

New Mexico Health Insurance Coverage, 2009-2013

Exploratory, Ordinary Least Squares, and Geographically Weighted Results Using ArcGIS Desktop Spatial Statistics

Larry Spear 4/16/2016 (Draft)

A dataset consisting of selected average statistics was derived from the U.S. Census Bureau's American Community Survey (ACS, 2009 – 2013) for New Mexico's census tracts (n=499). These data were originally processed and made available at the New Mexico Community Data Collaborative (NMCDC) ArcGIS Online web site as either feature layers or feature services. ArcGIS Desktop and SAS University Edition were used for further data preparation.

The results from ArcGIS exploratory regression suggest several possible models that can be used to describe the relationship between the percent of population without health insurance (Per_WO_Ins), the dependent variable and a series of candidate explanatory variables. A reasonable model (AdjR² = 0.64 and VIF 2.26 see below) was found that had four explanatory variables; per capita income (Per_CAPITA_INC), percent of population in poverty (PER_POVERTY), percent Hispanic and Latino (P_HISPLAT), and percent American Indian (P_AMINDIAN). It is important to note that some common sense is necessary when using exploratory regression. Selecting all or most of the variables as candidate explanatory variables will result in many potential models not being able to be estimated due to severe multicollinearity (data redundancy). This is why I chose only seven potential explanatory variables (see Appendix) and selected the model with just four of these.

More diagnostics from this model were produced using ArcGIS Ordinary Least Squares (OLS) regression. The [OLS results](#) show that this regression model is mostly a good fit to the data. But, given the Konker (BP) Statistic (P=0.000145) there is evidence of either heteroscedasticity or non-stationarity (although mostly random looking with some cone shape - see Residual vs. Predictor Plot). Further, the Jargue-Bera Statistic is also significant (p<0.01) indicating that the model predictions are biased and that the residuals are not normally distributed with inconsistent variance (although they are close to normal – see Histogram of Standardized Residuals).

Given these diagnostic results from ArcGIS OLS the ArcGIS Global Moran's I tool was run on the standardized residuals to evaluate the degree of spatial autocorrelation. These results (see below) confirm that there is significant (p<0.01) clustering of residual values. Additional analysis will be conducted using Geographically Weighted Regression (GWR) to see if a better fit can be obtained.

Exploratory Regression Output (portion only – selected model):

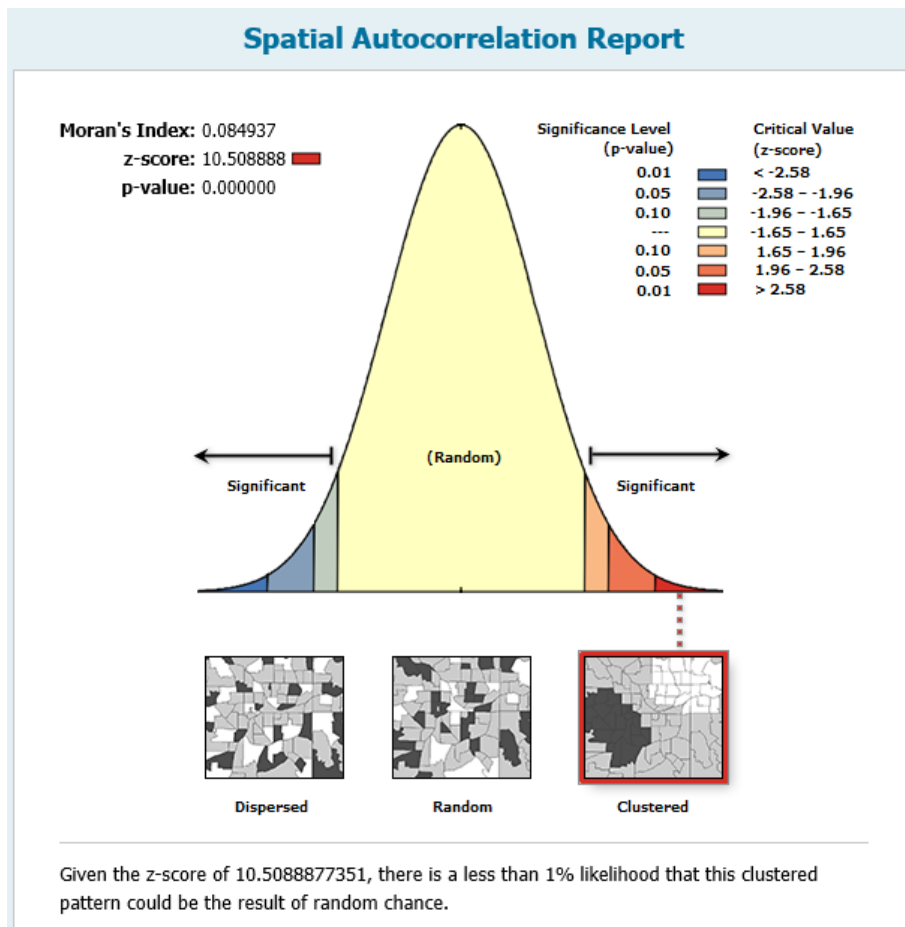
```
*****
Choose 4 of 7 Summary

Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF  SA   Model
0.64 3221.35 0.00 0.00 2.26 0.00 -PER_CAPITA_INC*** +PER_POVERTY*** +P_HISPLAT*** +P_AMINDIAN***
0.63 3228.45 0.00 0.00 2.85 0.00 -MEDIAN_HOUSE_INC*** +PER_POVERTY*** +P_HISPLAT*** +P_AMINDIAN***
0.63 3232.61 0.00 0.00 16.71 0.00 -MEDIAN_HOUSE_INC*** +P_HISPLAT*** +P_WHITE*** +P_AMINDIAN***

Passing Models
AdjR2 AICc JB K(BP) VIF SA Model
*****
```

