APPENDICES

APPENDIX 1:

1. Documentation of Data Edits, Queries, and Issues in Reconciliation

Jacqueline Hughes-Oliver and Tae-Young Heo

February 2002

Documentation of Data Edits, Queries, and Issues in Reconciliation

Jacqueline M. Hughes-Oliver and Tae-Young Heo
February 28, 2002

1 Purpose and Summary

The purpose of this document is to provide written explanations of edits and queries that have surfaced following detailed inspection of two sets of data received from NCDOT. Several edits have already occurred, with approval from NCDOT. For issues that are currently unresolved, the project has proceeded by using data obtained in November 2001. Queries that require responses are printed in bold. Almost all variables have at least one associated query.

2 Background of Data Files Received

In November 2001, Larry Wikoff delivered a file of PTC count data for the test area consisting of Chatham, Durham, Johnston, Orange and Wake counties; this file will be referred to as FILEL. FILEL contains non-attributed data for stations, including variables countyid (called the SIPS code in ArcView files), stationid, urbnum, routecode, routenum, description, atrgroup, comments, case, as well as raw (daily) and factored (aadt) counts for up to two different counting periods.

In December 2001, Erik Brun delivered three files, one of which is an attributed PTC count file for the entire state of NC; this file will be referred to as FILEE. FILEE contains several variables present in FILEL, namely sips, station, routecode, routenum, description, atrgroup, and four <count year,aadt> pairs. In addition, FILEE also contains station locations (x-, y- coordinates in the State Plane System), data on land use in areas surrounding stations (variables level2 and lucode), unique identifiers for linking station information to census data (through variables county90, tract90, group90), and segment attributes such as speed, number of lanes, and whether or not a segment is access controlled.

3 Issues

Three basic categories of data issues have arisen: those specific to FILEL, those specific to FILEE, and those obtained from merging the two files.

3.1 Issues Specific to FILEL

3.1.1 Stations Missing All Raw Counts or Other Key Variables

Some stations are missing raw (unfactored) counts or ATR groupings or month/day. Larry reported that "only the control stations in the Interstate system will have raw data associated with them, and therefore only they will have day, month, ATR group, and raw data counts." This is due to the impracticality of placing tube counters on the Interstate. Counts at "non-control stations" are

1

obtained by taking differences between numbers of entries and exits on the Interstate near the non-control stations. Counts on entry and exit ramps are obtained in the usual manner (with portable tube counters) placed at these so-called "control stations."

To keep track of this issue, when case="INTER", the variable comments was given a value of either "not control station" or "control station", depending on the data provided for that station. There were a total of 85 stations with case="INTER", as outlined below:

		comments= control station	comments= not control station	
routecode=	.	0	2*	2
	1	16	63	79
	2	0	4*	4
		16	69	85

The records in cells containing * are deemed questionable and are provided in full detail below:

```
     countyid stationid urb.num routecode routenum       description atrgroup
1638       91      2235       .      \"\"     \"\"              \"\"        .
2741      108       208     108     \"2\"   \"64\" \"BYP W SR 2000\"        .
2743      108       210     108     \"2\"   \"64\" \"BYP E SR 2000\"        .
2814      108       282     108     \"2\"   \"64\" \"BYP N SR 2921\"        .
2817      108       285     108     \"2\"   \"64\" \"BYP W SR 1012\"        .
3510      108      2234       .      \"\"     \"\"              \"\"        .

     year month day   aadt count1 count2 count3 year.2 month.2 day.2 aadt.2
1638 2000     .   .  19210     NA     NA     NA     NA      NA    NA     NA
2741 2000     .   . 124719     NA     NA     NA     NA      NA    NA     NA
2743 2000     .   . 115642     NA     NA     NA     NA      NA    NA     NA
2814 2000     .   .  93486     NA     NA     NA     NA      NA    NA     NA
2817 2000     .   .  95891     NA     NA     NA     NA      NA    NA     NA
3510 2000     .   .  87004     NA     NA     NA     NA      NA    NA     NA

     count1.2 count2.2 count3.2                comments      case  uniqid
1638       NA       NA       NA \"not control station\" \"INTER\"  912235
2741       NA       NA       NA \"not control station\" \"INTER\" 1080208
2743       NA       NA       NA \"not control station\" \"INTER\" 1080210
2814       NA       NA       NA \"not control station\" \"INTER\" 1080282
2817       NA       NA       NA \"not control station\" \"INTER\" 1080285
3510       NA       NA       NA \"not control station\" \"INTER\" 1082234
```

The stations (with uniqid=) 912235 and 1082234 have been deleted because this study is restricted to stations that were active in 1998 and these stations were added in 1999 or later. This information was provided by Larry.

2

Another fix that was provided by Larry is that the missing routecode for station 1080428 is 4:

```
     countyid stationid urb.num routecode routenum      description atrgroup
2957      108       428     108      \"\" \"1005\" \"N SANDY FORK\"        4

     year month day  aadt count1 count2 count3 year.2 month.2 day.2 aadt.2
2957 1998     4   3 28489  33154  32541     NA     NA      NA    NA     NA

     count1.2 count2.2 count3.2 comments    case  uniqid
2957       NA       NA       NA     \"\" \"3-4\" 1080428
```

Currently, the only stations with missing values for atrgroup (10 stations), day (67 stations), or count (67 stations) are those with comments=not control station. **Why do stations 1080208, 1080210, 1080282, and 1080285 have routecode=2 yet they are also classified with case=INTER?**

3.1.2 Check for AADT-Count Inconsistencies, Without Estimated

Kent Taylor and Larry provided a list of stations that had an apparently good aadt (i.e., not estimated), but whose aadt appeared to be either rather low or rather high compared to the raw counts. During a November 2001 meeting, Larry commented that the seemingly low aadts are probably a function of having many trucks on the particular road segment. This seems a likely explanation for the low aadts. Is there an explanation for the high aadts? **Are there any modifications needed, or can we proceed with the data as is?**

3.1.3 Check for AADT-Count Inconsistencies, Estimated Only

Kent and Larry provided a list of stations that had only estimated, or adjusted, counts, but whose aadt appeared to be either rather low or rather high compared to the raw counts. During a November 2001 meeting, Larry commented that Kent didn't think these should cause concern since all of them have only estimated aadts. **Are there any modifications needed, or can we proceed with the data as is?**

3.1.4 Stations With Multiple Records

Kent and Larry provided a list of stations having multiple records for the year that aadt was computed. These result from recounts within the same year and do not represent problems, except for the following:

```
     countyid stationid urb.num routecode routenum    description atrgroup
1046       91        30     108     \"2\"    \"1\" \"N OF NC 55\"        1
1047       91        30     108     \"2\"    \"1\" \"N OF NC 55\"        1

     year month day  aadt count1 count2 count3 year.2 month.2 day.2 aadt.2
1046 1998    10   2 12422  13067  14765  11566     NA      NA    NA     NA
1047 1998    10   4 12422  16685  16170     NA     NA      NA    NA     NA
```

3

```
     count1.2 count2.2 count3.2 comments    case uniqid
1046       NA       NA       NA     \"\" \"3-6\" 910030
1047       NA       NA       NA     \"\" \"3-6\" 910030
```

For this station, the recounting dates overlap with the original counting dates. Kent and Larry are currently investigating this station.

3.2 Issues Specific to FILEE

All queries found to date in FILEE are given in this subsection, but queries related to stations outside the 5-county test area are currently of *secondary* concern and do not need immediate attention. Queries slated with secondary priority will be so listed in this subsection.

3.2.1 Resolution of AADT

It appears that the four AADT variables have been rounded to the nearest hundredth or the nearest tenth (for AADTs less than 1000). Because this is too coarse a resolution for the proposed statistical analysis, the AADTs provided in FILEL will always be used; the AADTs in FILEE will never enter the statistical analysis. Is this okay? If not, is there a way to get the actual AADTs in FILEE and not rounded versions?

3.2.2 Stations Missing Key Variables or Variables Have Unusual Values

Variable sips has value 0 for 351 stations in Alamance county. Looking at the records of other stations in Alamance county, it appears the missing sips is 101. This value has been substituted. Is this okay? *(secondary)*

Variable co.num has value 0 for (all) 897 stations in Alamance county. What is this county number? *(secondary)*

Variable dist for station 480540 in Iredell county has value 543, while all others have value either 1 or 2 or 3. The dist=543 appears to be an error. What is the correct value? *(secondary)*

Variables routecode and routenum are missing for 4,337 and 4,265 stations across almost all counties. This is partly the reason Shannon McDonald, Erik, Kent, Larry, and Susan Cosper expended great effort in determining strategies for imputing missing values; details are provided in the December 2001 report "Recordings and Methodology to Calculating Capacity Variables for Unmatched Stations" written by Shannon McDonald. As a remedy, the original variables routecode and routenum have been overwritten using variable route1 with the following rule:

routecode is the first digit of route1
routenum is the last four digits of route1

With this change, only 70 records have routecode=9 (unknown). On the other hand, 10,356 records now have routenum=0 (unknown). Variable routenum is not expected to be used in the statistical modeling, so the fact that so many values are unknown should not be a problem. Are these adjustments and conclusions okay?

Variable `description` is missing for 2,417 stations across almost all counties. This variable is not expected to be important and so nothing will be done to remedy this omission. **Is this okay?** *(secondary)*

Variable `distflag` takes value 100 for 1,353 stations: 192 in Chatham, 248 in Durham, 242 in Johnston, 184 in Orange, and 487 in Wake counties. In all other cases, `distflag` takes value 0. This variable was discussed at the 2/21/02 meeting, at which time it was decided that it represents a successful "snap" (to within 30 meters) after the adjustments performed by Erik. Erik thought that no stations should have `distflag` values different from 0. **Will the x- and y-coordinates need adjustment for these stations?** A related variable is `distance`. At the 2/21/02 meeting, the conclusion was that `distance` should never exceed 30, but 858 stations have distance exceeding 30 and even achieving a maximum of 119.9. Moreover, these large values are not related in any obvious way to `distflag`>0; they are not limited to the 5-county test area but instead are spread over the entire state. **Any advice?**

Variables `level2` and `lucode` are either missing or have value 0 for 96 stations in Alexander (3), Bertie (1), Caldwell (61), Cleveland (1), Craven (1), Duplin (2), Forsyth (1), Gates (1), Jackson (11), Johnston (1), Montgomery (1), Onslow (2), Person (1), Polk (1), Surry (3), Watauga (1) and Wayne (4) counties. **Are these values available?** *(secondary)*

Variables `state90`, `county90`, `tract90`, `group90` are all missing for the following stations:

	uniq.id	sips	station	co.num	co.name	cycle.yr	div	dist	routecode	routenum	description	atr.grp
10944	940060	94	60	94	WATAUGA	A	11	2	3	88	SWOFSR1343	1
14545	850096	85	96	85	SURRY	A	11	1	3	104	NEOFSR1725	1
14570	850071	85	71	85	SURRY	A	11	1	3	0	NOFNC18	2
23260	720001	72	1	72	PERSON	A	5	2	2	0	NEOF1374	1

	axle.fac	count.yr1	aadt.yr1	count.yr2	aadt.yr2	count.yr3	aadt.yr3	count.yr4	aadt.yr4
10944	98	1999	940	1998	900	1997	910	1996	800
14545	99	1999	2300	1998	2300	1997	2200	1996	2900
14570	89	1999	1800	1998	1900	1997	1700	1996	1500
23260	85	1999	3700	1998	3300	1997	3200	1996	3100

	giscount	symbol	fips	distflag	x.coord	y.coord	distance	level2	lucode	state90	county90
10944	940	1	0	0	365039.5	293305.7	0.221		0	NA	NA
14545	2300	1	0	0	469616.9	312435.2	14.556		0	NA	NA
14570	1800	1	0	0	439631.7	313558.2	9.243		0	NA	NA
23260	3700	1	0	0	618409.6	309757.2	0.000		0	NA	NA

	tract90	group90	acntrl	speed	lanes	route1	level	color
10944	NA	NA	1	55	2	30000088	0	0
14545	NA	NA	1	55	2	30000104	0	0
14570	NA	NA	1	53	2	30000000	29	5
23260	NA	NA	1	53	3	20000000	28	3

We know `state90` must take value 37, and `county90` can be determined by comparing to other stations in the same county (189 for Watauga, 171 for Surry, and 145 for person), but the values for `tract90` and `group90` are not so obvious. **What are these values?** *(secondary)*

5

The process of interpreting and imputing values for `speed`, `lanes`, `route1`, `level`, and `color` has been addressed in "Recordings and Methodology to Calculating Capacity Variables for Unmatched Stations."

3.2.3 Stations With Multiple Records

Seven stations (none in the 5-county test area) have multiple records, as given below. The records differ because of variables `cycle.yr`, `(x.coord, y.coord)`, `distance`, `level2`, `lucode`, `county90`, `tract90`, `group90`, `acntrl`, `speed`, `lanes`, `route1`, `level`, or `color`. What are the correct values, and does this indicate a problem with these variables for other stations?

	uniq.id	sips	station	co.num	co.name	cycle.yr	div	dist	routecode	routenum	description
1837	1100067	110	67	33	FORSYTH	A	9	2	3	67	W OF PETERS
1838	1100067	110	67	33	FORSYTH	V	9	2	5	0	W OF PETERS
4274	1050566	105	566	35	GASTON	V	12	1	2	321	N OF SR 2406
4275	1050566	105	566	35	GASTON	V	12	1	4	1932	N OF SR 2406
13951	851788	85	1788	85	SURRY	E	11	1	4	1751	S OF SR 1753
13952	851788	85	1788	85	SURRY	O	11	1	4	1756	S OF SR 1753
32243	480635	48	635	48	IREDELL	E	12	2	4	1607	
32244	480635	48	635	48	IREDELL	E	12	2	4	1599	
32328	480540	48	540	48	IREDELL	E	12	543	4	0	E OF SR 2120
32329	480540	48	540	48	IREDELL	E	12	2	4	2124	E OF SR 2120
38216	330092	33	92	33	FORSYTH	A	9	2	4	1611	WOFSR1688
38217	330092	33	92	33	FORSYTH	O	9	2	4	2018	WOFSR1688
46522	101565	10	1565	10	BUNCOMBE	V	13	2	4	0	E OF SR 2139
46523	101565	10	1565	44	BUNCOMBE	V	13	2	4	1358	E OF SR 2139

	atr.grp	axle.fac	count.yr1	aadt.yr1	count.yr2	aadt.yr2	count.yr3	aadt.yr3	count.yr4	aadt.yr4
1837	1	99	1998	13000	1995	14000	1993	12000	1991	15000
1838	1	99	1998	13000	1995	14000	1993	12000	1991	15000
4274	4	90	1998	18000	1995	18000	1993	18000	1991	18000
4275	4	90	1998	18000	1995	18000	1993	18000	1991	18000
13951	1	99	1999	1700	1997	1500	1995	1400	1993	1500
13952	1	99	1999	1700	1997	1500	1995	1400	1993	1500
32243	1	99	1998	20	1996	30	1994	20	1992	10
32244	1	99	1998	20	1996	30	1994	20	1992	10
32328	1	99	1998	40	1996	60	1994	50	1992	60
32329	1	99	1998	40	1996	60	1994	50	1992	60
38216	1	98	1999	4500	1998	4500	1997	5600	1996	3800
38217	1	98	1999	4500	1998	4500	1997	5600	1996	3800
46522	1	99	1999	1200	1997	700	1995	930	1991	900
46523	1	99	1999	1200	1997	700	1995	930	1991	900

	giscount	symbol	fips	distflag	x.coord	y.coord	distance	level2	lucode
1837	13000	3	67	0	490275.7	268420.4	0.162	RESIDENTIAL	11
1838	13000	3	67	0	496244.8	259695.9	5.633	COMMERCIAL AND SERVICES	12
4274	18000	3	71	0	409308.9	165493.0	0.000	INDUSTRIAL	13
4275	18000	3	71	0	427156.5	176558.6	0.000	DECIDUOUS FOREST LAND	41
13951	1700	1	171	0	468862.6	306944.7	1.024	RESIDENTIAL	11
13952	1700	1	171	0	467937.6	306554.8	2.331	RESIDENTIAL	11

6

32243	20	1	97	0	431045.9	254233.5	3.253	CROPLAND AND PASTURE	21
32244	20	1	97	0	430741.7	254078.9	3.321	CROPLAND AND PASTURE	21
32328	40	1	97	0	452315.0	247438.4	7.722	CROPLAND AND PASTURE	21
32329	40	1	97	0	452214.7	248516.9	5.581	CROPLAND AND PASTURE	21
38216	4500	1	67	0	489025.2	270207.1	7.090	CROPLAND AND PASTURE	21
38217	4500	1	67	0	516161.3	269847.2	8.000	CROPLAND AND PASTURE	21
46522	1200	1	21	0	289318.6	226238.7	14.091	CROPLAND AND PASTURE	21
46523	1200	1	21	0	289235.6	192321.6	24.187	TRANS/COMM/UTIL	14

	state90	county90	tract90	group90	acntrl	speed	lanes	route1	level	color
1837	37	67	2701	2	1	35	4	30000067	0	0
1838	37	67	10	1	1	25	2	50000000	35	7
4274	37	71	334	1	1	45	4	20000321	0	0
4275	37	71	301	3	1	55	2	40001932	0	0
13951	37	171	9902	3	1	55	2	40001751	0	0
13952	37	171	9902	3	1	55	2	40001756	0	0
32243	37	97	609	1	1	55	2	40001607	0	0
32244	37	97	609	1	1	55	2	40001599	0	0
32328	37	97	608	2	1	50	2	40000000	32	62
32329	37	97	608	2	1	55	2	40002124	0	0
38216	37	67	2801	3	1	55	2	40001611	0	0
38217	37	67	3104	4	1	55	2	40002018	0	0
46522	37	21	28	4	1	50	2	40000000	32	62
46523	37	89	9906	2	1	35	2	40001358	0	0

3.3 Issues From Merging FILEL and FILEE

3.3.1 Stations Exist in FILEE but not in FILEL

Within the 5-county test area, 1,054 stations exist in FILEE but not in FILEL. Of these, 1,010 have AADTs (for the most recently counted year) less than 1,000 and 44 have AADTs at least 1,000. Early in this project, a decision was made to exclude from the statistical analysis all stations with AADT less than 1,000; this accounts for 1,010 of the 1,054 stations that could not be reconciled between FILEL and FILEE. Another early decision was to only include data for count years 1998 and later. Because 37 of the 44 stations with AADT at least 1,000 have most recent count years before 1998, this explains them not being included in FILEL. Data for the seven remaining stations is given below. Were they excluded because they are "soil" segments?

	uniq.id	sips	station	co.num	co.name	cycle.yr	div	dist	routecode	routenum	description
24374	670725	67	725	67	ORANGE	E	7	1	4	1395	N OF SR 1328
5788	1030554	103	554	31	DURHAM	V	5	2	4	2145	N OF NC 54
5794	1030548	103	548	31	DURHAM	V	5	2	4	1999	SWOF I 40
6022	1030301	103	301	31	DURHAM	V	5	2	1	85	EX173B TO X172
6118	1030183	103	183	31	DURHAM	V	5	2	5	0	EOFSR1445
6179	1030119	103	119	31	DURHAM	V	5	2	4	1127	E OF KENT ST
2261	1080895	108	895	91	WAKE	V	5	1	9	0	SW OF I-40

	atr.grp	axle.fac	count.yr1	aadt.yr1	count.yr2	aadt.yr2	count.yr3	aadt.yr3	count.yr4	aadt.yr4
24374	1	99	1998	1000	1996	1100	1994	1100	1992	1100
5788	1	91	1999	10000	1997	18000	1994	8200	1992	7500

5794	1	99	1999	1700	1994	15000	1992	11000	1990	9000
6022	1	82	1999	42000	1998	36000	1997	45000	1996	33000
6118	1	98	1999	5000	1994	5200	1992	6500	1990	6900
6179	4	98	1999	17000	1994	16000	1992	16000	1990	14000
2261	4	92	1998	68000	1996	52000	1993	65000	1991	47000

	giscount	symbol	fips	distflag	x.coord	y.coord	distance	level2	lucode
24374	1000	1	135	0	599358.3	259784.9	1.558	RESIDENTIAL	11
5788	10000	2	63	100	620776.2	238332.6	7.262	MIXED FOREST LAND	43
5794	1700	1	63	0	622095.2	237603.4	6.404	COMMERCIAL AND SERVICES	12
6022	42000	5	63	0	613065.8	252990.1	3.300	TRANS/COMM/UTIL	14
6118	5000	1	63	100	618190.2	250210.2	0.010	INDUST & COMMERC CMPLXS	15
6179	17000	3	63	0	617139.9	249243.9	0.175	RESIDENTIAL	11
2261	68000	6	183	0	633001.9	223481.9	5.796	TRANS/COMM/UTIL	14

	state90	county90	tract90	group90	acntrl	speed	lanes	route1	level	color
24374	37	135	111	1	1	35	2	40001395	0	0
5788	37	63	2097	3	1	55	4	40002145	0	0
5794	37	63	2097	3	1	55	4	40001999	0	0
6022	37	63	1707	9	3	55	3	10000085	0	0
6118	37	63	302	3	1	25	2	50000000	35	7
6179	37	63	5	3	2	35	3	40001127	0	0
2261	37	183	53001	2	1	30	1	95000000	47	68

3.3.2 Stations Existing in Both FILEL and FILEE Have Different Values for Variables

FILEL and FILEE were merged by variable uniqid, resulting in a total of 3508 records (corresponding to 3411 unique stations) each containing 63 variables. Given a particular station, however, some variables have equal values; for example, the routecode variables from FILEL and FILEE should match exactly. Unfortunately, the appropriate variables do not always match.

Cross-classification tables of routecode from FILEL and FILEE result in 339 unmatched cases, as shown below:

	routecode from FILEE					
	1	2	3	4	5	9
routecode from FILEL 1	78	0	0	0	0	1
2	4	358	5	13	11	1
3	1	14	308	21	2	0
4	1	35	24	2251	167	5
5	0	7	4	22	174	1

A single record that illustrates this is given below, where routecode is from FILEL and routecode1 is from FILEE:

	uniqid	countyid	stationid	urb.num	routecode	routenum	description	atrgroup	year	month	day
31	180038	18	38	-1	\"2\"	\"421\"	\"W OF SR 1105\"	1	1998	9	4

8

```
   aadt count1 count2 count3 year.2 month.2 day.2 aadt.2 count1.2 count2.2 count3.2 comments
31 2321   2970   2970     NA     NA      NA    NA     NA       NA       NA       NA     \"\"

      case sips station co.num co.name cycle.yr div dist routecode1 routenum2 description3 atr.grp
31 \"3-4\"   18      38     18 CHATHAM        A   8    1          4      1006    SOFSR1310       1

   axle.fac count.yr1 aadt.yr1 count.yr2 aadt.yr2 count.yr3 aadt.yr3 count.yr4 aadt.yr4 giscount
31       81      1998     2300      1997     2200      1996     2500      1995     2100     2300

   symbol fips distflag  x.coord  y.coord distance                level2 lucode state90 county90
31      1   37        0 563927.6 223326.5     5.81 DECIDUOUS FOREST LAND     41      37       37

   tract90 group90 acntrl speed lanes   route1 level color
31     202       1      1    55     2 40001006     0     0
```

Station 180038 (as shown above) also illustrates the problem of comparing routenum (FILEL) with routenum2 (FILEE) and comparing description (FILEL) with description3 (FILEE). These comparisons result in 650 and 3067 unmatched cases, respectively.

There are 83 unmatched cases for comparing variables atrgroup (FILEL) and atr.grp (FILEE), with an example provided below:

```
                 atr.grp
                  from
                  FILEE
               1 2   4  5
atrgroup  .    1 0   9  0
  from     1 2784 0   0  0
  FILEL  11   32 2  14  0
         12   12 0   0  0
         13    0 0   0 13
          4    0 0 641  0

    uniqid countyid stationid urb.num routecode routenum        description atrgroup year month
308 310011       31        11     103     \"1\"   \"40\" \"EX282 TO EX283\"        . 2000     .

    day   aadt count1 count2 count3 year.2 month.2 day.2 aadt.2 count1.2 count2.2 count3.2
308   . 139807     NA     NA     NA     NA      NA    NA     NA       NA       NA       NA

                   comments      case sips station co.num co.name cycle.yr div dist routecode1
308 \"not control station\" \"INTER\"   31      11     31  DURHAM        A   5    2          1

    routenum2   description3 atr.grp axle.fac count.yr1 aadt.yr1 count.yr2 aadt.yr2 count.yr3
308        40 EX282 TO EX283       1       82      1999   134000      1998   128000      1997

    aadt.yr3 count.yr4 aadt.yr4 giscount symbol fips distflag  x.coord  y.coord distance
308   131000         0        0   134000      7   63        0 624356.1 236148.6     9.57

             level2 lucode state90 county90 tract90 group90 acntrl speed lanes   route1 level
```

9

```
308 TRANS/COMM/UTIL     14      37       63    2097       4      3    65     3 10000040     0

    color
308     0
```

There are 257 unmatched cases for comparing variables year (FILEL) and count.yr1 (FILEE), with an example provided below:

```
              count.yr1
                from
                FILEE
            1997 1998 1999
year   1997    0    0   74
from   1998    0 1126    5
FILEL  1999    2    0 2125
       2000    0  104   72
  uniqid countyid stationid urb.num routecode routenum      description atrgroup year month day
9 180016       18        16      -1     \"3\"  \"902\" \"S OF SR 1176\"        1 2000     8   3

  aadt count1 count2 count3 year.2 month.2 day.2 aadt.2 count1.2 count2.2 count3.2 comments
9 2123   2255   2168     NA     NA      NA    NA     NA       NA       NA       NA     \"\"

     case sips station co.num co.name cycle.yr div dist routecode1 routenum2 description3 atr.grp
9 \"2-1\"   18      16     18 CHATHAM        A   8    1          3       902    NOFSR2302       1

  axle.fac count.yr1 aadt.yr1 count.yr2 aadt.yr2 count.yr3 aadt.yr3 count.yr4 aadt.yr4 giscount
9       99      1999     3000      1998     2900      1997     2900      1996     2900     3000

  symbol fips distflag  x.coord  y.coord distance                  level2 lucode state90 county90
9      1   37        0 574682.6 207100.6        0 OTHER URBAN OR BUILT-UP     17      37       37

  tract90 group90 acntrl speed lanes   route1 level color
9     205       2      1    55     2 30000902     0     0
```

Of all the variables mentioned in this subsection, routecode, atrgroup and year are probably the most important. Are adjustments needed?

APPENDIX 2:

2. Procedure and Corrections of PTC Data

Shannon McDonald

May 2002

Procedures and Corrections of PTC Data

This report is a procedural documentation and follow-up to the issues outlined in Jackie Hughes-Oliver's report *Documentation of Data Edits, Queries, and Issues in Reconciliation* (February 28, 2002).

1. Merging the mapping road linework with the LRS data.

Input data: Converted CADD road layer (mapping) – includes the entire statewide inventory of roads and an attribute that indicates route classification (LDGS_LEVEL).

LRS road coverage – state-maintained road network includes Universe data.

Output data: MERGEROADS.SHP

The input datasets were merged to create a statewide road network inventory. The LRS data has been joined to the segments along routes where they are available. The LDGS_LEVEL attribute from the converted road layer is also retained and is available for the entire dataset.

An attribute (LRS_MATCH) was added to MERGEROADS to indicate the records that are attributed with the LRS data. Forty-one percent of the road segments are assigned attributes from the LRS:

LRS Status	Segments (Count)	Length (km)
Matched	388712	125553
Unmatched	561483	153144

2. Assigning route classifications for the road network.

Input data: MERGEROADS.SHP

Output data: MERGEROADS.SHP – with route type attribute RTE_TYPE added

The route classification will be used for several steps in the data preparation process, as well as for use with the project analysis. Each road segment route is presented as an eight-digit number in the MERGEROADS database within the ROUTE1 attribute. The first digit of the number indicates the route classification. For example, where ROUTE1 = 10400077, the route classification is 1 (interstate).

Where available, the ROUTE1 values were used to calculate the route type (held within RTE_TYPE). The ROUTE1 field values are available for 415,492 segments (44%). For the others, the LDGS_LEVEL field was used:

LDGS_LEVEL	Assigned Route Classification (RTE_TYPE)
27	1 (interstate)
28	2 (US)
29	3 (NC)
32	4 (SR)
35	5 (Local)
38	9 (park roads)
All others	7 (miscellaneous)

3. Speed limit, access control, and number of lanes.

Input data: MERGEROADS.SHP

MB.SHP – municipal boundary shapefile for North Carolina

Output data: MERGEROADS.SHP

For the LRS-matched road segments, the speed limit and number of lanes are available within the SPEED and LANES fields, respectfully. The access control was not available at the time of this report.

The speed limits and lane values for the LRS-matched records were separated into two groups—urban and rural—for each route classification. The determinant for urban roads was the municipal boundary shapefile MB.SHP. If the road segment intersects with a feature of this shapefile, the road was tagged as urban. An attribute named URBANFLG was added to MERGEROADS to indicate when a road segment is urban. The values were analyzed by these sub-classifications for the assignment of the nonmatched segments. The assignments were based primarily on the average (mean) of the speed limit and lane values of the matched segments. For the speed limit, the average was rounded up to the nearest 5

At the time of this report, there exists no attribute to indicate a route segment is divided. The NCDOT GIS Unit is working on this presently. However, the interstate highway system (route class 1) has been adjusted to indicate the number of lanes per segment (i.e. LANES = 2 for a four-lane, divided highway). Other route classes are not guaranteed to be correct at this time.

Interstate routes

6160 of 6500 matched (94.8%)

	Urban	Rural
Matched	3030	3130
Nomatch	222	118

Statistic	Urban SL	Rural SL	Urban Lane	Rural Lane
Mean	60	64	3	2
Min	35	55	1	1
Max	70	70	6	4
Var	31	26	1	0
SD	6	5	1	1
Assignment	**60**	**65**	**3**	**2**

US routes

33618 of 40821 matched (82.4%)

	Urban	Rural
Matched	15049	18569
Nomatch	3404	3799

Statistic	Urban SL	Rural SL	Urban Lane	Rural Lane
Mean	41	53	3	3
Min	20	20	1	2
Max	65	70	8	6
Var	86	31	1	1
SD	9	6	1	1
Assignment	40	55	3	3

NC routes

42752 of 45413 matched (94.1%)

	Urban	Rural
Matched	13812	28940
Nomatch	1429	1232

Statistic	Urban SL	Rural SL	Urban Lane	Rural Lane
Mean	40	53	3	2
Min	20	20	2	1
Max	60	60	7	6
Var	72	22	1	0
SD	8	5	1	0
Assignment	40	55	3	2

SR routes

306182 of 324983 matched (94.2%)

	Urban	Rural
Matched	52295	253887
Nomatch	3185	15616

Statistic	Urban SL	Rural SL	Urban Lane	Rural Lane
Mean	39	53	2	2
Min	20	20	1	1
Max	55	65	7	6
Var	62	36	0	0
SD	8	6	1	0
Assignment	40	55	2	2

Local routes

No class 5 (local) routes were matched. All are **assigned a value of 25 mph and 2 lanes**, the same as the assignment in the original dataset (reported on 12/03/02). All other route types (7 and 9) are assigned

values of 25 mph and 2 lanes as well (these routes will not be used in this model, but may in the routing process).

4. Joining the TSU data to the PTC shapefile.

Input data: PTC.DBF – PTC station location attributes associated with PTC.SHP

TBL_STAT.DBF – TSU attribute data for traffic counts

Output data: PTC.DBF

These two tables have a one-to-one relationship and were joined using the PTC unique identifier UNIQ_ID (UNIQ_ID7 in TBL_STAT.DBF). The output includes both attributes from the original PTC.DBF and the joined table TBL_STAT.DBF. Of 51,025 stations in the PTC shapefile, **35,029** had matching records from TBL_STAT.DBF. This included the entire TBL_STAT.DBF dataset. **These matches represent the PTC stations that will be modeled for this project.**

5. Snapping PTC stations in PTC.SHP to the base road network MERGEROADS.

Input data: PTC.SHP – PTC station location and attributes for statewide inventory (51,025 stations)
MERGEROADS – entire NC inventory of road network, including theattributes of the Universe data where available.

Output data: PTC_SNAP.SHP

NO-RTESNAP.DBF

First routine:

Initially, a snapping routine was invoked that snapped stations to the road network where their road classifications were the same. For example, using this method, all stations located along interstates would only have an opportunity to snap to segments in the road network that represent interstates. This was done by selecting the road classifications from each dataset and then applying the snapping routine to the selected sets. For convenience, a separate PTC dataset (shapefile) for each classification was created.

The attributes used for selecting PTC station and road segments by route classification were ROUTECODE and RTE_TYPE, respectfully. (Note that the ROUTECODE attribute in the PTC shapefile and in the road network MERGEROADS is not the same). In each of these attributes, a single number (1-9) represents the road classification:

Route Classification	ROUTECODE (PTC shapefile)	RTE_TYPE (mergeroads)
Interstate	1	1
US	2	2
NC	3	3
SR	4	4
Local	5	5
Misc	NA	6-9

The following were selected from each data set during the routine:

Route Classification	PTC	Road Network
1	425	6500
2	4174	40821
3	4864	45413
4	24315	324983
5	17247	ALL

All stations with values within the ROUTECODE (i.e. ROUTECODE is not null) have been matched to the TSU file TBL_STAT.DBF and therefore are part of this model.

Class 5 includes all of the roads because the data is incomplete for PTC stations and inconsistent for the road network. Of the 17247 stations, 1251 are modeled (i.e. ROUTECODE = 5).

The routine created the following results:

Using a 120 meter snapping tolerance, 1361 stations did not match to the linework:

Route Classification	No Snap
1	1
2	261
3	316
4	768
5	15

Sixteen misses are due to spatial problems (outside the 120m tolerance). These include 1 interstate-class station and 15 class-5 stations. Of the 15 class-five stations, only 5 are to be modeled. A list of unsnapped stations were exported as a DBF file (NO-RTESNAP.DBF) and given to the TSU for review (NO-RTESNAP.DBF).

Second routine:

Because of the inconsistencies in data, the first routine was subject to errors, thus a second routine was invoked: The entire PTC station inventory was snapped to the nearest line segment regardless of route classification using a 200 meter tolerance. The distance the station moved (i.e. the distance from the nearest road segment) is recorded in the output file's DISTANCE attribute.

The output file for this routine is PTC_SNAP.SHP and included 51017 stations with 8 un-snapped points. Of these 8, none are modeled.

6. Post-processing of PTC_SNAP.

Input file: PTC_SNAP.SHP

Output file: PTC_MODEL.SHP

The next step was to eliminate the stations that are not modeled. The output file contains 35,029 stations.

7. Joining US Census data to the PTC station shapefile.

Input data: PTC_SNAP.SHP

DEMOGRAPHICS.SHP – statewide polygon coverage of the US Census Bureau block groups.

Output data: PTC_MODEL_INTERUM1.SHP

A spatial join was used to attach the US Census attributes to the PTC station shapefile. Most of the Census attributes were then stripped, leaving only the identifiers for joining to the rest of the table (STATE90, COUNTY90, TRACT90, and GROUP90).

8. Joining Landuse data to the PTC station shapefile.

Input data: PTC_MODEL_INTERUM1.SHP

LANDUSE.SHP – statewide polygon coverage and attributes at the Anderson II level.

Output data: PTC_MODEL_INTERUM1.SHP

A spatial join was used to attach the land use attributes to the PTC station shapefile. The attributes retained in PTC_MODEL_INTERUM1.SHP from LANDUSE.SHP include LUCODE (identifier) and LEVEL2.

9. Data corrections for US Census data

Input data: AREA.DBF – North Carolina county names and SIPS

FIPS_CNTY.DBF

PTC_MODEL_INTERUM1.SHP

Output data: PTC_MODEL_INTERUM1.SHP

Resource data: DEMOGRAPHICS.SHP

As noted in Dr. Jackie Hughes-Olivers' report (February 28, 2002 *Document of Data Edits, Queries, and Issues in Reconciliation* Section 3.2.2), several stations had conflicting information with the Census data. The data was checked thoroughly for consistencies in their county assignments to resolve these problems.

At this point, there is no relationship between the PTC data county assignment and the US Census Bureau county values (COUNTY90). The latter is coded with the federal code identifier (FIPS), while the PTC stations are assigned to counties by the state identifier (CO_NUM). These two numbers are not coincident.

To check consistencies, a table was joined to PTC_MODEL_INTERUM1.SHP. The AREA.DBF table contains the SIPS identifier attribute and the county name (AREA_NAME). The table was joined to the

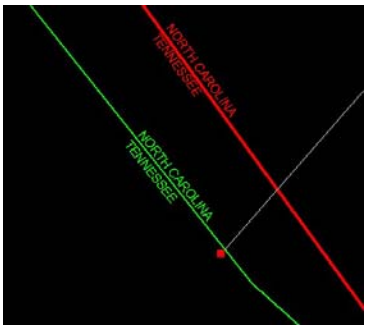
PTC shapefile by its CO_NUM value, and thus gives the county name for which the PTC station was digitized. A second table, FIPS_CNTY.DBF, was temporarily joined to the PTC_MODEL_INTERUM1.SHP file by the FIPS identifier. This table includes the county name as well (NAME), and indicates the county for which the census information is assigned.

By comparing the two county names within each record, discrepancies could easily be found:

SELECT WHERE [AREA_NAME] <> [NAME]

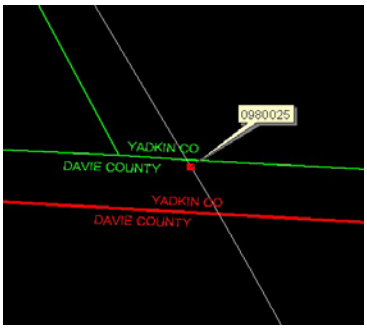
The result was 84 stations with differing county names. The problems fall into three categories, two of which are loosely related to the inconsistencies of the county line definitions between the NCDOT county maps and the US Census Bureau boundaries.

1. Stateline – four stations fell outside the NC state boundary. These stations had no data for the census identifiers. Identifiers from the census block group that is nearest to the station and intersects with the route segment for which the station is located are assigned to these stations.



UNIQ_ID	STATE90	COUNTY90	TRACT90	GROUP90
0720001	37	145	9801	1
0850096	37	171	9902	1
0850071	37	171	9908	1
0940060	37	189	9801	1

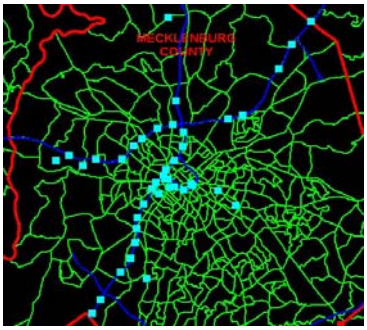
2. County line – 38 stations were assigned different census counties than their NCDOT assignments. The assignment was determined by selecting the closest block group within the assigned county (CO_NUM) that intersects with the route segment for which the station is located.



