Data Pooling for Wilmington
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

- Household size with sampling measurement of error (MOE),
- Household distribution of household vehicles with MOE,
- Household distribution of household income with MOE, and
- Population age distributions 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

Figure 2. Principal components plot for 2-cluster and 3-cluster solution.

These two components explain 66.49% of the point variability.
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

<table>
<thead>
<tr>
<th>Wilmington MPO Counties</th>
<th>Coefficients:</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.59</td>
<td>1.16</td>
<td>1.38</td>
<td></td>
</tr>
<tr>
<td>DRVRCNT</td>
<td>0.53</td>
<td>0.63</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>HHSIZE</td>
<td>1.71</td>
<td>0.29</td>
<td>5.88</td>
<td></td>
</tr>
<tr>
<td>WRKCOUNT</td>
<td>0.46</td>
<td>0.48</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td></td>
<td>208</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td></td>
<td>0.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wilmington MPO + Pooled Counties</th>
<th>Coefficients:</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.96</td>
<td>0.47</td>
<td>2.06</td>
<td></td>
</tr>
<tr>
<td>DRVRCNT</td>
<td>0.81</td>
<td>0.26</td>
<td>3.09</td>
<td></td>
</tr>
<tr>
<td>HHSIZE</td>
<td>1.87</td>
<td>0.13</td>
<td>14.14</td>
<td></td>
</tr>
<tr>
<td>WRKCOUNT</td>
<td>0.53</td>
<td>0.23</td>
<td>2.33</td>
<td></td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td></td>
<td>1145</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td></td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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