 1.536 |
| Std. Dev. | 3.952 |
| Minimum | 0.053 |
| Maximum | 15.48 |

The hypothetical Head Start facility only affected student origins that were nearer to the Winona site. Of the 239 student origins nearer to this site, 114 would benefit (gain closer proximity) from the addition of the hypothetical site. That is equal to 31.1% of all participants and applicants from the past three years and 47.7% of all Winona site student origins.

Of the 36 block groups that were nearer to the Winona site, 21 or 58.3% of them would be affected by the addition of the hypothetical site. This shows a significant effect on the service area examined.

This study does not assume the resources exist to build a new Head Start site in Winona County. It is meant to be used as tool to help acquire the funding and other necessary resources to do so.

**Linear Regression of Origins and Census Predictors**

Determining if the census data or which particular census variables correlated with the student origins data was done using SPSS statistics software. The census data demographic values per block group were used in a step-wise linear regression model as independent variables in SPSS. The number of student origins per block group was used as the dependent variable. The purpose of this model was to find which of the census variables could be used to best predict or explain the distribution of the student origins in Winona County.

Of the 14 variables arbitrarily chosen for this study, four were determined by the model to best explain the variability in the distribution of the student origins:

1) Children under 5 years old whose income is below the poverty level. (Ibplu5y)

2) Individuals whose income is equal to 50-74% of the poverty level income. (pnt50_pnt74)

3) Individuals whose income is less than 50% of the poverty level income. (U_pnt50)

4) Total number of households per block group. (Hh_ttl)

As each variable is added to the linear regression model, more of the variability is explained. The variables that held no significance were excluded from the model, leaving only four significant variables. This model was able to explain 72.5 % of the variability in the distribution of the student origins using the arbitrarily chosen independent census data variables (Table 3).

Table 3. Linear regression model steps and variables with R squared values. The closer the value is to 1, the greater the predictor.

| MODEL | Variable | R Square |
|---|---|---|
| 1 | Ibplu5y | 0.616 |
| 2 | pnt50_pnt74 | 0.654 |
| 3 | U_pnt50 | 0.692 |
| 4 | Hh_ttl | 0.725 |

The linear regression model was used to create a raster layer. This layer showed which block groups best fit the linear regression model for predicting the areas with high student origin concentrations. First, individual rasters were created for each of the four variables and then reclassified 1 to 5 using natural breaks. The reclassified rasters were combined using the raster calculator. The weights given to each layer during the raster calculation correspond to their significance in the model:

([Ibplu5y_rc] * 0.616) +
([pnt50_pnt74_rc] * 0.038) +
([u_pnt50_rc] * 0.038) +
([Hh_ttl_rc] * 0.033) = lin_regr raster

The resulting raster layer effectively depicts the linear regression

model visually for use in a spatial context. The raster was reclassified 1 to 5 (1 being a weak match and 5 being the strongest match) using natural breaks (Figure 9).

The hypothetical site location, which was chosen based upon the kernel density layer and the current location of a school, ended up in a block group of the linear regression raster layer with a value of 5. This seems to further reinforce the potential effectiveness of the hypothetical Head Start facility location.

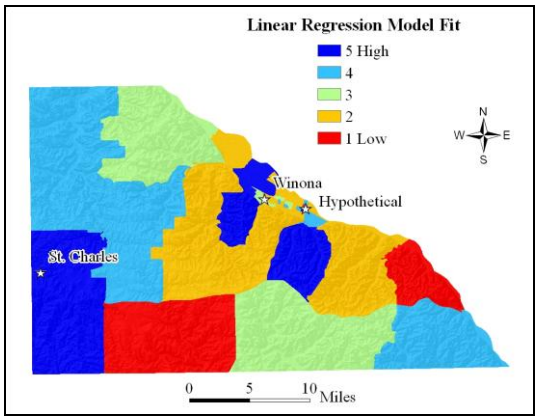


Figure 9. The linear regression model in the form of a weighted raster. The hypothetical site location is located in a block group with a value of 5.

### *Block Groups Fitting the Linear Regression Model*

The block groups of the linear regression layer with a value of 4 or 5 were chosen as the block groups with the highest potential need for Head Start facilities in Winona County. The distances to the nearest Head Start facility for these block groups were then examined in greater detail.

A total of 15 block groups were determined to contain the highest concentrations of populations that best fit the model (Figure 10). The student origin mean distance was used to create the summary statistic subset for these block groups as these are intended to provide a greater indication of the location of the most populated areas per block group where student origins existed. The block group centroid was used where there were no student origins.

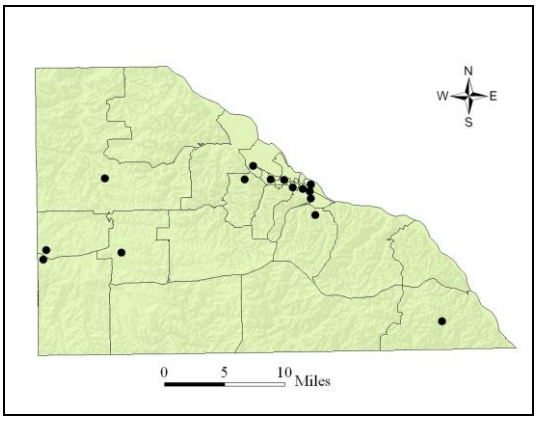


Figure 10. The student origin means representing block groups with a linear regression model value of 4 or 5.

The block group with the greatest average amount of distance separating it and a Head Start facility is included in this data subset. The distance value for this block group is the same as the block group centroid data set. This is because this block group did not contain any geocoded student origins. This is likely due to it being 18.49 miles from the nearest Winona County Head Start. The calculation of the distance to the hypothetical site reduces this distance to 15.48 miles, which is still a significant amount of separation.