UNIQ_ID	DOT CNTY	CENSUS CNTY	OLD C90	T90	G90	NEW C90	T90	G90
0370001	GRAHAM	MACON	113	9702	3	075	9802	2
0490006	JACKSON	SWAIN	173	9603	2	099	9501	3
0491543	JACKSON	SWAIN	173	960198	2	099	9501	3
0440532	HENDERSON	BUNCOMBE	021	002201	4	089	9906	2
0100077	BUNCOMBE	MCDOWELL	111	9707	2	021	003102	4
0050015	AVERY	CALDWELL	027	0310	1	011	9902	3
0020035	ALLEGHANY	ASHE	009	9706	1	005	9503	3
0960030	WILKES	SURRY	171	9908	6	193	9602	1
0541731	LINCOLN	GASTON	071	030598	1	109	0704	3
0980025	YADKIN	DAVIE	059	0801	1	197	0505	7

0280010	DAVIDSON	FORSYTH	067	0036	9	057	0602	7
0160039	CASWELL	ROCKINGHAM	157	0401	6	033	9903	3
0400072	GUILFORD	FORSYTH	067	00302	5	081	016201	2
1070379	GUILFORD	FORSYTH	067	003306	8	081	0163	1
1070590	DAVIDSON	RANDOLPH	151	031601	9	057	0610	2
0611522	MONTGOMERY	MOORE	125	951398	1	123	960198	1
0761712	RICHMOND	MONTGOMERY	123	9605	4	153	9702	1
0821633	SCOTLAND	RICHMOND	153	9711	3	165	0105	4
0621581	MOORE	LEE	105	0305	1	125	9504	1
0621582	MOORE	LEE	105	0305	1	125	9505	1
0910679	WAKE	DURHAM	063	002097	3	183	053698	2
0180082	CHATHAM	WAKE	183	053403	1	037	0207	1
0341524	FRANKLIN	VANCE	181	9609	4	069	0601	2
0921662	WARREN	HALIFAX	083	9907	4	185	9501	5
0420083	HARNETT	CUMBERLAND	051	0037	1	085	0706	2
0420085	HARNETT	CUMBERLAND	051	0026	1	085	0705	6
0231647	COLUMBUS	BLADEN	017	9506	6	047	9905	1
0500100	JOHNSTON	SAMPSON	163	9702	3	101	0413	4
0630149	NASH	EDGECOMBE	065	0206	1	127	0106	5
0711501	PERQUIMANS	GATES	073	9701	3	143	9801	6
0711504	PERQUIMANS	CHOWAN	041	9902	1	143	9801	6
0711506	PERQUIMANS	CHOWAN	041	9902	1	143	9801	6
0711518	PERQUIMANS	CHOWAN	041	9902	1	143	9801	6
0711519	PERQUIMANS	CHOWAN	041	9902	1	143	9802	5
0970086	WILSON	WAYNE	191	0002	3	195	0010	1
0950145	WAYNE	GREENE	079	9502	2	191	0003	2
0530090	LENOIR	JONES	103	9803	2	107	0114	3
0891796	UNION	MECKLENBURG	119	005804	3	179	020301	4

3. Incorrect county assignment – 42 stations have incorrect values for their CO_NUM field. All fell within Mecklenburg County but had an assignment of Cabarrus. The CO_NUM field was changed from 12 to 59. The AREA_NAME was also changed from "Cabarrus" to "Mecklenburg". Most of these stations appear to be along interstate highways.



UNIQ_ID	UNIQ_ID	UNIQ_ID
1020074	1020077	1020084
1020085	1020130	1020133
1020362	1020364	1020373
1020375	1020382	1020410
1020417	1020425	1020428
1020433	1020496	1020498
1020524	1020527	1020541
1020561	1020568	1020586
1020591	1020593	1020611
1020621	1020645	1020718
1020723	1020771	1020775
1020798	1020976	1020988
1021043	1021059	1021133
1021170	1021255	1021260

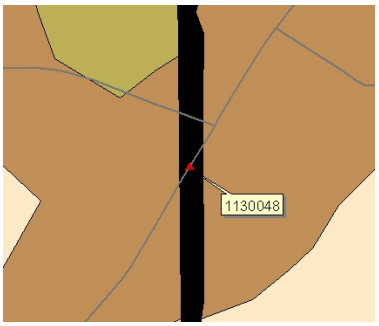
10. Data corrections for land use data.

Input data: PTC_MODEL_INTERUM1.SHP

Output data: PTC_MODEL_INTERUM1.SHP

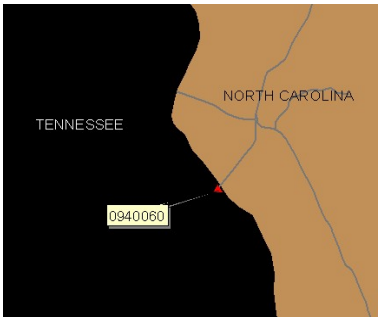
Resource data: LANDUSE.SHP

Similar problems are also present in the land use data assigned to the PTC stations. The statewide coverage was derived from separate tiles. Where these tiles are not perfectly aligned, polygons are formed that do not have any attribute information (called "slivers"). The solution was to determine the nearest polygon with data that intersects the route segment on which the PTC station is located and assign its attributes for LUCODE and LEVEL2:

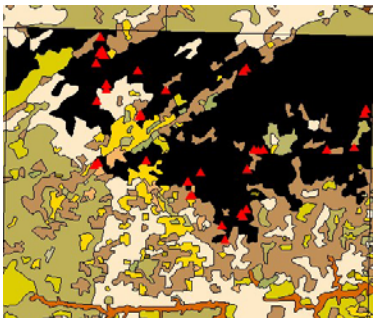


UNIQ_ID	LUCODE	LEVEL2
0070021	43	MIXED FOREST LAND
1130048	21	CROPLAND AND PASTURE
1130225	11	RESIDENTIAL
1130198	16	MXD URBAN OR BUILT-UP
0850096	21	CROPLAND AND PASTURE
0301601	43	MIXED FOREST LAND
0330019	11	RESIDENTIAL

In addition, there are four stations that lay outside the state boundary. (These are the same stations as those that lay outside the census data state boundary.) The data from the nearest LANDUSE polygon was assigned to these PTC stations:



UNIQ_ID	LUCODE	LEVEL2
0850071	41	DECIDUOUS FOREST LAND
0850096	21	CROPLAND AND PASTURE
0940060	21	CROPLAND AND PASTURE
0720001	43	MIXED FOREST LAND



One station (02401612) lay within an area without data (there appears to be a missing polygon in the land use data). The code assigned to this polygon was derived from locating the most abundant land use type within that particular area. The largest land use classification around the station was EVERGREEN FOREST LAND (LUCODE = 42).

Forty-eight stations have an LUCODE = 0 and a null LEVEL2 value. These stations reside on polygons that have no data or are missing. Since the areas are of missing data are large and classification cannot be determined from our dataset, no adjustments were made for these stations. (NOTE: most of these stations occur within a single tile in western NC).

11. Joining the MERGEROADS data to the PTC shapefile.

Input data: PTC_MODEL_INTERUM1.SHP

MERGEROADS.SHP

Output data: PTC_FINAL1.SHP

A spatial join was performed against the two input datasets to produce PTC_FINAL1.SHP. The database was stripped of attributes not relative to the model. All PTC stations modeled are snapped along a route segment, and thus all have attributes from this dataset (refer to Appendix A: Data Dictionaries for a complete list of fields).

Miscellaneous notes about data acquisition and integrity.

The field that contains the axle factor was missing from the original data acquired from the TSU (TBL_STAT.DBF). The axle factor data was joined to the PTC station database when it was discovered missing. The join, using TBL_UNIQ.DBF provided by TSU, was a one-to-one and complete match for the 35,029 stations.

Duplicate fields were stripped from the final dataset. A thorough check to ensure that the information was consistent for a give record was ensued before a field was deleted. The fields included two county identifiers (SIPS in PTC.SHP and COUNTYID in TBL_STAT.DBF), and two station identifiers (STATION in PTC.SHP and STATIONID in TBL_STAT.DBF). Duplicate fields with the same name were deleted during the join processes (within exporting the joined tables or creating a shapefile from a joined set). Also deleted was the urban identifier field URB_NO. The urban number can be derived using the SIPS field (where SIPS > 99).

Two fields, RTE_TYPE and ROUTECODE, have the same data definition, but do not contain the same values for every record. Both are single digit classifications for route type. For purposes related to statistical analysis of the TSU-provided data, the ROUTECODE values are valid. However, RTE_TYPE should be considered when accounting for speed limits and number of lanes (i.e. data derived from MERGEROADS). The inconsistencies are due to errors in both data sets. As reported above, stations that possess these discrepancies are being reviewed.

Appendix A: Data Dictionary for PTC_FINAL1.DBF

UNIQ_ID	Seven-digit unique identifier for PTC stations. Combination of the County SIPS code (SIPS) and the Station ID (STATION).
UNIQ_ID8	Eight-digit unique identifier for PTC stations. Combination of the County SIPS code (SIPS) and the Station ID (STATION). The eighth digit was placed after the SIPS (third digit) for future reference to the station. This identifier was not used during the process outlined in this report.
SIPS	State county and urban code. Counties are numbered from 0 to 99. SIPS values greater than 99 belong to an urban area and are assigned with the urban area code.
STATION	PTC station identifier. The identifiers are unique within each SIPS
CO_NUM	SIPS code for the county in which the station is located. The CO_NUM differs from the SIPS for those stations that are within urban areas.
CO_NAME	County name in which the PTC station was originally digitized (county from the NCDOT county maps).
CYCLE_YR	Classification for rotation of the traffic count data acquisition and analysis.
RD_SURFACE	Classification of road surface type.
ROUTECODE	Single-digit classification of route type assigned to the PTC stations. This classification reflects how the traffic data is acquired and analyzed.
ROUTENUM	Route number assigned to the segment on which the PTC station is located (originally in PTC.SHP, not MERGEROADS).
DESCRIPTIO	Description of the route on which the PTC station is located.
ATRGROUP	ATR group assignment for the PTC station
YEAR	
MONTH	
DAY	
AADT	
COUNT1	
COUNT2	
COUNT3	
YEAR_2	
MONTH_2	
DAY_2	
AADT_2	
COUNT1_2	
COUNT2_2	
COUNT3_2	
YEAR_NEW	
AADT_NEW	
MONTH_01	
DAY_01	
COUNT1_01	
COUNT2_01	
COMMENTS	
CASE	
AXLEFACTOR	Multiplier for accounting the variations in the number of axles per vehicle.
DISTANCE	Distance in meters the PTC station moved during the snapping process. Numbers are no greater than the snapping tolerance of 200 meters.

X_COORD	The X-coordinate location of the PTC station in NAD83 North Carolina State Plane coordinate system.
Y_COORD	The Y-coordinate location of the PTC station in NAD83 North Carolina State Plane coordinate system.
STATE90	State identifier for the US Census block groups. Used in conjunction with COUNTY90, TRACT90, and GROUP90 to join the entire block group dataset to the PTC database.
COUNTY90	County identifier for the US Census block groups. Used in conjunction with STATE90, TRACT90, and GROUP90 to join the block group dataset to the PTC database.
TRACT90	Census tract identifier for the US Census block groups. Used in conjunction with STATE90, COUNTY90, and GROUP90 to join the block group dataset to the PTC database.
GROUP90	Census block group identifier for the US Census block groups. Used in conjunction with STATE90, COUNTY90, and TRACT90 to join the block group dataset to the PTC database.
LUCODE	Land use classification (Anderson level 2)
LEVEL2	Land use description (Anderson level 2)
SPEED	Assigned speed limit to the route segment on which the PTC station is located.
LANES	Number of lanes represented by the route segment on which the PTC station is located. NCDOT GIS is in the process of incorporating a new attribute to indicate when this number references a divided highway drawn as multiple lines on the GIS coverage. At the time of this report, the LANES values are not consistent with the digitized segments and may or may not be representative of both lanes on a divided or undivided highway.
ROUTE1	9-digit Route classification and description of the segment on which the PTC station is located. The number is a code and contains several items of information including the route classification (interstate, US, NC, etc.) and route number. ROUTE1 was used to derive RTE_TYPE. ROUTE1 is not complete for the PTC dataset.
RTE_TYPE	Single-digit route classification for the route segment on which the PTC station is located. Derived from the values of ROUTE1 where ROUTE1 > 0. For cases when ROUTE1 = 0, the LGDS_LEVEL values were translated.
LRS_MATCH	Binary flag to indicate if the PTC station data is located on a segment that contains LRS attributes, and thus contain those values for LANES and SPEED. When LRS_MATCH = 1, the data is derived from the LRS system.
URBANFLG	Binary flag to indicate if the PTC station is located within an urban area as defined by the polygon features of MB.SHP (municipal boundary shapefile).

APPENDIX 3:

3. Data Description for Spatial Study

Larry Wikoff

May 2002

Data Description for Spatial Study
Larry Wikoff
May 17, 2002
Revision 1

This describes the data delivered to Dr. Jackie Hughes-Oliver in support of the Spatial Research Project for the Traffic Survey Unit of the North Carolina Department of Transportation.

General Approach
In general, this approach was used. There are some exceptions, as noted.

- The station list is the active station set as determined by the digitized station set as of 1998. However, any station whose most current count year was 1997 or earlier was eliminated.
- For primaries, which, for the purpose of this study includes all primary, secondary paved, and urban stations, the 1999 AADT was first examined, and if it was available and unadjusted, that calculated AADT and its associated raw data were delivered. If the 1999 count was unavailable or adjusted, a search was done for 2000 data, and then 1998 and 1997 in turn.
- Secondary soil stations were ignored if their AADT was less than 1000. If the AADT was 1000 or more, the 2000 data was searched, and if available, the AADT and raw data were delivered. If not, 1998 data was delivered.
- In some cases, where no other data was available, 2001 raw data and an AADT computed from that raw data were delivered.
- Case 99 was the catchall case, just in case the conditions were not captured by any of the previous case logic. All instances of Case 99 were examined to ensure that no serious error existed in the case selection logic.

Tables and files delivered

1. Spatial.Mdb – Access 97 database file with the following tables:
   - Area – a table of county names and county numbers.
   - TBL_STATE_DATA – the main table for data delivery.
   - TBL_SEASONAL – a special table containing the data that was collected for seasonal stations.
   - TBL_RECOUNTS – a special table containing only the stations that have multiple raw counts.
   - TblAddIntstate – a special table containing only the Interstate data. The Interstate is also contained in the main table, but was provided here for reference.
2. TBL_STATE_DATA.txt – space-delimited text version of TBL_STATE_DATA, explained above.
3. TBL_RECOUNTS.txt - space-delimited text version of TBL_RECOUNTS, explained above.
4. INTERSTATE.txt - space-delimited text version of tblAddInterstate, explained above.
5. TBL_SEASONAL.txt - space-delimited text version of TBL_SEASONAL, explained above.
6. TBL_UNIQ.bdf – dBase IV table of UNIQ_ID7, UNIQ_ID8, and AxleFactor. This was delivered after the initial delivery of the main table, when it was discovered that the axle factor was missing from the other tables.
7. TBL_UNIQ_AXLE.txt – space-delimited text version of TBL_UNIQ.dbf.

Main Table (TBL_STATE_DATA) Description
The table TBL_STATE_DATA is the main data set for the Spatial Project. Its content is described below.

Field Definitions - All numeric fields are integers, and the maximum field size is given on string fields. Within the text files, string data is enclosed in quotes.

Field Name	Type	**Description**
CountyID	Numeric	The county / Urban code, also called the SIPS code in the ArcView files.
StationID	Numeric.	The Station number.
URB_NUM	Numeric	The urban identifier if the station is in an urban area, -1 if not an urban station.
UNIQ_ID7	String(7)	7-character combination of County ID and Station ID, consisting of 3 characters for the county ID, and 4 characters for the Station ID; this is the same format as the corresponding ArcView field.
UNIQ_ID8	String(8)	8-character combination of County ID and Station ID, consisting of 3 characters for the county ID, and 5 characters for the Station ID; this is the format that is planned for future use within the Traffic Survey Unit.
RouteCode	String (1)	Route type identifier. 1 is Interstate, 2 is US, 3 is NC, 4 is secondary, and 5 is local.
RouteNum	String (25)	The road's route number, or, in the case of urban areas, the street name.
Description	String (30)	Text description of the station location
ATRGroup	Numeric	The ATR group assignment
Year	Numeric	The count year for which the raw data and resulting factored count were taken.
Month	Numeric	The month that the raw data was taken, if raw data is provided.
Day	Numeric	The day of week that the raw data was taken. 1 represents Sunday, 2 is Monday, and so on.
AADT	Numeric.	The Average Annual Daily Traffic count, as computed from the raw data, and factored by the axle correction factor and the seasonal factor, except when, based on case logic, an adjusted count is captured.
Count1	Numeric	First day's raw count, if raw data is provided
Count2	Numeric	Second day's raw count, if raw data is provided
Count3	Numeric	Third day's raw count, if raw data is provided
Year_2	Numeric	Count year for second set of data, if applicable.
Month_2	Numeric	Count month for second set of data, if applicable.
Day_2	Numeric	Count day for second set of data, if applicable
AADT_2	Numeric	The Average Annual Daily Count for the second set of data, if applicable.
Coun1_2		First day's raw count for second set of data, if applicable
Count2_2		Second day's raw count for second set of data, if applicable.
Count3_2		Third day's raw count for second set of data, if applicable.
Year_New	Numeric	The year for which a new AADT is computed, if applicable.
AADT_New	Numeric	New AADT for special circumstances, if existing AADT is

		not accepted.
Month_01	Numeric	Month for 2001 raw data if it is provided (for example, for stations with only adjusted counts).
Day_01	Numeric	Day for 2001 raw data if it is provided (for example, for stations with only adjusted counts).
Count1_01	Numeric	First day's count for 2001 raw data if it is provided (for example, for stations with only adjusted counts).
Count2_01	Numeric	Second day's count for 2001 raw data if it is provided (for example, for stations with only adjusted counts).
Comments	String (25)	General comments about the data records.
Case	String (5)	The select case number used in the Visual Basic code. These are specific to the logic that produces the data, and are explained in detail later.

Data Selection Logic
So that the source of the data can be identified, a case number is provided for each data record. Since the data for the pilot area was delivered in November 2001, the code that was used for data extraction has been carefully inspected, and some changes have been made. These changes are noted in the following paragraph.

To ensure that the logic that was used to derive the data was consistent for both the statewide data set and the pilot project data set, the entire data set was reassembled.

Notes about case numbers:
There have been some changes to the case numbers from the time of the delivery of the original pilot data set for the local area. To be specific:

- Cases 4-7 and 4-8 were redundant and were eliminated.
- Case 5-6 was changed to capture a slightly different situation. The situation that it originally captured was already captured by another logic test.
- Cases 5-6 and 5-7 were deleted because their specific combination of year and route type are trapped in another case.
- Cases 2-8 and 2-9 were added.
- Case 2001 was added – In cases where the only data available was from 2001, rather than omit the station from the data set, 2001 data was delivered if it was available and acceptable.
- Case 6-1 was added. These are cases where there was 2001 data, but the 2001 data was not acceptable, so a new factored count was generated, based on edit 2001 data. In most cases, the edits to the raw data were obvious, such as discarding an incomplete day's count, for example.
- Case "INTER" is special for the Interstate stations, which are calculated by use of a balancing algorithm. There is raw data only for the stations that are used as controls for those calculations.
- "AADT Accepted" implies an annual average daily count that was computed by factoring raw counts, and that has not been adjusted.
- During our quality control checks of the data, we noticed some anomalies in the day of week for some of the raw data records, which would have resulted in the use of incorrect factors for computing the AADT. For these, raw data was extracted from the special seasonal data counts, then factored into an AADT. The special case "SEAS" was assigned to these stations. Note that there is no raw data for the seasonal stations in the main table; the raw data is a different format, and the table which contains it has a separate explanation for its content.
- There were a number of stations of stations that had more than one set of raw data. In each of these cases, the raw data that resulted in the calculation of the stated AADT was delivered in the main table. All the raw data, including that delivered in the main table, is contained in a separate table called TBL_RECOUNTS.

Description of Case Numbers

Case	**Description**	**Results in Output File**
1-1	Primary or Secondary Station counted in 1999, AADT accepted, one 1999 raw data record	1999 Calculated AADT and associated raw data
1-2	Like 1-1, except no raw data	1999 Calculated AADT with no raw data
1-3	Like 1-1, except multiple 1999 raw data records	1999 Calculated AADT and associated raw data in multiple records
2-1	Primary counted in 1999, but it has an adjusted	2000 CALCULATED AADT

	AADT, and 2000 AADT is accepted, and has one 2000 raw data record	and raw data
2-2	Like 2-1, but no 2000 raw data	2000 Calculated AADT and no raw data
2-3	Like 2-1, but multiple 2000 raw data records	2000 Calculated AADT and raw data in multiple records
2-4	Primary counted in 1999. The 2000 AADT was adjusted or not taken, and the 1998 AADT is accepted, and one 1998 raw data record	1998 Calculated AADT and associated raw data
2-5	Like 2-4, except no 1998 raw data	1998 Calculated AADT and no raw data
2-6	Like 2-4, except multiple 1998 raw data records	1998 Calculated AADT and associated raw data in multiple records
2-7	Primary counted in 1999. The 2000 AADT was adjusted or not taken, AADT adjusted for 1998, and the 1997 AADT is accepted	1997 Calculated AADT and associate raw data
2-8	Like 2-7, but with no 1997 raw data	1997 Calculated AADT and no raw data
2-9	Like 2-7, but with multiple raw data records	1997 Calculated AADT and associated raw data in multiple records.
3-1	Primary counted in 1998, 2000 AADT accepted, and one 2000 raw data record	2000 Calculated AADT and associated raw data
3-2	Like 3-1, except no 2000 raw data	2000 Calculated AADT and no raw data
3-3	Like 3-1, except multiple 2000 raw data records	2000 Calculated AADT and associated raw data in multiple records
3-4	Primary counted in 1998, AADT accepted, adjusted or not counted in 2000, one 1998 raw data record	1998 Calculated AADT and associated raw data
3-5	Like 3-4, except no 1998 raw data	1998 Calculated AADT and no raw data
3-6	Like 3-4, except multiple 1998 raw data records	1998 Calculated AADT and associated raw data in multiple records
3-7	Primary counted in 1998, has adjusted AADT, adjusted or not counted in 2000, but has accepted 1997 AADT and single 1997 raw data record	1997 Calculated AADT and associated raw data
3-8	Like 3-7, but no raw 1997 raw data	1997 Calculated AADT and no raw data
3-9	Like 3-7, but multiple 1997 raw data records	1997 Calculated AADT and associated raw data in multiple records
4-1	Secondary counted in 1998, has AADT more than 999, counted 2000 AADT accepted, one 2000 raw data record	2000 Calculated AADT and associated raw data

4-2	Like 4-1, except no 2000 raw data record	2000 Calculated AADT and no raw data
4-3	Like 4-1, except multiple 2000 raw data records	2000 Calculated AADT and associated raw data in multiple records
4-4	Secondary with 1998 AADT accepted, has AADT more than 999, adjusted count or not counted in 2000, one 1998 raw data record	1998 Calculated AADT and associated raw data
4-5	Like 4-4, except no 1998 raw data records	1998 Calculated AADT and no raw data
4-6	Like 4-4, except multiple 1998 raw data records	1998 Calculated AADT and associated raw data in multiple records
5-4	Secondary counted in 1999, adjusted count, also has adjusted count in 1997	1999 and 1997 Calculated AADTs and associated 1999 and 1997 raw data
5-5	Primary count in 1999, adjusted count, adjust count or not counted in 2000, has adjusted count in 1998	1999 and 1998 adjusted AADT, and associated raw data
2001	Special case for 2001 data when all other sources had been searched and found lacking.	AADT computed from 2001 raw data, and associated raw data.
6-1	Special case for edited raw 2001 data	See Note 1.
INTER	Special case for Interstate station data	For control station, calculated AADT and associated raw data; for non-control stations, AADT only
SEAS	Special case for extracted seasonal data	AADT calculated from the seasonal raw data. See Note 2.
99	Anything else	Nothing

Notes:

1. In some cases, when no other data was available, 2001 data was sought. If the 2001 raw data was also not acceptable as is, and the raw data could be “fixed” by a simple edit, the edit was done, the edited raw data was used to compute a new factored AADT, and the raw data and the AADT were put into the table.
2. Seasonal data is included in both the main table and the seasonal data table. However, due to the special content of the seasonal data, the seasonal data records in the main table have no associated raw data.

Recounts Table (TBL_RECOUNTS) Description

This table contains records for stations which had multiple raw data records. As mentioned before, when more than one raw data record was found, the raw data that supported the stated AADT was delivered in the main table. In addition, all data records were delivered in the recounts table, and the AADT is left

blank for the records where the AADT does not appear to be associated with the raw data. The table format is the same as the main table TBL_STATE_DATA.

Seasonal Table (TBL_SEASONAL) Description
This table contains the data for the seasonal counts. Seasonal stations are PTC stations that are identified as part of a group of stations for which data will be collected in a special way for that count year. To be specific, data is collected for several days for four or five months in a year, resulting in a maximum of five raw data records for a given station. This data was inspected carefully, and some of it discarded, then each record was factored with the usual axle factor and seasonal factor, resulting in an average factored count for each set of raw data for the station, for the given month. Then, the averages of the factored counts for the station were themselves averaged to determine the AADT by station.

The data format is different from the main table, in that more raw data is provided for each station, and there is an additional field for the weekly values of the factored count.

Field Name	Type	**Description**
CountyID	Numeric	The county / Urban code, also called the SIPS code in the ArcView files.
StationID	Numeric	The Station number.
URB_NUM	Numeric	The urban identifier if the station is in an urban area, -1 if not an urban station.
UNIQ_ID7	String(7)	7-character combination of County ID and Station ID, consisting of 3 characters for the county ID, and 4 characters for the Station ID; this is the same format as the corresponding ArcView field.
UNIQ_ID8	String(8)	8-character combination of County ID and Station ID, consisting of 3 characters for the county ID, and 5 characters for the Station ID; this is the format that is planned for future use within the Traffic Survey Unit.
RouteCode	String (1)	Route type identifier. 1 is Interstate, 2 is US, 3 is NC, 4 is secondary, and 5 is local.
RouteNum	String (25)	The road's route number, or, in the case of urban areas, the street name.
Description	String (30)	Text description of the station location
ATRGroup	Numeric	The ATR group assignment
Year	Numeric	The count year for which the raw data and resulting factored count were taken.
Month	Numeric	The month that the raw data was taken, if raw data is provided.
Day	Numeric	The day number of the first raw data count. For this table, it is always 2, indicating a Monday.
FactCount	Numeric	The average of the factored daily counts, as computed from the raw data, and factored by the axle correction factor and the seasonal factor.
AADT	Numeric	The average of all the averaged daily counts for the station.
Count1	Numeric	First day's raw count.
Count2	Numeric	Second day's raw count.
Count3	Numeric	Third day's raw count.

Count4	Numeric	Fourth day's raw count.
Count5	Numeric	Fifth day's raw count.
Comments	String (25)	General comments about the data records.
Case	String (5)	Set to "SEAS"

Axle Factor Table (TBL_UNIQ.dbf) Description
As mentioned, this table was delivered after the main table was delivered, so that the axle factor could be included in the ArcView data.

Field Name	Type	**Description**
UNIQ_ID7	String(7)	7-character combination of County ID and Station ID, consisting of 3 characters for the county ID, and 4 characters for the Station ID; this is the same format as the corresponding ArcView field.
UNIQ_ID8	String(8)	8-character combination of County ID and Station ID, consisting of 3 characters for the county ID, and 5 characters for the Station ID; this is the format that is planned for future use within the Traffic Survey Unit.
AxleFactor	Numeric	Axle correction factor expresses as an integer.

Case Number Distribution
The main table TBL_STATE_DATA contains 35,029 records. The distribution of the data by case number follows.

Case Number	**Number of Occurrences**
1-1	27,583
1-3	348
2-1	236
2-3	4
2-4	11
2-7	940
2-9	12
3-1	2,921
3-3	73
3-4	1,666
3-6	102
3-7	8
4-1	207
4-4	61
5-4	101
5-5	1
6-1	4
2001	166

INTER	411
SEAS	174

APPENDIX 4:

4. FY 2002 GIS Project Role and Assessment of Accomplishment and Budget Standing

Shannon McDonald

June 2002

FY 2002 GIS Project Role and Assessment of Accomplishment and Budget Standing

A Spatial Editing and Validation Process for Short Count Traffic Data

Author: Shannon M. McDonald, GIS Contractor
Subject: GIS Assessment of Accomplishments and Budget Status for FY 2002

The purpose of this report is to inform the project team of the present advancements toward meeting the project goals and how the GIS team stands in regard to the budget and to offer recommendations for completing the project within the budgeted scope. The role of the GIS contractor, Shannon McDonald (WEI), and the NCDOT GIS department will be recounted from the Research Proposal submitted June 21, 2001. The scope of work referred to, as well as any other specificities of the project within this report will also be constructed from the Research Proposal unless otherwise stated. The interpretations of these items are those of the author's understanding and opinion and are subject to project team's evaluation.

GIS Contractor's Budget Allocation for FY 2002

GIS activities outlined in this report are provided by Shannon McDonald (henceforth referred to as the GIS Contractor) and Erik Brun of the NCDOT in cooperation with the TSU project staff. The budget for FY 2002 accounted for 296 hours from the GIS Contractor, while the budget for FY 2003 accounts for 665 hours (as stated in the NCSU Subcontract # 2001-1931-01, henceforth referred to as the NCSU Subcontract). The hours by task accounted for as of May 31, 2002 is as follows:

Function	Hours
GIS Programming and Modeling (GIS Application)	11.0
Meetings & Reports	74.0
Data Research	37.0
Data Preparation*	148.0
Total Hours	**270.0**

The GIS Contractor has engaged heavily in the preparation of data provided by the NCDOT GIS. This has resulted in time spent in other areas of the above breakdown (such as meetings and reports). Most hours under the GIS Programming and Modeling section is directly associated with the cleaning, entering, and formatting of data for use within this project (refer to the *Products, Problems, and Deviations Undertaken during FY 2002* section of this report for details).

Project Role of GIS Team

The primary roles of the GIS team (WEI and NCDOT) for the fiscal year 2002 was to 1) aid in research in the computing power available within the TSU, NCDOT, and local government agencies, 2) make available the AADT data in a GIS environment and explore other data pertinent to the modeling phases, and 3) begin research and studies in customizing a GIS application to allow analysts a thematic visualization of traffic count stations and data with other existing GIS layers.

The primary roles of the GIS Contractor, as outlined in the NCSU Subcontract, was to

1) Develop, test, and implement a modified GIS application as outlined in the *Task Descriptions* section of *Task 3: Automation and GIS Implementation*

2) Develop, test, and implement a distribution GIS as described in the *Task Description (B)* section of *Task 4: Implementation Guidelines*
3) Assist in ongoing research in available GIS software, data, and hardware alternatives to enhance productivity and publication.

The GIS Contractor has provided recommendations for software development and data usage. He has also provided assistance in the delivery, interpretation, and modeling of GIS data. These preliminary tasks have involved research into and cleaning of GIS data maintained by NCDOT GIS and other entities. He has been an integral part of each GIS-related task undertaken during the 2002 fiscal year, as discussed later in this report.

Erik Brun's (NCDOT GIS) primary role has been to facilitate in the delivery and explanation of data. His proximity to the data and his understanding of the GIS network structure at NCDOT is essential for the team's acquisition of data and implementation of the system. His involvement in the project has spanned from producing straight-line distances between stations for use with statistical modeling and assisting the initial stages of the station-to-station routing procedure. Another product of the GIS department at NCDOT was the initial delivery of the AADT/GIS layer data to NCSU for their study in S-PLUS software.

Products, Problems, and Deviations Undertaken during FY 2002

The work undergone during the 2002 fiscal year has been concentrated mostly on the research and acquisition of data. A product of this work has been cleaned data in a usable format for proceeding with the project. The following section describes the GIS efforts undertaken by the project team, and includes deviations and problems encountered thus far.

Task 1-D of the research proposal states that the team will obtain and explore data for use with this project. Data to be reviewed includes the output AADT data, the PTC station GIS data, and any other data that may be imperative for later phases of modeling and characterization. This includes data in a GIS-compatible format that can be easily utilized in creating base data for analyzing statistical anomalies of the PTC stations as well as data used as reference in the GIS application.

The GIS-related data that has been defined as pertinent to this project so far is as follows:

1) Entire statewide road network (local routes included) with data from the LRS (Universal file) merged
2) PTC feature locations (point shapefile)
3) Latest AADT count data
4) Municipal boundaries (polygon shapefile)
5) U.S. Census data (for demographic/economic attributes)
6) Land-use (polygon shapefile)

A GIS program was chosen as a tool for joining this data because of its ease of use and flexibility to produce analysis of a spatial manner. However, a primary problem with data within a GIS, especially with data originating from different sources, is its spatial inconsistencies. This has proven to be a major obstacle for providing timely, accurate data for analysis.

Several problems existed with the first iteration of the merged GIS and AADT count data as it was given to Dr. Jackie Hughes-Oliver (NCSU Statistics) resulting in the largest delay of the project thus far. This iteration was produced by the TSU and GIS units at NCDOT and contained inconsistencies between the

two data and with other GIS layers. In February, the GIS Contractor reviewed the problems with the GIS data. Simultaneously, the TSU group studied and corrected their data. Problems with the GIS data included spatial inconsistencies and error between the PTC station locations and the US Census Bureau polygons, NCDOT county boundaries, and the Anderson III land-use polygons. In addition, there were inconsistencies in the AADT data from the TSU and the GIS layers. These problems have been resolved and a new dataset was delivered to Dr. Jackie Hughes-Oliver in early May of 2002 (outlined in the *Procedures and Corrections of PTC Data*, May 12, 2002).

Another obstacle has been acquiring data with null attributes (attributes without values). The NCDOT GIS unit is presently undergoing production of the LRS road network. The network is an integral part of this project, especially with the preliminary data preparation (specifically, the routing procedure for deriving distances between stations along a desirable path uses the network as a base layer). In addition, data from the LRS road network layer (and the Universal dataset for which it is associated) is being used for Dr. Jackie Hughes-Oliver's statistical analysis.

Because of the importance and multi-use of this base data, the project is dependent upon its completion. Delay in the delivery of an LRS layer with a complete set of data has required the project team to compensate for missing data. Other missing data, such as data for local roads, has also impeded progress.

The routing process to find the distances between PTC stations via the *most likely traveled* routes has also been troublesome. The difficulty is largely due to no capacity variables being present within the current LRS. The presence of such a variable would allow the GIS team to use out-of-the-box network software to derive the distances required. The GIS Contractor and TSU are undergoing a process to create an algorithm that will provide a *cost variable* for deriving route distances.

In review, the project has been delayed due to errors and inconsistencies in data. In addition, the allocated budget for the GIS contractor, Shannon McDonald, for FY 02 has been reached behind the targeted scheduled task.

The next section outlines concerns for further delay in the project for FY 03, followed by recommendations for reaching the project goals while staying within the project budget and time constraints.

Problems Expected for Fiscal Year 2003

Ongoing problems with data will most likely continue to encumber the project. The delivery of a more complete LRS road network will greatly enhance the accuracy and efficiency of the GIS modeling procedures. It is not known at this time whether the data will be ready on time for statewide modeling. If so, values for local roads will still be derived through methodologies derived and presented during FY 2002 (the LRS network will not include such facilities).

Changes in data or the delivery of new data may also require revisiting methodologies or implementing new attributes. Although efforts have been made to create procedures of data production that allow seamless integration, this is not always possible.

The routing procedure is being executed for a subset of five counties (called the test area) and will need to be implemented on a statewide basis. Depending on Dr. Jackie Hughes-Oliver's findings, this procedure may require a vast amount of processing time. Alternatives to capturing the large amount of data produced from such a procedure has been and continues to be researched.

A large concern for the GIS group is the upgraded ESRI software, ArcGIS. The latest version (8.1 as of the date of this report) has proven to be problematic with spatial applications. In addition, the architecture of the software is vastly different from its predecessor ArcView 3.2. Programming ArcGIS v.8x will require some training of the GIS team. Although the training of personnel will not directly impact the project budget, the difference in programming efficiency may.

A detailed Use Case study has not been outlined in the project scope. The GIS Contractor feels that such a study is crucial to completing the project within the constraints of the proposal (refer to the following section for recommendations).

As of date, the GIS tasks for FY 2002 are behind schedule. Software, such as the new ArcGIS or alternatives have not begun testing. In addition, the routing procedure is still being derived and no data will result until its completion. This delay will carry into the new fiscal year. In association with the schedule status, the entire funding for the GIS Contractor for FY 2002 has been billed (as of June 15th, 2002).

Recommendations for Reaching Project Goals while Maintaining within Budget Constraints

Incomplete data, error within data, and inconsistencies among data continue to be the biggest setback to this project. Most often, waiting for data to be produced that best fits the procedures is not feasible. In these cases, where it is predicted that new data will emerge, efforts have been made to derive a methodology that will easily, if not seamlessly, integrate the updated data.

However, this is not always possible. For example, the routing procedure uses an algorithm that is derived from a defined set of attributes. Entering new attributes (constraints), although potentially beneficial to better define routes, will require further modeling. In such a case, it is better to conduct the project and integrate data only when proven significant benefit results. In this particular example, the routing procedure will be rerun on a statewide basis; it is recommended that the algorithm (once established) only be changed if a single attribute can account for a cost variable (such as capacity values), thus replacing the algorithm.

Alternatives to using ArcGIS are being researched. A decision will be made once a Use Case, or some alternative study, has been conducted for the application by the GIS team and TSU. It is possible, for example, a MapObjects application will suffice for the application. New developments in software, including enhancements and bug fixes, may affect the project and should be researched and scrutinized by the GIS Contractor. Although not part of the budgeted scope, continued consultation from the NCDOT GIS programmers and testers of the licensed GIS software may increase the efficiency in the design and creation of the GIS application. This will require the cooperation of the TSU and GIS units as well as the GIS Contractor.

In addition, close and careful planning must be taken to ensure that the applications apply to their target users to reduce multiple iterations of the GIS applications. This will include preliminary research and documentation of the proposed applications conducted between the GIS Contractor, TSU, and NCDOT GIS.

The role of the NCDOT GIS in the implementation is crucial to meeting the project goals. Cooperation between the TSU and GIS units must continue to include an assigned person with knowledge of the GIS data and infrastructure. Furthermore, having this project as a priority to the assigned person or persons may benefit the efficiency of the implementation process of the GIS application.

In relation to the last statement, it would be prudent to revisit the goals and duties of each of the project staff.

Summary

This document is aimed at outlining the accomplishments made and to report the current budget and completion status regarding GIS efforts toward this project. It also is intended to make recommendations for completing the GIS tasks within the allocated time and budget constraints. The overall status of the project is behind schedule and under budget. This report may not have covered all of the aspects of or causes for the current status, but do highlight those concerning the GIS efforts.

APPENDIX 5:

5. Data Edits Determined From "List of 55"

Shannon McDonald
Susan Cosper

November 2002

Data Edits Determined From "List of 55" ***November 2002***

Background

On August 21, 2002, a list of 55 stations was presented to NCDOT with a request to explain, if possible, their unusual behavior. These 55 stations, subsequently labeled "outliers," were flagged from a regression analysis that estimated their transformed AADT counts with associated error in excess of two on the transformed scale.

The primary goal of the project is to obtain a statistical model for relating AADT counts to systematic variations (e.g., route type, speed, number of lanes, land use in surrounding area, census information, etc.) and spatial variations (i.e., connectivity of road segments on which PTC stations are located). Prior to understanding spatial variation, we must learn about systematic variation. To this end, a regression analysis that assumes independence across AADT counts was performed.

A common assumption in regression analysis is that the response variable should be approximately normally distributed. AADT counts were noticeably non-normal, but a power transformation (AADT^0.2) resulted in an acceptable level of normality. Therefore the regression analysis actually modeled AADT^0.2 as a function of systematic variants.

Census information from 1990 was first compressed using principal components analysis. Almost 200 census attributes were compressed to seven components, which are simply weighted linear combinations of the original census attributes. Station-specific attributes entered the regression primarily nested within route type; for example, a speed limit of 55 miles per hour on an interstate has a different effect in the model than a speed limit of 55 miles per hour on a US highway. The resulting model provides estimates (a.k.a. predictions) of transformed AADT, and these estimates are compared to actual (a.k.a. observed) transformed AADT's to get residuals (= actual - estimated). A residual that is large in absolute value indicates that the model had difficulty capturing the phenomena at work for that particular station. The station in question may simply have erroneous data or the model may not be sufficiently specific.