Summary of Variable Significance			
Variable	% Significant	% Negative	% Positive
PER_CAPITA_INC	100.00	100.00	0.00
PER_POVERTY	100.00	0.00	100.00
P_WHITE	100.00	80.70	19.30
P_AMINDIAN	100.00	0.00	100.00
P_HISPLAT	84.21	50.88	49.12
MEDIAN_HOUSE_INC	80.70	89.47	10.53
PER_EMPLOYED	7.02	61.40	38.60

Global Moran's I Output (portion only):



Geographically Weighted Regression (GWR) :

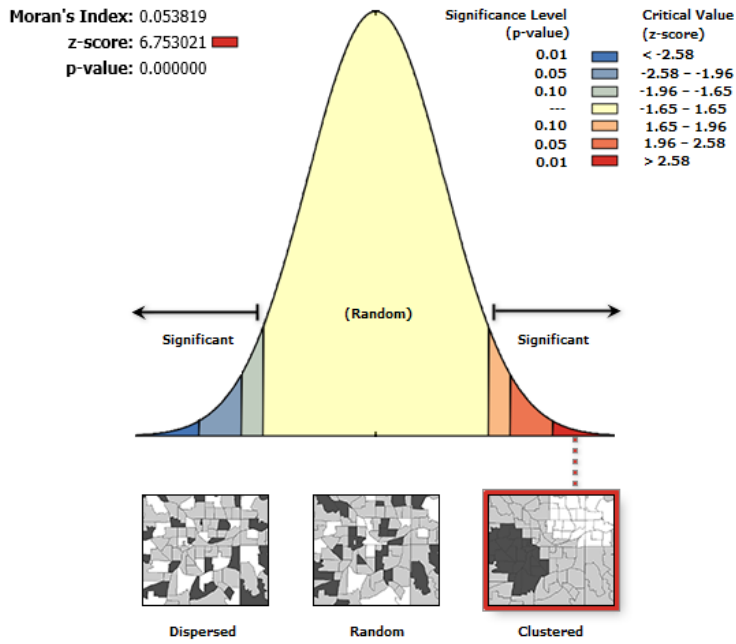
The GWR results (see below) showed a slight improvement ($\text{Adj } R^2 = 0.686$ and lower $\text{AICc} = 3174.56$) over the OLS results. Also all the condition numbers for the output feature class locations are less than 30, indicating no major problems with local multicollinearity. However, the standardized residuals are still clustered (see std. map below, $p < 0.1$) indicating that this model is still missing one or more explanatory variables. Apart from these shortcoming, the GWR Moran's Index (0.053819) and z-score (6.735021) show a noticeably decrease, an improvement, from the OLS results (Moran's I 0.084937, z-score 10.50888). Also very informative is the map of local R2 (see below), which shows where this model has performed well (red and orange), and did poorly (blues).

