Identifying Differences in a Rochester Neighborhood’s Subsectors

Joseph Blubaugh, Anne Geraci, Rachael Glazner, Shannon Nitroy.

Class: STAT653 First Project Report

Date: 20 Feb 2017 (SP17 semester)

Background

|  |  |
| --- | --- |
| Location of The Triangle in Rochester, New York | Sub-sectors of The Triangle |

NeighborWorks® Rochester (New York) seeks to stabilize city neighborhoods through home ownership education, community-focused home improvement lending, and energy conservation services to families and individuals. We are working with community survey data collected by the NeighborWorks’ Healthy Blocks initiative. HealthyBlocks focuses on improving the physical conditions of homes, resident engagement and leadership, and neighborhood image and identity.

The current focus is a neighborhood in Rochester called the “Triangle of North Winton Village” (“The Triangle”), located between Culver Rd., Merchants Rd, and East Main St. in the Northeast quadrant of the city. The agency is particularly interested in detecting and quantifying differences between the four sub-sectors of the neighborhood, as pictured below.

Project Objectives:

The primary objective for this project is to help the agency understand any statistically significant differences between the four geographical sectors, labeled 1 through 4, within the neighborhood, in an attempt to understand the different dynamics to better target outreach efforts. We hypothesize that sector 4 will be different in nature from the other three sectors, due to its proximity to the relatively affluent Browncroft area. The main variables of interest to the community group are whether the subject would recommend the neighborhood to a friend and how the subject thinks the community will change in the next three years.

Part I of this project models the relationship between the response variable, Years of residence of the head of household, and various other factors, including police and trash ratings, as well as home ownership, in order to gain insight into which factors influence length of time spent in the area.

Data – Source and Variables

The results of a mid-term survey (Year 3 of 5) are provided in the form of a spreadsheet. File includes 200 responses from a door-to-door survey of 450 households selected randomly from the 1200 households in the neighborhood. This response rate is very good for this type of survey, so the data may provide a fairly representative sample of the population.

(I’m going to leave all the raw data variables here – we’ll prune them once we have the modeling done. We also need to add the recoded and computed variables to this table.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Data Type: | Description of Variable |  |
| HH ID | Identifier | Household ID |  |
| Neighborhood | Categorical | Sub-section of | 1,2,3,4 represents neighborhood sub-section |
| Years | Numerical | Number of years you’ve lived here | Years |
| OwnRent | Categorical | Do you own, rent, live with family and friends, or other? | Own, Rent, Live with family/friends, Other |
| HHSize | Numerical | How many people 18 or over live in your household? | # People |
| HHChildren | Numerical | How many children under 18 live in your household? | # Children |
| BornYear | Numerical | In what year were you born? | Year |
| Age | Numerical Computed | Age of person completing Survey \* We will assume this is the head of household | Today() – BornYear (rounded to nearest year) |
| Gender | Categorical | What is your gender? | Male, Female |
| Race | Categorical | To which race do you identify? | Various races |
| SatLevel | Categorical Likert-4 | How Satisfied are you with the neighborhood? | Dissatisfied, Somewhat Dissatisfied Somewhat Satisfied Very Satisfied |
| Recommend | Categorical Likert-5 | How likely are you to recommend living in this neighborhood? | Definitely would not, Probably would not, Probably would, Definitely would |
| Participation Score | Computed Categorical | Computed from eight“community participation” factors. as follows:   * Have you participated in the resident association in the past year? * Did you volunteer to help others in the community in the past year? * Did you participate in community improvement, such as gardening or beautification, in the past year? * Have you supported local business events in the past year * Did you participate in any community social event or block party in the past year? * Did you support a local political candidate or initiative in the past year? * Did you participate in an advocacy group in the past year * Did you report a hazard or contact authorities about an incident in the last year? | LOW = Answered YES on 0,1, or 2 of any of the 8 questions MED = Answered YES on 3,4, or 5 HIGH = Answered YES on 6,7 or 8 |
| Willingness | Categorical Likert-4 | How willing are you to become involved with others in the community? | Not very willing, SW willing, willing, Very willing |
| MakeDifference | Categorical Likert-4 | How much of a difference can you make as an individual? | A little or none, Some, A fair amount, A great deal |
| GetRide | Categorical Likert-4 | How likely for neighbors to help if you needed a ride? | Very unlikely, SW Unlikely, SW Likely, Very Likely |
| GetFavor | Categorical Likert-4 | How likely for neighbors to help if you needed a favor (borrow a tool, pick up mail, etc.)? | Very unlikely, SW Unlikely, SW Likely, Very Likely |
| CheckElderly | Categorical Likert-4 | How likely for neighbors to check on an elderly neighbor? | Very unlikely, SW Unlikely, SW Likely, Very Likely |
| ChildCare Emergnecy | Categorical Likert-4 | How likely for neighbors to provide emergency child care? (4 Variables) | Very unlikely, SW Unlikely, SW Likely, Very Likely |
| PoliceRating | Categorical Likert-5 | How would you rate police response? | Very Poor, Poor, Fair, Good, Very good, NA |
| FireRating | Categorical Likert-5 | How would you rate fire department response? | Very Poor, Poor, Fair, Good, Very good, NA |
| EMSRating | Categorical Likert-5 | How would you rate ambulance response? | Very Poor, Poor, Fair, Good, Very good, NA |
| TrashRating | Categorical Likert-5 | How would you rate trash collection? | Very Poor, Poor, Fair, Good, Very good, NA |
| SnowRemoval Rating | Categorical Likert-5 | How would you rate snow removal/street cleaning? | Very Poor, Poor, Fair, Good, Very good, NA |
| FeelSafeDay | Categorical Likert-4 | How safe do you feel walking during the day? | Very unsafe, SW unsafe, SW safe, Very safe |
| FeelSafeNight | Categorical Likert-4 | How safe do you feel walking at night? | Very unsafe, SW unsafe, SW safe, Very safe |
| FeelSafeChildren | Categorical Likert-4 | How safe are children going to and from school? | Very unsafe, SW unsafe, SW safe, Very safe |
| FeelSafeElderly | Categorical Likert-4 | How safe are senior citizens? | Very unsafe, SW unsafe, SW safe, Very safe |
| ChangePast3Y | Categorical Likert-5 | How has the community changed vs. three years ago? | (has) Declined a lot, declined some, about the same, improved some, improved a lot |
| ChangeNext3Y | Categorical Likert-5 | How will the community change in the next three years? | (will) Decline a lot, decline some, about the same, improve some, improve a lot |