Absolute residuals larger than three indicate that a close investigation is warranted for those stations. Absolute residuals between two and three are considered less extreme, and absolute residuals less than two are acceptable. The aforementioned list of 55 stations is all stations (among the 3434 stations in the test area) for which absolute residuals exceeded two. Even if explanations are not obtained for all 55 stations, the model is already very successful in that only 1.6% of the stations have been flagged and we have explained 70% of the variability observed in transformed AADT counts by accounting only for systematic variation.

ArcView 3.2, a desktop geographic information system (GIS), was used to aid in the explanation for the 55 stations' peculiarity. This report outlines the techniques used to recognize the problems or conditions of the station data that resulted in the high residuals as well as demonstrates the use of GIS within this project.

The GIS platform gives analysts the opportunity to see multiple data (all of which is incorporated within the model that produced the list of outlier stations) as a geographic representation. The GIS interface allows the analysts a wide array of flexibility for viewing data unlike that of a standard database. Not only can the data joined to a particular station in question be observed, but also the environment in which the station is located. For example, a station's attributes can easily be observed in a table, and likewise—with a little research—that data of its neighboring stations. However, with the required data in place, the GIS interface offers its users the freedom of exploring data immediately surrounding the station. In addition, it expedites the process of observing station data when comparing it to neighboring stations and

stations within the same given class. In short, the data we are working with is spatial—represented by geographic entities—and GIS is the best available tool for viewing, querying and relating such data in a single interface.

The Process

By observing a station's data with that of surrounding stations, and together with the station's proximity to municipal boundaries, county lines, census tracks, land use, and other entities, we determined—or at least deducted—a reason for each station's residual value. The results are listed in the section below.

A brief observation of our findings is input into the PTC shapefile's database for each station. In addition, when it is determined that the primary cause of the station residual value is due to an incorrect route type, an adjusted value is given. The explanations listed by the station identifiers are in Appendix A of this report.

By analyzing the findings of this process, some traits in the data presented themselves allowing the possibility for automated fixes for stations with error, and giving an insight into how other procedures within the proposed PTC process will affect the final product.

Our Findings

Although the explanations are based on rational data observations, they do not necessarily represent all of the factors present for having a high residual count. For example, the explanation for a station may point to the fact that the station is in a high-traffic area (such as near an airport or industrial area). This would account for a high AADT on an SR or local road, but does not conclude that there is an absence of error for the station. In fact, the count recorded may have been in error, or the station may have been recorded at the wrong site. Such factors are not easily identified with the data at hand. However, all noticeable justifications are given. Those that point to error in data are verified by appropriate means and corrected.

Our recorded observations can be grouped into two distinct categories: 1) high residuals due to error in the data, and 2) high residuals due to the environment in which the station is located. The explanations for the second set are far more varied and not as easily identifiable. For the first group, most are easily identified and corrected.

<u>Error found in the data</u>

Of the 55 stations, 22 have been noted as having error in the data. Sixteen of the 22 have conflicting route types (between TSU and the LRS). It was these stations that were first reviewed. (A check for other stations within the test area revealed a larger population of stations exists with this type of error—380 stations total. An independent process of checking these stations will has been undertaken. A process to correct the error on a statewide basis [presently 3003 stations] is planned.) Six of the 16 stations have been snapped to the wrong road, thus populated with incorrect LRS data. (The process of determining stations with error in their location by the snapping process is outlined in Appendix B, while the snapping procedure is described in the report *Procedures and Corrections of PTC* Data dated 05/12/02.) A corrected route type for these stations was calculated in the shapefile's RTE_ADJUST field.

Eight of the 16 stations with conflicting route type values are assigned to incorrect route types. Specifically, the stations are assigned to road segments that have no LRS values and are designated as a

local route type. Also, each of these stations has a TSU route type assignment of state route (4). Each station was referenced to the station and AADT map to ensure the TSU assignment is correct. This resulted in a planned systematic approach to correcting some of the error on a statewide basis by logical query:

Query: ([LRS_MATCH] = 0) and ([ROUTECODE] = 4) and ([RTE_TYPE = 5])

Where LRS_MATCH indicates the availability of LRS data in the segment on which the station is located (0 means no LRS data present), and ROUTECODE and RTE_TYPE is the route type assignment from the TSU data and GIS linework, respectfully. At present, by selecting this set and calculating an adjusted route type of '4', the routine will eliminate 57 % of this type of error within the test area, and 51 % statewide.

Incorrect unique identifiers account for 6 of the 22 stations in the outlier list. In addition, station 0910565 has an incorrect identifier and is snapped to the wrong road. Because the unique identifiers are the key for which the TSU data is assigned, error in their values are of concern. Having the wrong identifier means having incorrect TSU data. This type of error is difficult to detect without checking each station independently.

Investigation of these stations indicates the reason for most of the incorrect identifiers is due to the SIPS assignment. For example, station 1080541 is located at station 0910541's position (and vice versa). The potential for incorrect assignment most likely occurs only in those counties with the presence of stations with urban SIPS. Even more specifically, stations can be narrowed down to include those near or in urban areas with a county SIPS, and conversely, urban SIPS outside the urban boundaries. Unfortunately, the assignment of SIPS by the TSU is not bounded by any geographic entity. A process was therefore created that utilizes available GIS representations of urban areas and municipal boundaries for selecting stations that are most likely in jeopardy of incorrect SIPS assignments. This process was undertaken for the test area and resulted in a possible systematic approach for correcting the statewide PTC station inventory. These processes and findings are detailed in the report *Quality Check Process for Station Unique Identifiers* dated 09/24/02.

The remaining stations varied in their error. Stations 1080599 and 1080369 have been deleted from the model. These two stations presented conflicting data raising uncertainty to their validity. The TSU location description for 1080599 is inconsistent, resulting in ambiguity in the data collected. Data collected for station 1080369 cannot be verified with the data available in the model. The station was found to be located on a ramp, and no return data is available (data from stations on other ramps for the same intersection).

Initial investigation of station 1080641 revealed that its location outside the Raleigh municipal boundary was the only difference in this station and its nearest neighbor located along the same route. The neighboring station (1080006) is not on the list of outliers, though its AADT is also high. This prompted an investigation and process development for the need of additional classifications for urban and municipal assignments (refer to the report entitled *New Urban and Municipal Field Value Assignments and their Relation to Station Locations and the PTC Data Model Process* dated 08/22/02). Further investigation revealed another error in the station data derived from a problem noted during the creation of the statewide road network: The LRS assignment is only present on one of the two non-converging segments that represent divided roads in the GIS linework (explained in detail in the report *Procedures and Corrections of PTC Data* dated 05/12/02). The station was snapped to the road segment that had no LRS attribute values. The GIS Unit at NCDOT is currently resolving this problem.

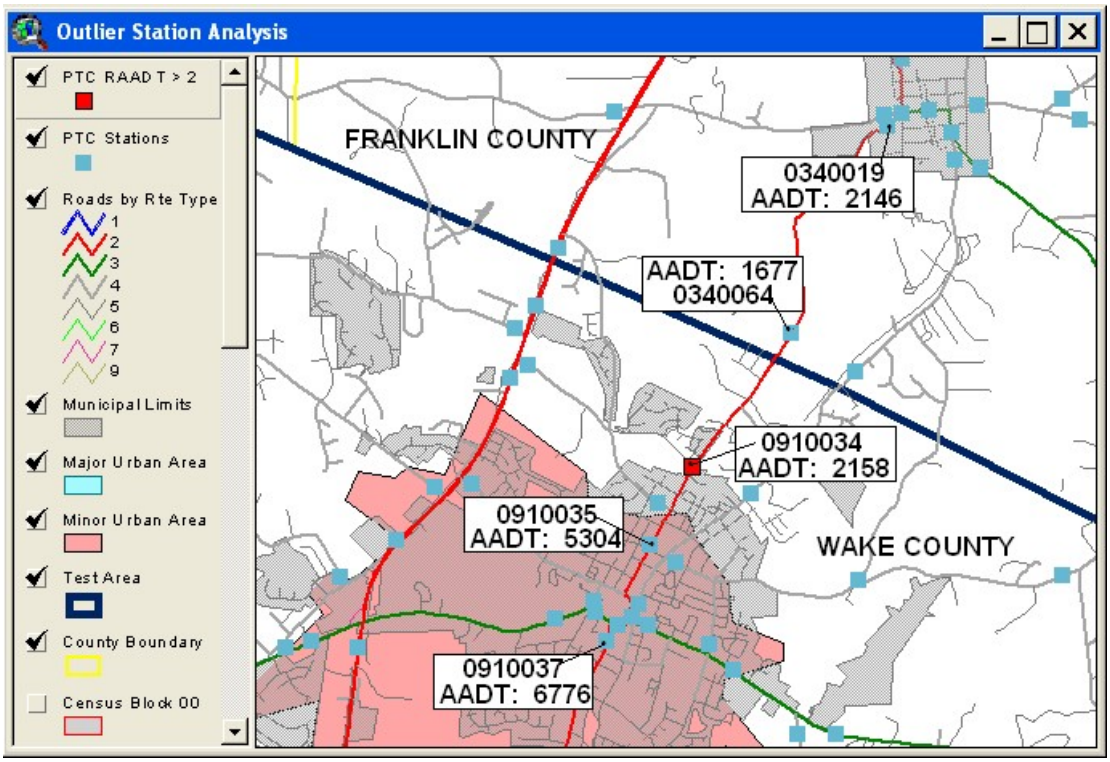
Environmental factors

Most of the stations on the outlier list (33 of 55) are not there because of error, but instead most likely contain high residual values because of the nature of their environment. (They may, however, contain error that is not evident in the process of this report—such as error in the count data collection process.) Most stations in this category can be reasonably explained when viewed with its surroundings. A few were not as evident, and are noted as such: these stations have a supposition as to why the data has a high residual value, but bares uncertainty in its wording. (For example: the given explanation for station 0910673 states "high volume area, but unsure of why neighboring station is not on list". The neighboring station has all of the same attributes with a higher AADT but does not appear on the list of outliers.) Further investigation beyond the scope of this exercise may reveal reasons within the regression analysis process or error in the data.

Most stations, however, fall within well-reasoned explanations. Many stations are located on state or local routes but in high-traffic areas. Stations along SR 1007 in eastern Wake County (Poole Road) have higher AADT counts than most other stations on routes of the same type in the area. This is because Poole Road is an alternative to the over-capacitated facility US 64 between Raleigh and the outlying towns of Knightdale, Wendell and Zebulon. Another example is that of stations 1080809, 1080815, and 1080819—all with high AADT counts located on state routes. These stations are also located on the three main entrances to Raleigh-Durham International Airport. Malls, parks, and other attractions may also account to high counts. (Including a station's proximity to major airports and other such facilities for modeling is a possibility presently being researched.)

Low volume areas are also often associated with high residual values in the test area. This is especially true for stations located on roads that have more desirable alternatives, and thus service only local traffic. Another cause noted are stations located on routes with no outlets (dead end roads).

Other observations associated with business routes and edge effects associated with municipal and county boundaries were made. Both are demonstrated in station 0910034. Located in Wake Forest, this station has a much lower count than that of other stations in the area located along the same route type. Although the station has a US route type assignment, and in fact is located on a US highway, it is located on US 1A, a business route. Predictably, in a small town such as Wake Forest, stations along a business route have a lower AADT than that of stations located on the nearby bypass, a major US route. In addition (and perhaps more importantly), this station has a lower count than even those along the same route to its south. Stations north of Wake Forest (those in Franklin County) have a lower count than 0910034. Franklin County is outside of the test area and was not used in the model. The lack of this data in the test set, and without the road network attributes in the model, the AADT volume appears to be too low.



Although this is the only station noted in the list of 55, it is safely assumed that the occurrence happens elsewhere in the state.

Other errors noted are mainly due to the road network geometry. These include factors such as limited access to a roadway and multiple entrances to residential neighborhoods.

Conclusions

The exercise was aimed to explain phenomena in the data of stations with high residuals. In addition, the results spawned other exercises aimed to find patterns in data error and develop systematic approaches to cleaning the data on a statewide basis. The exercise has given an insight to how important the GIS is as a tool to the PTC process, and at the same time witnessing how this procedure is dependent on and will improve with the introduction of other procedures outlined in the proposed PTC process.

High residuals were associated with numerous explanations. Conclusions should be viewed with caution as it is suspected, and with some cases confirmed, there exists more than one error in the data for a station with a high residual. The exercise has revealed patterns for stations with high residuals. These patterns give analysts a direction for researching the problems further. For example, the high number high residuals for stations on low volume roads prompted research into their data collection routines. It was found that a irregular collection practice has been in place for stations: it has been the practice of the TSU to only one 24-hour count has been used when a 2-day count proves to have a high variance. The single 24-hour count is thus used to represent 2 days, resulting in a possible miscalculation for day-to-day variance.

This stage of the project revealed the significance of boundary (edge) effects. Station data near the edges of municipalities, for example, often resemble those inside the municipalities. Additional classifications to differentiate stations within urban areas (fringes) have been added to the PTC data to compensate. Other information, such as distance from malls, airports, and other high-traffic generating facilities may

improve predictions. An algorithm to determine the number of intersections along a route where a station is located, thus identifying alternate route, may also improve the model.

The module designed in this project is not a stand-alone procedure. It has become apparent that with the introduction of other modules, the overall process will improve. An example of this is with the introduction of historical data: once the problems are resolved, they will be less likely to be problematic during the concurrent data collection cycle.

The spatial model continues to point to other nonsystematic errors related to the TSU dataset and data collection/processing currently in practice. Error associated with station mapping has been evident and problematic. Additional error introduced through the data capture of PTC stations has also become evident, though small in number. A field validation inventory and capture of X-Y locations using GPS may reduce error in the GIS dataset and aid in improving the consistency of data collection.

Appendix A: Explanation for "List of 55" outliers

UNIQ_ID	AADT	TAADT	PAADT	RAADT	Why
0910544	202.0000	2.8911	5.0327	-2.1415	low volume road with alternates
0501608	9082.0000	6.1892	3.9841	2.2051	high vol. route in western Johnston Co. near I-40
0501613	8153.0000	6.0571	3.7941	2.2630	high vol. route in western Johnston Co. near I-40
0501816	160.0000	2.7595	4.8542	-2.0947	GIS has as local
0910706	195.0000	2.8708	7.5228	-4.6520	WRONG ID's (sent via. e-mail 082302)
1030220	45040.0000	8.5255	6.5083	2.0172	snapped to wrong road (GIS)
1030130	49093.0000	8.6737	6.3799	2.2938	snapped to wrong road
0910034	2158.0000	4.6431	6.8536	-2.2105	low count north of Wake Forest (city boundary) and near county line, next aadt north is 1
1030120	60160.0000	9.0336	6.0525	2.9811	snapped to wrong road (GIS)
1080541	24703.0000	7.5605	5.0534	2.5071	WRONG ID'S (swap with 0910541)
1080539	26135.0000	7.6462	5.0577	2.5885	WRONG ID'S (swap with 09180539)
1080820	21067.0000	7.3235	5.2005	2.1230	route connects RDU Airport and US 70
1080819	17257.0000	7.0371	4.9679	2.0692	GIS has as local
1080870	5075.0000	5.5092	7.8183	-2.3091	OK-services small portion of res. area
1080815	36575.0000	8.1778	5.0739	3.1039	one of two main entrances to airport from I-40
1080809	20950.0000	7.3154	4.8481	2.4673	one of two main entrances to airport from I-40
1080512	925.0000	3.9195	6.1084	-2.1889	GIS has as local
1080264	483.0000	3.4418	5.4834	-2.0416	low volume, other alternatives
1080511	788.0000	3.7958	6.3744	-2.5786	no outlet
1080812	25800.0000	7.6265	4.6877	2.9388	mall and hotels off of I-40 near airport
1080698	10510.0000	6.3727	4.2070	2.1656	high vol. traffic (neighboring stations loose traffic at nearby intersection)
1080168	85405.0000	9.6894	5.7949	3.8945	snapped to wrong road
1080272	48283.0000	8.6449	6.5251	2.1198	GIS has as local
1080174	245.0000	3.0049	5.0944	-2.0895	low volume road near umstead park
1080175	228.0000	2.9620	4.9761	-2.0141	low volume road near umstead park
0910673	10436.0000	6.3637	3.2188	3.1448	high vol, but not sure why neighbor is not outlying
0910539	3041.0000	4.9728	7.3218	-2.3490	WRONG ID'S (swap with 1080539)
1080641	66102.0000	9.2054	5.9990	3.2064	high volume divided rd, lanes = 4, not 2, just outside of Raleigh city limits
0910541	4490.0000	5.3759	7.5678	-2.1920	WRONG ID'S (swap with 1080541)
1080915	24018.0000	7.5181	5.4250	2.0931	high volume station, nearby st 1080861 also high, but in different landuse
0910565	80.0000	2.4022	5.6859	-3.2836	WRONG ID'S, snapped to wrong road
1080450	336.0000	3.2009	5.6585	-2.4576	low vol. street, alt. outlets
1080599	296.0000	3.1207	6.0941	-2.9734	DELETE FROM DATASET
1080451	618.0000	3.6157	5.8608	-2.2451	low vol. street with other outlets
1080715	17801.0000	7.0809	4.9826	2.0983	high volume road (Poole Rd)-just outside Raleigh CL
1080714	15188.0000	6.8596	4.8131	2.0465	high volume road (Poole Rd)-just outside Raleigh CL
1080111	597.0000	3.5908	5.7816	-2.1908	GIS has as local
1080152	5606.0000	5.6199	8.5698	-2.9498	snapped to wrong road
0910736	132.0000	2.6553	4.6839	-2.0286	low vol, but Jackie should check with neighboring stations
1030544	36.0000	2.0477	4.4290	-2.3814	dead-end road (into new I-540)
1030576	167.0000	2.7832	5.1176	-2.3344	short connector road with only two outlets; other connector alternates available
1030705	95.0000	2.4862	4.9288	-2.4425	near major alternate route

1030599	118.0000	2.5964	4.7445	-2.1480	compare with 1030597
1030644	343.0000	3.2141	5.4619	-2.2478	GIS has as local road
1030654	462.0000	3.4114	5.5604	-2.1490	GIS has as local
1030658	352.0000	3.2308	5.4356	-2.2049	side entrance to residential development
1030619	85.0000	2.4316	4.8328	-2.4012	GIS has wrong road type for this portion of road
1030604	163.0000	2.7697	5.0956	-2.3259	low vol. road. neigjbor is higher, but could be due to a boat ramp or other facility on Falls
1030053	169.0000	2.7898	5.1447	-2.3548	geometry of road network, parallels major one-way road (limits access to this road)
0310619	194.0000	2.8679	4.9506	-2.0827	station should not have an city flg = 1
1080369	1447.0000	4.2864	6.9911	-2.7047	DELETE FROM DATASET
0910584	52.0000	2.2039	4.3931	-2.1892	dead end road (no outlet)
1080949	21171.0000	7.3307	5.1047	2.2260	high volume road (Poole Rd)-just outside Raleigh CL
0910842	95.0000	2.4862	4.6719	-2.1857	low vol with alt. routes
1030768	70.0000	2.3389	4.6830	-2.3441	low vol. area with alternate outlet to main route

Appendix B: Incorrect snapping

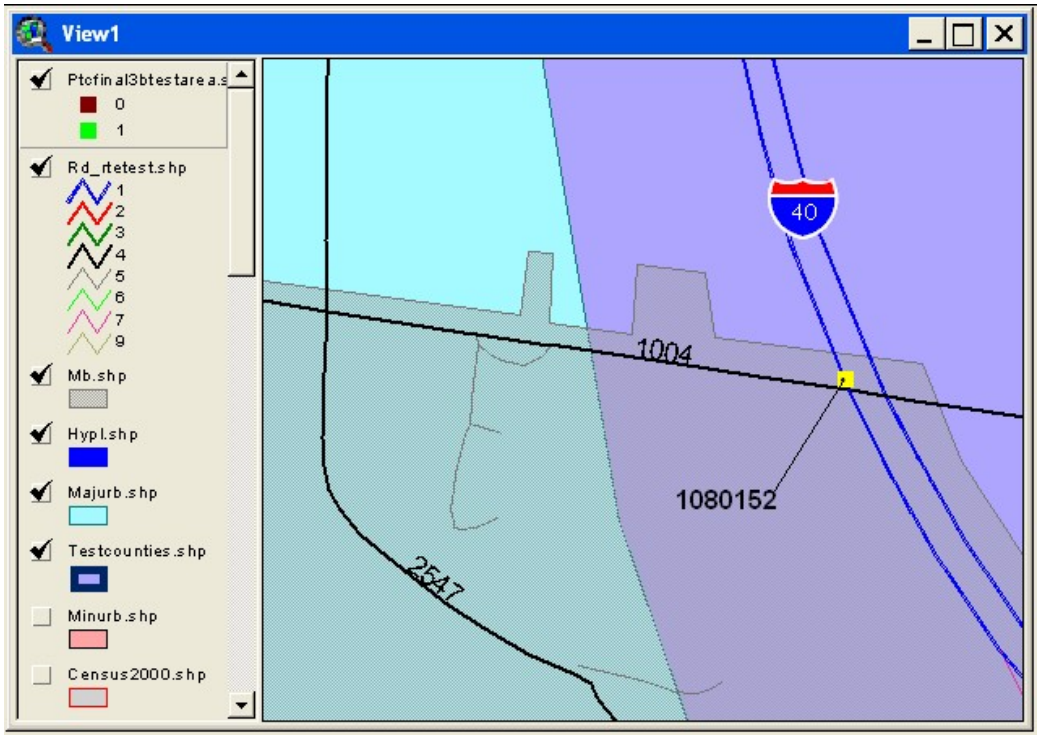
The process of snapping stations to the NCDOT linework was inexact in that the only attribute used for the process was the stations' proximity to the linework. In other words, each station was snapped to the closest road segment—given the segment was within a specified tolerance—even if that segment belonged to the incorrect route (the snapping process is explained in the report *Procedures and Corrections of PTC Data* dated 05/12/02).

Example: Station 1080152 (residual AADT = -2.9498) contains the following data:

TSU				LRS	
ROUTECODE	ROUTENUM	DESCRIPTION	AADT	ROUTE1	RTE_TYPE
4	1004	E OF SR 2547	5606	10000040	1

The above table indicates that station 1080152 is digitized closest to—and thus was snapped to—Interstate 40. However, the TSU data designates the station as being located on a state route (ROUTECODE = 4). Specifically, its location is on SR 1004, east of SR 2547. The AADT count suggests that the station is in fact located on an SR road rather than an interstate highway.

When viewing the station in the GIS interface, we can see that the station is located near the intersection of SR 1004 and I-40 east of SR 2547 as the TSU data indicates. We also see that the station is digitized on a segment of I-40, hence the LRS designations.



During the snapping process, the station snapped to the closest road segment—that of I-40. The incorrect snapping was easily verified by checking the station location on the station and AADT maps.

APPENDIX 6:

6. New Urban and Municipal Field Value Assignments and Their Relation to Station Locations and the PTC Data Model Process

Shannon McDonald

November 2002

NEW URBAN AND MUNICIPAL FIELD VALUE ASSIGNMENTS AND THEIR RELATION TO STATION LOCATIONS AND THE PTC DATA MODEL PROCESS

During the project staff meeting held on August 21, 2002, the possibility of adding an additional class to the URBAN_FLG attribute was discussed. At present, the attribute is binary: a station with an URBAN_FLG value of 1 is within a municipal boundary; a station with an URBAN_FLG value of 0 is outside of any municipal boundary. Being that the condition of AADT evaluation in Jackie's statistical model weighs the URBAN_FLG value in determining if the count is an outlier, it was decided that more variability in the form of additional classifications would be favorable.

This decision was made after examining the case of two stations: 1080006 and 1080641. Station 1080641 is on the list of stations whose values resulted in high residuals (absolute value greater than 2); but not its closest neighbor, station 1080006. Both stations have unusually high AADT counts for being on a local road. Both are on the same road in the same area, only a few miles apart. One difference is that station 1080641 is just outside the city limits (i.e. URBAN_FLG = 0) while the other is located inside the Raleigh municipal boundary (i.e. URBAN_FLG = 1). Thought this proved not to be the sole or primary factor for the difference in the two stations, it is a potential problem for stations near municipal boundaries. (Later investigation revealed that station 1080641 had no LRS assignment, and thus was assigned the default speed limit, number of lanes, and access control for local roads: 25, 2, and 1, respectfully. The number-of-lanes value is the primary cause of the high residuals. The lack of LRS attributes was due to the station being snapped to a divided road, represented as two non-converging segments, where only one side of the road is assigned. This problem is currently being resolved by the NCDOT GIS).

First, I proposed adding another classification to the URBAN_FLG field giving more flexibility in how the values are handled in the model. The idea was to set up a buffer around the city limits (municipal boundaries) that would include stations that were still effected by the cities, but are not within the limits. In other words, a classification is needed for stations that currently have a zero (0) value for URBAN_FLG, but may actually be traversed more like a station within the city limits. Instead of an arbitrary distance for the buffer, I suggested using an existing coverage that depicted a metro-area type of boundary.

My research found that the coverage I was actually suggesting does not depict metropolitan boundaries, rather the urban areas around and within major and minor cities. Furthermore, the urban areas are based on drawings rendered/digitized from 1990 data. Therefore, the boundaries offer decent buffers for cities and should suffice. However, I came up with a few alterations that require a couple of changes to the dataset.

First, to clarify some terminology: I'll refer to the stations within the municipal boundaries as those within city limits instead of within urban areas, since the new coverage actually depicts urban areas (those stations currently with an URBAN_FLG = 1). Next, the stations within the urban areas depicted by the new coverage will be considered 'within an urban area'. Those outside both will be just that: 'outside'.

This would give us three classifications:

1. Stations within the city limits
2. Stations within the urban areas, but outside the city limits
3. Stations outside both urban areas and city limits

In fact, we can further break down this classification, since we have data for both major urban areas (such as the cities within the RTP, Charlotte, Triad area, etc.) and for minor urban areas (such as Kannapolis, Wake Forest, Sanford, etc.). Now we have four classifications:

1. Stations within the city limits
2. Stations within major urban areas, but outside the city limits
3. Stations within minor urban areas, but outside the city limits
4. Stations outside both urban areas and city limits

Next, I found that the urban areas might be best used as a separate attribute/classification. There are several reasons for this; a few are obvious when looking at them visually (with GIS).

First, there are far fewer urban areas than municipal areas, so you don't always get a buffer effect. This is especially true with smaller cities and towns. That may be acceptable, since there would be less edge effect from such municipalities. However, even for the larger areas the municipal boundaries are often outside the urban boundaries. Therefore, it is possible to have a station that is inside the city limits, but outside the urban area. In the model above, the station would have an URBAN_FLG = 1 instead of 2 (if major) or 3 (if minor). The question is whether that station would act more like a '1' (being inside the city limits) or a '2' or '3' (within an urban area, but outside the city limits). It is presumed that more often the case would be the station should act more like the latter (2 or 3), since although the station is within the city limits, it is far outside the central or business district. The only way to be sure is to study each station case-by-case. Of course that is not feasible, so another alternative would probably serve best:

The creation of a new field in addition to URBAN_FLG would add flexibility to the model. Instead of four possible solutions for stations in or around urban areas, adding a new field will give six:

First, the current field, URBAN_FLG, should be renamed MUNICIPAL_FLG (or something intuitive) for the aforementioned sake of clarity. The field value will indicate whether the station is within or outside a city's limits (as it presently does):

MUNICIPAL_FLG	Condition
0	outside municipal boundary
1	inside municipal boundary

Next, we'll recreate a field called URBAN_FLG, but this time populate it according to the station's location as it relates to urban areas:

URBAN_FLG	Condition
0	outside urban areas
1	inside major urban area
2	inside minor urban area

The combination of these two fields gives us six possible classifications. For the example above, we could consider the station being inside the city limits but outside the urban area.

These assignments also give us a clearer picture of what stations truly lay within urban areas. In other words, stations within the city limits of a small town outside of an urban area could be treated differently than those within larger cities. In fact, stations within medium-sized cities within minor urban areas (such

as Sanford or Salisbury) can now be distinguished from those in smaller towns but in major urban areas (such as Morrisville). The possibilities of manipulating this part of the model is much more flexible with these two separate classifications.

There exists data—urban areas—that should be included to help better understand the AADT counts. Since the data is different and not directly related to the current URBAN_FLG determinant (i.e. municipal boundaries) the data should be treated as a separate entity used in conjunction with the municipal data, rather than forcing them in the same classification. Third, introducing the new urban data has created some need for clarification (and in fact, correctness) in our terminology (URBAN vs. MUNCIPAL AREAS). The result will be two fields classifying the stations by their spatial location as it relates to municipalities, and major and minor urban areas as defined by the shapefiles mb.shp, majurb.shp, and minurb.shp, respectfully.

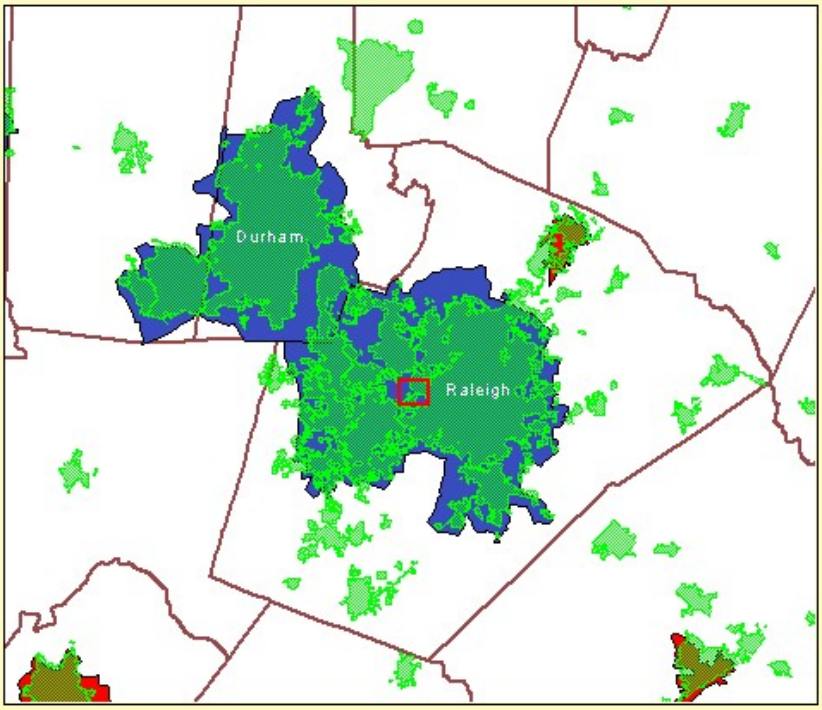
I have included a digital map for clarification:

The top map section shows the urban areas (translucent green) overlaying the major and minor urban areas (blue and red, respectfully) for the RTP area. The county boundaries are displayed as brown lines.

Notice that municipal boundaries sometimes extend beyond the urban areas. There are several possible reasons for this (for example some cities annex places far outside their centers, and the municipal boundaries have been updated and are more current than the urban areas created using 1990 data). Notice also that there are municipal boundaries (and PTC stations) that are not in urban areas.

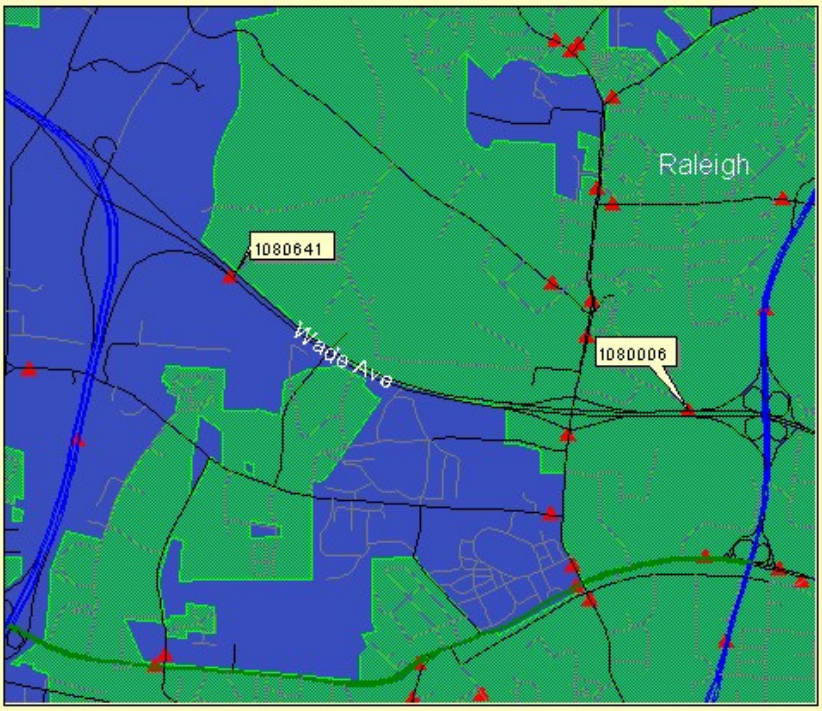
The bottom map shows a close-up of the Wade Avenue stations (stations 1080006 and 1080641) and how they relate to the municipal and urban data.

A Comparison of Urban and Municipal Boundaries for the RTP Area and their Potential Impact for Attribute Assignment of PTC Stations



Major and minor urban areas as they relate to the municipal limits for the RTP and surrounding area. This comparison signifies the differences in the two data as they are not directly related. The municipal boundaries are not entirely within urban boundaries, even for the larger cities of Raleigh and Durham.

Municipality
Minor Urban Area
Major Urban Area
County Boundary



Stations 1080641 and 1080006 are currently assigned different values for URBAN_FLG. This may be a determining factor for why the AADT counts for the former station are being rejected as an outlier while 1080006 is not, even though both are very close in their characteristics and count. With the new method of using both urban and municipal flags, the stations would have the following value assignemnt:

STATION	URB_FLG	MUN_FLG
1080641	1	0
1080006	1	1

The new assignment offer more flexibility in the statistical model

Interstate Highway
US Highway
NC Highway
SR Road/Ramp
Local Road
PTC Station
Municipality
Major Urban Area

Source: NCDOT GIS 1990 (urban boundaries), 1997 (municipal boundaries); NCDOT TSU 2002 (PTC stations).

APPENDIX 7:

7. Recordings and Methodology To Calculating Capacity Variables For Unmatched Stations

Shannon McDonald

June 2002

RECORDINGS AND METHODOLOGY TO CALCULATING CAPACITY VARIABLES FOR UNMATCHED STATIONS

Summary:

Highway capacities are needed in order to calculate, among other things, travel times between traffic count stations along routes. The variables needed for calculating these capacities (i.e. access control, posted speed limit, and number of lanes) are present in the NCDOT Universal database. The variables were brought into the GIS by linking the Universal records to highway segments in the NCDOT road layer. The line segments making up the road layer are then spatially joined to the traffic count stations via the statewide PTC shapefile.

Roughly 20% (10,000) traffic count sites lie on road segments that have no Universal database ID, and thus are not matched. Procedures were derived to calculate these variables using GIS and simple statistics. The sources used for the methodology were the existing matched values for the capacity variables and a list of values derived by observations of the Traffic Survey Unit (TSU), NCDOT. The list was used for quality control and as a source of information where there was an insufficient amount of data derived by the Universal file match (Appendix A). In addition, the mb.1 (municipal boundary) shapefile was used as a selector for indicating which stations were rural and which were urban.

The calculated values are recorded as shown in below. The *Route Type* categories coincide with the breakdown requested by NCDOT.

Route Type	AC (Access Control)	SL (Speed Limit)	NL (No. of Lanes)
Urban Interstate*	3	58	3
Rural Interstate*	3	63	2
Urban US*	1	41	3
Rural US*	1	53	3
Urban NC*	1	39	3
Rural NC*	1	53	2
Urban SR 2-Lane	1	35	2
Rural SR 2-Lane	1	50	2
Urban SR 4(+)-Lane	1	45	4
Rural SR 4(+)-Lane	1	55	4
Urban Local	1	25	2
Rural Local	1	25	2

* values for *Speed Limit* and *No. of Lanes* are averages derived from the matched Universal data (see Methodology section)

In addition, the Traffic Survey Unit also requested the Database Ranges:

Field	Min	Max	Range	Mean

AC	1	3	2	1
SL	20	70	50	49
NL	1	8	7	2

Methodology:

For the 10,000± stations that did not match with the Universal data, a method was derived to populate the variables by using simple GIS spatial selections and statistical mean calculations. The matched record values were used as input data for the statistical calculations. Where this data was insufficient, other factors, such as the TSU projections and the min/max values, were observed and used in determining the assignments. The goal was not to skew the entire dataset by assigning outlying values to the unmatched dataset. The following accounts for the methodology used in the valuation procedure.

To better understand this methodology, it is important to note a few characteristics of the PTC shapefile database:

- Records that were not matched had a posted speed limit, access control, and number-of-lanes value of –999
- Records that were not matched had a route code value of -99999999
- There is no level or color value for records that are matched, but they do exist for unmatched records
- The level field values classify the route type of each unmatched station
- The station data was derived from the NCDOT road layer (see Appendix B for the road layer leveling chart used for this procedure)

Record statistics from stations with LRS values (Universal data):

1. Select urban stations by using select-by-theme method with the PTC station shapefile as the selection theme and the municipal boundary shapefile (mb.1) as the selector theme.
2. Select appropriate route code (i.e. interstate, US, NC, or SR)
3. Select and view statistics for posted speed limit and number-of-lanes
4. Record average, minimum, and maximum values
5. Repeat procedures 1-4 for all route codes
6. Repeat procedures 1-5 for rural stations

Note: Access control values were not calculated statistically. These numbers were either 1 (unlimited access) or 3 (limited access). Only the interstates were assigned 3. The US routes were assigned a value of 1. This decision was made after observing the values of the matched set.

Populate variables (fields) in the PTC station shapefile:

Although the unmatched records had no route code from which to select, they did have populated level and color fields inherent from the road layer. When cross-referenced with the road layer leveling chart (Appendix B), the stations can be classified by their route type and selected accordingly.

The average posted speed limits and number-of-lanes for the interstate, US, and NC routes were entered using the average values calculated from the procedures outlined above (with both rounded to the nearest integer).

The SR routes proved less simple: there is no distinction between 2- and 4(+)-lane segments in the Universal file. Thus, the statistical values recorded for the SR routes are representative of all lane categories. The posted speed limit values were therefore determined by comparing the average, minimum, and maximum (range) for all lane-types to the projections submitted by the TSU:

Route Type	Min	Max	Ave	NCDOT	Assigned
Urban SR 2-Lane*	20	50	38	35	35
Rural SR 2-Lane+	20	55	54	45	50
Urban SR 4-Lane*	20	50	38	45	45
Rural SR 4-Lane*	20	55	54	55	55

* TSU values were within the range, and close to the mean; values assigned are the same as those given by TSU.

+ TSU values were within range, but far from the average.

For populating the unmatched stations, the color field values were used for distinguishing between the 2- and 4(+)-lane categories (Note: for 4(+)-lane distinction, colors 6 and 36 were selected—see the road layer leveling chart for descriptions).

For the rural SR 2-Lane routes, the assigned estimate was based on the statistical mean and total number of stations represented within the rural SR class. Since a high percentage of the rural SR class is made up of the 2-lane category, and the average posted speed limit for matched rural 2-lane stations was higher than the NCDOT values, a higher value was given. Note, however, that the difference is only 5 mph.

Note: The rural SR-2 Lane category account for roughly half (5000+) of the unmatched stations.

There are no LRS values for the local roads, therefore all local road speed limits were assigned the values submitted by TSU. The number of lanes and access code were input as 2 and 1, respectfully.