A slightly better model was suggested from the exploratory regression (see Appendix) that included a fifth explanatory variable, percent white (P_WHITE). I did not use this model for GWR as the VIF was somewhat large (17.46), indicating more global multicollinearity (Note: ArcGIS GWR failed to compute this model due to explanatory variable redundancy). However, I decided to use this model as an example for a comparison of results that can be obtained from GeoDa –GWR, R, and QGIS (being prepared). These facilities will compute this model and the results may still prove informative regardless of the variable redundancy.

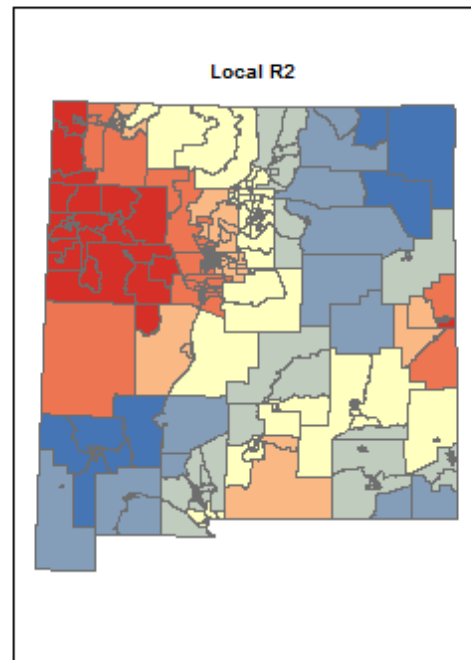
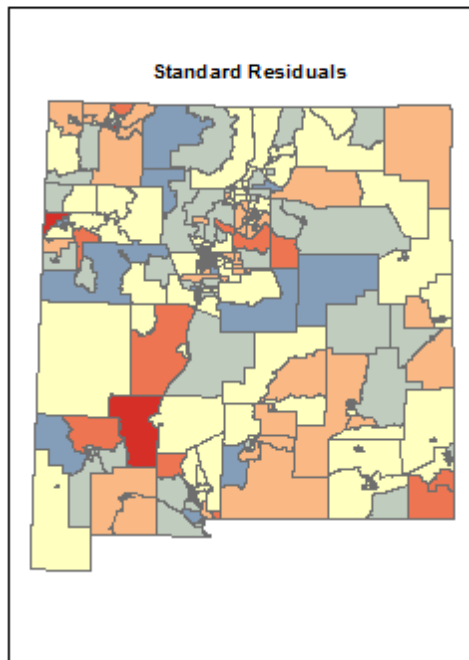
```
Executing: GeographicallyWeightedRegression NM_ACS13_Ex_UTM_Per_WO_Ins_Per_Capita_Inc;Per_Poverty;P_HispLat;P_AmIndian C:\gis\NMDOH
\GWR_T1.shp FIXED AICc # 30 # # 2273.130822 # # #
Start Time: Sat Apr 02 15:04:53 2016
Bandwidth      : 97385.952258826597
ResidualSquares : 14231.221313180327
EffectiveNumber : 50.757191266212715
Sigma         : 5.6346172547464244
AICc         : 3174.5684643636732
R2           : 0.71746808616952096
R2Adjusted   : 0.68610563215718812
Succeeded at Sat Apr 02 15:04:57 2016 (Elapsed Time: 3.70 seconds)
```

GWR_T1_supp				
	OID	VARNAME	VARIABLE	DEFINITION
	0	Bandwidth	97385.952259	
	1	ResidualSquares	14231.221313	
	2	EffectiveNumber	50.757191	
	3	Sigma	5.634617	
	4	AICc	3174.568464	
	5	R2	0.717468	
	6	R2Adjusted	0.686106	
	7	Dependent Field	0	Per_WO_Ins
	8	Explanatory Field	1	Per_Capita_Inc
	9	Explanatory Field	2	Per_Poverty
	10	Explanatory Field	3	P_HispLat
	11	Explanatory Field	4	P_AmIndian

Spatial Autocorrelation Report



Given the z-score of 6.75302054368, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.



