Data Pooling for Wilmington

Roberto Miquel, AICP

Krishnan Viswanathan

November 19, 2014

CDM Smith ®

Data pooling

- Objective: Identify counties with similar socio-demographics (SEDs) with the Wilmington MSA counties
  - Brunswick, New Hanover, and Pender
- Data: American Community Survey (ACS), 2007-2011 estimates and margin of error (MOE)
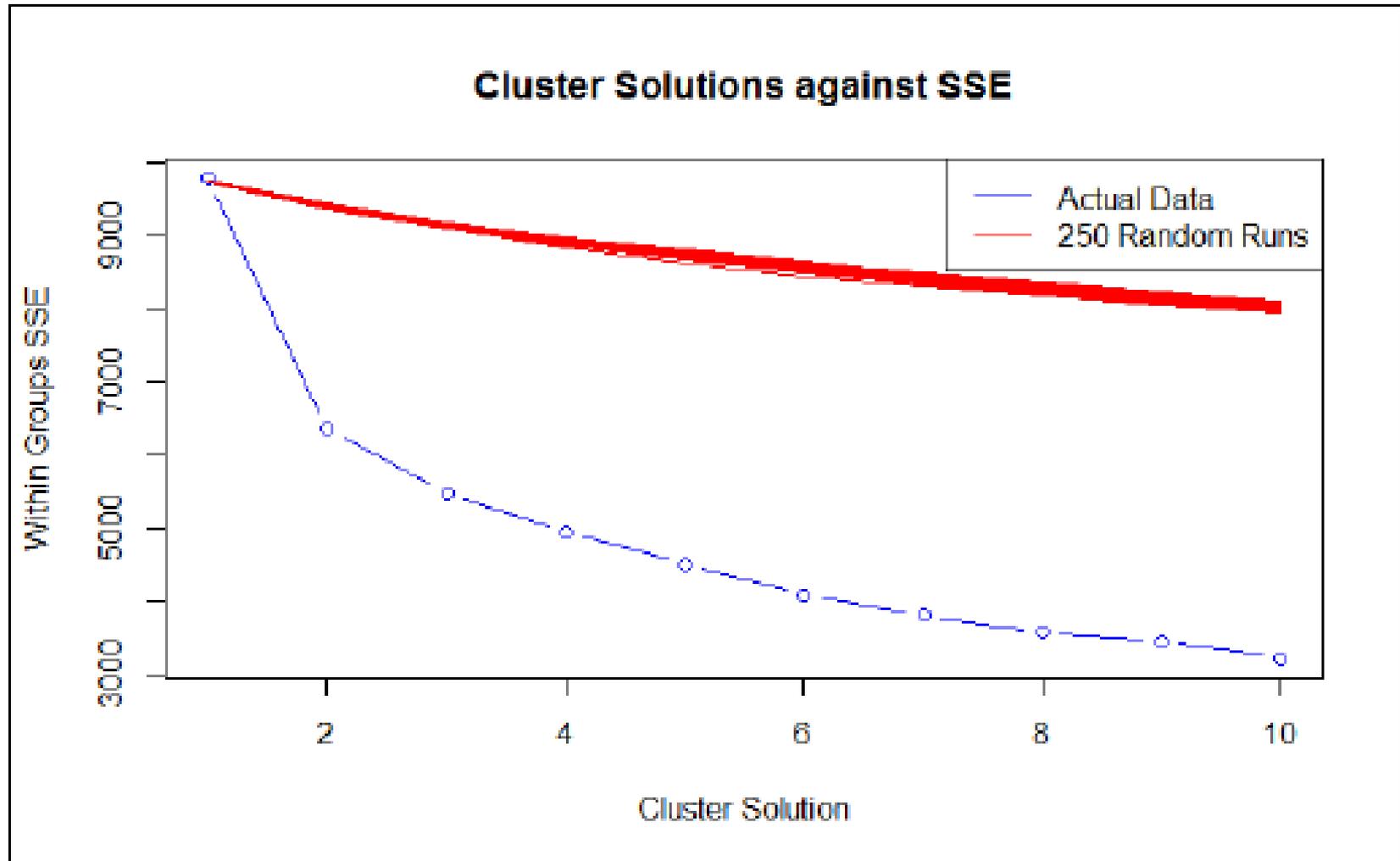- External counties: all counties from FL, NC, NY, SC, and VT

SEDs for data pooling

- <u>Household size</u> with sampling measurement of error (MOE),
- Household distribution of <u>household vehicles</u> with MOE,
- Household distribution of <u>household income</u> with MOE, and
- Population <u>age distributions</u> with MOE
- Total: 47 categories

Data pooling methods: Clustering

- Clustering Method
  - K-means clustering
  - Identify "nearest neighbors" based on proximate data
  - Attempts to organize counties based on distinct groupings
  - Clusters were based on principal counties (Brunswick, New Hanover, Pender)
  - Two scenarios emerged:
    - Two clusters
    - Three Clusters
  - Clusters fail to provide satisfactory explanatory power and the method is discarded

Data pooling methods: Clustering

Data pooling methods: Clustering

Data pooling methods: Sorting

- Sorting Method
  - Organize counties by descriptive variables
  - Index variables with respect to similar values
  - Rank order counties with the most similar values
  - Similarity determined by a comparison of standard errors in demographic characteristics between base counties and target counties
  - Original analysis was done using absolute values for demographic characteristics
  - Ultimately, using relative values proved more effective at identifying similarities

Data pooling methods: Sorting

Pooled Counties

- Sample size increased from 212 to 1,194 households

- **Brunswick**: Clay County, NC, Cherokee County, NC, Henderson County, NC, Graham County, NC, Tyrrell County, NC, Pamlico County, NC, Alleghany County, NC, Swain County, NC, Dixie County, FL

- **New Hanover**: Okaloosa County, FL, Pasquotank County, NC, Jefferson County, FL, Madison County, FL, Graham County, NC, Swain County, NC, Tyrrell County, NC, Broome County, NY, Taylor County, FL

- **Pender**: Stokes County, NC, Walton County, FL, Granville County, NC, Suwannee County, FL, Chenango County, NY, Kershaw County, SC, Lee County, NC, McDowell County, NC, Person County, NC

Travel Behavior Analysis

Wilmington MPO Counties			
Coefficients:	Estimate	Std. Error	t-value
(Intercept)	1.59	1.16	1.38
DRVRCNT	0.53	0.63	0.85
HHSIZE	1.71	0.29	5.88
WRKCOUNT	0.46	0.48	0.97
Degrees of Freedom	208		
Adjusted R^2	0.20		
Wilmington MPO + Pooled Counties			
Coefficients:	Estimate	Std. Error	t-value
(Intercept)	0.96	0.47	2.06
DRVRCNT	0.81	0.26	3.09
HHSIZE	1.87	0.13	14.14
WRKCOUNT	0.53	0.23	2.33
Degrees of Freedom	1145		
Adjusted R^2	0.27		

Conclusion

- Inclusion of additional samples makes the dataset more robust

- Most pooled counties came from NC

- Pooled data were used for developing the Wilmington Model trip generation rates

THANK YOU!!!

Roberto Miquel, miquelro@cdmsmith.com, 919-325-3605