Also assigned were the route codes for the unmatched records. Originally given a value of –99999999, these were recalculated as follows:

Route Type	New Route Code
Interstate	10000000
US	20000000
NC	30000000
SR	40000000
Local	50000000

Because no routes end in 7 zeros (0000000), these are distinguishable from the other, *true* route codes. The first number also indicates the route type for which their level values indicated.

Analyzing and Assigning Values to Other (Unknown) Route Types:

At this point, all but 70 stations were accounted for. These stations still had erroneous values for their access control, posted speed limit, number of lanes, and route code fields. The stations were categorically

selected (by their level value), cross-referenced with the leveling chart, and their type noted. The following outlines the findings and assignments:

No. of Stations	Route Type	AC	SL	NL	RC	
27	None*		999	999	999	99999999
3	Blue Ridge Pkwy		3	55	2	90000000
28	Ramps		3	30	1	95000000
12	Service Roads		1	45	2	96000000
70						

* These records were not matched with the road layer due to their spatial properties (i.e., they were not close enough to any segment—within 30 meters—to be referenced).

Although the values for these stations are objective, it is believed that their low number should not have a significant impact on our research.

Appendix A: TSU Projections for Speed Limits

The following projections are from the Traffic Survey Unit, NCDOT in response to a request by Shannon M. McDonald, GIS, Wetherill Engineering, Inc. The numbers are objective opinions by the staff.

Route Type	**Urban Speed Limit**	**Rural Speed Limit**
Interstate	60	70
US	45	55
NC	45	55
SR 2-Lane	35	45
SR 4-Lane	45	55
Local	25	25

Appendix B: Linework (Road Layer) Leveling Chart

Linework Leveling Chart

Linework Leveling Chart						
Level	**Feature Description**	**Color**	**Weight**	**LC**	**Line Scale**	**Polygon**
	Interstate Highway					
27	Divided C/L	2	1	0		
27	4-Lane Undivided	2	11	GIS Rds Und Lan 4 (4)	30000	
27	5-Lane Undivided	16	11	GIS Rds Und Lan 5 (5)	30000	
27	6-Lane Undivided	39	11	GIS Rds Und Lan 6 (6)	30000	
27	7-Lane Undivided	40	11	GIS Rds Und Lan 7 (7)	30000	
27	8-Lane Undivided	63	11	GIS Rds Und Lan 8	30000	
27	9-Lane Undivided	64	11	GIS Rds Und Lan 9	30000	
27	10-Lane Undivided	87	11	GIS Rds Und Lan 10 (0)	30000	
	Bridge		MDL			
	US Highway					
28	Divided C/L	3	1	0		
28	2-Lane Undivided	3	11	GIS Rds Und Lan 2 (2)	30000	
28	3-Lane Undivided	22	11	GIS Rds Und Lan 3 (3)	30000	
28	4-Lane Undivided	23	11	GIS Rds Und Lan 4 (4)	30000	
28	5-Lane Undivided	24	11	GIS Rds Und Lan 5 (5)	30000	
28	6-Lane Undivided	25	11	GIS Rds Und Lan 6 (6)	30000	
28	7-Lane Undivided	26	11	GIS Rds Und Lan 7 (7)	30000	
28	8-Lane Undivided	48	11	GIS Rds Und Lan 8	30000	
28	9-Lane Undivided	49	11	GIS Rds Und Lan 9	30000	
28	10-Lane Undivided	50	11	GIS Rds Und Lan 10 (0)	30000	
28	US Dirt Road	120	3	GIS Rds Sec Type 41 (4)	30000	
	Bridge		MDL			
	NC Highway					
29	Divided C/L	5	1	0		
29	2-Lane Undivided	5	11	GIS Rds Und Lan 2 (2)	30000	
29	3-Lane Undivndivided	55	11	GIS Rds Und Lan 6 (6)	30000	
29	7-Lane Undivided	56	11	GIS Rds Und Lan 7 (7)	30000	
29	8-Lane Undivided	57	11	GIS Rds Und Lan 8	30000	
29	9-Lane Undivided	79	11	GIS Rds Und Lan 9	30000	

29	10-Lane Undivided	80	11	GIS Rds Und Lan 10 (0)	30000	
29	NC Dirt Road	81	3	GIS Rds Sec Type 41 (4)	30000	
	Bridge		MDL			
30	Database Problem					
31	Database Problem					
	Secondary Roads					
32	Divided C/L	6	1	0		
32	4-Lane Undivided	36	11	GIS Rds Und Lan 4 (4)	30000	
32	Type 51	62	1	0		
32	Type 41	136	3	GIS Rds Sec Type 41 (4)	30000	
32	Type 20	19	3	GIS Rds Sec Type 20 (2)	30000	
32	Type 10	25	3	GIS Rds Sec Type 10 (1)	30000	
	Bridge		MDL			
	Urban Roads					
35	Divided C/L	7	4	0		
35	2-Lane Undivided	7	0	0		
35	4-Lane Undivided	29	11	GIS Rds Und Lan 4 (3)	30000	
35	Urban Dirt Road	149	3	GIS Rds Sec Type 41 (4)	30000	
	Bridge		MDL			
	County Roads (Not System)					
38	Type 51	18	11	0		
38	Type 41	42	3	GIS Rds Sec Type 41 (4)	30000	
38	Type 20	43	3	GIS Rds Sec Type 20 (2)	30000	
38	Type 10	19	3	GIS Rds Sec Type 10 (1)	30000	
39	Service Road	0	0	0		
40	Trails Or Greenways	9	0	0		
	Building					
41	Shopping Center	171	2	0		Yes
41	Warehouse	187	2	0		Yes
41	Manufacturing Plant	188	2	0		Yes

	Railroads					
44	Mainline	24	0	GIS Mis Rr (1)	30000	
44	Spurs	31	0	GIS Mis Rr (2)	30000	
	Abandoned Railroad					
44	Without Tracks	14	0	GIS Mis Rr (3)	30000	
44	With Tracks	13	0	GIS Mis Rr (4)	30000	
47	Ramp	68	0	0		
	Boundary (Political)					
48	Old City Limits.	4	3	3		Yes
48	Out	22	3	3		
48	Inside Out	5	3	3		
49	New City Limits.	0	3	3		Yes
49	Out	22	3	3		
49	Inside Out	5	3	3		
50	U.E.B	181	2	0		Yes
52	State Line	0	6	6		
52	County Line	0	5	5		
	Boundary (Property)					
53	Military Reserv (Mrb)	0	3	0		Yes
53	Reservation (Rpb)	3	3	0		Yes
53	Park Boundary (Pb)	2	3	0		Yes
53	Fed. Game Land (Fgl)	4	3	0		Yes
53	Fire Districts (Fdb)	5	3	0		Yes
53	National Forest (Nfb)	6	3	0		Yes
53	State Forest (Sfb)	7	3	0		Yes
53	State Game Land (Sgl)	8	3	0		Yes
53	Nature Preserve	11	3	0		Yes
53	National Park	10	3	0		Yes
53	Industrial Park	9	3	0		Yes

54	Blue Ridge Parkway	20	11	0	
55	Airport Parameter	136	2	0	
56	Powerline	4	3	GIS Mis Pow	30000
56	Telephone Line	5	3	GIS Mis Tel	30000
56	Gasline	6	3	GIS Mis Gas	30000
56	Water Pipe	7	3	GIS Mis Wat	30000
57	Forest	120	1	3	
	Hydrography Structures				
58	Dam	99	3	0	
58	Locks	99	2	0	
58	Fords	99	1	0	
	Hydrography Single Line				
59	Creek (Single Line)	99	0	6	
59	Stream (Single Line)	100	0	6	
59	Canal (Single Line)	101	0	0	
59	Urban	0	0	0	
59	Single Line River	98	1	0	
	Shoreline				
60	Ocean	99	6	0	
60	Sound	99	5	0	
60	Intracoastal Waterway	99	7	0	
60	Inlet Or Bay	99	3	1	
60	River (Double Line)	99	1	0	
60	Creek (Double Line)	99	4	0	
60	Lake	99	3	0	
60	Pond	99	3	2	
60	Reservoir	99	3	1	
60	Ferry Lines	99	1	1	
60	Island Inside Shoreline	117	0	0	
60	Urban Pond	0	0	0	
	Water Body				
61	Buffers	99	0	0	
61	Ocean	99	1	0	

61	Sound	99	2	0
61	Intracoastal Waterway	99	3	0
61	Inlet Or Bay	99	4	0
61	River (Double Line)	98	0	0
61	Creek (Double Line)	101	0	0
61	Lake	99	0	1
61	Pond	99	0	2
61	Reservoir	99	0	3
61	Urban Pond	0	0	4
61	Island	102	0	0
61	Shown In Inset Only	100	0	5
63	Inset Cut Lines	29	0	0
	Leader lines			
5	I - Leadline Arrow	2	3	
6	US - Leadline Arrow	3	3	
7	NC - Leadline Arrow	5	3	
10	Urban - Leadline Arrow	30	3	
11	SR - Leadline Arrow	38	3	
14	Rail - Leadline Arrow	24	3	
17	Miles - Leadline Arrow	0	3	
17	Miles - Leadline Arrow	0	0	
- LINEWORK VERSION 1.1 Date: 5/08/00				

APPENDIX 8:

8. Determining the Final Mean Model for the 5-County Test Area

Jacqueline M. Hughes-Oliver
Tae-Young Heo

July 2003

Determining the Final Mean Model for the 5-County Test Area
Jacqueline M. Hughes-Oliver and Tae-Young Heo
July 18, 2003

1. Introduction

Modeling was based on two broad categories of data for the 5-county test area, "station data" and "census data." First, these data groups were individually summarized with respect to consolidating the information they contain. Validation of the census summary was accomplished by randomly splitting the data in approximately equal parts. Second, the summarized versions of both data groups were merged and variable selection performed to create the final model. This document provides a record of all steps performed, with a corresponding SAS file included in Appendix A. In Section 2, we describe the two data categories. In Section 3, we explain the steps used to summarize each category. Section 4 details model selection and Section 5 provides concluding remarks.

It is important to note that the final mean model does not account for correlations induced by spatial connectivity of road segments. Rather, the final mean model will be used as input to the spatial model, and these details will be provided in a later document.

2. Data Description

2.1 Station Data

In October 2002, following an intense period of data edits motivated by a preliminary mean model, we received a "final" dataset containing what we will refer to as "station data." Attributes/variables in this file include

- unique station identifiers,
- station location information (county, urban area, flags of being within cities and major or minor urban areas, and North Carolina state plane coordinates),
- route segment classifiers (categories such as interstate and US route, surface type, speed limit, number of lanes, access control status, route number, and description),
- PTC count information (actual counts, year, month and day for up to three counting periods),
- methods for converting PTC counts to annual average daily traffic (AADT) counts (axle factoring/conversion and ATR group identifier),
- land use in the area surrounding stations, and
- attributes for linking to both 1990 and 2000 census information.

An old listing of attributes in this file is given in Appendix B.

Data edits leading to the current file included changing some unique identifiers and swapping several records. Details are provided in the two documents *Data Edits Determined From "List of 55"* and *New Urban and Municipal Field Value Assignments and Their Relation to Station Locations and the PTC Data Model Process*, both authored by Shannon McDonald in November 2002.

These documents make it clear that attribute RTE_ADJUST supersedes all other versions of route segment classification. They also fully document the creation and anticipated use of attributes MUNICIPAL_F and URBAN_FLG. Finally, they question the process of SIPS assignment to stations in counties with both rural and urban assignments. Unfortunately, the documents do not explain which of the three surrogate attributes, SIPS, CO_NUM with URB_NUM, or COUNTYID, or a derived attribute should be used. Appendix C contains two-way tables showing current relationships between these

attributes. We use the second of these four options. Several facts motivated our decision. First, several data dictionaries suggest that attributes SIPS and COUNTYID should be the same, but they are not. Second, in a meeting on May 15, 2002 NCDOT personnel explained that attributes SIPS and URB_NUM are quite different in the information they provide. Kent Taylor went on to suggest that SIPS should be eliminated from the analysis, and we should use attributes CO_NUM and URB_NUM instead; we combined these attributes using

CO_URB=CO_NUM*10000 + URB_NUM*(URB_NUM>0).

But the aforementioned record swaps due to incorrect unique identifiers were not complete swaps and because of this it is not clear that attributes CO_NUM and URB_NUM correctly correspond to UNIQ_ID. For this reason, an attribute was derived as a replacement for SIPS based on unique identifiers. The two "unique" station identifiers UNIQ_ID and UNIQ_ID8 should match completely, except for an additional zero in the fourth digit of UNIQ_ID8. There is, however, one of the 3431 stations for which there is no match; it has UNIQ_ID= 1030955 and UNIQ_ID8= 03100633. We concluded that the attribute UNIQ_ID is probably the most accurate, and so we based our derivation of SIPS_NEW on this attribute:

SIPS_NEW=FLOOR(UNIQ_ID/10000).

In other words, our SIPS_NEW is the first three digits of attribute UNIQ_ID. SIPS_NEW is actually very similar to SIPS and COUNTYID, so it was also abandoned. As expected, CO_URB is a very important variable in the final mean model; this is discussed in Section 5.

Another edit was applied to attribute LANE, namely to double its value for interstates. This comes from the long-standing recognition that NCDOT's Line Referencing System (LRS) does not count both directions of lanes on divided highways. There may be other non-interstate divided highways for which this adjustment is needed, but this would take a case-by-case investigation that is not currently possible. We only apply the adjustment for interstates, using

LANES=LANES + LANES*(RTE_CLASS=1);

We applied one additional edit to the station data. Station data is matched to 1990 census data (described below) by the four attributes STATE90, COUNTY90, TRACT90, and GROUP90. Each record in the 1990 census file is merged with multiple records in the station data file. If a single record in the station data file is not matched to a census record, it is quite reasonable to conclude there is an error in the station data record. Station UNIQ_ID=1080682 had TRACT90=54009, but there was no such tract in the census file. We changed this station to have TRACT90=54005.

2.2 Census Data

In May 2002, in response to data queries in February and March of 2002, an updated file of 1990 census data was delivered. Shortly after, we also received a file of 2000 census data. The 2000 census data was pursued because the preliminary model seemed inadequate for the data, and it was hoped that the 2002 census data, being available on a much finer scale, would help explain more variability in the data and thus improve the model. Unfortunately, the 2000 census data offered no improvement beyond the 1990 census data. The problem is that while the 2000 block-level census data is available on a much finer scale, it has far fewer attributes than the 1990 block-group census data, thus resulting in a canceling effect that results in no improvement. We then decided to return to the 1990 census data.

The 1990 census data was cleaned before we initially received it from NCDOT. A data dictionary is included as Appendix D. Unfortunately, some attributes in the file are not included in the data dictionary. These include PERSON2, HOUSHOLD2, HOUSING2, AGELT_18_2, etc. Actually, summing the attribute PERSON2 yields the reported 1990 population of North Carolina as given in Table 4, Land Area, Population, and Housing Units: 1990 of document CPH-2 for North Carolina, available at http://www.census.gov/prod/cen1990/cph2/cph-2-35.pdf. Attribute PERSON does not sum to the reported North Carolina population. On the other hand, detailed age, racial, and other demographic attributes sum

to attribute PERSON, not PERSON2. For this reason, we use only the attributes listed in the data dictionary of Appendix D. All other attributes, with the exception of LAND_KM and WATER_KM, were ignored. Attribute MILIQUAR was also deleted because all values for the 5-county test area were zero. This attribute should be reinstated when developing the model for the entire state, since it will then take non-zero values.

3. Category Summaries

3.1 Census Data

Principal components analysis (PCA) was performed on 146 census attributes to reduce them to a handful of principal components. Principal components are linear combinations of the original variables selected in such a way that all principal components are orthogonal to each other and they capture most of the variability in the original variables. The primary benefit of PCA is data reduction—use fewer variables but retain the predictive power of the original set of variables. Varimax orthogonal rotation was performed to improve interpretation of the principal components.

The 1990 census file actually contains almost 200 attributes. Preprocessing to reduce this number to 146 was accomplished by noting that the attributes had many singularities. For example, attributes MALE and FEMALE sum to PERSONS, so there is no need to keep all three attributes; FEMALE was deleted. Some attributes were also scaled relative to size of the census block. Specifically, population density and housing density were obtained as

POPDEN=PERSONS/(SHAPE_AREA/1000000)

HOUSEDEN=HOUSING/(SHAPE_AREA/1000000),

where SHAPE_AREA is derived by the GIS software ArcView and is measured in units of square meters. Attribute LAND_KM, also a measure of census block size (sq km), is available in the census file, but there were many instances where its value was zero and could not be used for density calculations. PERSONS and HOUSING were excluded from the PCA, instead being replaced by POPDEN and HOUSEDEN.

We also investigated PCA where all counts were made relative to either PERSONS or HOUSING. For example, the count of MALES was replaced by the proportion of males in the census block

MALES=MALES/PERSONS.

This investigation was motivated by the belief that relative counts might be more informative. This was not the case, however. PCA where all count variables were replaced by proportion variables were not as effective in using a small number of principal components to capture variability. For example, the first principal components using counts captured about 50% of the variability while the first principal components using proportions captured less than 13% of the variability. All further analysis used counts, except for the attributes POPDEN and HOUSEDEN.

PCAs are most stable when they are developed on independent datasets that lead to the same conclusions. For this purpose, the 3431 records were separated into approximately equal groups, and PCA was performed on both groups. The resulting principal components were compared across groups to determine whether or not they represent the same information. More specifically, correlation coefficients were obtained for the group of principal components from dataset #1 compared to dataset #2. PCAs for the two datasets are comparable if each component from one dataset is highly correlated with only one component from the other dataset. If correlations are weak or a component is equally correlated with two or more components from another group, then fewer principal components should probably be retained. Splitting the dataset into two smaller datasets was accomplished by checking if a standard normal pseudo-random variate was negative or not.

In both datasets, 21 principal components accounted for more than 89% of the variability in the 146 census attributes. These two sets of 21 principal components were also highly correlated with each other. Upon further investigation of the 21 components, however, we decided to reduce to only seven principal components. The 21 components included some that were functions of only one of two attributes, and this does not accomplish the amount of reduction we were seeking. The seven principal components only account for approximately 77% of the variability in the census attributes but the r-square value for the overall mean model was practically unchanged, dropping from .78 for the model using all 21 components to .75 for the model with only seven components. Furthermore, correlations between the sets of seven principal components were always at least 0.98, indicating strong agreement. Details of the PCA with retaining seven principal components are given in Appendix E. Even though the number of components retained was determined using the smaller datasets, the entire dataset was used to obtain the principal components to be included in refining the mean model. A summary of the seven principal components is given in Table 1.

Table 1: Summary of Rotated Principal Components Retained

Principal Component	Attributes in Factor (- means negative impact on component)
2. Single	APARTMNT, ROOM_1_3, RNT_2550, ONEPERHH, RNT_5075, VEHICL_1, SEASONAL, PUBSEWER, INC_1525, ATTACHED, DUPLEX
3. College	AGEIS_20, COLLEGE, AGE18_19, DORMITOR, SINGLE_ SAMESTAT, AGE21_24, ASIAN, ABLEENGL, AMIND, OTH_HEAT, O_NOINST, INSTREET, MENTAL
4. Poor	CHILDPOV, NOVEHICL RNT_LT25, PUBL_INC, INPOVRTY, BLACK, SINGWICH, INC_LT15, NODIPLOM, PUBTRANS, VAL_2550, INCPRCAP(-), INC_MEDN (-)
5. Farm Life	LAND_KM, COALWOOD, BOTL_GAS, FARM_INC, VAL_LT25, F_MIGRNT, WATER_KM, RNT_MEDI (-), POPDEN (-), HOUSEDEN(-), HOUSHOLD (-), FAMILIES (-)
6. Elderly	AGE65_74, AGE75_84, SOCS_INC, WIDOWED, RETI_INC, FUELKERO, BLTBFR70, NURSHOME, MEDYRBLT
7. Wealthy	VAL_2C3C, VAL_GT3C, RNT_751K, RNT_GT1K, VAL_MEDI
1. Other	This component includes all other attributes, and includes middle income families with and without children who are employed.

3.2 Station Data

Summaries obtained for the station data are based on less formal analyses than those for the census data. All summaries are simple merges of related variables, also considered to be interaction terms in statistical jargon. Five basic merges were performed.

LANES and RTE_CLASS (renamed from RTE_ADJUST) were merged to produce

LANES_RTE=LANES*100 + RTE_CLASS.

In general, higher classes of routes often have more lanes. For example, 13 of the 14 stations on routes with eight lanes are interstate stations, and the last station is on a US route. By grouping number of lanes with route classifications, we hope to reduce the complexity of the mean model. Indeed, we expect that when LANES_RTE=801 (meaning 8 lanes on an interstate) the model will predict higher AADT than when LANES_RTE=202 (meaning 2 lanes on a US route).

Similarly, speed limit and route classification are combined to yield

SPEED_RTE=SPEED*100 + RTE_CLASS,

while number of lanes, speed limit, and route classification are all combined to yield

LANES_SPEED_RTE=LANES*100000 + SPEED*100 + RTE_CLASS.

Actually, all three variables might not always be needed, because information may be gleaned from two of them to imply the third. We will return to this issue later in the document.

Attributes CYCLE_YR and RD_SURFACE are highly related, thus motivating their merge. The code is

CYCLE_SURFACE=100*(CYCLE_YR="A") + 200*(CYCLE_YR="O") + 300*(CYCLE_YR="E") + 400*(CYCLE_YR="V") + RD_SURFACE

"Primary" routes are counted annually, so that CYCLE_YR=A, and they are always asphalt surfaces, so that RD_SURFACE=1, resulting in the only cases of CYCLE_SURFACE=101. Dirt roads are only counted in even years, resulting in CYCLE_SURFACE=303. Former dirt road that were recently paved but have not yet been transferred to an odd-year counting schedule constitute the cases CYCLE_SURFACE=302. It is also note-worthy that annually counted stations are usually on either interstates, US routes, or NC routes, odd-year counting is typically for SR and local routes, while variable counting is reserved for urban routes.

Finally, the merging of MUNICPAL_F and URBAN_FLG results in six designations of station location, depending on existence inside/outside city limits and inside/outside major or minor urban areas:

CITY_URB=100*(MUNICPAL_F=1) + URBAN_FLG;

In preparation for using station data in the mean model, the newly created attributes, plus attributes RTE_CLASS, CO_URB, A_CONTROL, YEAR, MONTH, DAY, and LUCODE, were all converted to class variables. We take a rather conservative approach in fitting a near-saturated model to avoid missing any major effects.

4. Model Selection

Combining the seven principal components (called FACTOR1, FACTOR2, …, FACTOR7) from the census data with the 114 attributes from the station data, a full regression model for AADT yields an r-square value of 0.7786 and well-behaved residuals. We are, however, in search of a simpler model that uses fewer attributes. We accomplish this using a stepwise selection procedure with significance-level-to-enter of 0.50 and significance-level-to-stay of 0.10. The selected model is based on 60 attributes, with r-square value of 0.7766 and Mallow's Cp of 31.1. A summary of the model is given in Table 2 below and details are given in Appendix F. Because of linear dependencies, the INTERCEPT term actually also corresponds to the following attributes:

RTE_CLASS5 (Local), CO_URB (Wake, 108), LANES_RTE18 (8 lanes, US routes), SPEED_RTE26 (70 mph, INTER), LANES_SPEED_RTE58 (8 lanes, 65 mph, INTER), CYCLE_SURFACE7 (variable cycle, concrete surface), A_CONTROL3 (interstate), YEAR5 (2001), MONTH13 (missing month), DAY5 (missing day), LUCODE18 (transitional areas), CITY_URB6 (in city, in minor urban area).

Table 2: Summary of the Selected Model

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	6.23009	0.18230	34.17	<.0001
Factor1	Other	1	0.18055	0.01431	12.62	<.0001

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Factor2	Single	1	0.06189	0.01426	4.34	<.0001
Factor3	College	1	0.02196	0.01424	1.54	0.1232
Factor5	Farm	1	-0.22093	0.02482	-8.90	<.0001
Factor7	Wealthy	1	0.02437	0.01353	1.80	0.0719
rte_class1	INTER	1	1.58235	0.19193	8.24	<.0001
co_urb4	Durham, 103	1	-0.27598	0.05143	-5.37	<.0001
co_urb5	Johnston	1	-0.10479	0.05705	-1.84	0.0663
co_urb6	Orange	1	0.23604	0.06945	3.40	0.0007
lanes_rte1	2, US	1	-0.72604	0.08761	-8.29	<.0001
lanes_rte2	2, NC	1	-0.68759	0.08106	-8.48	<.0001
lanes_rte3	2, SR	1	-1.40233	0.14900	-9.41	<.0001
lanes_rte4	2, local	1	-2.14652	0.15592	-13.77	<.0001
lanes_rte11	4, SR	1	-0.54427	0.15121	-3.60	0.0003
lanes_rte17	8, INTER	1	0.67292	0.31274	2.15	0.0315
speed_rte6	25, SR	1	-1.99556	0.22916	-8.71	<.0001
speed_rte19	55, INTER	1	-0.78039	0.22330	-3.49	0.0005
speed_rte20	55, US	1	0.61472	0.08981	6.84	<.0001
lanes_speed_rte5	2, 25, SR	1	1.16671	0.29628	3.94	<.0001
lanes_speed_rte9		1	-0.12178	0.05291	-2.30	0.0214
lanes_speed_rte17	2, 55, SR	1	-0.36403	0.05036	-7.23	<.0001
lanes_speed_rte20	3, 30, SR	1	-1.12314	0.36648	-3.06	0.0022
lanes_speed_rte21	3, 35, US	1	-0.43645	0.27367	-1.59	0.1109
lanes_speed_rte23	3, 35, SR	1	-1.07605	0.26215	-4.10	<.0001
lanes_speed_rte33	4, 35, NC	1	-0.41579	0.16305	-2.55	0.0108
lanes_speed_rte36	4, 45, US	1	0.29614	0.17145	1.73	0.0842
lanes_speed_rte41	4, 55, NC	1	1.21275	0.34416	3.52	0.0004

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lanes_speed_rte47	5, 45, US	1	1.15271	0.53774	2.14	0.0321
lanes_speed_rte49	6, 35, SR	1	-0.64290	0.34152	-1.88	0.0599
lanes_speed_rte50	6, 45, US	1	0.55461	0.21203	2.62	0.0089
lanes_speed_rte55	6, 65, INTER	1	0.58097	0.24362	2.38	0.0171
lanes_speed_rte56	8, 45, US	1	1.16814	0.76019	1.54	0.1245
lanes_speed_rte57	8, 60, INTER	1	-1.05003	0.42207	-2.49	0.0129
cycle_surface1	Ann, Asph	1	1.12672	0.12889	8.74	<.0001
cycle_surface2	Odd, Conc	1	0.66526	0.09340	7.12	<.0001
cycle_surface3	Odd, Soil	1	1.90623	0.76120	2.50	0.0123
cycle_surface4	Even, Conc	1	1.09071	0.21282	5.13	<.0001
cycle_surface5	Even, Soil	1	0.78204	0.14952	5.23	<.0001
cycle_surface6	Vari, Asph	1	1.51836	0.10935	13.89	<.0001
cycle_surface7	Vari, Conc	1	1.05681	0.08294	12.74	<.0001
a_control2	Medium	1	-0.36254	0.13969	-2.60	0.0095
year2	1998	1	-0.22280	0.05931	-3.76	0.0002
year3	1999	1	-0.15888	0.05608	-2.83	0.0046
month1	Jan	1	-0.28200	0.08586	-3.28	0.0010
month2	Feb	1	-0.19963	0.09564	-2.09	0.0369
month5	May	1	-0.33902	0.04747	-7.14	<.0001
month6	June	1	-0.23031	0.06234	-3.69	0.0002
month11	Nov	1	0.22131	0.10669	2.07	0.0381
day4	Thu	1	-0.41573	0.12556	-3.31	0.0009
lucode1	Residential	1	-0.12207	0.04154	-2.94	0.0033
lucode2	Comm & Services	1	0.16336	0.05495	2.97	0.0030
lucode4	Trans, Comm, Util	1	0.40724	0.06583	6.19	<.0001
lucode5	Indust & Comm Complxs	1	0.45578	0.14543	3.13	0.0017

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lucode6	Mxd Urban or Built-up	1	0.13388	0.04741	2.82	0.0048
lucode10	Other Ag Land	1	0.67339	0.25360	2.66	0.0080
lucode11	Shrub & Brush Rangeland	1	-0.97617	0.43900	-2.22	0.0262
city_urb1	out, out	1	-0.69996	0.05688	-12.31	<.0001
city_urb3	out, minor	1	-0.39946	0.22589	-1.77	0.0771
city_urb4	city, out	1	-0.24726	0.07410	-3.34	0.0009
city_urb5	city, minor	1	0.21557	0.05360	4.02	<.0001

Stations with absolute residuals larger than 2 are listed in Table 3 below. Many of these stations have already been fully investigated by NCDOT personnel and have been labeled "unusual but okay" for a variety of reasons; see *Data Edits Determined From "List of 55"*. By identifying these stations, the model is working as it should.

Table 3: Outlier Stations From the Selected Model

Obs	UNIQ_ID	AADT	baadt	paadt	raadt
1	0181759	6532	5.79441	3.77547	2.01894
2	1030304	699	3.70591	5.73771	-2.03179
3	1030658	352	3.23079	5.64238	-2.41159
4	1030657	525	3.49971	5.64238	-2.14267
5	1030053	169	2.78983	4.99464	-2.20481
6	1030644	343	3.21410	5.58444	-2.37035
7	1030561	18302	7.12032	5.05602	2.06430
8	1030544	36	2.04767	4.67883	-2.63115
9	1030705	95	2.48625	5.07038	-2.58413
10	1030434	695	3.70166	5.78271	-2.08105
11	1030654	462	3.41137	5.82793	-2.41656
12	1030235	8009	6.03553	8.04917	-2.01364
13	0501816	160	2.75946	4.78944	-2.02998

Obs	UNIQ_ID	AADT	baadt	paadt	raadt
14	0501608	9082	6.18923	3.99492	2.19430
15	0501613	8153	6.05708	3.86104	2.19604
16	1030768	70	2.33894	4.48748	-2.14854
17	1080861	27731	7.73737	5.60554	2.13183
18	1080915	24018	7.51809	5.14976	2.36833
19	1080168	85405	9.68939	7.16847	2.52092
20	1080175	228	2.96201	5.23324	-2.27123
21	1080641	66102	9.20541	6.32483	2.88058
22	1080264	483	3.44183	5.47290	-2.03107
23	1080498	1756	4.45559	6.55738	-2.10180
24	0910584	52	2.20394	4.41330	-2.20936
25	1080698	10510	6.37266	4.20687	2.16578
26	1080616	469	3.42164	5.49988	-2.07824
27	0912026	10429	6.36280	4.23721	2.12560
28	1080512	925	3.91948	6.30312	-2.38364
29	1080514	2135	4.63318	6.81135	-2.17817
30	1080450	336	3.20087	5.56312	-2.36225
31	1080819	17257	7.03709	5.03433	2.00276
32	1080815	36575	8.17781	5.44157	2.73624
33	1080812	25800	7.62648	5.25713	2.36934
34	0910673	10436	6.36366	3.57772	2.78593
35	1080451	618	3.61574	6.00247	-2.38673
36	1080715	17801	7.08091	4.60084	2.48006
37	0910034	2158	4.64312	6.95505	-2.31192
38	0910842	95	2.48625	4.51263	-2.02638
39	1080473	847	3.85103	6.00840	-2.15737
40	1080714	15188	6.85962	4.70427	2.15535
41	1080111	597	3.59083	6.07494	-2.48411

Obs	UNIQ_ID	AADT	baadt	paadt	raadt
42	1080879	39077	8.28675	6.08340	2.20335
43	1080174	245	3.00492	5.34986	-2.34494
44	1080511	788	3.79582	6.12820	-2.33238

To ensure validity of the selected attributes, we ran the regression on one part of the dataset and used the results to predict for the other dataset. Residuals were all less than three in absolute value and more than 95% had absolute values less than two. An assumption of normality is also well supported.

5. Conclusions

The selected model is effective in capturing more than 77% of the variability in AADT counts (after a power transformation) in the 5-county test area. It is based on only 60 attributes instead of the more than 300 possible attributes. The model has been validated by splitting the test area data into approximately two equal parts.

For future use in developing the spatial model for the entire state, we will perform the following steps.

1. Run PCA on the statewide training 1990 census data. Output 7 components.
2. Create the additional attributes for the statewide training station data and obtain the associated class variables.
3. Perform model selection and validate the model using the statewide validation set.