Appendix :

Additional Exploratory and GWR Output

```

*****
Choose 5 of 7 Summary
Highest Adjusted R-Squared Results
AdjR2  AICc  JB  K(BP)  VIF  SA  Model
0.65  3206.72  0.00  0.00  17.46  0.00  -PER_CAPITA_INC*** +PER_POVERTY*** +P_HISPLAT*** +P_WHITE*** +P_AMINDIAN***
0.64  3219.44  0.00  0.00  16.73  0.00  -MEDIAN_HOUSE_INC*** +PER_POVERTY*** +P_HISPLAT*** +P_WHITE*** +P_AMINDIAN***
0.64  3220.12  0.00  0.00  17.44  0.00  -MEDIAN_HOUSE_INC*** -PER_CAPITA_INC*** +P_HISPLAT*** +P_WHITE*** +P_AMINDIAN***
Passing Models
AdjR2  AICc  JB  K(BP)  VIF  SA  Model
*****

```

Data (From Census ACS, 2009-2013) Note: variable/item names changed by me in SAS or by NMCDC.

<u>Variable</u>	<u>Census Name</u>	<u>Description</u>
Pop_Avg	HC01_VC130	Civilian noninstitutionalized population
Pop_Avg_W_Ins	HC01_VC131	Civilian noninstitutionalized population - With health insurance coverage
Per_W_Ins	HC03_VC131	Percent - Civilian noninstitutionalized population – With health insurance coverage
Pop_AVG_WO_Ins	HC02_VC134	Civilian noninstitutionalized population - Without health insurance coverage
Per_WO_Ins	HC03_VC134	Percent - Civilian noninstitutionalized population – Without health insurance coverage
Household_Avg	HC01_VC74	INCOME AND BENEFITS (IN 2013 INFLATION-ADJUSTED DOLLARS) - Total households
Median_House_Inc	HC01_VC85	INCOME AND BENEFITS (IN 2013 INFLATION-ADJUSTED DOLLARS) - Total households - Median household income (dollars)
Mean_House_Inc	HC01_VC86	INCOME AND BENEFITS (IN 2013 INFLATION-ADJUSTED DOLLARS) - Total households - Mean household income (dollars)
Per_Capita_Inc	HC01_VC118	INCOME AND BENEFITS (IN 2013 INFLATION-ADJUSTED DOLLARS) - Per capita income (dollars)
Employed_Avg	HC01_VC03	EMPLOYMENT STATUS - Population 16 years and over
Per_Employed	HC03_VC04	Percent - EMPLOYMENT STATUS - Population 16 years and over - In labor force
Employed_Not	HC01_VC09	EMPLOYMENT STATUS - Population 16 years and over – Not in labor force
Per_Employed_Not	HC03_VC09	Percent - EMPLOYMENT STATUS - Population 16 years and over – Not in labor force
Per_Family_Poverty	HC03_VC161	PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL - All families

<u>Variable</u>	<u>Census Name</u>	<u>Description</u>
Per_Poverty	HC03_VC171	Percent - PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL - All people
Per_GE18_Poverty	HC03_VC176	Percent - PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL - 18 years and over
Total_	B01001e1	Total: Total population -- (Estimate)
Hisp_Lat_Not	B01001He1	White alone, not Hispanic or Latino population
White	B01001Ae1	People who are White alone
Black	B01001Be1	Black or African American alone
AmIndian	B01001Ce1	People who are American Indian and Alaska Native alone
Asian	B01001De1	People who are Asian alone
HispLat	B01001Ie1	Hispanic or Latino population -- (Estimate)
P_Hisp_Lat		
P_HispLat_Not		
P_White		
P_Black		
P_AmIndian		
P_Asian		

**Note: These data derived from the NMCDC ArcGIS Online web mapping applications. Not sure of the original variable/item name from the census ACS 2009-2013. For instance Total population estimate could be B01001e1 or B01003e1. Will update when I know which census variables were used.