The summary statistics for this subset of block groups possesses the same range of values as the full student origin means but the mean, median, and standard deviation values are slightly higher.

The distance calculations after the addition of the hypothetical site improved the summary statistics (Table 4). This dataset showed the greatest amount of improvement in summary statistics of all the datasets (student origins, block group centroids, and student origin means) where change occurred between two and three site distance calculations.

Table 4. Summary statistics for the 15 student origin means that best fit the linear regression model. All values refer to miles.

| Subset Student Origin Means (2 sites) | |
|---|---|
| N= | 15.00 |
| Mean | 3.916 |
| Median | 2.816 |
| Std. Dev. | 4.59 |
| Minimum | 0.053 |
| Maximum | 18.49 |

| Subset Student Origin Means (3 sites) | |
|---|---|
| N= | 15.00 |
| Mean | 2.758 |
| Median | 1.166 |
| Std. Dev. | 4.184 |
| Minimum | 0.053 |
| Maximum | 15.48 |

**Conclusion**

The calculation of distance and summary statistics of student origins, block group centroids, and student origin means has led to the creation of a dataset that can be useful for comparing and examining the Winona County Head Start service area in a spatial context. While the student origins represent points of actual need, the block group centroids and, more so the student origin means, represent centers of potential need. Knowing their respective proximities to current Head Starts will help to gain an understanding of how to best reach them in the future.

The correlation of U.S. census data to the number of student origin means per block group has allowed for the creation of a model that could be used in other counties in order to predict need areas. This model, along with a preexisting knowledge of need could be a powerful tool for determining the location of future Head Start sites.

Further statistical analysis of the distance values calculated could possibly answer more questions. If a national standard or average was determined or available for distance to Head Starts, this would be something that could be used to further judge the equity and utility of Head Start sites in Winona County and all counties in need of examination or analysis.

As has been shown by this study, sites have been added and would potentially be added to areas with high population densities and high need populations. This makes areas with lower concentrations of population in general (rural areas) automatically at a disadvantage when it comes to close proximity to a variety of needs and services.

Fortunately, a bus service is in place for both Winona County Head Start sites. Unfortunately, it does not run on all days classes are held due to limited access to resources such as personnel and buses. That is why, for the purposes of this study, distances from the student origins to the nearest Head Start facilities were used rather than bus route distances.

The bus service is further complicated by the fact that a student may remain on the bus for no more than one hour. For this reason, students are placed into classes ultimately based

upon their respective proximities to the nearest Head Start. This can lead to the need for students to be shuffled around from class to class if they change residence during the school year.

Examination of the Winona County Head Start service area has provided a variety of useful insights, but this study is just a starting point for determining the importance or value of this information. It is up to the stakeholders of Head Start in Winona County to determine its potential and if further analyses are required.

**Acknowledgements**

I would like to thank everyone at the Department of Resource Analysis for their support and guidance. I would like to thank Barb Gunderson for seeing the potential of my project and for perpetuating my cause. I would like to thank Beth Stanford for her support and for providing me with the data that made it all possible.

**References**

Administration for Children and Families (ACF), Head Start Program Fact Sheet. Retrieved October 8, 2007 from http://www.acf.hhs.gov /programs/ohs/about/fy2007.html.

2003 Community Assessment. Child Care Resources, Incorporated Head Start/Early Head Start Division Serving Dixie, Gilchrist and Levy Counties, 2003. Retrieved October 8, 2007 from http://www.afn.org/~ beladona/headstart/.

Butler, A. and Gish, M., 2003. Head Start: Background and Funding. *Report for Congress, Feb. 5, 2003.* Retrieved October 8, 2007 from http://opencrs .cdt.org/rpts /RL30952_20030205.pdf.

Federal Register. 2008. Notices. Vol. 73, No. 15. Retrieved April 8, 2008 from http://aspe.hhs. gov/ POVERTY/ 08fedreg.htm.

Pearce, J., Witten, K. and Bartie, P., 2006. Neighbourhoods and health: a GIS approach to measuring community resource accessibility. *Journal of Epidemiol Community Health*, 60, 389-395. Retrieved October 8, 2007 from jech.bmj.com.

Appendix A. U.S. Census Data 2000 Summary File 3 Variables. Variables found to be significant by the stepwise linear regression model are in bold.

1) Households: Income in 1999 below poverty level (Hh_Ibpl)

**2) Households: Total (Hh_ttl)**

3) Population for whom poverty status is determined: Total (ttl_pop)

**4) Population for whom poverty status is determined: Ratio of income in 1999 to poverty level; Under .50 (U_pnt50)**

**5) Population for whom poverty status is determined: Ratio of income in 1999 to poverty level; 50 to .74 (pnt50_pnt74)**

6) Population for whom poverty status is determined: Ratio of income in 1999 to poverty level; 75 to .99 (pnt75_pnt99)

**7) Population for whom poverty status is determined: Income in 1999 below poverty level; Under 5 years (Ibplu5y)**

8) Population for whom poverty status is determined: Income in 1999 below poverty level; 5 years (Ibpl5y)

9) Families: Total (ttl_fams)

10) Families: Income in 1999 below poverty level; Married-couple family; With related children under 5 years only (FIbplMCcu5)

11) Families: Income in 1999 below poverty level; Male householder; no wife present; With related children under 5 years only (FIbplMcu5)

12) Families: Income in 1999 below poverty level; Female householder; no husband present; With related children under 5 years only (FIbplFcu5)

13) Families: Income in 1999 below poverty level (ttl_FIbpl)

14) Population for whom poverty status is determined: Income in 1999 below poverty level (ttl_Ibpl)