APPENDIX A: SAS Program

```
** Last updated: July 17, 2003;
** Data used:  Delivered by Shannon October 2002;
**----------------------------------------------;
ods trace on;
*OPTIONS PAGENO=1 NODATE NONUMBER PS=44 LS=120; *landscape printing;
OPTIONS PAGENO=1 LS=120;      *landscape printing;

*libname ptcfinal "C:\Documents and Settings\hughesol\My
Documents\Research\DOT\SASLibraries\ptcdemo";
DATA PTCdemo; SET ptcfinal.ptcdemo; id=_n_; *proc print;
*The following variables did not come from DOT/Shannon, but were created by
Tae-Young. Omit them.;
      drop urb_no aadt_up lanes_new rte_new;
*The most recently created routecode variable is rte_adjust;
      rte_class=rte_adjust;
*Combine co_num & urb_num to create "replacement" for sips;
      co_urb=co_num*10000 + urb_num*(urb_num>0);
*As of 7/14/03, it's not clear what I should use for sips, so for now I'll
just take the first 3 digits of uniq_id
      *--this assumes the correct id is uniq_id and not uniq_id8 ...;
      sips_new=floor(uniq_id/10000);
/*Leave the following data out of the model. Use model to
      predict at these sites, thus validating the model ...
*TSU updated some stations according to 2001 counts;
      if year_new>0 then do;
            year=year_new;
            aadt=aadt_new;
            month=month_01;
            day=day_01;
            count1=count1_01;
            count2=count2_01;
            end;
      *AADT_UP=AADT+(AADT_NEW-AADT)*(AADT_NEW>0);
*/
*Double LRS-based lanes for interstates;
      lanes=lanes + lanes*(rte_class=1);
      run;

***************************** BEGIN: Summaries for report
**************************************;
ods html file='C:\Documents and Settings\hughesol\My
Documents\Research\DOT\SASprograms\MeanModel\summaries.html';
ods select Freq.CO_URB_by_SIPS.CrossTabFreqs2
Freq.COUNTYID_by_SIPS.CrossTabFreqs3
      Freq.SIPS_NEW_by_SIPS.CrossTabFreqs4
Freq.COUNTYID_by_CO_URB.CrossTabFreqs7
      Freq.SIPS_NEW_by_CO_URB.CrossTabFreqs8
Freq.SIPS_NEW_by_COUNTYID.CrossTabFreqs12;
proc freq data=ptcdemo; tables (sips co_urb countyid sips_new)*(sips co_urb
countyid sips_new);
title 'Frequency Tables for "Surrogate" Attributes'; run;
***************************** END: Summaries for report
**************************************;
ods html close;
```

```
***************************** BEGIN: PCA of 1990 Census Data
*************************************;
ods html file='C:\Documents and Settings\hughesol\My
Documents\Research\DOT\SASprograms\MeanModel\pca_census.html';
*Prepare data for principal components analysis of census information ...;
data ptcdemo; set ptcdemo;
      drop miliquar;                      *This only takes 0 values;
*Need population density (persons per area), but don't really know what
census variables to use.
*Here are my "guesses" of variable meanings as of 7/15/03:
*     persons=total/actual number of persons in region      **will use this
for numerator**
*     agege_03 < persons2 < persons  !!NOT TRUE!! PERSONS2 ACTUALLY SUMS TO
REPORTED VALUE! But I don't have things like male & female to sum to
persons2, only for summing to persons.
*     samp_pop=persons
*     per_samp ??? people in region/people in county ??? -- takes values like
12, 14, 9, ...
*     land_km=area of region, in sq km (?), obtained from census  BUT IS
SOMETIMES 0!!
*     shape_area=area of region, in sq m (?), obtained from ArcView     THIS
IS NEVER 0;
      /*if popden=0 then delete;    *Won't have census values; NOT TRUE!! */
      if persons=0 then persons=.0001;    *if persons=0 then delete;
      if housing=0 then housing=.0001;    *if housing=0 then delete;
      popden=persons/(shape_area/1000000); *per sq km;
      houseden=housing/(shape_area/1000000);
*Adjust some other census variables to be relative to "shape_area";
      families=families/(shape_area/1000000);
      houshold=houshold/(shape_area/1000000);
*Ignore census variables that cause singularities;
      drop female o_ethnic age15_19 age20_24 agegt_84 agege_05 agege_25
divorced singnoch
      juvenile occupied vacant rent_occ room_gt6 prunitg4 agg_valu persons2--
per_samp cantengl
      noncitiz born_for frm_abrd com_gt44 gradschl notinwrk public_ manual
famiwork inc_gt75
      inc_aggr prvwater prvsewer vehiclg2 aggvehcl;
      run;

*Create random numbers that will allow record selection for training &
validating models;
data ptcdemo; set ptcdemo; myran=rannor(id); run;

OPTIONS PAGENO=1 NODATE NONUMBER PS=44 LS=120;
ods select eigenvalues screeplot orthrotfactpat;
/*prin comp anal for full data--demo*/ PROC FACTOR DATA=ptcdemo
SIMPLE METHOD=PRIN priors=one score
nfactor=7 SCREE ROTATE=VARIMAX outstat=spc7 reorder;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
title1 'Principal Components Analysis of PTCDEMO';
RUN;
proc score data=ptcdemo score=spc7 out=pc7;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
run;
```

```
data ptcdemo; set pc7; /*This adds new output from PCA to previous PCA
output*/
data ptcdemo1; set ptcdemo; if myran<0 then delete; *if id>1715 then delete;
*if (id/2=floor(id/2)) then delete; run;
ods select eigenvalues screeplot orthrotfactpat;
/*prin comp anal for 1st part of data--demo*/ PROC FACTOR DATA=ptcdemo1
METHOD=PRIN priors=one score
nfactor=7 SCREE ROTATE=VARIMAX outstat=spc7 reorder;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
title1 'Principal Components Analysis of PTCDEMO1';
RUN;
proc score data=ptcdemo score=spc7 out=pc7;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
run;

data ptcdemo; set pc7; /*This adds new output from PCA to previous PCA
output*/
data ptcdemo2; set ptcdemo; if myran ge 0 then delete; *if id le 1715 then
delete; *if (id/2 ne floor(id/2)) then delete; run;
ods select eigenvalues screeplot orthrotfactpat;
/*prin comp anal for 2nd part of data--demo*/ PROC FACTOR DATA=ptcdemo2
METHOD=PRIN priors=one score
nfactor=7 SCREE ROTATE=VARIMAX outstat=spc7 reorder;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
title1 'Principal Components Analysis of PTCDEMO2';
RUN;
proc score data=ptcdemo score=spc7 out=pc7;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
run;

data ptcdemo; set pc7; run;
****************************** END: PCA of 1990 Census Data
*************************************;
ods html close; run;

****************************** BEGIN: Model selection
*************************************;
*OPTIONS PAGENO=1 NODATE NONUMBER PS=40 LS=120;
ods html file='C:\Documents and Settings\hughesol\My
Documents\Research\DOT\SASprograms\MeanModel\selection.html';
data ptcdemo; set ptcdemo;    /* Create new variables, not class variables */
*Create interactions between lanes, route classification, and speed;
      lanes_rte=lanes*100 + rte_class;
      speed_rte=speed*100 + rte_class;
      lanes_speed_rte=lanes*100000 + speed*100 + rte_class;
*Create interactions between cycle year and route surface;
      cycle_surface=100*(cycle_yr="A") + 200*(cycle_yr="O") +
                          300*(cycle_yr="E") + 400*(cycle_yr="V") +
                          rd_surface;
*Create interactions between municipality and urban flags;
      city_urb=100*(municpal_f=1) + urban_flg;
*Change missing values for month and day to zeros;
      if month=. then month=0;
      if day=. then day=0;
      run;
```

```
proc freq data=ptcdemo; tables rte_class sips_new co_urb lanes_rte speed_rte
lanes_speed_rte cycle_surface a_control year month day lucode city_urb;
title1 'Frequency tables for all class variables of station data'; run;
/*Class variables created using IML*/ proc iml;
      use ptcdemo;
      read all var{rte_class} into rte_class;
rte_class=design(rte_class)[,1:4];
      varnames='rte_class1':'rte_class99';
      create temp1 from rte_class[colname=varnames]; append from rte_class;
      use ptcdemo;
      read all var{co_urb} into co_urb; co_urb=design(co_urb)[,1:8];
      varnames='co_urb1':'co_urb99';
      create temp2 from co_urb[colname=varnames]; append from co_urb;
      use ptcdemo;
      read all var{lanes_rte} into lanes_rte;
lanes_rte=design(lanes_rte)[,1:17];
      varnames='lanes_rte1':'lanes_rte99';
      create temp3 from lanes_rte[colname=varnames]; append from lanes_rte;
      use ptcdemo;
      read all var{speed_rte} into speed_rte;
speed_rte=design(speed_rte)[,1:25];
      varnames='speed_rte1':'speed_rte99';
      create temp4 from speed_rte[colname=varnames]; append from speed_rte;
      use ptcdemo;
      read all var{lanes_speed_rte} into lanes_speed_rte;
lanes_speed_rte=design(lanes_speed_rte)[,1:57];
      varnames='lanes_speed_rte1':'lanes_speed_rte99';
      create temp5 from lanes_speed_rte[colname=varnames]; append from
lanes_speed_rte;
      use ptcdemo;
      read all var{cycle_surface} into cycle_surface;
cycle_surface=design(cycle_surface)[,1:7];
      varnames='cycle_surface1':'cycle_surface99';
      create temp6 from cycle_surface[colname=varnames]; append from
cycle_surface;
      use ptcdemo;
      read all var{a_control} into a_control;
a_control=design(a_control)[,1:2];
      varnames='a_control1':'a_control99';
      create temp7 from a_control[colname=varnames]; append from a_control;
      use ptcdemo;
      read all var{year} into year; year=design(year)[,1:4];
      varnames='year1':'year99';
      create temp8 from year[colname=varnames]; append from year;
      use ptcdemo;
      read all var{month} into month; month=design(month)[,2:13];
      varnames='month1':'month99';
      create temp9 from month[colname=varnames]; append from month;
      use ptcdemo;
      read all var{day} into day; day=design(day)[,2:5];
      varnames='day1':'day99';
      create temp10 from day[colname=varnames]; append from day;
      use ptcdemo;
      read all var{lucode} into lucode; lucode=design(lucode)[,1:17];
      varnames='lucode1':'lucode99';
      create temp11 from lucode[colname=varnames]; append from lucode;
```

```
      use ptcdemo;
      read all var{city_urb} into city_urb; city_urb=design(city_urb)[,1:5];
      varnames='city_urb1':'city_urb99';
      create temp12 from city_urb[colname=varnames]; append from city_urb;
/*Merge class variables*/data ptcdemo;
merge ptcdemo temp1 temp2 temp3 temp4 temp5 temp6 temp7 temp8 temp9 temp10
temp11 temp12;
label
factor1='Other' factor2='Single' factor3='College' factor4='Poor'
factor5='Farm' factor6='Elderly' factor7='Wealthy'
rte_class1='INTER' rte_class2='US' rte_class3='NC' rte_class4='SR'
co_urb1='Chatham' co_urb2='Chatham, 103' co_urb3='Durham' co_urb4='Durham,
103' co_urb5='Johnston'
co_urb6='Orange' co_urb7='Orange, 101' co_urb8='Orange, 103'
lanes_rte1='2, US' lanes_rte2='2, NC' lanes_rte3='2, SR' lanes_rte4='2,
local' lanes_rte5='3, US'
lanes_rte6='3, NC' lanes_rte7='3, SR' lanes_rte8='4, INTER' lanes_rte9='4,
US' lanes_rte10='4, NC'
lanes_rte11='4, SR' lanes_rte12='5, US' lanes_rte13='5, SR' lanes_rte14='6,
INTER' lanes_rte15='6, US'
lanes_rte16='6, SR' lanes_rte17='8, INTER'
speed_rte1='20, US'
speed_rte6='25, SR' speed_rte11='35, SR' speed_rte19='55, INTER'
speed_rte20='55, US'
speed_rte23='60, INTER' speed_rte24='65, INTER' lanes_speed_rte5='2, 25, SR'
lanes_speed_rte17='2, 55, SR' lanes_speed_rte20='3, 30, SR'
lanes_speed_rte21='3, 35, US'
lanes_speed_rte23='3, 35, SR' lanes_speed_rte24='3, 40, US'
lanes_speed_rte31='4, 25, SR'
lanes_speed_rte33='4, 35, NC' lanes_speed_rte36='4, 45, US'
lanes_speed_rte41='4, 55, NC'
lanes_speed_rte42='4, 55, SR' lanes_speed_rte46='5, 35, SR'
lanes_speed_rte47='5, 45, US'
lanes_speed_rte49='6, 35, SR' lanes_speed_rte50='6, 45, US'
lanes_speed_rte53='6, 55, NC'
lanes_speed_rte55='6, 65, INTER' lanes_speed_rte56='8, 45, US'
lanes_speed_rte57='8, 60, INTER'
cycle_surface1='Ann, Asph' cycle_surface2='Odd, Conc' cycle_surface3='Odd,
Soil' cycle_surface4='Even, Conc'
cycle_surface5='Even, Soil' cycle_surface6='Vari, Asph' cycle_surface7='Vari,
Conc'
a_control1='Low' a_control2='Medium'
year1='1997' year2='1998' year3='1999' year4='2000'
month1='Jan' month2='Feb' month3='Mar' month4='Apr' month5='May'
month6='June' month7='July' month8='Aug'
month9='Sept' month10='Oct' month11='Nov' month12='Dec'
day1='Mon' day2='Tue' day3='Wed' day4='Thu'
lucode1='Residential' lucode2='Comm & Services' lucode4='Trans, Comm, Util'
lucode5='Indust & Comm Complxs'
lucode6='Mxd Urban or Built-up' lucode10='Other Ag Land' lucode11='Shrub &
Brush Rangeland'
city_urb1='out, out' city_urb2='out, major' city_urb3='out, minor'
city_urb4='city, out' city_urb5='city, major';
run;

ods select boxcox;
```

```
/*Run full model and determine Box-Cox transformation*/PROC TRANSREG
DATA=ptcdemo SS2 DETAILS OUTTEST=B;
MODEL BOXCOX(AADT/LAMBDA=-2 -1 -0.5 TO 0.5 BY 0.05 1 2 CONVENIENT
ALPHA=0.00001) =IDENTITY(
      factor1--factor7 rte_class1--city_urb5);
OUTPUT OUT=OUT5 RDPREFIX=R PPREFIX=P;
title 'Box-Cox regression using full model (PCA of census + all station
data)';
RUN;
/*Best Box-Cox power is .2*/ data ptcdemo; set ptcdemo; baadt=aadt**(.2);
run;
/*Run full model using aadt^.2*/proc reg data=ptcdemo;
model baadt=factor1--factor7 rte_class1--city_urb5 ;
output out=out1 p=pred1 r=raadt;
title1 "aadt^.2 regression using full model (PCA of census + all station
data)"; run;
/* NORMALITY check*/ PROC UNIVARIATE DATA=OUT1 NORMAL PLOT; VAR RAADT;
title2 'Normality check on residuals'; RUN;
data mycopy; set ptcdemo; run;
data ptcdemo; set mycopy; run;
ods select step66.anova step66.selparmest selectionsummary;
/*Do model selection using aadt^.2*/proc reg data=ptcdemo;
model baadt=factor1--factor7 rte_class1--city_urb5 / selection=stepwise
details=steps;
output out=out1 p=pred1 r=raadt;
title1 "Stepwise Selection for aadt^.2 regression from full model (PCA of
census + all station data)"; run; quit;
/*Fit final model using aadt^.2*/proc reg data=ptcdemo outest=est;
model baadt=factor1 factor2 factor3 factor5 factor7 rte_class1 co_urb4--
co_urb6 lanes_rte1--lanes_rte4
lanes_rte11 lanes_rte17 speed_rte6 speed_rte19 speed_rte20
lanes_speed_rte5 lanes_speed_rte9 lanes_speed_rte17 lanes_speed_rte20
lanes_speed_rte21 lanes_speed_rte23
lanes_speed_rte33 lanes_speed_rte36 lanes_speed_rte41 lanes_speed_rte47
lanes_speed_rte49 lanes_speed_rte50
lanes_speed_rte55--lanes_speed_rte57 cycle_surface1--cycle_surface7
a_control2 year2 year3
month1 month2 month5 month6 month11 day4
lucode1 lucode2 lucode4 lucode5 lucode6 lucode10 lucode11 city_urb1 city_urb3
city_urb4 city_urb5;
output out=out1 p=paadt r=raadt;
title1 "Fit final model for aadt^.2 regression from full model (PCA of census
+ all station data)"; run; quit;
/* NORMALITY check*/ PROC UNIVARIATE DATA=OUT1 NORMAL PLOT; VAR RAADT;
title2 'Normality check on residuals'; RUN;
data check; set out1; keep uniq_id aadt baadt paadt raadt; if (abs(raadt) le
2) then delete; run;
proc print data=check; title2 'abs(residual)>2'; run;
/*Try the final set of variables with only ptcdemo1 data*/
data ptcdemo1; set ptcdemo; baadt_orig=baadt; if myran ge 0 then baadt=.;
run;
/*Fit final model using aadt^.2*/proc reg data=ptcdemo1 outest=est;
model baadt=factor1 factor2 factor3 factor5 factor7 rte_class1 co_urb4--
co_urb6 lanes_rte1--lanes_rte4
lanes_rte11 lanes_rte17 speed_rte6 speed_rte19 speed_rte20
```

```
lanes_speed_rte5 lanes_speed_rte9 lanes_speed_rte17 lanes_speed_rte20
lanes_speed_rte21 lanes_speed_rte23
lanes_speed_rte33 lanes_speed_rte36 lanes_speed_rte41 lanes_speed_rte47
lanes_speed_rte49 lanes_speed_rte50
lanes_speed_rte55--lanes_speed_rte57 cycle_surface1--cycle_surface7
a_control2 year2 year3
month1 month2 month5 month6 month11 day4
lucode1 lucode2 lucode4 lucode5 lucode6 lucode10 lucode11 city_urb1 city_urb3
city_urb4 city_urb5;
output out=out1 p=paadt r=raadt;
title1 "Fit final model to ptcdemo1"; run; quit;
data out1; set out1; raadt=baadt_orig-paadt; if myran<0 then delete; run;
/* NORMALITY check*/ PROC UNIVARIATE DATA=OUT1 NORMAL PLOT; VAR RAADT;
title2 'Normality check on residuals for ptcdemo1'; RUN;
/*Try the final set of variables with only ptcdemo2 data*/
data ptcdemo2; set ptcdemo; baadt_orig=baadt; if myran<0 then baadt=.; run;
/*Fit final model using aadt^.2*/proc reg data=ptcdemo2 outest=est;
model baadt=factor1 factor2 factor3 factor5 factor7 rte_class1 co_urb4--
co_urb6 lanes_rte1--lanes_rte4
lanes_rte11 lanes_rte17 speed_rte6 speed_rte19 speed_rte20
lanes_speed_rte5 lanes_speed_rte9 lanes_speed_rte17 lanes_speed_rte20
lanes_speed_rte21 lanes_speed_rte23
lanes_speed_rte33 lanes_speed_rte36 lanes_speed_rte41 lanes_speed_rte47
lanes_speed_rte49 lanes_speed_rte50
lanes_speed_rte55--lanes_speed_rte57 cycle_surface1--cycle_surface7
a_control2 year2 year3
month1 month2 month5 month6 month11 day4
lucode1 lucode2 lucode4 lucode5 lucode6 lucode10 lucode11 city_urb1 city_urb3
city_urb4 city_urb5;
output out=out1 p=paadt r=raadt;
title1 "Fit final model to ptcdemo2"; run; quit;
data out1; set out1; raadt=baadt_orig-paadt; if myran ge 0 then delete; run;
/* NORMALITY check*/ PROC UNIVARIATE DATA=OUT1 NORMAL PLOT; VAR RAADT;
title2 'Normality check on residuals for ptcdemo2'; RUN;

***************************** END: Model selection
*************************************;
ods html close;
```

APPENDIX B: Old Data Dictionary for Station Data (fits file name PTC_FINAL1.DBF)

UNIQ_ID	Seven-digit unique identifier for PTC stations. Combination of the County SIPS code (SIPS) and the Station ID (STATION).
UNIQ_ID8	Eight-digit unique identifier for PTC stations. Combination of the County SIPS code (SIPS) and the Station ID (STATION). The eighth digit was placed after the SIPS (third digit) for future reference to the station. This identifier was not used during the process outlined in this report.
SIPS	State county and urban code. Counties are numbered from 0 to 99. SIPS values greater than 99 belong to an urban area and are assigned with the urban area code.
STATION	PTC station identifier. The identifiers are unique within each SIPS
CO_NUM	SIPS code for the county in which the station is located. The CO_NUM differs from the SIPS for those stations that are within urban areas.
CO_NAME	County name in which the PTC station was originally digitized (county from the NCDOT county maps).
CYCLE_YR	Classification for rotation of the traffic count data acquisition and analysis.
RD_SURFACE	Classification of road surface type.
ROUTECODE	Single-digit classification of route type assigned to the PTC stations. This classification reflects how the traffic data is acquired and analyzed.
ROUTENUM	Route number assigned to the segment on which the PTC station is located (originally in PTC.SHP, not MERGEROADS).
DESCRIPTIO	Description of the route on which the PTC station is located.
ATRGROUP	ATR group assignment for the PTC station
YEAR	
MONTH	
DAY	
AADT	
COUNT1	
COUNT2	
COUNT3	
YEAR_2	
MONTH_2	
DAY_2	
AADT_2	
COUNT1_2	
COUNT2_2	
COUNT3_2	
YEAR_NEW	
AADT_NEW	
MONTH_01	
DAY_01	
COUNT1_01	
COUNT2_01	
COMMENTS	
CASE	
AXLEFACTOR	Multiplier for accounting the variations in the number of axles per vehicle.
DISTANCE	Distance in meters the PTC station moved during the snapping process. Numbers are no greater than the snapping tolerance of 200 meters.

X_COORD	The X-coordinate location of the PTC station in NAD83 North Carolina State Plane coordinate system.
Y_COORD	The Y-coordinate location of the PTC station in NAD83 North Carolina State Plane coordinate system.
STATE90	State identifier for the US Census block groups. Used in conjunction with COUNTY90, TRACT90, and GROUP90 to join the entire block group dataset to the PTC database.
COUNTY90	County identifier for the US Census block groups. Used in conjunction with STATE90, TRACT90, and GROUP90 to join the block group dataset to the PTC database.
TRACT90	Census tract identifier for the US Census block groups. Used in conjunction with STATE90, COUNTY90, and GROUP90 to join the block group dataset to the PTC database.
GROUP90	Census block group identifier for the US Census block groups. Used in conjunction with STATE90, COUNTY90, and TRACT90 to join the block group dataset to the PTC database.
LUCODE	Land use classification (Anderson level 2)
LEVEL2	Land use description (Anderson level 2)
SPEED	Assigned speed limit to the route segment on which the PTC station is located.
LANES	Number of lanes represented by the route segment on which the PTC station is located. NCDOT GIS is in the process of incorporating a new attribute to indicate when this number references a divided highway drawn as multiple lines on the GIS coverage. At the time of this report, the LANES values are not consistent with the digitized segments and may or may not be representative of both lanes on a divided or undivided highway.
ROUTE1	9-digit Route classification and description of the segment on which the PTC station is located. The number is a code and contains several items of information including the route classification (interstate, US, NC, etc.) and route number. ROUTE1 was used to derive RTE_TYPE. ROUTE1 is not complete for the PTC dataset.
RTE_TYPE	Single-digit route classification for the route segment on which the PTC station is located. Derived from the values of ROUTE1 where ROUTE1 > 0. For cases when ROUTE1 = 0, the LGDS_LEVEL values were translated.
LRS_MATCH	Binary flag to indicate if the PTC station data is located on a segment that contains LRS attributes, and thus contain those values for LANES and SPEED. When LRS_MATCH = 1, the data is derived from the LRS system.
URBANFLG	Binary flag to indicate if the PTC station is located within an urban area as defined by the polygon features of MB.SHP (municipal boundary shapefile).

APPENDIX C: Two-way Tables Showing Current Relationships Between SIPS, CO_NUM with URB_NUM, and COUNTYID

Frequency Tables for "Surrogate" Attributes
The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of co_urb by SIPS

co_urb	SIPS(SIPS) 18	31	50	67	91	101	103	108	Total
180000	305 8.89 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	305 8.89
180103	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	15 0.44 100.00 1.77	0 0.00 0.00 0.00	15 0.44
310000	0 0.00 0.00 0.00	26 0.76 100.00 66.67	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	26 0.76
310103	0 0.00 0.00 0.00	13 0.38 1.90 33.33	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	670 19.53 98.10 78.92	0 0.00 0.00 0.00	683 19.91
500000	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25
670000	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71
670101	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	21 0.61 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	21 0.61
670103	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	164 4.78 100.00 19.32	0 0.00 0.00 0.00	164 4.78
910108	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	601 17.52 39.23 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	931 27.13 60.77 100.00	1532 44.65
Total	305 8.89	39 1.14	489 14.25	196 5.71	601 17.52	21 0.61	849 24.74	931 27.13	3431 100.00

Frequency
Percent
Row Pct
Col Pct

Table of COUNTYID by SIPS

COUNTYID(COUNTYID)	SIPS(SIPS) 18	31	50	67	91	101	103	108	Total
18	305 8.89 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	305 8.89
31	0 0.00 0.00 0.00	39 1.14 97.50 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.03 2.50 0.12	0 0.00 0.00 0.00	40 1.17
50	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25
67	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71
91	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	600 17.49 99.67 99.83	0 0.00 0.00 0.00	0 0.00 0.00 0.00	2 0.06 0.33 0.21	602 17.55
101	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	21 0.61 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	21 0.61
103	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	848 24.72 100.00 99.88	0 0.00 0.00 0.00	848 24.72
108	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.03 0.11 0.17	0 0.00 0.00 0.00	0 0.00 0.00 0.00	929 27.08 99.89 99.79	930 27.11
Total	305 8.89	39 1.14	489 14.25	196 5.71	601 17.52	21 0.61	849 24.74	931 27.13	3431 100.00

Frequency
Percent
Row Pct
Col Pct

Table of sips_new by SIPS

sips_new	SIPS(SIPS) 18	31	50	67	91	101	103	108	Total
18	305 8.89 100.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	305 8.89

	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
31	0 0.00 0.00 0.00	39 1.14 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	39 1.14
50	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25
67	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71
91	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	600 17.49 99.67 99.83	0 0.00 0.00 0.00	0 0.00 0.00 0.00	2 0.06 0.33 0.21	602 17.55
101	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	21 0.61 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	21 0.61
103	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	849 24.74 100.00 100.00	0 0.00 0.00 0.00	849 24.74
108	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.03 0.11 0.17	0 0.00 0.00 0.00	0 0.00 0.00 0.00	929 27.08 99.89 99.79	930 27.11
Total	305 8.89	39 1.14	489 14.25	196 5.71	601 17.52	21 0.61	849 24.74	931 27.13	3431 100.00

Frequency
Percent
Row Pct
Col Pct

Table of COUNTYID by co_urb

COUNTYID(COUNTYID)	co_urb									Total
	180000	180103	310000	310103	500000	670000	670101	670103	910108	
18	305 8.89 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	305 8.89
31	0 0.00 0.00 0.00	0 0.00 0.00 0.00	26 0.76 65.00 100.00	14 0.41 35.00 2.05	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	40 1.17
50	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25

67	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71
91	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	602 17.55 100.00 39.30	602 17.55
101	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	21 0.61 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	21 0.61
103	0 0.00 0.00 0.00	15 0.44 1.77 100.00	0 0.00 0.00 0.00	669 19.50 78.89 97.95	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	164 4.78 19.34 100.00	0 0.00 0.00 0.00	848 24.72
108	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	930 27.11 100.00 60.70	930 27.11
Total	305 8.89	15 0.44	26 0.76	683 19.91	489 14.25	196 5.71	21 0.61	164 4.78	1532 44.65	3431 100.00

Frequency
Percent
Row Pct
Col Pct

Table of sips_new by co_urb

sips_new	co_urb									Total
	180000	**180103**	**310000**	**310103**	**500000**	**670000**	**670101**	**670103**	**910108**	
18	305 8.89 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	305 8.89
31	0 0.00 0.00 0.00	0 0.00 0.00 0.00	26 0.76 66.67 100.00	13 0.38 33.33 1.90	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	39 1.14
50	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25
67	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71
91	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	602 17.55 100.00 39.30	602 17.55
101	0	0	0	0	0	0	21	0	0	21

	0.00 0.00 0.00	0.00 0.00 0.00	0.00 0.00 0.00	0.00 0.00 0.00	0.00 0.00 0.00	0.00 0.00 0.00	0.61 100.00 100.00	0.00 0.00 0.00	0.00 0.00 0.00	0.61
103	0 0.00 0.00 0.00	15 0.44 1.77 100.00	0 0.00 0.00 0.00	670 19.53 78.92 98.10	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	164 4.78 19.32 100.00	0 0.00 0.00 0.00	849 24.74
108	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	930 27.11 100.00 60.70	930 27.11
Total	305 8.89	15 0.44	26 0.76	683 19.91	489 14.25	196 5.71	21 0.61	164 4.78	1532 44.65	3431 100.00

Frequency
Percent
Row Pct
Col Pct

Table of sips_new by COUNTYID

sips_new	COUNTYID(COUNTYID) 18	31	50	67	91	101	103	108	Total
18	305 8.89 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	305 8.89
31	0 0.00 0.00 0.00	39 1.14 100.00 97.50	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	39 1.14
50	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	489 14.25
67	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	196 5.71
91	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	602 17.55 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	602 17.55
101	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	21 0.61 100.00 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	21 0.61
103	0 0.00 0.00 0.00	1 0.03 0.12 2.50	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	848 24.72 99.88 100.00	0 0.00 0.00 0.00	849 24.74
108	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	930 27.11	930 27.11

	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	100.00 100.00	
Total	305 8.89	40 1.17	489 14.25	196 5.71	602 17.55	21 0.61	848 24.72	930 27.11	3431 100.00

APPENDIX D: Data Dictionary for 1990 Census

Basic Demographics

This data table is an extract from the complete Summary Tape Files 1A and 3A, reflecting the most commonly used demographic characteristics. It is theoretically available for all summary levels except Census Blocks.

Except where otherwise noted, all fields represent absolute counts of persons or housing units as they fall within particular categories. For example, persons can be counted according to ethnicity, and housing units can be grouped according to the number of rooms in the unit. To determine the percent of persons or housing units that fall within a category relative to a larger group, simply divide by the appropriate universe field and multiply by 100. (i.e., Percent Hispanic = Hispanics / Total Population * 100).

Field names are limited to 8 characters to ensure compatibility with all applications.

Population Characteristics

Total Population	
Persons	Total Population
Families	Total Families
Houshold	Total Households
Samp_pop	Sample Population
Per_samp	Percent of Persons in Sample
Sex	
Male	Male
Female	Females
Ethnicity	
White	Whites
Black	Blacks
Amind	American Indians
Asian	Asian and Pacific
O_ethnic	Other Ethnicity
Hispanic	Hispanics of All Races
Age	
Agelt_05	Age Less Than 5
Age05_09	Age 5 to 9
Age10_14	Age 10 to 14
Age15_17	Age 15 to 17
Age18_19	Age 18 to 19
Age15_19	Age 15 to 19
Ageis_20	Age 20 Years
Age21_24	Age 21 to 24

Age20_24	Age 20 to 24
Age25_34	Age 25 to 34
Age35_44	Age 35 to 44
Age45_54	Age 45 to 54
Age55_64	Age 55 to 64
Age65_74	Age 65 to 74
Age75_84	Age 75 to 84
Agegt_84	Age Over 84
Agege_03	Age 3 or Older
Agege_05	Age 5 or Older
Agege_16	Age 16 or Older
Agege_25	Age 25 or Older
Marital Status	
Single	Never Married
Married	Married
Separate	Married and Separated
Widowed	Widowed
Divorced	Divorced
Family Type	
Oneperhh	One Person Households
Marrwich	Married with Child Families
Marrnoch	Married no Child Families
Singwich	Single with Child Families
Singnoch	Single no Child Families
Persons in Families	
Perinfam	In Families
Persons in Institutions	
Corrinst	In Correctional Facilities
Nurshome	In Nursing Homes
Mental	In Mental Institutions
Juvenile	In Juvenile Detention
O_instit	In Other Institutions
O_noinst	In Other Non-Institutional Group Quarters
Dormitor	In Dormitories

Miliquar	In Military Quarters
Shelters	In Homeless Shelters
Instreet	On the Street
Ability to Speak English	
Onlyengl	Speak English Only
Ableengl	Able to Speak English
Cantengl	Cannot Speak English
Citizenship	
Native	Native Citizens
Natural	Naturalized Citizens
Noncitiz	Non Citizens
Region of Birth	
Born_ins	Born in Same State
Bornnort	Born in Northeast USA
Bornmidw	Born in Midwest USA
Bornsout	Born in South USA
Bornwest	Born in West USA
Born_for	Born Outside USA
Migratory Status	
Samehous	At Same House 1985
Samecnty	In Same County 1985
Samestat	In Same State 1985
Oth_stat	In Other State 1985
Frm_abrd	In Other Country 1985
Workers	
Workers	Total Workers Over 16
Place of Work	
Wrkincty	Worked in County of Residence
Wrkexcty	Worked outside County of Residence
Type of Commute	
Drvalone	Drove Alone to Work
Carpool	Carpooled to Work
Pubtrans	Public Transit to Work
Wrk_home	Worked at Home
Duration of Commute	

Com_lt15	Commute Under 15 Min
Com_1529	Commute 15 to 29 Min
Com_3044	Commute 30 to 44 Min
Com_gt44	Commute Over 44 Min
Student Population	
Preprima	Preprimary Students
Elemscnd	Elem and High Students
College	College Students
Educational Attainment	
Nodiplom	No High School Diploma
Highschl	High School Graduates
Somecolg	Attended College
Colggrad	College Graduates
Gradschl	Attended Graduate School
Employment Status	
Armdforc	In Armed Forces
Employed	Total Employed
Unemploy	Total Unemployed
Notinwrk	Total Not in Workforce
Industry of Employment	
Primary	In Primary Industries
Manufact	In Manufacturing
Utility	In Transport and Utilities
Trade	In Trade
Fire	In Finance and Insurance
Services	In Commercial Services
Profserv	In Professional Services
Public	In Public Administration
Occupation	
Execprof	As Managers and Professionals
Techsale	As Technical-Sales-Clerical
Service	As Service Employee
Manual	As Operator-Assembler-Laborer
Employment Sector	
Privwork	In Private Sector

Publwork	In Public Sector
Selfwork	Self Employed
Famiwork	Unpaid Family Work
Household Income	
Inc_lt15	Income Under $15,000 in Household
Inc_1525	Income $15-25,000 in Household
Inc_2535	Income $25-35,000 in Household
Inc_3550	Income $35-50,000 in Household
Inc_5075	Income $50-75,000 in Household
Inc_gt75	Income Over $75,000 in Household
Inc_medn	Median Household Income
Inc_aggr	Aggregate Household Income
Households by Type of Income	
Wage_sal	Wages or Salary in Household
Self_inc	Self Employed Income in Household
Farm_inc	Farm Income in Household
Inte_inc	Interest Income in Household
Socs_inc	Social Security in Household
Publ_inc	Public Assistance in Household
Reti_inc	Retirement Income in Household
Per Capita Income	
Incprcap	Per Capita Income
Poverty	
Childpov	Children in Poverty
Inpovrty	Persons in Poverty

Housing Characteristics

Housing Units	
Housing	Housing Units
Occupancy	
Occupied	Occupied Units
Vacant	Vacant Units
Tenure	
Ownr_occ	Owner Occupied Units
Rent_occ	Renter Occupied Units
Occasional Units	
Seasonal	Seasonal Use Units

F_migrnt	Migratory Worker Units
Rooms in Unit	
Room_1_3	Units with 1 to 3 Rooms
Room_4_6	Units with 4 to 6 Rooms
Room_gt6	Units with Over 6 Rooms
Persons in Unit	
Prunit12	Units with 1 to 2 Persons
Prunit34	Units with 3 to 4 Persons
Prunitg4	Units with Over 4 Persons
Housing Value	
Val_lt25	Units Valued Under $25,000
Val_2550	Units Valued $25-50,000
Val_501c	Units Valued $50-100,000
Val_1c2c	Units Valued $100-200,000
Val_2c3c	Units Valued $200-300,000
Val_gt3c	Units Valued Over $300,000
Val_medi	Median Value of Units
Agg_valu	Aggregate Value of Units
Housing Contract Rent	
Rnt_lt25	Unit Rent Under $250
Rnt_2550	Unit Rent $250-500
Rnt_5075	Unit Rent $500-750
Rnt_751k	Unit Rent $750-1000
Rnt_gt1k	Unit Rent Over $1000
Rnt_medi	Median Rent
Units in Structure	
Detached	Single Detached Units
Attached	Single Attached Units
Duplex	Duplex Units
Apartmnt	Apartment Units
Source of Water	
Pubwater	Water from Public Source
Prvwater	Water from Private Source
Type of Sewer	
Pubsewer	Sewer with Public Utility

Prvsewer	Sewer by Private Means
Age of Housing Units	
Bltbfr70	Built before 1970
Blt_7079	Built 1970-1979
Blt_8084	Built 1980-1984
Bltaft84	Built after 1984
Medyrblt	Median Year Built
Source of Heat	
Publ_gas	Heat from Gas Utility
Botl_gas	Heat from Bottled Gas
Electric	Heat from Electricity
Fuelkero	Heat from Oil or Kerosene
Coalwood	Heat from Coal or Wood
Oth_heat	Heat from Other Source
Vehicles in Household	
Novehicl	No Vehicles in Household
Vehicl_1	One Vehicle in Household
Vehicl_2	Two Vehicles in Household
Vehiclg2	Over Two Vehicles in Household
Aggvehcl	Aggregate Vehicles

APPENDIX E: Principal Components Analysis of Census Attributes

Principal Components Analysis of PTCDEMO
The FACTOR Procedure
Initial Factor Method: Principal Components

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	70.4444708	57.2656486	0.4825	0.4825
2	13.1788222	2.7647778	0.0903	0.5728
3	10.4140444	3.0439548	0.0713	0.6441
4	7.3700895	2.8429347	0.0505	0.6946
5	4.5271548	0.8454394	0.0310	0.7256
6	3.6817154	1.0694537	0.0252	0.7508
7	2.6122616	0.5410474	0.0179	0.7687
8	2.0712142	0.2351467	0.0142	0.7829
9	1.8360675	0.0938535	0.0126	0.7955
10	1.7422139	0.1853511	0.0119	0.8074
11	1.5568628	0.3100533	0.0107	0.8180
12	1.2468095	0.0262880	0.0085	0.8266
13	1.2205215	0.0110814	0.0084	0.8349
14	1.2094401	0.0299697	0.0083	0.8432
15	1.1794704	0.0636414	0.0081	0.8513
16	1.1158290	0.0526132	0.0076	0.8590
17	1.0632158	0.0806946	0.0073	0.8662
18	0.9825212	0.0212228	0.0067	0.8730
19	0.9612984	0.0302229	0.0066	0.8795
20	0.9310755	0.0711996	0.0064	0.8859
21	0.8598760	0.0435559	0.0059	0.8918

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
22	0.8163201	0.0627173	0.0056	0.8974
23	0.7536028	0.0379089	0.0052	0.9026
24	0.7156939	0.0433148	0.0049	0.9075
25	0.6723791	0.0187842	0.0046	0.9121
26	0.6535948	0.0688606	0.0045	0.9166
27	0.5847343	0.0201119	0.0040	0.9206
28	0.5646223	0.0391121	0.0039	0.9244
29	0.5255103	0.0257020	0.0036	0.9280
30	0.4998082	0.0430893	0.0034	0.9314
31	0.4567190	0.0335105	0.0031	0.9346
32	0.4232085	0.0200687	0.0029	0.9375
33	0.4031398	0.0117710	0.0028	0.9402
34	0.3913688	0.0132030	0.0027	0.9429
35	0.3781658	0.0201546	0.0026	0.9455
36	0.3580112	0.0101033	0.0025	0.9480
37	0.3479079	0.0166154	0.0024	0.9503
38	0.3312925	0.0084015	0.0023	0.9526
39	0.3228910	0.0202100	0.0022	0.9548
40	0.3026810	0.0087918	0.0021	0.9569
41	0.2938892	0.0157040	0.0020	0.9589
42	0.2781852	0.0161950	0.0019	0.9608
43	0.2619902	0.0056899	0.0018	0.9626
44	0.2563002	0.0170388	0.0018	0.9644
45	0.2392614	0.0083080	0.0016	0.9660
46	0.2309534	0.0107143	0.0016	0.9676

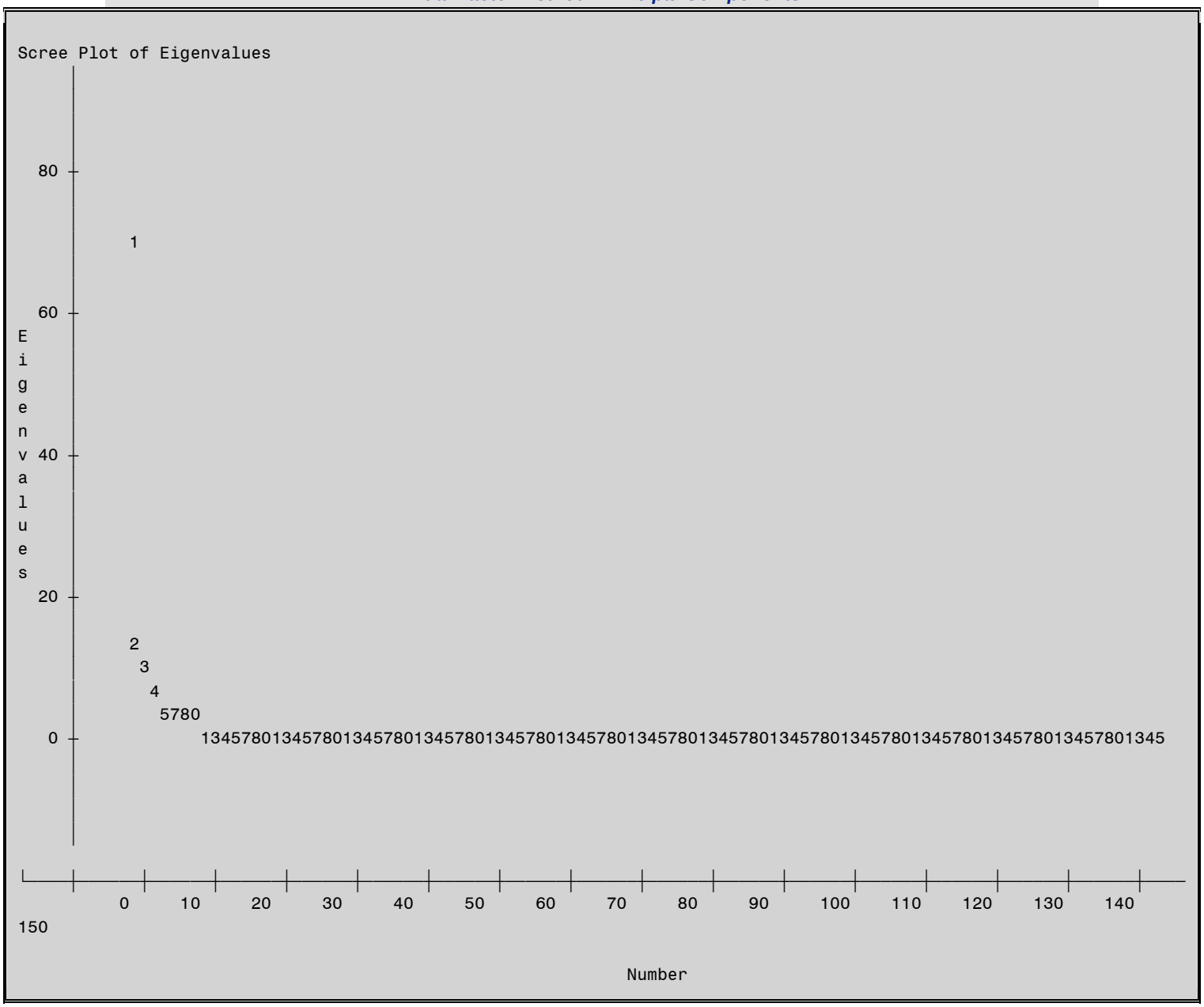
Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
47	0.2202391	0.0110598	0.0015	0.9691
48	0.2091793	0.0056689	0.0014	0.9705
49	0.2035104	0.0190991	0.0014	0.9719
50	0.1844113	0.0031423	0.0013	0.9732
51	0.1812690	0.0065698	0.0012	0.9744
52	0.1746992	0.0045973	0.0012	0.9756
53	0.1701019	0.0108224	0.0012	0.9768
54	0.1592795	0.0029457	0.0011	0.9779
55	0.1563338	0.0133457	0.0011	0.9789
56	0.1429881	0.0069767	0.0010	0.9799
57	0.1360114	0.0055883	0.0009	0.9809
58	0.1304231	0.0034204	0.0009	0.9818
59	0.1270026	0.0029173	0.0009	0.9826
60	0.1240853	0.0034021	0.0008	0.9835
61	0.1206832	0.0051902	0.0008	0.9843
62	0.1154930	0.0089317	0.0008	0.9851
63	0.1065614	0.0028985	0.0007	0.9858
64	0.1036629	0.0039072	0.0007	0.9865
65	0.0997557	0.0044557	0.0007	0.9872
66	0.0953000	0.0026896	0.0007	0.9879
67	0.0926103	0.0027507	0.0006	0.9885
68	0.0898596	0.0037428	0.0006	0.9891
69	0.0861169	0.0032213	0.0006	0.9897
70	0.0828956	0.0061362	0.0006	0.9903
71	0.0767594	0.0017196	0.0005	0.9908

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
72	0.0750399	0.0028971	0.0005	0.9913
73	0.0721428	0.0061335	0.0005	0.9918
74	0.0660093	0.0021251	0.0005	0.9923
75	0.0638842	0.0036074	0.0004	0.9927
76	0.0602768	0.0020352	0.0004	0.9931
77	0.0582416	0.0016122	0.0004	0.9935
78	0.0566294	0.0019917	0.0004	0.9939
79	0.0546376	0.0037738	0.0004	0.9943
80	0.0508638	0.0029792	0.0003	0.9946
81	0.0478846	0.0014290	0.0003	0.9949
82	0.0464556	0.0045025	0.0003	0.9953
83	0.0419531	0.0014718	0.0003	0.9956
84	0.0404814	0.0031190	0.0003	0.9958
85	0.0373624	0.0011454	0.0003	0.9961
86	0.0362169	0.0009791	0.0002	0.9963
87	0.0352378	0.0013374	0.0002	0.9966
88	0.0339004	0.0020771	0.0002	0.9968
89	0.0318233	0.0016910	0.0002	0.9970
90	0.0301323	0.0004529	0.0002	0.9972
91	0.0296794	0.0040806	0.0002	0.9974
92	0.0255989	0.0004092	0.0002	0.9976
93	0.0251897	0.0016969	0.0002	0.9978
94	0.0234928	0.0010556	0.0002	0.9979
95	0.0224372	0.0008948	0.0002	0.9981
96	0.0215424	0.0015646	0.0001	0.9982