Methods

Because our data come from a survey rather than a designed experiment, we use our initial analysis to construct an ANCOVA regression on the number of years the participant has resided in the area. It is an ANCOVA specification because our response variable is continuous and our explanatory variables are either discrete or continuous. We want to determine if the slopes of our linear model differ for each subsector, so we will examine the coefficient of NeighborhoodID in particular. This is one dimension into determining overall differences in the four subsectors, but is also interesting because a naïve assumption may be that people who are happy with their location tend to stay there longer. If we can identify factors that lead to people staying in an area longer, we gain some information about how to make improvements, another goal of the NeighborWorks program. Our initial model, including all variables we believe may contribute to years of residence, is as follows:

log(Years) = NeighborhoodID + Age + RecommendCat + Gender + FeelSafeNightCat + SatLevelCat + Race + SnowRemovalCat + ParticipationScoreCat + OwnRent + PoliceRating + FeelSafeDayCat + TrashRatingCat

After removing insignificant covariates, but keeping NeighborhoodID because it is of interest, we are left with the following specification:

log(Years) = NeighborhoodID + Age + OwnRent + PoliceRating + FeelSafeDayCat + TrashRatingCat

We used the software JMP to statistically analyze our data.

Results



First, we discuss the covariates left in the model and the implications their significance have. Age of course has a positive coefficient, which we expect. The “own” category of OwnRent has a positive coefficient and the “rent” category has a negative one, both of which we also expect, since renters are less likely to stay in one place than homeowners. FeelSafeDayCat has a positive coefficient, which is also in line with expectations. The coefficients for PoliceRatingCat high, and both levels of TrashRatingCat are negative, which is surprising, given that we would expect better police and trash services ratings to contribute to more years in the area. Neighborhood ID is not significant, but the years of residence is lower for sector 1 relative to sector 4, but higher for sectors 2 and 3 relative to sector 4.



Next, we look more into the differences between our four subsectors. Even though their coefficients were not significant, indicating that years of residence is not a statistically significant component of the differences in neighborhoods, we can still see whether there were significant differences between the pairs of subsectors. We can see this is not the case, based on the Tukey HSD. No pairs of subsectors have significantly different years of residence, and the least squares means for each subsector are very close in value.

Discussion

This simple ANCOVA model provides insight into one of the many facets of this dataset. We find that people tend, on average, to live in each of the four subsectors for the same amount of time. We also find that factors such as age, home ownership, and police and trash ratings contribute to the number of years of residence, but the unexpected signs on some of the coefficients may indicate we have omitted variable bias.

Team Members:

* Anne Geraci – teaches Math and Statistics and run the Math Center at St. John Fisher College.
* Shannon Nitroy – BA in Economics & Math. Works as a financial analyst for the Federal Reserve in Washington, DC
* Rachel Glazner – second year PhD student in Ecology and Evolutionary Biology. Working on a project this semester using a drone to survey wading birds along the Texas coast.
* Joseph Blubaugh (Online) – works as a data scientist for Caterpiller. Has a BA in Economics.