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
97	0.0199777	0.0012443	0.0001	0.9984
98	0.0187334	0.0009921	0.0001	0.9985
99	0.0177413	0.0003931	0.0001	0.9986
100	0.0173482	0.0013355	0.0001	0.9987
101	0.0160128	0.0009915	0.0001	0.9989
102	0.0150213	0.0006339	0.0001	0.9990
103	0.0143873	0.0017933	0.0001	0.9991
104	0.0125941	0.0009528	0.0001	0.9991
105	0.0116412	0.0004736	0.0001	0.9992
106	0.0111676	0.0013576	0.0001	0.9993
107	0.0098100	0.0007693	0.0001	0.9994
108	0.0090407	0.0003497	0.0001	0.9994
109	0.0086910	0.0008132	0.0001	0.9995
110	0.0078778	0.0005013	0.0001	0.9995
111	0.0073764	0.0008573	0.0001	0.9996
112	0.0065191	0.0009490	0.0000	0.9996
113	0.0055702	0.0004074	0.0000	0.9997
114	0.0051628	0.0003077	0.0000	0.9997
115	0.0048551	0.0004289	0.0000	0.9997
116	0.0044262	0.0001132	0.0000	0.9998
117	0.0043131	0.0010973	0.0000	0.9998
118	0.0032158	0.0001237	0.0000	0.9998
119	0.0030920	0.0003898	0.0000	0.9999
120	0.0027023	0.0003824	0.0000	0.9999
121	0.0023199	0.0000923	0.0000	0.9999

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
122	0.0022276	0.0001038	0.0000	0.9999
123	0.0021237	0.0003455	0.0000	0.9999
124	0.0017782	0.0001008	0.0000	0.9999
125	0.0016774	0.0002196	0.0000	0.9999
126	0.0014578	0.0001746	0.0000	0.9999
127	0.0012832	0.0002427	0.0000	1.0000
128	0.0010405	0.0000696	0.0000	1.0000
129	0.0009710	0.0001770	0.0000	1.0000
130	0.0007940	0.0000478	0.0000	1.0000
131	0.0007462	0.0000230	0.0000	1.0000
132	0.0007232	0.0002053	0.0000	1.0000
133	0.0005180	0.0001044	0.0000	1.0000
134	0.0004136	0.0000633	0.0000	1.0000
135	0.0003502	0.0001723	0.0000	1.0000
136	0.0001780	0.0000616	0.0000	1.0000
137	0.0001164	0.0000053	0.0000	1.0000
138	0.0001111	0.0000045	0.0000	1.0000
139	0.0001066	0.0000273	0.0000	1.0000
140	0.0000793	0.0000115	0.0000	1.0000
141	0.0000678	0.0000383	0.0000	1.0000
142	0.0000294	0.0000027	0.0000	1.0000
143	0.0000267	0.0000137	0.0000	1.0000
144	0.0000130	0.0000021	0.0000	1.0000
145	0.0000109	0.0000030	0.0000	1.0000
146	0.0000079		0.0000	1.0000

Principal Components Analysis of PTCDEMO
The FACTOR Procedure
Initial Factor Method: Principal Components



Principal Components Analysis of PTCDEMO

The FACTOR Procedure
Rotation Method: Varimax

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
MARRWICH	MARRWICH	0.96606	-0.03717	0.00509	-0.00362	0.10218	0.07191	0.15176
PERINFAM	PERINFAM	0.96261	0.06034	0.01843	0.11830	0.06060	0.18980	0.10076
PRUNIT34	PRUNIT34	0.95900	0.10046	0.04831	0.11405	0.05720	0.13253	0.09489
MARRIED	MARRIED	0.95173	0.09014	0.04510	-0.02176	0.08813	0.20701	0.14837
AGE35_44	AGE35_44	0.94820	0.13507	0.02286	0.04960	0.00382	0.10173	0.20921
AGE05_09	AGE05_09	0.94614	-0.01348	0.00173	0.19134	0.03140	0.08791	0.13132
AGE10_14	AGE10_14	0.93434	-0.06612	-0.03118	0.20229	0.06039	0.08956	0.15070
AGELT_05	AGELT_05	0.92847	0.08217	0.04526	0.18654	-0.01467	0.09014	0.01812
ELEMSCND	ELEMSCND	0.92822	-0.06408	-0.01099	0.21143	0.06931	0.07616	0.12461
OWNR_OCC	OWNR_OCC	0.92273	-0.03475	-0.03263	-0.03849	0.14770	0.27456	0.12146
DRVALONE	DRVALONE	0.91706	0.34612	0.09285	-0.00739	-0.03648	0.11223	0.05336
AGE15_17	AGE15_17	0.91284	-0.08866	-0.01499	0.19825	0.06623	0.10081	0.16185
SOMECOLG	SOMECOLG	0.90954	0.28355	0.01445	-0.00550	-0.08392	0.12113	-0.03173
PRIVWORK	PRIVWORK	0.90415	0.31922	0.19363	0.03814	-0.03493	0.09886	0.00880
AGE45_54	AGE45_54	0.90018	0.02446	-0.00775	0.03209	0.05803	0.21362	0.23741
VEHICL_2	VEHICL_2	0.89163	0.36350	0.07258	-0.01176	-0.02623	0.15200	0.11506
COM_1529	COM_1529	0.89155	0.33066	0.16777	0.00062	-0.07945	0.07870	0.04163
EMPLOYED	EMPLOYED	0.88484	0.32751	0.29004	0.05481	-0.01908	0.12230	0.03530
MARRNOCH	MARRNOCH	0.88063	0.20111	0.06383	-0.06018	0.07451	0.33154	0.14612
INC_5075	INC_5075	0.88011	0.20355	0.01573	-0.13436	-0.13077	0.08556	0.22123
DETACHED	DETACHED	0.87431	-0.08903	-0.02650	0.01480	0.08976	0.35427	0.17957
MALE	MALE	0.86791	0.22261	0.38271	0.11086	0.00958	0.14640	0.07745
UTILITY	UTILITY	0.85998	0.22604	0.03017	-0.01913	-0.03878	-0.05558	-0.05905
COM_3044	COM_3044	0.85895	0.00524	-0.01642	0.07445	0.24519	-0.02203	-0.05865

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
NATIVE	NATIVE	0.85705	0.18959	0.41395	0.13029	0.00497	0.18042	0.05454
WAGE_SAL	WAGE_SAL	0.85426	0.46230	0.11308	0.08813	-0.04796	0.14150	0.06724
INC_3550	INC_3550	0.85293	0.36581	0.03877	-0.02570	0.00459	0.18811	-0.13432
TRADE	TRADE	0.85203	0.33039	0.25989	0.06431	-0.04264	0.08947	-0.03369
SAMECNTY	SAMECNTY	0.85040	0.30094	0.04883	0.15559	-0.10257	0.17244	-0.08584
WRKINCTY	WRKINCTY	0.84611	0.31085	0.30502	0.05609	-0.14571	0.11558	-0.04314
ONLYENGL	ONLYENGL	0.84572	0.19472	0.43411	0.11962	0.00491	0.18282	0.05952
FIRE	FIRE	0.84412	0.28276	0.06089	-0.06077	-0.17942	0.03012	0.05029
TECHSALE	TECHSALE	0.84046	0.30500	0.38141	-0.00966	-0.10078	0.07385	-0.02690
AGEGE_03	AGEGE_03	0.83822	0.21462	0.44301	0.11999	-0.01337	0.17367	0.08194
WHITE	WHITE	0.82996	0.20648	0.38321	-0.10821	0.06716	0.16359	0.17454
AGE25_34	AGE25_34	0.81592	0.49488	0.15511	0.08999	-0.06522	0.04867	-0.06572
MANUFACT	MANUFACT	0.81275	0.14010	0.00351	0.09343	0.24263	0.07884	-0.02780
SAMEHOUS	SAMEHOUS	0.80691	-0.07828	0.01342	0.18396	0.21206	0.38283	0.03103
WORKERS	WORKERS	0.80330	0.28669	0.46129	0.08299	0.00180	0.16093	0.06266
EXECPROF	EXECPROF	0.80124	0.36507	0.20836	-0.07577	-0.16856	0.06940	0.32112
SERVICES	SERVICES	0.79338	0.32319	0.32364	0.05872	-0.12585	0.05403	-0.04805
SELF_INC	SELF_INC	0.79220	0.24990	0.09423	-0.00379	0.02834	0.27315	0.25882
BORNSOUT	BORNSOUT	0.79067	0.24203	0.44536	-0.02102	-0.16379	0.04678	0.16820
PREPRIMA	PREPRIMA	0.79023	0.07237	0.07604	-0.00375	-0.10809	-0.02217	0.37288
COLGGRAD	COLGGRAD	0.78734	0.38034	0.13270	-0.06651	-0.20772	0.05331	0.34639
AGEGE_16	AGEGE_16	0.78176	0.25728	0.51865	0.09944	-0.02327	0.18398	0.06833
BORN_INS	BORN_INS	0.78033	0.04846	0.31159	0.26751	0.18512	0.30876	-0.20623
BLTAFT84	BLTAFT84	0.77117	0.35067	0.02146	-0.08247	0.01624	-0.11681	0.13710
VAL_1C2C	VAL_1C2C	0.76894	0.03657	0.01356	-0.15126	-0.14898	0.03201	0.46489

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
CARPOOL	CARPOOL	0.76846	0.24137	0.11123	0.25988	0.20435	0.15936	-0.16971
INTE_INC	INTE_INC	0.75868	0.43569	0.12953	-0.08558	-0.12322	0.19708	0.35412
ELECTRIC	ELECTRIC	0.75717	0.55959	0.11652	0.08235	-0.05461	0.00070	0.03779
BLT_8084	BLT_8084	0.74927	0.46255	0.03426	0.02373	-0.01094	-0.12407	0.09719
AGE55_64	AGE55_64	0.74224	0.08985	-0.00762	0.08072	0.07365	0.53161	0.10230
ROOM_4_6	ROOM_4_6	0.74206	0.51736	0.09307	0.18578	0.07162	0.26075	-0.16269
HIGHSCHL	HIGHSCHL	0.73169	-0.02474	-0.08790	0.14015	0.28063	0.35567	-0.33239
SERVICE	SERVICE	0.72781	0.20725	0.23451	0.28184	0.25484	0.20471	-0.26508
VAL_501C	VAL_501C	0.72356	0.05165	-0.02383	-0.00574	-0.11991	0.36282	-0.34710
BORNMIDW	BORNMIDW	0.72207	0.30128	0.36138	-0.07486	-0.16135	-0.03311	0.35943
SELFWORK	SELFWORK	0.71987	0.08551	0.04533	0.02647	0.29627	0.24883	0.14502
BLT_7079	BLT_7079	0.68887	0.42480	0.08696	0.13127	-0.00758	0.09584	0.05684
BORNNORT	BORNNORT	0.68613	0.29041	0.39949	-0.01515	-0.19509	0.01086	0.39468
INC_2535	INC_2535	0.68276	0.54986	0.09338	0.10677	0.01045	0.23247	-0.18069
PRUNIT12	PRUNIT12	0.67815	0.64770	0.13461	0.10793	-0.08639	0.24902	0.08021
PRIMARY_	PRIMARY_	0.64060	0.06014	-0.00765	0.10516	0.48131	0.16182	-0.22293
SEPARATE	SEPARATE	0.63164	0.39021	0.05218	0.45952	-0.11921	0.19434	-0.19483
OTH_STAT	OTH_STAT	0.62930	0.29830	0.48484	-0.02226	-0.16747	-0.08703	0.28637
BORNWEST	BORNWEST	0.62798	0.31758	0.45055	-0.04149	-0.12756	0.00217	0.29636
PUBWATER	PUBWATER	0.62680	0.59503	0.14638	0.16825	-0.32871	0.16452	0.12999
PUBLWORK	PUBLWORK	0.62034	0.31983	0.55233	0.09669	-0.04846	0.12384	0.07466
PROFSERV	PROFSERV	0.60964	0.36186	0.50163	0.04174	-0.15072	0.19799	0.22697
NATURAL_	NATURAL_	0.60637	0.35683	0.32207	-0.03438	-0.16073	0.02971	0.35333
PUBL_GAS	PUBL_GAS	0.57270	0.30791	0.09985	0.07537	-0.44526	0.28799	0.24434
WRKEXCTY	WRKEXCTY	0.56803	0.22650	0.09263	0.01826	0.35107	0.08319	0.24211

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
COM_LT15	COM_LT15	0.56669	0.43432	0.54831	0.06561	-0.18304	0.21437	0.07446
HISPANIC	HISPANIC	0.53563	0.39966	0.39411	0.11196	-0.01399	-0.01408	0.07924
WRK_HOME	WRK_HOME	0.51250	0.04847	0.40100	-0.04684	0.08716	0.21007	0.28146
UNEMPLOY	UNEMPLOY	0.46786	0.22909	0.44377	0.40522	-0.11306	0.15912	-0.02750
ARMDFORC	ARMDFORC	0.39005	0.28993	0.04050	-0.01742	-0.00789	0.14651	-0.09097
APARTMNT	APARTMNT	0.36144	0.81948	0.19388	0.16559	-0.24798	-0.02755	0.04588
ROOM_1_3	ROOM_1_3	0.34861	0.78804	0.19059	0.22832	-0.17311	0.00040	0.09801
RNT_2550	RNT_2550	0.39133	0.77323	0.21281	0.18340	-0.25409	0.05666	-0.04515
ONEPERHH	ONEPERHH	0.52809	0.75159	0.11959	0.15274	-0.14046	0.22896	0.06622
RNT_5075	RNT_5075	0.43013	0.74638	0.11661	-0.08079	-0.16287	-0.02889	0.12580
VEHICL_1	VEHICL_1	0.57017	0.70832	0.16745	0.16713	-0.15737	0.22545	-0.00679
SEASONAL	SEASONAL	0.29389	0.66168	0.08004	-0.02281	0.18995	0.10889	0.10099
PUBSEWER	PUBSEWER	0.54182	0.64466	0.15905	0.18657	-0.36129	0.18489	0.09815
INC_1525	INC_1525	0.60052	0.62289	0.14344	0.23826	-0.00776	0.22991	-0.14855
ATTACHED	ATTACHED	0.48132	0.53409	0.08439	-0.03386	-0.20843	0.03010	0.14968
DUPLEX	DUPLEX	0.08697	0.36452	0.15075	0.35401	-0.32275	0.32577	0.08195
AGEIS_20	AGEIS_20	0.05639	-0.02594	0.98038	0.01705	-0.04741	-0.02785	-0.05325
COLLEGE	COLLEGE	0.10585	0.06830	0.97909	0.01623	-0.06790	-0.01969	0.01400
AGE18_19	AGE18_19	0.03520	-0.16575	0.95672	-0.01037	-0.03543	-0.02691	-0.05763
DORMITOR	DORMITOR	-0.07221	-0.18543	0.94704	-0.04024	-0.04465	-0.04855	-0.04884
SINGLE_	SINGLE_	0.33271	0.25729	0.88363	0.12226	-0.10583	0.01577	0.00453
SAMESTAT	SAMESTAT	0.24618	0.17532	0.88277	0.02170	-0.01037	0.01866	-0.01978
AGE21_24	AGE21_24	0.35367	0.49548	0.71001	0.13963	-0.06539	-0.01045	-0.01383
ASIAN	ASIAN	0.39131	0.28145	0.67477	-0.00298	-0.19146	-0.07086	0.27653
ABLEENGL	ABLEENGL	0.52525	0.36650	0.57085	0.07617	-0.15330	0.03828	0.24256

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
AMIND	AMIND	0.44569	0.17489	0.49182	0.10133	0.02499	0.03660	-0.13739
OTH_HEAT	OTH_HEAT	0.04536	-0.01911	0.39750	0.18601	0.00170	0.05225	0.21089
O_NOINST	O_NOINST	0.02128	0.11283	0.22697	0.14967	-0.02238	0.04378	0.03493
INSTREET	INSTREET	-0.08620	0.07446	0.15708	0.03996	-0.00101	-0.03702	0.11473
MENTAL	MENTAL	-0.02389	0.01318	0.11485	-0.01557	-0.02343	-0.01375	-0.04398
CHILDPOV	CHILDPOV	0.16740	0.03004	0.03796	0.81941	-0.02266	-0.04095	-0.04810
NOVEHICL	NOVEHICL	0.02670	0.17710	0.02397	0.80486	-0.15090	0.25145	0.07804
RNT_LT25	RNT_LT25	-0.06275	-0.05979	0.04496	0.80383	0.00684	0.17520	-0.03055
PUBL_INC	PUBL_INC	0.07690	-0.03870	-0.03058	0.79915	0.08902	0.24188	-0.07398
INPOVRTY	INPOVRTY	0.18231	0.44504	0.21667	0.69788	0.05300	0.04900	0.05155
BLACK	BLACK	0.32856	0.05305	0.19131	0.64881	-0.19395	0.09901	-0.25074
SINGWICH	SINGWICH	0.57279	0.21816	0.02491	0.62933	-0.16355	0.15526	-0.16312
INC_LT15	INC_LT15	0.24813	0.51038	0.19219	0.62261	0.10692	0.28144	-0.00268
NODIPLOM	NODIPLOM	0.34822	-0.13547	-0.08812	0.49762	0.41454	0.38184	-0.34454
PUBTRANS	PUBTRANS	0.08698	0.32132	0.45409	0.47680	-0.19857	0.01971	0.12549
VAL_2550	VAL_2550	0.06595	-0.12856	-0.09472	0.44701	0.35697	0.40050	-0.34925
INCPRCAP	INCPRCAP	0.33250	0.05813	-0.15996	-0.45790	-0.27093	0.21465	0.42615
INC_MEDN	INC_MEDN	0.48844	-0.23961	-0.11775	-0.50209	-0.15813	0.06783	0.32476
LAND_KM	LAND_KM	0.01665	-0.07127	-0.06866	-0.02513	0.82112	0.01818	0.02259
COALWOOD	COALWOOD	0.11728	-0.15396	-0.03677	0.04657	0.75190	0.09248	-0.04535
BOTL_GAS	BOTL_GAS	0.24014	-0.19648	-0.11030	0.12021	0.75087	0.10333	-0.14253
FARM_INC	FARM_INC	0.14020	0.12060	-0.01517	0.06900	0.67412	0.06336	0.04419
VAL_LT25	VAL_LT25	-0.08512	-0.10319	-0.08878	0.31480	0.59442	0.14958	-0.10136
F_MIGRNT	F_MIGRNT	-0.07852	-0.00051	-0.04451	0.09655	0.47363	-0.05405	0.04906
WATER_KM	WATER_KM	-0.04240	0.01461	-0.01882	-0.04670	0.25790	0.00318	0.11071

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
RNT_MEDI	RNT_MEDI	0.35369	0.07926	-0.02187	-0.35982	-0.45089	0.06976	0.33255
popden		-0.06182	0.14216	0.37076	0.39643	-0.64361	-0.05780	0.07375
houseden		-0.06736	0.30451	0.04583	0.40179	-0.64774	-0.02059	0.10798
HOUSHOLD	HOUSHOLD	-0.06448	0.30016	0.04702	0.39425	-0.65178	-0.01207	0.10933
FAMILIES	FAMILIES	0.04205	0.17107	0.03341	0.39088	-0.71064	-0.00009	0.09687
AGE65_74	AGE65_74	0.45639	0.17418	-0.00608	0.16666	0.07040	0.78622	0.10357
AGE75_84	AGE75_84	0.20252	0.22151	-0.00589	0.21000	0.05797	0.77730	0.22425
SOCS_INC	SOCS_INC	0.42017	0.16971	-0.03198	0.27073	0.08386	0.74822	0.12750
WIDOWED	WIDOWED	0.33842	0.19600	-0.02617	0.33252	0.05033	0.72712	0.03251
RETI_INC	RETI_INC	0.55644	0.20587	0.02367	0.04448	-0.12055	0.64260	0.19015
FUELKERO	FUELKERO	0.23374	-0.07522	-0.05736	0.15887	0.39577	0.60380	-0.24508
BLTBFR70	BLTBFR70	0.28180	0.26283	0.20426	0.31337	-0.01619	0.58913	0.00389
NURSHOME	NURSHOME	0.01107	0.19700	0.00550	0.02728	-0.06195	0.22061	0.08862
MEDYRBLT	MEDYRBLT	0.16385	-0.04033	0.02333	0.04946	0.03123	0.20864	0.05793
O_INSTIT	O_INSTIT	-0.01715	-0.00529	0.00451	-0.07666	-0.09529	0.10033	0.08089
SHELTERS	SHELTERS	-0.11575	0.02230	0.01700	0.05603	-0.06861	-0.13258	0.03509
VAL_2C3C	VAL_2C3C	0.46200	-0.15475	0.02529	-0.09011	-0.04580	-0.00018	0.66275
VAL_GT3C	VAL_GT3C	0.24176	-0.01858	0.04061	-0.12063	-0.08815	0.17379	0.65317
RNT_751K	RNT_751K	0.37059	0.29420	0.09195	-0.01603	-0.10978	0.08233	0.52573
RNT_GT1K	RNT_GT1K	0.08889	0.09896	0.01623	0.02377	0.01240	0.15283	0.51427
VAL_MEDI	VAL_MEDI	0.35712	-0.08443	0.24228	-0.32517	-0.24477	0.10840	0.51392
CORRINST	CORRINST	-0.00919	-0.02907	0.01949	0.00735	-0.07326	0.01640	-0.10723

Principal Components Analysis of PTCDEMO1

The FACTOR Procedure
Initial Factor Method: Principal Components

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	70.5718054	57.3528250	0.4834	0.4834
2	13.2189804	2.5046545	0.0905	0.5739
3	10.7143259	3.5259695	0.0734	0.6473
4	7.1883565	2.7576349	0.0492	0.6965
5	4.4307216	0.6394795	0.0303	0.7269
6	3.7912420	1.2088643	0.0260	0.7528
7	2.5823777	0.4682525	0.0177	0.7705
8	2.1141252	0.2976247	0.0145	0.7850
9	1.8165005	0.1141169	0.0124	0.7975
10	1.7023836	0.1351683	0.0117	0.8091
11	1.5672153	0.2702216	0.0107	0.8198
12	1.2969936	0.0287292	0.0089	0.8287
13	1.2682645	0.0784061	0.0087	0.8374
14	1.1898583	0.0354695	0.0081	0.8456
15	1.1543888	0.0575807	0.0079	0.8535
16	1.0968081	0.0140511	0.0075	0.8610
17	1.0827570	0.0566320	0.0074	0.8684
18	1.0261250	0.0495444	0.0070	0.8754
19	0.9765806	0.0065802	0.0067	0.8821
20	0.9700004	0.1351656	0.0066	0.8888
21	0.8348347	0.0294586	0.0057	0.8945
22	0.8053762	0.0583918	0.0055	0.9000
23	0.7469844	0.0243720	0.0051	0.9051
24	0.7226124	0.0512830	0.0049	0.9101

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
25	0.6713294	0.0564623	0.0046	0.9147
26	0.6148671	0.0213831	0.0042	0.9189
27	0.5934840	0.0547735	0.0041	0.9229
28	0.5387104	0.0182351	0.0037	0.9266
29	0.5204753	0.0222759	0.0036	0.9302
30	0.4981994	0.0435005	0.0034	0.9336
31	0.4546990	0.0116490	0.0031	0.9367
32	0.4430500	0.0565787	0.0030	0.9398
33	0.3864713	0.0089809	0.0026	0.9424
34	0.3774903	0.0204873	0.0026	0.9450
35	0.3570031	0.0123172	0.0024	0.9474
36	0.3446858	0.0076784	0.0024	0.9498
37	0.3370075	0.0238434	0.0023	0.9521
38	0.3131640	0.0151648	0.0021	0.9542
39	0.2979993	0.0140174	0.0020	0.9563
40	0.2839819	0.0034557	0.0019	0.9582
41	0.2805261	0.0044359	0.0019	0.9602
42	0.2760902	0.0154504	0.0019	0.9620
43	0.2606398	0.0031969	0.0018	0.9638
44	0.2574430	0.0253299	0.0018	0.9656
45	0.2321131	0.0099016	0.0016	0.9672
46	0.2222115	0.0088761	0.0015	0.9687
47	0.2133354	0.0174376	0.0015	0.9702
48	0.1958978	0.0005876	0.0013	0.9715
49	0.1953102	0.0091253	0.0013	0.9728

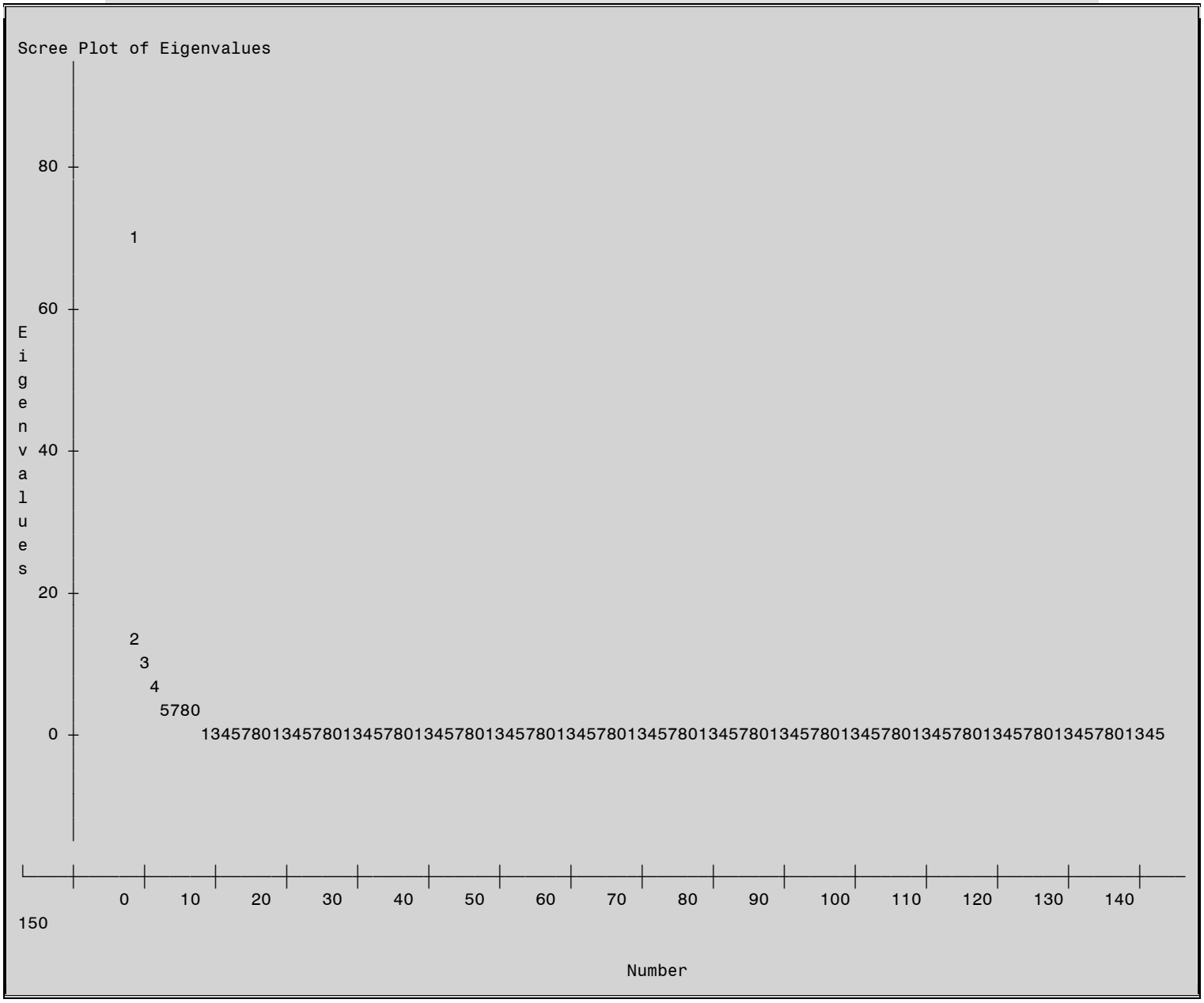
Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
50	0.1861850	0.0079285	0.0013	0.9741
51	0.1782565	0.0103518	0.0012	0.9753
52	0.1679047	0.0035592	0.0012	0.9765
53	0.1643455	0.0038498	0.0011	0.9776
54	0.1604957	0.0130153	0.0011	0.9787
55	0.1474804	0.0071785	0.0010	0.9797
56	0.1403019	0.0062933	0.0010	0.9807
57	0.1340085	0.0055951	0.0009	0.9816
58	0.1284134	0.0080933	0.0009	0.9825
59	0.1203201	0.0022871	0.0008	0.9833
60	0.1180330	0.0085115	0.0008	0.9841
61	0.1095215	0.0019718	0.0008	0.9849
62	0.1075497	0.0044989	0.0007	0.9856
63	0.1030508	0.0045287	0.0007	0.9863
64	0.0985221	0.0039190	0.0007	0.9870
65	0.0946031	0.0007583	0.0006	0.9876
66	0.0938448	0.0065160	0.0006	0.9883
67	0.0873288	0.0037879	0.0006	0.9889
68	0.0835409	0.0003668	0.0006	0.9894
69	0.0831741	0.0048457	0.0006	0.9900
70	0.0783284	0.0048591	0.0005	0.9906
71	0.0734693	0.0044326	0.0005	0.9911
72	0.0690367	0.0020192	0.0005	0.9915
73	0.0670175	0.0028109	0.0005	0.9920
74	0.0642066	0.0009218	0.0004	0.9924

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
75	0.0632848	0.0028842	0.0004	0.9929
76	0.0604006	0.0029798	0.0004	0.9933
77	0.0574208	0.0025180	0.0004	0.9937
78	0.0549027	0.0022875	0.0004	0.9940
79	0.0526152	0.0024850	0.0004	0.9944
80	0.0501302	0.0047169	0.0003	0.9947
81	0.0454133	0.0009751	0.0003	0.9951
82	0.0444382	0.0004809	0.0003	0.9954
83	0.0439573	0.0023004	0.0003	0.9957
84	0.0416569	0.0047483	0.0003	0.9960
85	0.0369086	0.0009329	0.0003	0.9962
86	0.0359756	0.0011646	0.0002	0.9965
87	0.0348110	0.0014216	0.0002	0.9967
88	0.0333894	0.0009567	0.0002	0.9969
89	0.0324327	0.0019547	0.0002	0.9971
90	0.0304781	0.0020551	0.0002	0.9973
91	0.0284230	0.0032158	0.0002	0.9975
92	0.0252071	0.0009761	0.0002	0.9977
93	0.0242310	0.0004734	0.0002	0.9979
94	0.0237576	0.0021884	0.0002	0.9980
95	0.0215692	0.0007236	0.0001	0.9982
96	0.0208456	0.0007455	0.0001	0.9983
97	0.0201001	0.0022631	0.0001	0.9985
98	0.0178370	0.0011252	0.0001	0.9986
99	0.0167118	0.0003248	0.0001	0.9987

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
100	0.0163869	0.0006764	0.0001	0.9988
101	0.0157105	0.0008969	0.0001	0.9989
102	0.0148136	0.0024016	0.0001	0.9990
103	0.0124120	0.0002626	0.0001	0.9991
104	0.0121494	0.0009594	0.0001	0.9992
105	0.0111900	0.0004686	0.0001	0.9993
106	0.0107214	0.0018031	0.0001	0.9993
107	0.0089183	0.0002281	0.0001	0.9994
108	0.0086902	0.0006884	0.0001	0.9995
109	0.0080018	0.0008442	0.0001	0.9995
110	0.0071576	0.0004844	0.0000	0.9996
111	0.0066732	0.0004077	0.0000	0.9996
112	0.0062655	0.0010912	0.0000	0.9997
113	0.0051742	0.0003822	0.0000	0.9997
114	0.0047920	0.0002471	0.0000	0.9997
115	0.0045449	0.0001391	0.0000	0.9998
116	0.0044058	0.0003830	0.0000	0.9998
117	0.0040228	0.0008017	0.0000	0.9998
118	0.0032211	0.0001282	0.0000	0.9998
119	0.0030929	0.0007586	0.0000	0.9999
120	0.0023343	0.0001975	0.0000	0.9999
121	0.0021368	0.0001119	0.0000	0.9999
122	0.0020249	0.0000235	0.0000	0.9999
123	0.0020014	0.0003596	0.0000	0.9999
124	0.0016418	0.0001518	0.0000	0.9999

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
125	0.0014900	0.0001893	0.0000	0.9999
126	0.0013007	0.0000913	0.0000	1.0000
127	0.0012095	0.0002126	0.0000	1.0000
128	0.0009968	0.0001706	0.0000	1.0000
129	0.0008262	0.0000328	0.0000	1.0000
130	0.0007934	0.0000894	0.0000	1.0000
131	0.0007040	0.0000343	0.0000	1.0000
132	0.0006697	0.0001999	0.0000	1.0000
133	0.0004697	0.0000717	0.0000	1.0000
134	0.0003981	0.0000694	0.0000	1.0000
135	0.0003287	0.0001714	0.0000	1.0000
136	0.0001572	0.0000431	0.0000	1.0000
137	0.0001141	0.0000124	0.0000	1.0000
138	0.0001017	0.0000043	0.0000	1.0000
139	0.0000974	0.0000235	0.0000	1.0000
140	0.0000739	0.0000105	0.0000	1.0000
141	0.0000635	0.0000351	0.0000	1.0000
142	0.0000284	0.0000044	0.0000	1.0000
143	0.0000240	0.0000117	0.0000	1.0000
144	0.0000123	0.0000019	0.0000	1.0000
145	0.0000105	0.0000033	0.0000	1.0000
146	0.0000071		0.0000	1.0000

The FACTOR Procedure
Initial Factor Method: Principal Components



Principal Components Analysis of PTCDEMO1
The FACTOR Procedure
Rotation Method: Varimax

Rotated Factor Pattern

		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
MARRWICH	MARRWICH	0.96931	-0.04389	0.00812	-0.00312	0.10248	0.07938	0.12659
PERINFAM	PERINFAM	0.96272	0.05365	0.01767	0.12280	0.06182	0.19440	0.08975
PRUNIT34	PRUNIT34	0.95825	0.10209	0.04588	0.12389	0.06003	0.14296	0.08054
MARRIED	MARRIED	0.95458	0.08237	0.04503	-0.02346	0.08965	0.20930	0.13193
AGE35_44	AGE35_44	0.95219	0.13101	0.01646	0.04905	0.00631	0.09783	0.19514
AGE05_09	AGE05_09	0.94628	-0.01666	0.00139	0.20138	0.02928	0.09648	0.11393
AGELT_05	AGELT_05	0.93225	0.06626	0.04332	0.18820	-0.01590	0.08553	0.00226
AGE10_14	AGE10_14	0.93062	-0.06894	-0.03131	0.21342	0.06479	0.09931	0.14296
OWNR_OCC	OWNR_OCC	0.92706	-0.02876	-0.02974	-0.03336	0.14244	0.27833	0.10003
ELEMSCND	ELEMSCND	0.92490	-0.06651	-0.01163	0.22042	0.07146	0.08919	0.10814
DRVALONE	DRVALONE	0.92176	0.33627	0.07838	-0.00639	-0.03929	0.12261	0.03689
SOMECOLG	SOMECOLG	0.91461	0.26313	0.00891	-0.00542	-0.08165	0.13798	-0.04236
PRIVWORK	PRIVWORK	0.91281	0.31451	0.16211	0.03561	-0.03551	0.10839	-0.00835
AGE15_17	AGE15_17	0.91106	-0.08751	-0.01827	0.20542	0.06817	0.11603	0.14849
VEHICL_2	VEHICL_2	0.89784	0.35724	0.05965	-0.01083	-0.02153	0.15057	0.09460
COM_1529	COM_1529	0.89510	0.32399	0.15419	0.00262	-0.07708	0.09546	0.02124
AGE45_54	AGE45_54	0.89475	0.02959	-0.00898	0.03396	0.06417	0.22950	0.23287
INC_5075	INC_5075	0.89073	0.19326	0.00663	-0.13775	-0.13145	0.07764	0.19233
EMPLOYED	EMPLOYED	0.88935	0.32940	0.26902	0.06018	-0.01936	0.13265	0.02309
MARRNOCH	MARRNOCH	0.88316	0.19425	0.06120	-0.06371	0.07668	0.33066	0.14031
MALE	MALE	0.87938	0.22843	0.35119	0.11456	0.01412	0.15876	0.06634
DETACHED	DETACHED	0.87790	-0.09404	-0.02368	0.02512	0.09820	0.33575	0.17927
UTILITY	UTILITY	0.86934	0.18706	0.02921	-0.01642	-0.02723	-0.03369	-0.08994
NATIVE	NATIVE	0.86550	0.19809	0.38895	0.13888	0.00704	0.18548	0.04216
WAGE_SAL	WAGE_SAL	0.85633	0.46126	0.09977	0.09284	-0.04596	0.14499	0.05659
WRKINCTY	WRKINCTY	0.85428	0.30540	0.28349	0.06211	-0.13646	0.12877	-0.05453

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
ONLYENGL	ONLYENGL	0.85414	0.20407	0.40930	0.12834	0.00857	0.18964	0.04749
COM_3044	COM_3044	0.85358	-0.00185	-0.01079	0.09473	0.24534	0.00520	-0.07798
INC_3550	INC_3550	0.85087	0.33528	0.02980	-0.03478	-0.00985	0.22988	-0.16600
SAMECNTY	SAMECNTY	0.85051	0.26929	0.03810	0.15550	-0.10103	0.19874	-0.11549
TECHSALE	TECHSALE	0.84976	0.30814	0.35994	-0.00604	-0.09595	0.07532	-0.04592
TRADE	TRADE	0.84928	0.34040	0.24603	0.06303	-0.03459	0.10444	-0.04506
AGEGE_03	AGEGE_03	0.84743	0.22453	0.41807	0.12628	-0.00928	0.17892	0.07245
FIRE	FIRE	0.84631	0.28256	0.05090	-0.06639	-0.18468	0.05464	0.02479
WHITE	WHITE	0.84266	0.21422	0.35829	-0.10819	0.07081	0.17022	0.15083
MANUFACT	MANUFACT	0.82456	0.09443	-0.00351	0.07776	0.22065	0.08962	-0.04335
AGE25_34	AGE25_34	0.82225	0.48531	0.14232	0.08990	-0.06726	0.04865	-0.08433
WORKERS	WORKERS	0.81905	0.29462	0.42612	0.08922	0.00380	0.17947	0.05006
BORNSOUT	BORNSOUT	0.81673	0.25589	0.39766	-0.02047	-0.15816	0.05298	0.14909
EXECPROF	EXECPROF	0.81289	0.36487	0.18363	-0.07348	-0.17042	0.06326	0.31086
PREPRIMA	PREPRIMA	0.79697	0.05821	0.06920	0.00134	-0.10656	-0.03596	0.35515
COLGGRAD	COLGGRAD	0.79659	0.38649	0.11491	-0.06767	-0.21295	0.03690	0.32446
BLTAFT84	BLTAFT84	0.79608	0.32165	0.00439	-0.07784	-0.00509	-0.11364	0.09233
SELF_INC	SELF_INC	0.79558	0.26131	0.08224	0.00143	0.02680	0.25858	0.25693
VAL_1C2C	VAL_1C2C	0.79509	0.07150	0.00439	-0.15577	-0.15102	0.00940	0.42242
SERVICES	SERVICES	0.79379	0.33474	0.31027	0.06311	-0.12822	0.05392	-0.05188
AGEGE_16	AGEGE_16	0.79363	0.27103	0.49156	0.10472	-0.01892	0.18915	0.05980
SAMEHOUS	SAMEHOUS	0.79133	-0.07928	0.01493	0.19290	0.22392	0.39266	0.03302
BLT_8084	BLT_8084	0.77334	0.41338	0.01013	0.01157	-0.01989	-0.08824	0.08183
BORN_INS	BORN_INS	0.77129	0.04017	0.32228	0.27923	0.18459	0.32237	-0.20991
INTE_INC	INTE_INC	0.76792	0.44171	0.11121	-0.08949	-0.12217	0.18805	0.33637

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
ELECTRIC	ELECTRIC	0.76464	0.54883	0.10062	0.09378	-0.05482	0.02522	0.01195
CARPOOL	CARPOOL	0.75833	0.23810	0.10525	0.27061	0.20790	0.18527	-0.16599
BORNMIDW	BORNMIDW	0.74344	0.33583	0.30932	-0.07284	-0.16142	-0.04876	0.34784
ROOM_4_6	ROOM_4_6	0.73668	0.51725	0.08346	0.19302	0.07291	0.26931	-0.16883
AGE55_64	AGE55_64	0.73409	0.09458	-0.00878	0.07851	0.08078	0.53932	0.10471
VAL_501C	VAL_501C	0.72186	0.01593	-0.01973	0.00442	-0.11628	0.37294	-0.35203
BORNNORT	BORNNORT	0.71997	0.31692	0.34290	-0.01335	-0.19716	-0.00272	0.38100
SERVICE	SERVICE	0.71891	0.22182	0.22217	0.29404	0.24724	0.23795	-0.26356
SELFWORK	SELFWORK	0.71830	0.09070	0.03682	0.02243	0.28012	0.25660	0.12653
HIGHSCHL	HIGHSCHL	0.71661	-0.04554	-0.07269	0.13935	0.29936	0.37189	-0.33795
OTH_STAT	OTH_STAT	0.69574	0.34672	0.39832	-0.01996	-0.17081	-0.10277	0.28363
INC_2535	INC_2535	0.68380	0.54464	0.08071	0.10577	0.03069	0.23466	-0.17370
PRUNIT12	PRUNIT12	0.68205	0.65195	0.11860	0.10929	-0.08136	0.23942	0.07285
BLT_7079	BLT_7079	0.65676	0.46056	0.08231	0.13690	0.00838	0.14500	0.04631
BORNWEST	BORNWEST	0.64014	0.36370	0.40637	-0.03678	-0.11517	0.00177	0.29637
PUBWATER	PUBWATER	0.63115	0.59759	0.12421	0.17046	-0.33097	0.14325	0.12786
PRIMARY_	PRIMARY_	0.62874	0.04930	0.00106	0.10058	0.47283	0.19579	-0.24778
PROFSERV	PROFSERV	0.62241	0.39637	0.46110	0.06553	-0.14841	0.18015	0.22912
NATURAL_	NATURAL_	0.62178	0.33909	0.26727	-0.04151	-0.16741	0.04844	0.37737
SEPARATE	SEPARATE	0.61903	0.36544	0.03692	0.47282	-0.11472	0.19367	-0.20042
PUBLWORK	PUBLWORK	0.60165	0.33885	0.56894	0.12728	-0.04394	0.13270	0.08400
WRKEXCTY	WRKEXCTY	0.57913	0.25515	0.08793	0.02082	0.32832	0.08253	0.23856
COM_LT15	COM_LT15	0.57802	0.45269	0.51513	0.06425	-0.18940	0.20362	0.09103
PUBL_GAS	PUBL_GAS	0.57199	0.32213	0.08424	0.05992	-0.44599	0.24791	0.26437
HISPANIC	HISPANIC	0.55743	0.43808	0.33294	0.12030	-0.00027	-0.00393	0.05890

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
ABLEENGL	ABLEENGL	0.55055	0.37193	0.53620	0.05813	-0.16245	0.05564	0.24225
WRK_HOME	WRK_HOME	0.51461	0.09619	0.39223	-0.04356	0.09467	0.17809	0.26454
INC_MEDN	INC_MEDN	0.48534	-0.24166	-0.08767	-0.45904	-0.15075	0.08834	0.37053
UNEMPLOY	UNEMPLOY	0.44203	0.27336	0.43526	0.42314	-0.09875	0.14683	-0.02863
ARMDFORC	ARMDFORC	0.34878	0.25958	0.02331	-0.03374	0.00055	0.19898	-0.08132
APARTMNT	APARTMNT	0.34603	0.82715	0.17496	0.16770	-0.24625	-0.04142	0.04732
RNT_2550	RNT_2550	0.36392	0.79414	0.20047	0.19918	-0.23936	0.02806	-0.01955
ROOM_1_3	ROOM_1_3	0.34772	0.79226	0.17452	0.23612	-0.17531	-0.01755	0.08429
ONEPERHH	ONEPERHH	0.53304	0.75990	0.10145	0.15268	-0.13577	0.20963	0.05691
RNT_5075	RNT_5075	0.43745	0.72464	0.07830	-0.10248	-0.18503	-0.00696	0.09916
VEHICL_1	VEHICL_1	0.56206	0.71758	0.15147	0.17225	-0.15778	0.22457	-0.00686
SEASONAL	SEASONAL	0.29449	0.65662	0.06509	-0.03587	0.14575	0.14428	0.05749
PUBSEWER	PUBSEWER	0.54460	0.64527	0.13648	0.18535	-0.36954	0.16527	0.09668
INC_1525	INC_1525	0.57953	0.63689	0.13407	0.24785	-0.00491	0.22598	-0.14612
ATTACHED	ATTACHED	0.48951	0.52561	0.05810	-0.05536	-0.23062	0.05614	0.11739
DUPLEX	DUPLEX	0.06471	0.39376	0.12457	0.35766	-0.32947	0.30810	0.08010
AGEIS_20	AGEIS_20	0.07322	0.00370	0.98078	0.02965	-0.04792	-0.02085	-0.04972
COLLEGE	COLLEGE	0.12427	0.11156	0.97253	0.03330	-0.06668	-0.01992	0.02063
AGE18_19	AGE18_19	0.05162	-0.14313	0.96116	-0.00436	-0.03443	-0.02940	-0.05258
DORMITOR	DORMITOR	-0.05819	-0.16328	0.95383	-0.03646	-0.04403	-0.05282	-0.04279
SAMESTAT	SAMESTAT	0.24385	0.19931	0.90595	0.02980	-0.01708	0.01252	-0.01422
SINGLE_	SINGLE_	0.34793	0.30086	0.86112	0.14203	-0.10481	0.02261	0.00890
ASIAN	ASIAN	0.43458	0.30443	0.67205	-0.03092	-0.19549	-0.06490	0.29600
AGE21_24	AGE21_24	0.36155	0.54684	0.66098	0.16652	-0.06056	0.01425	-0.01546
AMIND	AMIND	0.41016	0.15681	0.51575	0.09352	0.03726	0.02733	-0.11559

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
OTH_HEAT	OTH_HEAT	0.04084	-0.02226	0.40532	0.16644	0.01372	0.04851	0.19663
O_NOINST	O_NOINST	0.01874	0.04631	0.23674	0.16608	-0.04461	0.10980	0.04446
INSTREET	INSTREET	-0.06807	0.06415	0.12858	0.02688	-0.00591	-0.04818	0.07993
MENTAL	MENTAL	-0.01596	0.01620	0.12172	-0.01861	-0.02590	-0.01772	-0.04356
CHILDPOV	CHILDPOV	0.12837	0.03888	0.03550	0.84519	-0.02160	-0.02563	-0.02998
NOVEHICL	NOVEHICL	0.01251	0.16951	0.02712	0.83440	-0.14067	0.19969	0.06816
RNT_LT25	RNT_LT25	-0.04970	-0.08844	0.04643	0.82293	0.00629	0.18673	-0.07082
PUBL_INC	PUBL_INC	0.03987	-0.02810	-0.02889	0.81755	0.09953	0.20165	-0.04519
INPOVRTY	INPOVRTY	0.15471	0.46850	0.20112	0.71043	0.05927	0.05973	0.05293
BLACK	BLACK	0.30909	0.05402	0.18369	0.67212	-0.19445	0.09175	-0.22186
SINGWICH	SINGWICH	0.54372	0.22212	0.01317	0.65826	-0.16479	0.15138	-0.14799
INC_LT15	INC_LT15	0.22375	0.51347	0.18311	0.63568	0.12366	0.26080	-0.01224
PUBTRANS	PUBTRANS	0.06639	0.37889	0.38760	0.51874	-0.17259	0.01472	0.13720
NODIPLOM	NODIPLOM	0.35165	-0.15131	-0.06757	0.47798	0.41356	0.40129	-0.36122
VAL_2550	VAL_2550	0.05187	-0.12850	-0.08571	0.45904	0.35539	0.39825	-0.33536
LAND_KM	LAND_KM	0.00819	-0.07277	-0.06322	-0.02712	0.82561	0.02754	0.02412
COALWOOD	COALWOOD	0.09703	-0.14247	-0.02160	0.06438	0.75639	0.10805	-0.04204
BOTL_GAS	BOTL_GAS	0.24394	-0.19678	-0.09133	0.11835	0.74469	0.15069	-0.15877
FARM_INC	FARM_INC	0.15067	0.13661	-0.00863	0.05175	0.65533	0.09175	0.00860
VAL_LT25	VAL_LT25	-0.09248	-0.10266	-0.08319	0.31263	0.57702	0.16458	-0.10604
F_MIGRNT	F_MIGRNT	-0.08354	0.00500	-0.04376	0.08654	0.47346	-0.05866	0.03109
WATER_KM	WATER_KM	-0.04717	0.01360	-0.02052	-0.03658	0.27335	-0.00110	0.13758
RNT_MEDI	RNT_MEDI	0.34791	0.08060	0.00202	-0.33839	-0.44515	0.08621	0.36034
popden		-0.05964	0.16054	0.35249	0.38094	-0.64934	-0.06824	0.04281
houseden		-0.06646	0.32035	0.04514	0.37956	-0.65622	-0.03191	0.06921

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
HOUSHOLD	HOUSHOLD	-0.06240	0.31796	0.04728	0.37172	-0.66176	-0.02405	0.07159
FAMILIES	FAMILIES	0.02670	0.17883	0.03309	0.39154	-0.71513	-0.00906	0.08142
AGE75_84	AGE75_84	0.21581	0.21175	-0.00056	0.19528	0.06981	0.78727	0.19981
AGE65_74	AGE65_74	0.45595	0.17228	-0.00409	0.16641	0.09199	0.77480	0.12175
WIDOWED	WIDOWED	0.35677	0.18186	-0.01902	0.31029	0.05727	0.74307	-0.00912
SOCS_INC	SOCS_INC	0.43206	0.16800	-0.02905	0.26287	0.10676	0.73820	0.12257
RETI_INC	RETI_INC	0.56134	0.19466	0.01650	0.05068	-0.10463	0.61376	0.22943
FUELKERO	FUELKERO	0.24238	-0.10692	-0.04963	0.18653	0.39769	0.58217	-0.24285
BLTBFR70	BLTBFR70	0.23624	0.29750	0.21493	0.32584	0.00001	0.58152	0.02885
NURSHOME	NURSHOME	0.02422	0.19917	0.00904	-0.00317	-0.07081	0.26702	0.06168
MEDYRBLT	MEDYRBLT	0.17092	-0.03551	0.02966	0.07337	0.03956	0.25536	0.09262
SHELTERS	SHELTERS	-0.09936	0.01373	0.01305	0.02377	-0.07063	-0.14773	-0.00948
VAL_GT3C	VAL_GT3C	0.23984	-0.00728	0.03802	-0.10543	-0.08267	0.18078	0.69402
VAL_2C3C	VAL_2C3C	0.45906	-0.14805	0.02151	-0.07534	-0.03666	-0.01276	0.67083
VAL_MEDI	VAL_MEDI	0.36263	-0.07603	0.27248	-0.31005	-0.23083	0.11748	0.55047
RNT_GT1K	RNT_GT1K	0.08756	0.08009	0.02721	0.01772	0.01377	0.16248	0.49363
RNT_751K	RNT_751K	0.42148	0.31047	0.07467	-0.03591	-0.13065	0.05406	0.47544
INCPRCAP	INCPRCAP	0.31408	0.04936	-0.13605	-0.43977	-0.27219	0.22251	0.45627
O_INSTIT	O_INSTIT	-0.02557	-0.01546	0.00000	-0.08703	-0.09061	0.08542	0.12838
CORRINST	CORRINST	0.01087	-0.03696	0.02329	-0.02506	-0.08226	0.02867	-0.16100

Principal Components Analysis of PTCDEMO2
The FACTOR Procedure
Initial Factor Method: Principal Components

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	70.4219955	57.2092441	0.4823	0.4823
2	13.2127513	3.0476879	0.0905	0.5728
3	10.1650634	2.6175377	0.0696	0.6425
4	7.5475257	2.8745458	0.0517	0.6942
5	4.6729799	1.0570998	0.0320	0.7262
6	3.6158801	0.9351096	0.0248	0.7509
7	2.6807705	0.6092264	0.0184	0.7693
8	2.0715441	0.1701145	0.0142	0.7835
9	1.9014296	0.1401855	0.0130	0.7965
10	1.7612441	0.1793045	0.0121	0.8086
11	1.5819396	0.2492056	0.0108	0.8194
12	1.3327339	0.0743539	0.0091	0.8285
13	1.2583801	0.0362079	0.0086	0.8372
14	1.2221722	0.0790402	0.0084	0.8455
15	1.1431321	0.0691360	0.0078	0.8534
16	1.0739961	0.0256311	0.0074	0.8607
17	1.0483650	0.0619170	0.0072	0.8679
18	0.9864480	0.0392136	0.0068	0.8746
19	0.9472344	0.0550143	0.0065	0.8811
20	0.8922201	0.0183657	0.0061	0.8872
21	0.8738545	0.0706612	0.0060	0.8932
22	0.8031933	0.0401544	0.0055	0.8987
23	0.7630389	0.0319501	0.0052	0.9040
24	0.7310888	0.0209887	0.0050	0.9090
25	0.7101001	0.0524869	0.0049	0.9138

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
26	0.6576132	0.0599015	0.0045	0.9183
27	0.5977117	0.0447844	0.0041	0.9224
28	0.5529273	0.0392409	0.0038	0.9262
29	0.5136864	0.0170139	0.0035	0.9297
30	0.4966725	0.0418366	0.0034	0.9331
31	0.4548359	0.0422591	0.0031	0.9363
32	0.4125768	0.0102769	0.0028	0.9391
33	0.4023000	0.0047463	0.0028	0.9418
34	0.3975536	0.0163443	0.0027	0.9446
35	0.3812094	0.0148551	0.0026	0.9472
36	0.3663542	0.0196766	0.0025	0.9497
37	0.3466777	0.0091454	0.0024	0.9520
38	0.3375322	0.0020948	0.0023	0.9544
39	0.3354374	0.0342113	0.0023	0.9567
40	0.3012261	0.0218319	0.0021	0.9587
41	0.2793942	0.0106739	0.0019	0.9606
42	0.2687203	0.0043330	0.0018	0.9625
43	0.2643873	0.0193117	0.0018	0.9643
44	0.2450756	0.0085531	0.0017	0.9660
45	0.2365225	0.0186142	0.0016	0.9676
46	0.2179083	0.0046982	0.0015	0.9691
47	0.2132101	0.0073986	0.0015	0.9705
48	0.2058115	0.0093372	0.0014	0.9719
49	0.1964743	0.0154615	0.0013	0.9733
50	0.1810127	0.0016261	0.0012	0.9745

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
51	0.1793867	0.0113388	0.0012	0.9758
52	0.1680479	0.0065790	0.0012	0.9769
53	0.1614688	0.0047452	0.0011	0.9780
54	0.1567237	0.0132466	0.0011	0.9791
55	0.1434771	0.0045338	0.0010	0.9801
56	0.1389433	0.0032305	0.0010	0.9810
57	0.1357128	0.0025816	0.0009	0.9820
58	0.1331311	0.0037088	0.0009	0.9829
59	0.1294223	0.0071406	0.0009	0.9838
60	0.1222817	0.0077020	0.0008	0.9846
61	0.1145797	0.0050047	0.0008	0.9854
62	0.1095750	0.0059514	0.0008	0.9861
63	0.1036236	0.0065317	0.0007	0.9868
64	0.0970918	0.0017981	0.0007	0.9875
65	0.0952937	0.0039464	0.0007	0.9882
66	0.0913473	0.0035753	0.0006	0.9888
67	0.0877720	0.0032956	0.0006	0.9894
68	0.0844764	0.0042342	0.0006	0.9900
69	0.0802422	0.0051447	0.0005	0.9905
70	0.0750975	0.0015900	0.0005	0.9910
71	0.0735075	0.0027392	0.0005	0.9915
72	0.0707682	0.0024611	0.0005	0.9920
73	0.0683071	0.0060586	0.0005	0.9925
74	0.0622485	0.0005219	0.0004	0.9929
75	0.0617267	0.0065409	0.0004	0.9933

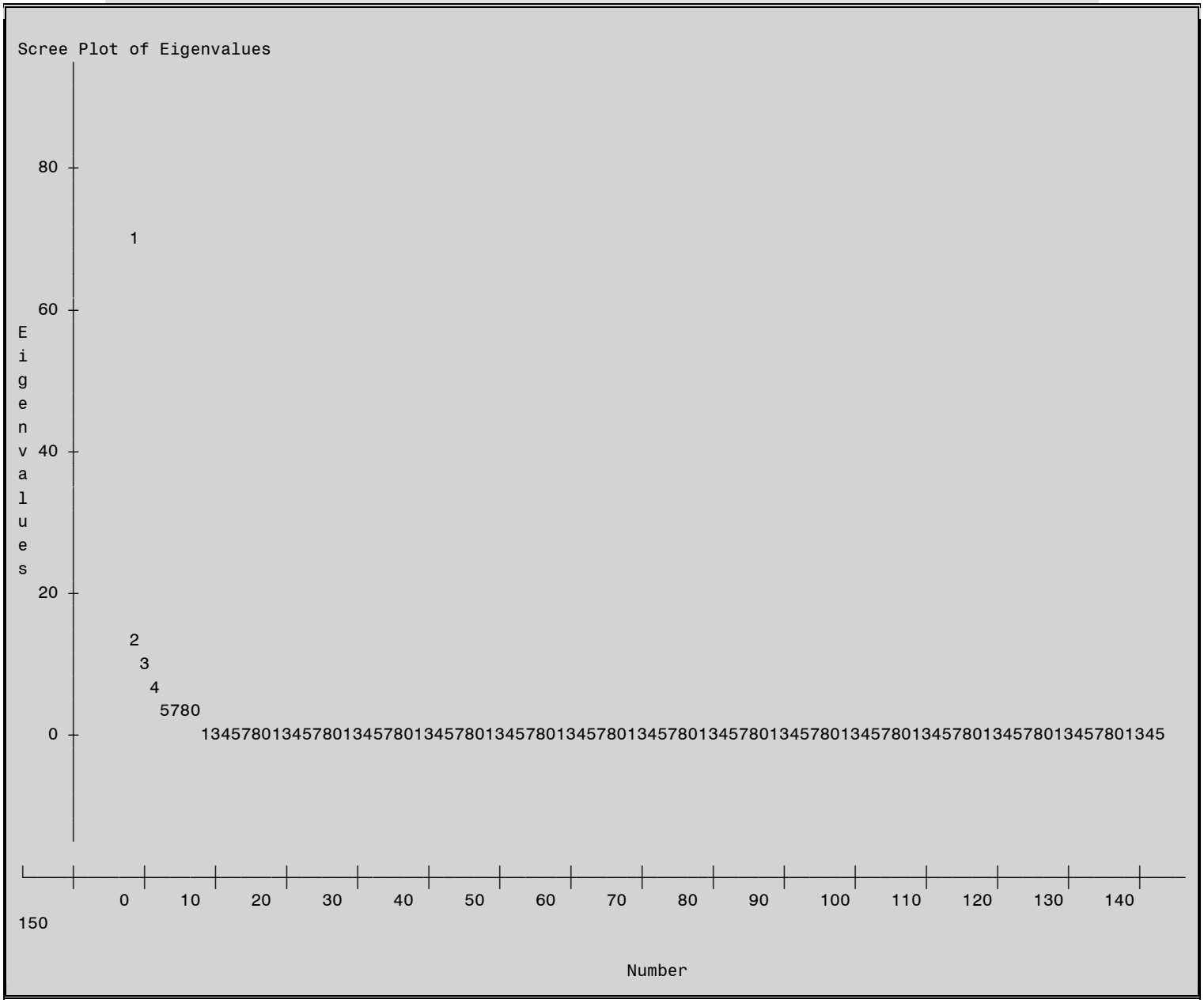
Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
76	0.0551857	0.0017557	0.0004	0.9937
77	0.0534300	0.0024498	0.0004	0.9941
78	0.0509802	0.0005618	0.0003	0.9944
79	0.0504184	0.0035155	0.0003	0.9948
80	0.0469029	0.0034824	0.0003	0.9951
81	0.0434204	0.0037655	0.0003	0.9954
82	0.0396549	0.0005598	0.0003	0.9957
83	0.0390951	0.0007881	0.0003	0.9959
84	0.0383070	0.0012283	0.0003	0.9962
85	0.0370787	0.0032727	0.0003	0.9964
86	0.0338060	0.0024120	0.0002	0.9967
87	0.0313940	0.0013851	0.0002	0.9969
88	0.0300089	0.0012843	0.0002	0.9971
89	0.0287245	0.0014379	0.0002	0.9973
90	0.0272867	0.0014465	0.0002	0.9975
91	0.0258402	0.0017724	0.0002	0.9977
92	0.0240678	0.0022215	0.0002	0.9978
93	0.0218463	0.0011814	0.0001	0.9980
94	0.0206649	0.0002812	0.0001	0.9981
95	0.0203837	0.0012294	0.0001	0.9983
96	0.0191544	0.0011703	0.0001	0.9984
97	0.0179841	0.0003679	0.0001	0.9985
98	0.0176162	0.0016685	0.0001	0.9986
99	0.0159476	0.0002613	0.0001	0.9987
100	0.0156864	0.0008427	0.0001	0.9988

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
101	0.0148437	0.0006452	0.0001	0.9989
102	0.0141985	0.0012227	0.0001	0.9990
103	0.0129757	0.0011373	0.0001	0.9991
104	0.0118384	0.0013164	0.0001	0.9992
105	0.0105221	0.0006794	0.0001	0.9993
106	0.0098426	0.0005759	0.0001	0.9994
107	0.0092667	0.0008612	0.0001	0.9994
108	0.0084055	0.0002438	0.0001	0.9995
109	0.0081617	0.0007390	0.0001	0.9995
110	0.0074227	0.0006665	0.0001	0.9996
111	0.0067562	0.0006441	0.0000	0.9996
112	0.0061121	0.0008176	0.0000	0.9997
113	0.0052945	0.0006291	0.0000	0.9997
114	0.0046654	0.0002438	0.0000	0.9997
115	0.0044216	0.0001495	0.0000	0.9998
116	0.0042720	0.0004509	0.0000	0.9998
117	0.0038211	0.0007132	0.0000	0.9998
118	0.0031079	0.0002931	0.0000	0.9998
119	0.0028148	0.0003253	0.0000	0.9999
120	0.0024895	0.0003489	0.0000	0.9999
121	0.0021406	0.0000262	0.0000	0.9999
122	0.0021143	0.0002066	0.0000	0.9999
123	0.0019077	0.0001497	0.0000	0.9999
124	0.0017580	0.0002632	0.0000	0.9999
125	0.0014948	0.0001040	0.0000	0.9999

Eigenvalues of the Correlation Matrix: Total = 146 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
126	0.0013908	0.0002533	0.0000	1.0000
127	0.0011375	0.0000987	0.0000	1.0000
128	0.0010388	0.0001465	0.0000	1.0000
129	0.0008923	0.0001588	0.0000	1.0000
130	0.0007335	0.0000585	0.0000	1.0000
131	0.0006750	0.0001115	0.0000	1.0000
132	0.0005634	0.0000826	0.0000	1.0000
133	0.0004808	0.0001060	0.0000	1.0000
134	0.0003748	0.0000470	0.0000	1.0000
135	0.0003278	0.0001649	0.0000	1.0000
136	0.0001629	0.0000570	0.0000	1.0000
137	0.0001059	0.0000032	0.0000	1.0000
138	0.0001027	0.0000034	0.0000	1.0000
139	0.0000993	0.0000277	0.0000	1.0000
140	0.0000717	0.0000107	0.0000	1.0000
141	0.0000610	0.0000336	0.0000	1.0000
142	0.0000275	0.0000012	0.0000	1.0000
143	0.0000262	0.0000137	0.0000	1.0000
144	0.0000125	0.0000025	0.0000	1.0000
145	0.0000100	0.0000023	0.0000	1.0000
146	0.0000077		0.0000	1.0000

Principal Components Analysis of PTCDEMO2
The FACTOR Procedure

Initial Factor Method: Principal Components



Principal Components Analysis of PTCDEMO2
The FACTOR Procedure
Rotation Method: Varimax

Rotated Factor Pattern							
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
PERINFAM	PERINFAM	0.96328	0.07295	0.02173	0.11105	0.06307	0.17978	0.10915
MARRWICH	MARRWICH	0.96251	-0.02147	0.00669	-0.00387	0.10966	0.05600	0.17490
PRUNIT34	PRUNIT34	0.95926	0.10748	0.05235	0.10159	0.05875	0.11778	0.10639
MARRIED	MARRIED	0.94900	0.10321	0.04764	-0.02266	0.09386	0.19908	0.16101
AGE05_09	AGE05_09	0.94653	-0.00036	0.00536	0.18021	0.03718	0.07310	0.14797
AGE35_44	AGE35_44	0.94443	0.14815	0.03039	0.04488	0.00889	0.10098	0.21833
AGE10_14	AGE10_14	0.93931	-0.05355	-0.02729	0.18974	0.05958	0.07184	0.15742
ELEMSCND	ELEMSCND	0.93215	-0.05176	-0.00634	0.20218	0.07013	0.05617	0.14024
AGELT_05	AGELT_05	0.92472	0.10577	0.04993	0.18283	-0.01169	0.08967	0.03330
OWNR_OCC	OWNR_OCC	0.91942	-0.03836	-0.03188	-0.04411	0.15938	0.26525	0.13978
AGE15_17	AGE15_17	0.91618	-0.08059	-0.00806	0.18916	0.06704	0.07868	0.17308
DRVALONE	DRVALONE	0.90979	0.36110	0.10588	-0.01659	-0.03014	0.10556	0.06272
AGE45_54	AGE45_54	0.90689	0.02557	-0.00421	0.02590	0.05879	0.19262	0.23894
SOMECOLG	SOMECOLG	0.90231	0.30502	0.02087	-0.01208	-0.08364	0.10683	-0.02106
PRIVWORK	PRIVWORK	0.89241	0.32957	0.22282	0.03358	-0.03282	0.09472	0.01956
COM_1529	COM_1529	0.88462	0.34504	0.18028	-0.00915	-0.07837	0.06583	0.05715
VEHICL_2	VEHICL_2	0.88395	0.37483	0.08307	-0.02139	-0.02518	0.15489	0.12740
MARRNOCH	MARRNOCH	0.87889	0.20879	0.06681	-0.06203	0.07926	0.32939	0.14685
EMPLOYED	EMPLOYED	0.87746	0.33403	0.30846	0.04286	-0.01654	0.11541	0.04161
DETACHED	DETACHED	0.87435	-0.08224	-0.02573	0.00334	0.08992	0.36085	0.18110
INC_5075	INC_5075	0.86861	0.22011	0.02510	-0.13881	-0.11869	0.08921	0.24857
COM_3044	COM_3044	0.86051	0.02105	-0.01778	0.05764	0.24789	-0.05260	-0.04143
MALE	MALE	0.85435	0.22712	0.41108	0.10354	0.00803	0.13551	0.08549
TRADE	TRADE	0.85120	0.32972	0.27156	0.05955	-0.04961	0.07897	-0.02603
INC_3550	INC_3550	0.85101	0.39224	0.04817	-0.02312	0.01507	0.15604	-0.10875

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
WAGE_SAL	WAGE_SAL	0.85021	0.46978	0.12218	0.07327	-0.04686	0.14221	0.06913
SAMECNTY	SAMECNTY	0.84696	0.33490	0.05970	0.14768	-0.10731	0.15411	-0.05825
NATIVE	NATIVE	0.84655	0.19176	0.43689	0.11880	0.00542	0.17564	0.06537
UTILITY	UTILITY	0.84619	0.26859	0.03444	-0.02656	-0.04686	-0.07421	-0.02893
FIRE	FIRE	0.83948	0.28725	0.07052	-0.06423	-0.17239	0.01054	0.07272
ONLYENGL	ONLYENGL	0.83511	0.19632	0.45668	0.10795	0.00410	0.17655	0.07014
WRKINCTY	WRKINCTY	0.83465	0.32290	0.32428	0.04326	-0.15371	0.10707	-0.03115
TECHSALE	TECHSALE	0.82746	0.31000	0.40085	-0.01911	-0.10250	0.07551	-0.00909
AGEGE_03	AGEGE_03	0.82716	0.21610	0.46513	0.11002	-0.01437	0.16882	0.08964
SAMEHOUS	SAMEHOUS	0.82475	-0.07671	0.01506	0.17591	0.20139	0.36596	0.02955
WHITE	WHITE	0.81396	0.20879	0.40661	-0.11129	0.07277	0.15698	0.19373
AGE25_34	AGE25_34	0.80638	0.51015	0.16411	0.08080	-0.06344	0.05441	-0.05638
MANUFACT	MANUFACT	0.80030	0.18603	0.01335	0.10745	0.26212	0.06790	-0.02360
SELF_INC	SELF_INC	0.79122	0.24368	0.10330	-0.01802	0.03753	0.28338	0.25059
SERVICES	SERVICES	0.78980	0.32118	0.33491	0.04792	-0.12243	0.05814	-0.04572
BORN_INS	BORN_INS	0.78846	0.06134	0.30338	0.25885	0.17953	0.29254	-0.20074
EXECPROF	EXECPROF	0.78838	0.37652	0.22888	-0.08940	-0.15580	0.07571	0.32414
WORKERS	WORKERS	0.78493	0.28885	0.49147	0.07252	0.00210	0.14733	0.07096
PREPRIMA	PREPRIMA	0.78285	0.10102	0.08479	-0.01367	-0.09503	-0.01639	0.39066
CARPOOL	CARPOOL	0.77799	0.24626	0.11631	0.24622	0.19280	0.13755	-0.18375
COLGGRAD	COLGGRAD	0.77704	0.38646	0.14645	-0.07718	-0.18980	0.06927	0.36159
AGEGE_16	AGEGE_16	0.76777	0.25494	0.54164	0.09016	-0.02476	0.18080	0.07490
BORNSOUT	BORNSOUT	0.76227	0.24006	0.48796	-0.02852	-0.16261	0.04439	0.18525
AGE55_64	AGE55_64	0.75404	0.08019	-0.00638	0.07779	0.06815	0.52185	0.09844
INTE_INC	INTE_INC	0.74884	0.43797	0.14228	-0.09408	-0.11189	0.20709	0.36314

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
HIGHSCHL	HIGHSCHL	0.74711	-0.01318	-0.09763	0.14515	0.25556	0.33605	-0.32410
ROOM_4_6	ROOM_4_6	0.74543	0.51646	0.09771	0.17002	0.06465	0.26156	-0.16788
VAL_1C2C	VAL_1C2C	0.74439	0.01405	0.02239	-0.15272	-0.12954	0.04613	0.50478
ELECTRIC	ELECTRIC	0.74386	0.57945	0.12742	0.06189	-0.05133	-0.01317	0.05466
BLTAFT84	BLTAFT84	0.74119	0.39055	0.03759	-0.09320	0.04550	-0.11581	0.17015
SERVICE	SERVICE	0.73348	0.19689	0.24530	0.27135	0.25298	0.17775	-0.27262
VAL_501C	VAL_501C	0.72530	0.07485	-0.02349	-0.01817	-0.12861	0.35328	-0.33078
SELFWORK	SELFWORK	0.72313	0.08290	0.05370	0.02911	0.31456	0.23658	0.14902
BLT_8084	BLT_8084	0.72136	0.51681	0.05538	0.02561	-0.00076	-0.14818	0.09958
BLT_7079	BLT_7079	0.71639	0.39811	0.08729	0.11713	-0.02246	0.05582	0.06202
BORNMIDW	BORNMIDW	0.69902	0.28241	0.40615	-0.08568	-0.14865	-0.01668	0.36518
INC_2535	INC_2535	0.68183	0.55095	0.10113	0.09774	-0.01227	0.23797	-0.19575
PRUNIT12	PRUNIT12	0.67289	0.64764	0.14227	0.09325	-0.08765	0.26544	0.07817
BORNNORT	BORNNORT	0.65108	0.28027	0.44836	-0.02567	-0.18067	0.02647	0.40297
PRIMARY_	PRIMARY_	0.64909	0.07165	-0.01297	0.11462	0.48504	0.13103	-0.21015
SEPARATE	SEPARATE	0.64460	0.41160	0.06390	0.43594	-0.13576	0.20327	-0.19502
PUBLWORK	PUBLWORK	0.63554	0.31710	0.53407	0.06240	-0.04904	0.11337	0.06571
PUBWATER	PUBWATER	0.62252	0.59844	0.15965	0.14982	-0.32398	0.19063	0.12773
INC_1525	INC_1525	0.61870	0.61115	0.14516	0.21843	-0.01483	0.24296	-0.16025
BORNWEST	BORNWEST	0.61300	0.28798	0.48752	-0.05336	-0.12945	0.00614	0.29157
SINGWICH	SINGWICH	0.60380	0.21778	0.03337	0.59276	-0.17540	0.16248	-0.17770
PROFSERV	PROFSERV	0.59635	0.33941	0.53349	0.00971	-0.14598	0.21722	0.21997
NATURAL_	NATURAL_	0.58989	0.38399	0.37172	-0.03970	-0.15060	0.01924	0.31516
PUBL_GAS	PUBL_GAS	0.57885	0.29708	0.10957	0.07411	-0.43881	0.32065	0.23005
OTH_STAT	OTH_STAT	0.56583	0.26804	0.55603	-0.03113	-0.15576	-0.06632	0.28265

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
WRKEXCTY	WRKEXCTY	0.55795	0.20773	0.09452	0.01305	0.37759	0.08273	0.22383
HISPANIC	HISPANIC	0.51212	0.37299	0.44294	0.09484	-0.02770	-0.01224	0.08453
WRK_HOME	WRK_HOME	0.51150	0.01276	0.40773	-0.05106	0.09104	0.23314	0.29529
UNEMPLOY	UNEMPLOY	0.49331	0.19880	0.44730	0.38170	-0.13037	0.17081	-0.02598
AMIND	AMIND	0.47725	0.19911	0.47249	0.10700	0.01152	0.04200	-0.15477
ARMDFORC	ARMDFORC	0.42780	0.30982	0.05653	-0.01054	-0.02341	0.10680	-0.10446
MEDYRBLT	MEDYRBLT	0.15519	-0.04127	0.01457	0.03455	0.02488	0.15129	0.03728
SHELTERS	SHELTERS	-0.13037	0.03134	0.02108	0.07876	-0.07106	-0.11300	0.06562
APARTMNT	APARTMNT	0.37199	0.82054	0.20126	0.14784	-0.24885	-0.00393	0.03650
ROOM_1_3	ROOM_1_3	0.34543	0.79708	0.19557	0.20845	-0.16751	0.02521	0.10429
RNT_5075	RNT_5075	0.41764	0.76840	0.14523	-0.07637	-0.14093	-0.03140	0.13584
RNT_2550	RNT_2550	0.41486	0.76031	0.21415	0.15305	-0.26792	0.09050	-0.07498
ONEPERHH	ONEPERHH	0.52202	0.74814	0.12719	0.13760	-0.14181	0.25627	0.06653
VEHICL_1	VEHICL_1	0.57632	0.70210	0.17429	0.14756	-0.15844	0.23669	-0.01525
SEASONAL	SEASONAL	0.28810	0.66673	0.08708	-0.02015	0.23079	0.09330	0.11871
PUBSEWER	PUBSEWER	0.53924	0.64773	0.17190	0.17034	-0.35346	0.21173	0.09487
ATTACHED	ATTACHED	0.47357	0.55442	0.10695	-0.02709	-0.18843	0.02190	0.16546
NURSHOME	NURSHOME	-0.00131	0.19102	-0.00095	0.05323	-0.05726	0.18655	0.11268
COLLEGE	COLLEGE	0.08481	0.04558	0.98236	0.00397	-0.06745	-0.01651	0.01344
AGEIS_20	AGEIS_20	0.03700	-0.03889	0.97997	0.01101	-0.04807	-0.03088	-0.04929
AGE18_19	AGE18_19	0.01704	-0.17375	0.95425	-0.00851	-0.03713	-0.02610	-0.05253
DORMITOR	DORMITOR	-0.08817	-0.19435	0.94339	-0.03612	-0.04585	-0.04483	-0.04453
SINGLE_	SINGLE_	0.31460	0.23354	0.89889	0.10215	-0.10739	0.01571	0.00186
SAMESTAT	SAMESTAT	0.24508	0.16987	0.85857	0.01604	-0.00258	0.02179	-0.02210
AGE21_24	AGE21_24	0.33986	0.46339	0.74503	0.10897	-0.07067	-0.02028	-0.01773

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
ASIAN	ASIAN	0.35280	0.27914	0.67441	0.01652	-0.18316	-0.07577	0.25810
ABLEENGL	ABLEENGL	0.50012	0.37390	0.59727	0.08399	-0.14372	0.02856	0.23482
COM_LT15	COM_LT15	0.55554	0.42302	0.57272	0.05685	-0.17690	0.23029	0.05439
PUBTRANS	PUBTRANS	0.10535	0.28209	0.50627	0.42776	-0.22500	0.03095	0.11347
OTH_HEAT	OTH_HEAT	0.04958	-0.00129	0.39097	0.20685	-0.00716	0.05205	0.22807
O_NOINST	O_NOINST	0.01903	0.17733	0.21887	0.13531	-0.00735	-0.00543	0.03058
INSTREET	INSTREET	-0.10270	0.08485	0.17703	0.04775	0.00217	-0.02143	0.13611
MENTAL	MENTAL	-0.03213	0.01108	0.10806	-0.01180	-0.02230	-0.00890	-0.04379
RNT_LT25	RNT_LT25	-0.07363	-0.02434	0.04143	0.79065	-0.00325	0.16669	0.01301
CHILDPOV	CHILDPOV	0.20863	0.03368	0.03750	0.79045	-0.03906	-0.05548	-0.06948
PUBL_INC	PUBL_INC	0.11983	-0.04301	-0.03611	0.78053	0.06832	0.27445	-0.09957
NOVEHICL	NOVEHICL	0.04248	0.19560	0.01534	0.77760	-0.16133	0.29364	0.10011
INPOVRTY	INPOVRTY	0.20888	0.43984	0.22128	0.67817	0.03962	0.04510	0.04028
BLACK	BLACK	0.35078	0.05690	0.19672	0.62327	-0.20914	0.10612	-0.27208
INC_LT15	INC_LT15	0.27133	0.51959	0.19048	0.60317	0.08744	0.30628	0.00106
NODIPLOM	NODIPLOM	0.34727	-0.12563	-0.10517	0.52477	0.39859	0.36405	-0.33207
VAL_2550	VAL_2550	0.08571	-0.13990	-0.10261	0.44071	0.34135	0.40288	-0.37021
INCPRCAP	INCPRCAP	0.35314	0.06648	-0.18303	-0.47946	-0.24925	0.19683	0.40835
INC_MEDN	INC_MEDN	0.49176	-0.23320	-0.14227	-0.53974	-0.14440	0.02942	0.29767
LAND_KM	LAND_KM	0.02498	-0.07262	-0.07204	-0.01260	0.81632	0.00669	-0.00242
BOTL_GAS	BOTL_GAS	0.23471	-0.19684	-0.12407	0.13510	0.75060	0.05610	-0.14208
COALWOOD	COALWOOD	0.13705	-0.16523	-0.04866	0.04054	0.74618	0.07140	-0.06487
FARM_INC	FARM_INC	0.12976	0.10571	-0.02217	0.08816	0.68601	0.04301	0.04733
VAL_LT25	VAL_LT25	-0.07565	-0.10673	-0.09434	0.32303	0.59984	0.13701	-0.11706
F_MIGRNT	F_MIGRNT	-0.07326	-0.00560	-0.04530	0.10971	0.47099	-0.04812	0.04707

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
WATER_KM	WATER_KM	-0.03686	0.01592	-0.01859	-0.05315	0.24688	0.00367	0.07753
CORRINST	CORRINST	-0.02970	-0.02485	0.01849	0.03757	-0.07131	0.00961	-0.05836
RNT_MEDI	RNT_MEDI	0.35930	0.08352	-0.04444	-0.38558	-0.43924	0.04428	0.32230
houseden		-0.06881	0.29995	0.04017	0.40946	-0.64175	-0.00190	0.15663
popden		-0.06459	0.13792	0.38449	0.40010	-0.64272	-0.04076	0.11641
HOUSHOLD	HOUSHOLD	-0.06695	0.29383	0.04052	0.40235	-0.64406	0.00648	0.15733
FAMILIES	FAMILIES	0.05918	0.17163	0.03017	0.37478	-0.71100	0.01440	0.12392
AGE65_74	AGE65_74	0.46396	0.16561	-0.01093	0.16196	0.05294	0.79363	0.08892
AGE75_84	AGE75_84	0.19576	0.22378	-0.01445	0.22084	0.05476	0.77020	0.25190
SOCS_INC	SOCS_INC	0.41547	0.16536	-0.03823	0.27580	0.06736	0.75401	0.13797
WIDOWED	WIDOWED	0.32599	0.20111	-0.03511	0.35103	0.04288	0.71616	0.07569
RETI_INC	RETI_INC	0.55895	0.20994	0.02699	0.03153	-0.12623	0.66192	0.16287
FUELKERO	FUELKERO	0.23077	-0.06139	-0.06301	0.13881	0.38851	0.62166	-0.24998
BLTBFR70	BLTBFR70	0.32888	0.22816	0.18943	0.29615	-0.03335	0.59410	-0.01639
DUPLEX	DUPLEX	0.11258	0.33633	0.16720	0.33602	-0.32234	0.35290	0.07994
O_INSTIT	O_INSTIT	-0.00669	0.00447	0.00921	-0.05734	-0.09402	0.11032	0.04154
VAL_2C3C	VAL_2C3C	0.46795	-0.14362	0.03103	-0.10883	-0.03433	-0.00064	0.65381
VAL_GT3C	VAL_GT3C	0.24772	-0.02014	0.04157	-0.14216	-0.07834	0.16003	0.60847
RNT_751K	RNT_751K	0.32573	0.29478	0.10202	-0.00478	-0.07336	0.10386	0.56904
RNT_GT1K	RNT_GT1K	0.08764	0.12649	0.00690	0.02943	0.03051	0.14061	0.54380
VAL_MEDI	VAL_MEDI	0.35145	-0.07705	0.21418	-0.33587	-0.23447	0.08191	0.49813

Correlations between principal components from ptcdemo1 (end in 2) and ptcdemo2 (end in 3):

Correlation Matrix

	Factor 13	Factor 23	Factor 33	Factor 43	Factor 53	Factor 63	Factor 73
Factor 12	0.9977	-0.0002	0.0166	-0.0272	0.0079	-0.0495	0.0406
Factor 22	0.0084	0.9950	0.0598	0.0091	0.0047	0.0203	0.0197
Factor 32	-0.0121	-0.0518	0.9964	-0.0078	-0.0078	0.0178	-0.0205
Factor 42	0.0348	-0.0150	0.0100	0.9941	-0.0215	0.0067	0.0012
Factor 52	0.0020	-0.0158	0.0091	0.0109	0.9955	-0.0077	-0.0257
Factor 62	0.0473	-0.0062	-0.0310	0.0062	0.0101	0.9898	-0.0327
Factor 72	-0.0188	-0.0305	0.0202	-0.0265	0.0073	0.0459	0.9862

APPENDIX F: Details of Model Selection

Frequency tables for all class variables of station data

The FREQ Procedure

rte_class	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	83	2.42	83	2.42
2	383	11.16	466	13.58
3	322	9.39	788	22.97
4	2444	71.23	3232	94.20
5	199	5.80	3431	100.00

sips_new	Frequency	Percent	Cumulative Frequency	Cumulative Percent
18	305	8.89	305	8.89
31	39	1.14	344	10.03
50	489	14.25	833	24.28
67	196	5.71	1029	29.99
91	602	17.55	1631	47.54
101	21	0.61	1652	48.15
103	849	24.74	2501	72.89
108	930	27.11	3431	100.00

co_urb	Frequency	Percent	Cumulative Frequency	Cumulative Percent
180000	305	8.89	305	8.89
180103	15	0.44	320	9.33
310000	26	0.76	346	10.08
310103	683	19.91	1029	29.99
500000	489	14.25	1518	44.24
670000	196	5.71	1714	49.96
670101	21	0.61	1735	50.57

co_urb	Frequency	Percent	Cumulative Frequency	Cumulative Percent
670103	164	4.78	1899	55.35
910108	1532	44.65	3431	100.00

lanes_rte	Frequency	Percent	Cumulative Frequency	Cumulative Percent
202	135	3.93	135	3.93
203	244	7.11	379	11.05
204	2247	65.49	2626	76.54
205	199	5.80	2825	82.34
302	131	3.82	2956	86.16
303	32	0.93	2988	87.09
304	23	0.67	3011	87.76
401	48	1.40	3059	89.16
402	98	2.86	3157	92.01
403	46	1.34	3203	93.35
404	160	4.66	3363	98.02
502	2	0.06	3365	98.08
504	3	0.09	3368	98.16
601	22	0.64	3390	98.81
602	16	0.47	3406	99.27
604	11	0.32	3417	99.59
801	13	0.38	3430	99.97
802	1	0.03	3431	100.00

speed_rte	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2002	5	0.15	5	0.15
2003	1	0.03	6	0.17
2004	5	0.15	11	0.32

speed_rte	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2502	5	0.15	16	0.47
2503	2	0.06	18	0.52
2504	29	0.85	47	1.37
2505	199	5.80	246	7.17
3004	5	0.15	251	7.32
3502	83	2.42	334	9.73
3503	74	2.16	408	11.89
3504	555	16.18	963	28.07
4002	93	2.71	1056	30.78
4003	29	0.85	1085	31.62
4004	253	7.37	1338	39.00
4502	65	1.89	1403	40.89
4503	41	1.19	1444	42.09
4504	286	8.34	1730	50.42
5004	1	0.03	1731	50.45
5501	17	0.50	1748	50.95
5502	128	3.73	1876	54.68
5503	175	5.10	2051	59.78
5504	1310	38.18	3361	97.96
6001	10	0.29	3371	98.25
6501	48	1.40	3419	99.65
6502	4	0.12	3423	99.77
7001	8	0.23	3431	100.00

lanes_speed_rte	Frequency	Percent	Cumulative Frequency	Cumulative Percent
202002	5	0.15	5	0.15
202003	1	0.03	6	0.17

lanes_speed_rte	Frequency	Percent	Cumulative Frequency	Cumulative Percent
202502	2	0.06	8	0.23
202503	2	0.06	10	0.29
202504	17	0.50	27	0.79
202505	199	5.80	226	6.59
203502	45	1.31	271	7.90
203503	46	1.34	317	9.24
203504	448	13.06	765	22.30
204004	253	7.37	1018	29.67
204502	26	0.76	1044	30.43
204503	25	0.73	1069	31.16
204504	238	6.94	1307	38.09
205004	1	0.03	1308	38.12
205502	57	1.66	1365	39.78
205503	170	4.95	1535	44.74
205504	1290	37.60	2825	82.34
302502	1	0.03	2826	82.37
302504	4	0.12	2830	82.48
303004	5	0.15	2835	82.63
303502	8	0.23	2843	82.86
303503	3	0.09	2846	82.95
303504	12	0.35	2858	83.30
304002	88	2.56	2946	85.86
304003	29	0.85	2975	86.71
304504	1	0.03	2976	86.74
305502	34	0.99	3010	87.73
305504	1	0.03	3011	87.76
402004	5	0.15	3016	87.90

lanes_speed_rte	Frequency	Percent	Cumulative Frequency	Cumulative Percent
402502	2	0.06	3018	87.96
402504	8	0.23	3026	88.20
403502	30	0.87	3056	89.07
403503	25	0.73	3081	89.80
403504	87	2.54	3168	92.33
404002	5	0.15	3173	92.48
404502	22	0.64	3195	93.12
404503	16	0.47	3211	93.59
404504	41	1.19	3252	94.78
405501	12	0.35	3264	95.13
405502	35	1.02	3299	96.15
405503	5	0.15	3304	96.30
405504	19	0.55	3323	96.85
406501	28	0.82	3351	97.67
406502	4	0.12	3355	97.78
407001	8	0.23	3363	98.02
503504	2	0.06	3365	98.08
504502	2	0.06	3367	98.13
504504	1	0.03	3368	98.16
603504	6	0.17	3374	98.34
604502	14	0.41	3388	98.75
604504	5	0.15	3393	98.89
605501	5	0.15	3398	99.04
605502	2	0.06	3400	99.10
606001	4	0.12	3404	99.21
606501	13	0.38	3417	99.59
804502	1	0.03	3418	99.62

lanes_speed_rte	Frequency	Percent	Cumulative Frequency	Cumulative Percent
806001	6	0.17	3424	99.80
806501	7	0.20	3431	100.00

cycle_surface	Frequency	Percent	Cumulative Frequency	Cumulative Percent
101	248	7.23	248	7.23
202	707	20.61	955	27.83
203	1	0.03	956	27.86
302	15	0.44	971	28.30
303	41	1.19	1012	29.50
401	650	18.94	1662	48.44
402	1636	47.68	3298	96.12
403	133	3.88	3431	100.00

A_CONTROL				
A_CONTROL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	2663	77.62	2663	77.62
2	685	19.97	3348	97.58
3	83	2.42	3431	100.00

YEAR				
YEAR	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1997	74	2.16	74	2.16
1998	1009	29.41	1083	31.57
1999	2067	60.24	3150	91.81
2000	206	6.00	3356	97.81
2001	75	2.19	3431	100.00

MONTH

MONTH	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	102	2.97	102	2.97
1	102	2.97	204	5.95
2	107	3.12	311	9.06
3	686	19.99	997	29.06
4	868	25.30	1865	54.36
5	627	18.27	2492	72.63
6	242	7.05	2734	79.69
7	21	0.61	2755	80.30
8	63	1.84	2818	82.13
9	321	9.36	3139	91.49
10	100	2.91	3239	94.40
11	55	1.60	3294	96.01
12	137	3.99	3431	100.00

DAY				
DAY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	102	2.97	102	2.97
2	1526	44.48	1628	47.45
3	1308	38.12	2936	85.57
4	439	12.80	3375	98.37
5	56	1.63	3431	100.00

LUCODE				
LUCODE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
11	672	19.59	672	19.59
12	341	9.94	1013	29.52
13	28	0.82	1041	30.34

LUCODE				
LUCODE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
14	197	5.74	1238	36.08
15	29	0.85	1267	36.93
16	333	9.71	1600	46.63
17	28	0.82	1628	47.45
21	757	22.06	2385	69.51
23	2	0.06	2387	69.57
24	9	0.26	2396	69.83
32	3	0.09	2399	69.92
41	281	8.19	2680	78.11
42	147	4.28	2827	82.40
43	566	16.50	3393	98.89
53	5	0.15	3398	99.04
61	1	0.03	3399	99.07
75	2	0.06	3401	99.13
76	30	0.87	3431	100.00

city_urb	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1546	45.06	1546	45.06
1	322	9.39	1868	54.44
2	12	0.35	1880	54.79
100	260	7.58	2140	62.37
101	1255	36.58	3395	98.95
102	36	1.05	3431	100.00

Box-Cox regression using full model (PCA of census + all station data)
The TRANSREG Procedure

Transformation Information for BoxCox(AADT)				
Lambda		R-Square	Log Like	
-2.000		0.12	-45919.9	
-1.000		0.35	-34083.8	
-0.500		0.56	-29753.0	
-0.450		0.59	-29412.0	
-0.400		0.61	-29091.0	
-0.350		0.63	-28790.8	
-0.300		0.64	-28512.2	
-0.250		0.66	-28256.3	
-0.200		0.68	-28023.7	
-0.150		0.70	-27815.4	
-0.100		0.71	-27631.9	
-0.050		0.73	-27474.1	
0.000		0.74	-27342.3	
0.050		0.75	-27237.1	
0.100		0.76	-27158.6	
0.150		0.77	-27107.1	
0.200		0.78	-27082.4	<
0.250		0.79	-27084.6	*
0.300		0.79	-27113.3	
0.350		0.80	-27168.3	
0.400		0.80	-27249.0	
0.450		0.81	-27354.9	
0.500		0.81	-27485.5	
1.000		0.82	-29989.6	
2.000		0.76	-39019.0	
< - Best Lambda * - Confidence Interval				

Transformation Information for BoxCox(AADT)				
Lambda		R-Square	Log Like	
+ - Convenient Lambda				

aadt^.2 regression using full model (PCA of census + all station data)
The REG Procedure
Model: MODEL1
Dependent Variable: baadt

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	121	6689.59207	55.28588	96.19	<.0001
Error	3309	1901.83650	0.57475		
Corrected Total	3430	8591.42857			

Root MSE	0.75812	R-Square	0.7786
Dependent Mean	5.39260	Adj R-Sq	0.7705
Coeff Var	14.05854		

NOTE: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.

NOTE: The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

lanes_rte4 =	Intercept - rte_class1 - rte_class2 - rte_class3 - rte_class4
lanes_rte10 =	rte_class3 - lanes_rte2 - lanes_rte6
lanes_rte16 =	rte_class4 - lanes_rte3 - lanes_rte7 - lanes_rte11 - lanes_rte13
lanes_rte17 =	rte_class1 - lanes_rte8 - lanes_rte14
speed_rte7 =	Intercept - rte_class1 - rte_class2 - rte_class3 - rte_class4
speed_rte21 =	rte_class3 - speed_rte2 - speed_rte5 - speed_rte10 - speed_rte13 - speed_rte16
speed_rte22 =	rte_class4 - speed_rte3 - speed_rte6 - speed_rte8 - speed_rte11 - speed_rte14 - speed_rte17 - speed_rte18

speed_rte25 =	rte_class2 - speed_rte1 - speed_rte4 - speed_rte9 - speed_rte12 - speed_rte15 - speed_rte20
lanes_speed_rte1 =	speed_rte1
lanes_speed_rte2 =	speed_rte2
lanes_speed_rte4 =	speed_rte5
lanes_speed_rte6 =	Intercept - rte_class1 - rte_class2 - rte_class3 - rte_class4
lanes_speed_rte10 =	speed_rte14
lanes_speed_rte14 =	speed_rte18
lanes_speed_rte15 =	lanes_rte1 - speed_rte1 - lanes_speed_rte3 - lanes_speed_rte7 - lanes_speed_rte11
lanes_speed_rte16 =	lanes_rte2 - speed_rte2 - speed_rte5 - lanes_speed_rte8 - lanes_speed_rte12
lanes_speed_rte17 =	lanes_rte3 - speed_rte14 - speed_rte18 - lanes_speed_rte5 - lanes_speed_rte9 - lanes_speed_rte13
lanes_speed_rte20 =	speed_rte8
lanes_speed_rte22 =	lanes_rte6 - speed_rte13
lanes_speed_rte25 =	speed_rte13
lanes_speed_rte27 =	lanes_rte5 - lanes_speed_rte18 - lanes_speed_rte21 - lanes_speed_rte24
lanes_speed_rte28 =	lanes_rte7 - speed_rte8 - lanes_speed_rte19 - lanes_speed_rte23 - lanes_speed_rte26
lanes_speed_rte29 =	speed_rte3
lanes_speed_rte30 =	speed_rte4 - lanes_speed_rte3 - lanes_speed_rte18
lanes_speed_rte31 =	speed_rte6 - lanes_speed_rte5 - lanes_speed_rte19
lanes_speed_rte32 =	speed_rte9 - lanes_speed_rte7 - lanes_speed_rte21
lanes_speed_rte33 =	-lanes_rte6 + speed_rte10 + speed_rte13 - lanes_speed_rte8
lanes_speed_rte35 =	speed_rte12 - lanes_speed_rte24
lanes_speed_rte37	speed_rte16 - lanes_speed_rte12

=	
lanes_speed_rte38 =	-rte_class4 + lanes_rte3 + lanes_rte7 + lanes_rte11 + speed_rte11 + speed_rte17 - lanes_speed_rte9 - lanes_speed_rte13 - lanes_speed_rte23 - lanes_speed_rte26 - lanes_speed_rte34
lanes_speed_rte40 =	-rte_class2 + lanes_rte9 + speed_rte1 + speed_rte15 + speed_rte20 + lanes_speed_rte3 + lanes_speed_rte7 + lanes_speed_rte18 + lanes_speed_rte21 + lanes_speed_rte24 - lanes_speed_rte36
lanes_speed_rte41 =	rte_class3 - lanes_rte2 - speed_rte10 - speed_rte13 - speed_rte16 + lanes_speed_rte8 + lanes_speed_rte12
lanes_speed_rte42 =	rte_class4 - lanes_rte3 - lanes_rte7 - speed_rte3 - speed_rte6 - speed_rte11 - speed_rte17 + lanes_speed_rte5 + lanes_speed_rte9 + lanes_speed_rte13 + lanes_speed_rte19 + lanes_speed_rte23 + lanes_speed_rte26
lanes_speed_rte43 =	-rte_class1 + lanes_rte8 + speed_rte19 + speed_rte23 + speed_rte24 - lanes_speed_rte39
lanes_speed_rte44 =	rte_class2 - speed_rte1 - speed_rte4 - speed_rte9 - speed_rte12 - speed_rte15 - speed_rte20
lanes_speed_rte45 =	rte_class1 - speed_rte19 - speed_rte23 - speed_rte24
lanes_speed_rte47 =	lanes_rte12
lanes_speed_rte48 =	lanes_rte13 - lanes_speed_rte46
lanes_speed_rte49 =	speed_rte11 - lanes_speed_rte9 - lanes_speed_rte23 - lanes_speed_rte34 - lanes_speed_rte46
lanes_speed_rte50 =	-rte_class2 + lanes_rte1 + lanes_rte5 + lanes_rte9 + lanes_rte15 + speed_rte15 - lanes_speed_rte11 - lanes_speed_rte36
lanes_speed_rte51 =	rte_class4 - lanes_rte3 - lanes_rte7 - lanes_rte11 - lanes_rte13 - speed_rte11 + lanes_speed_rte9 + lanes_speed_rte23 + lanes_speed_rte34 + lanes_speed_rte46
lanes_speed_rte52 =	speed_rte19 - lanes_speed_rte39
lanes_speed_rte53 =	rte_class2 - lanes_rte1 - lanes_rte5 - lanes_rte9 - speed_rte15 + lanes_speed_rte11 + lanes_speed_rte36
lanes_speed_rte55 =	lanes_rte14 - speed_rte19 + lanes_speed_rte39 - lanes_speed_rte54
lanes_speed_rte56 =	rte_class2 - lanes_rte1 - lanes_rte5 - lanes_rte9 - lanes_rte12 - lanes_rte15
lanes_speed_rte57 =	speed_rte23 - lanes_speed_rte54
a_control2 =	Intercept - rte_class1 - a_control1
day4 =	month1 + month2 + month3 + month4 + month5 + month6 + month7 + month8 + month9 + month10 + month11 + month12 - day1 - day2 - day3

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	B	3.92169	0.35579	11.02	<.0001
Factor1	Other	1	0.17836	0.01506	11.85	<.0001
Factor2	Single	1	0.06343	0.01479	4.29	<.0001
Factor3	College	1	0.02982	0.01547	1.93	0.0539
Factor4	Poor	1	-0.01154	0.01433	-0.81	0.4206
Factor5	Farm	1	-0.21999	0.02575	-8.54	<.0001
Factor6	Elderly	1	-0.00696	0.01426	-0.49	0.6254
Factor7	Wealthy	1	0.02917	0.01449	2.01	0.0442
rte_class1	INTER	B	4.76976	0.49191	9.70	<.0001
rte_class2	US	B	2.22528	1.06897	2.08	0.0374
rte_class3	NC	B	3.44479	0.40985	8.41	<.0001
rte_class4	SR	B	1.86685	0.40941	4.56	<.0001
co_urb1	Chatham	1	0.06233	0.12485	0.50	0.6176
co_urb2	Chatham, 103	1	-0.01347	0.21583	-0.06	0.9502
co_urb3	Durham	1	-0.00722	0.19436	-0.04	0.9704
co_urb4	Durham, 103	1	-0.36950	0.07276	-5.08	<.0001
co_urb5	Johnston	1	-0.08652	0.12182	-0.71	0.4776
co_urb6	Orange	1	0.27779	0.12688	2.19	0.0286
co_urb7	Orange, 101	1	-0.09197	0.22039	-0.42	0.6765
co_urb8	Orange, 103	1	-0.15766	0.09932	-1.59	0.1125
lanes_rte1	2, US	B	-0.79818	0.97447	-0.82	0.4128
lanes_rte2	2, NC	B	-1.88018	0.35219	-5.34	<.0001
lanes_rte3	2, SR	B	-1.48380	0.40739	-3.64	0.0003
lanes_rte4	2, local	0	0	.	.	.
lanes_rte5	3, US	B	0.03566	0.97500	0.04	0.9708
lanes_rte6	3, NC	B	0.10546	0.46606	0.23	0.8210

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lanes_rte7	3, SR	B	0.41722	0.86261	0.48	0.6286
lanes_rte8	4, INTER	B	-0.66178	0.32710	-2.02	0.0431
lanes_rte9	4, US	B	0.00903	0.97511	0.01	0.9926
lanes_rte10	4, NC	0	0	.	.	.
lanes_rte11	4, SR	B	-0.44593	0.36406	-1.22	0.2207
lanes_rte12	5, US	B	-0.07890	0.93779	-0.08	0.9330
lanes_rte13	5, SR	B	-0.38609	0.83582	-0.46	0.6442
lanes_rte14	6, INTER	B	-0.09173	0.35864	-0.26	0.7982
lanes_rte15	6, US	B	-0.59474	0.79171	-0.75	0.4526
lanes_rte16	6, SR	0	0	.	.	.
lanes_rte17	8, INTER	0	0	.	.	.
speed_rte1	20, US	B	0.59694	0.54669	1.09	0.2750
speed_rte2		B	-0.17501	0.76690	-0.23	0.8195
speed_rte3		B	0.25083	0.39116	0.64	0.5214
speed_rte4		B	-0.06716	0.67011	-0.10	0.9202
speed_rte5		B	0.24195	0.56229	0.43	0.6670
speed_rte6	25, SR	B	-1.82777	0.32683	-5.59	<.0001
speed_rte7		0	0	.	.	.
speed_rte8		B	-1.28615	0.83779	-1.54	0.1248
speed_rte9		B	0.12896	0.41097	0.31	0.7537
speed_rte10		B	-1.60835	0.37673	-4.27	<.0001
speed_rte11	35, SR	B	-0.35640	0.51114	-0.70	0.4857
speed_rte12		B	0.34031	0.51907	0.66	0.5121
speed_rte13		B	-1.39698	0.60024	-2.33	0.0200
speed_rte14		B	0.31450	0.07614	4.13	<.0001
speed_rte15		B	1.21429	0.71738	1.69	0.0906

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
speed_rte16		B	-1.25278	0.39210	-3.20	0.0014
speed_rte17		B	0.26170	0.21383	1.22	0.2211
speed_rte18		B	0.04770	0.79190	0.06	0.9520
speed_rte19	55, INTER	B	-1.57292	0.50733	-3.10	0.0019
speed_rte20	55, US	B	0.68355	0.40580	1.68	0.0922
speed_rte21		0	0	.	.	.
speed_rte22		0	0	.	.	.
speed_rte23	60, INTER	B	-1.15810	0.53061	-2.18	0.0291
speed_rte24	65, INTER	B	-0.01998	0.30987	-0.06	0.9486
speed_rte25		0	0	.	.	.
lanes_speed_rte1		0	0	.	.	.
lanes_speed_rte2		0	0	.	.	.
lanes_speed_rte3		B	0.48714	0.78585	0.62	0.5354
lanes_speed_rte4		0	0	.	.	.
lanes_speed_rte5	2, 25, SR	B	1.36446	0.37636	3.63	0.0003
lanes_speed_rte6		0	0	.	.	.
lanes_speed_rte7		B	-0.01722	0.24787	-0.07	0.9446
lanes_speed_rte8		B	1.63373	0.39955	4.09	<.0001
lanes_speed_rte9		B	0.60198	0.51395	1.17	0.2416
lanes_speed_rte10		0	0	.	.	.
lanes_speed_rte11		B	-1.09520	0.61915	-1.77	0.0770
lanes_speed_rte12		B	1.36927	0.42669	3.21	0.0013
lanes_speed_rte13		B	0.14075	0.22147	0.64	0.5251
lanes_speed_rte14		0	0	.	.	.
lanes_speed_rte15		0	0	.	.	.
lanes_speed_rte16		0	0	.	.	.

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lanes_speed_rte17	2, 55, SR	0	0	.	.	.
lanes_speed_rte18		B	-0.10341	0.95394	-0.11	0.9137
lanes_speed_rte19		B	-0.28995	0.90933	-0.32	0.7499
lanes_speed_rte20	3, 30, SR	0	0	.	.	.
lanes_speed_rte21	3, 35, US	B	-0.58200	0.35929	-1.62	0.1054
lanes_speed_rte22		0	0	.	.	.
lanes_speed_rte23	3, 35, SR	B	-0.86443	0.94419	-0.92	0.3600
lanes_speed_rte24	3, 40, US	B	-0.37434	0.39813	-0.94	0.3472
lanes_speed_rte25		0	0	.	.	.
lanes_speed_rte26		B	-1.25476	1.09782	-1.14	0.2531
lanes_speed_rte27		0	0	.	.	.
lanes_speed_rte28		0	0	.	.	.
lanes_speed_rte29		0	0	.	.	.
lanes_speed_rte30		0	0	.	.	.
lanes_speed_rte31	4, 25, SR	0	0	.	.	.
lanes_speed_rte32		0	0	.	.	.
lanes_speed_rte33	4, 35, NC	0	0	.	.	.
lanes_speed_rte34		B	0.54518	0.48734	1.12	0.2634
lanes_speed_rte35		0	0	.	.	.
lanes_speed_rte36	4, 45, US	B	-0.88925	0.62758	-1.42	0.1566
lanes_speed_rte37		0	0	.	.	.
lanes_speed_rte38		0	0	.	.	.
lanes_speed_rte39		B	0.86404	0.48231	1.79	0.0733
lanes_speed_rte40		0	0	.	.	.
lanes_speed_rte41	4, 55, NC	0	0	.	.	.
lanes_speed_rte42	4, 55, SR	0	0	.	.	.

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lanes_speed_rte43		0	0	.	.	.
lanes_speed_rte44		0	0	.	.	.
lanes_speed_rte45		0	0	.	.	.
lanes_speed_rte46	5, 35, SR	B	1.32216	1.04274	1.27	0.2049
lanes_speed_rte47	5, 45, US	0	0	.	.	.
lanes_speed_rte48		0	0	.	.	.
lanes_speed_rte49	6, 35, SR	0	0	.	.	.
lanes_speed_rte50	6, 45, US	0	0	.	.	.
lanes_speed_rte51		0	0	.	.	.
lanes_speed_rte52		0	0	.	.	.
lanes_speed_rte53	6, 55, NC	0	0	.	.	.
lanes_speed_rte54		B	0.44061	0.60926	0.72	0.4696
lanes_speed_rte55	6, 65, INTER	0	0	.	.	.
lanes_speed_rte56	8, 45, US	0	0	.	.	.
lanes_speed_rte57	8, 60, INTER	0	0	.	.	.
cycle_surface1	Ann, Asph	1	1.11106	0.15554	7.14	<.0001
cycle_surface2	Odd, Conc	1	0.53080	0.14526	3.65	0.0003
cycle_surface3	Odd, Soil	1	1.77038	0.77533	2.28	0.0225
cycle_surface4	Even, Conc	1	1.07701	0.25295	4.26	<.0001
cycle_surface5	Even, Soil	1	0.76001	0.17995	4.22	<.0001
cycle_surface6	Vari, Asph	1	1.48457	0.11328	13.11	<.0001
cycle_surface7	Vari, Conc	1	1.03371	0.08672	11.92	<.0001
a_control1	Low	B	0.45251	0.19618	2.31	0.0211
a_control2	Medium	0	0	.	.	.
year1	1997	1	0.13326	0.14914	0.89	0.3716
year2	1998	1	-0.25164	0.12198	-2.06	0.0392

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
year3	1999	1	-0.10433	0.12166	-0.86	0.3912
year4	2000	1	-0.06677	0.16724	-0.40	0.6897
month1	Jan	B	-0.91403	0.21227	-4.31	<.0001
month2	Feb	B	-0.85600	0.18970	-4.51	<.0001
month3	Mar	B	-0.68553	0.20052	-3.42	0.0006
month4	Apr	B	-0.67931	0.20093	-3.38	0.0007
month5	May	B	-1.03657	0.20016	-5.18	<.0001
month6	June	B	-0.91804	0.20626	-4.45	<.0001
month7	July	B	-0.72313	0.26600	-2.72	0.0066
month8	Aug	B	-0.64075	0.21979	-2.92	0.0036
month9	Sept	B	-0.81423	0.21440	-3.80	0.0001
month10	Oct	B	-0.76752	0.22885	-3.35	0.0008
month11	Nov	B	-0.43390	0.22252	-1.95	0.0513
month12	Dec	B	-0.59589	0.20883	-2.85	0.0044
day1	Mon	B	0.46363	0.13248	3.50	0.0005
day2	Tue	B	0.45966	0.13203	3.48	0.0005
day3	Wed	B	0.41723	0.13721	3.04	0.0024
day4	Thu	0	0	.	.	.
lucode1	Residential	1	-0.22703	0.15132	-1.50	0.1336
lucode2	Comm & Services	1	0.04108	0.15630	0.26	0.7927
lucode3		1	-0.08364	0.20988	-0.40	0.6903
lucode4	Trans, Comm, Util	1	0.28731	0.16074	1.79	0.0740
lucode5	Indust & Comm Complxs	1	0.34144	0.20557	1.66	0.0968
lucode6	Mxd Urban or Built-up	1	0.01510	0.15441	0.10	0.9221
lucode7		1	-0.17108	0.20691	-0.83	0.4084
lucode8		1	-0.13810	0.15265	-0.90	0.3657

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lucode9		1	-0.61624	0.56291	-1.09	0.2737
lucode10	Other Ag Land	1	0.55677	0.29515	1.89	0.0593
lucode11	Shrub & Brush Rangeland	1	-1.11227	0.46554	-2.39	0.0169
lucode12		1	-0.07776	0.15628	-0.50	0.6188
lucode13		1	-0.07307	0.16144	-0.45	0.6509
lucode14		1	-0.12960	0.15183	-0.85	0.3934
lucode15		1	-0.04718	0.37638	-0.13	0.9002
lucode16		1	-0.29095	0.77607	-0.37	0.7078
lucode17		1	-0.78781	0.56136	-1.40	0.1606
city_urb1	out, out	1	-0.61629	0.14741	-4.18	<.0001
city_urb2	out, major	1	0.09680	0.15609	0.62	0.5352
city_urb3	out, minor	1	-0.30770	0.26085	-1.18	0.2382
city_urb4	city, out	1	-0.17757	0.14430	-1.23	0.2186
city_urb5	city, major	1	0.32572	0.14950	2.18	0.0294

aadt^.2 regression using full model (PCA of census + all station data)
Normality check on residuals
The UNIVARIATE Procedure
Variable: raadt (Residual)

Moments			
N	3431	Sum Weights	3431
Mean	0	Sum Observations	0
Std Deviation	0.74462829	Variance	0.55447128
Skewness	-0.1001229	Kurtosis	0.31388923
Uncorrected SS	1901.8365	Corrected SS	1901.8365
Coeff Variation	.	Std Error Mean	0.01271245

Basic Statistical Measures			
Location		Variability	
Mean	0.00000	Std Deviation	0.74463
Median	0.00000	Variance	0.55447
Mode	-0.02520	Range	5.40682
		Interquartile Range	0.98017

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	0.5	Pr >= |M|	1.0000
Signed Rank	S	28532	Pr >= |S|	0.6230

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.01526	Pr > D	0.0507
Cramer-von Mises	W-Sq	0.148241	Pr > W-Sq	0.0249
Anderson-Darling	A-Sq	1.205648	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	2.774732
99%	1.689638
95%	1.188366
90%	0.920294
75% Q3	0.511077
50% Median	0.000000
25% Q1	-0.469092
10%	-0.937391
5%	-1.210478

Quantiles (Definition 5)	
Quantile	Estimate
1%	-1.871888
0% Min	-2.632089

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-2.63209	811	2.46484	2675
-2.51123	805	2.52790	2006
-2.45980	3039	2.68830	2607
-2.44221	2674	2.74248	2039
-2.40749	3117	2.77473	2629

aadt^.2 regression using full model (PCA of census + all station data)
Normality check on residuals
The UNIVARIATE Procedure
Variable: raadt (Residual)

```
                                        Histogram                        #
Boxplot
                    2.75+*                                               4
0
                        .*                                              12
0
                        .****                                           58
|
                        .***********                                   201
|
                        .*******************************               601          +--
---+
                        .*********************************************  842          *--
+--*
                        .************************************************* 896          +--
---+
                        .***************************                    520
|
                        .***********                                   204
|
                        .****                                           69
```

```
0
                    .**                                                22
0
                -2.75+*                                                 2
0
                     ----+----+----+----+----+----+----+----+----+---
                     * may represent up to 19 counts


                                      Normal Probability Plot
                          2.75+                                                 *
                              |                                                 *
                              |                                            ******
                              |                                       *******
                              |                                   ********
                              |                            ********
                              |                       ********
                              |                  ********
                              |       +******
                              |+******
                              |*
                         -2.75+*
                                +----+----+----+----+----+----+----+----+----+----+
                                    -2        -1         0        +1        +2
```

Stepwise Selection for aadt^.2 regression from full model (PCA of census + all station data)
The REG Procedure
Model: MODEL1
Dependent Variable: baadt
Stepwise Selection: Step 66
Variable lanes_rte17 Entered: R-Square = 0.7766 and C(p) = 31.1230

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	60	6671.70425	111.19507	195.20	<.0001
Error	3370	1919.72432	0.56965		
Corrected Total	3430	8591.42857			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	6.23009	0.18230	665.31101	1167.93	<.0001

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Factor1	0.18055	0.01431	90.71052	159.24	<.0001
Factor2	0.06189	0.01426	10.72558	18.83	<.0001
Factor3	0.02196	0.01424	1.35413	2.38	0.1232
Factor5	-0.22093	0.02482	45.13384	79.23	<.0001
Factor7	0.02437	0.01353	1.84684	3.24	0.0719
rte_class1	1.58235	0.19193	38.72101	67.97	<.0001
co_urb4	-0.27598	0.05143	16.40388	28.80	<.0001
co_urb5	-0.10479	0.05705	1.92228	3.37	0.0663
co_urb6	0.23604	0.06945	6.57954	11.55	0.0007
lanes_rte1	-0.72604	0.08761	39.11888	68.67	<.0001
lanes_rte2	-0.68759	0.08106	40.98565	71.95	<.0001
lanes_rte3	-1.40233	0.14900	50.45780	88.58	<.0001
lanes_rte4	-2.14652	0.15592	107.96722	189.53	<.0001
lanes_rte11	-0.54427	0.15121	7.38005	12.96	0.0003
lanes_rte17	0.67292	0.31274	2.63736	4.63	0.0315
speed_rte6	-1.99556	0.22916	43.19720	75.83	<.0001
speed_rte19	-0.78039	0.22330	6.95735	12.21	0.0005
speed_rte20	0.61472	0.08981	26.68783	46.85	<.0001
lanes_speed_rte5	1.16671	0.29628	8.83322	15.51	<.0001
lanes_speed_rte9	-0.12178	0.05291	3.01757	5.30	0.0214
lanes_speed_rte17	-0.36403	0.05036	29.76596	52.25	<.0001
lanes_speed_rte20	-1.12314	0.36648	5.35034	9.39	0.0022
lanes_speed_rte21	-0.43645	0.27367	1.44880	2.54	0.1109
lanes_speed_rte23	-1.07605	0.26215	9.59818	16.85	<.0001
lanes_speed_rte33	-0.41579	0.16305	3.70421	6.50	0.0108
lanes_speed_rte36	0.29614	0.17145	1.69958	2.98	0.0842
lanes_speed_rte41	1.21275	0.34416	7.07350	12.42	0.0004

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
lanes_speed_rte47	1.15271	0.53774	2.61757	4.60	0.0321
lanes_speed_rte49	-0.64290	0.34152	2.01873	3.54	0.0599
lanes_speed_rte50	0.55461	0.21203	3.89770	6.84	0.0089
lanes_speed_rte55	0.58097	0.24362	3.23953	5.69	0.0171
lanes_speed_rte56	1.16814	0.76019	1.34509	2.36	0.1245
lanes_speed_rte57	-1.05003	0.42207	3.52568	6.19	0.0129
cycle_surface1	1.12672	0.12889	43.53141	76.42	<.0001
cycle_surface2	0.66526	0.09340	28.90001	50.73	<.0001
cycle_surface3	1.90623	0.76120	3.57244	6.27	0.0123
cycle_surface4	1.09071	0.21282	14.96246	26.27	<.0001
cycle_surface5	0.78204	0.14952	15.58285	27.36	<.0001
cycle_surface6	1.51836	0.10935	109.82620	192.80	<.0001
cycle_surface7	1.05681	0.08294	92.48408	162.35	<.0001
a_control2	-0.36254	0.13969	3.83698	6.74	0.0095
year2	-0.22280	0.05931	8.03956	14.11	0.0002
year3	-0.15888	0.05608	4.57254	8.03	0.0046
month1	-0.28200	0.08586	6.14560	10.79	0.0010
month2	-0.19963	0.09564	2.48206	4.36	0.0369
month5	-0.33902	0.04747	29.05146	51.00	<.0001
month6	-0.23031	0.06234	7.77501	13.65	0.0002
month11	0.22131	0.10669	2.45105	4.30	0.0381
day4	-0.41573	0.12556	6.24501	10.96	0.0009
lucode1	-0.12207	0.04154	4.91890	8.63	0.0033
lucode2	0.16336	0.05495	5.03468	8.84	0.0030
lucode4	0.40724	0.06583	21.80268	38.27	<.0001
lucode5	0.45578	0.14543	5.59474	9.82	0.0017
lucode6	0.13388	0.04741	4.54256	7.97	0.0048

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
lucode10	0.67339	0.25360	4.01636	7.05	0.0080
lucode11	-0.97617	0.43900	2.81660	4.94	0.0262
city_urb1	-0.69996	0.05688	86.26317	151.43	<.0001
city_urb3	-0.39946	0.22589	1.78144	3.13	0.0771
city_urb4	-0.24726	0.07410	6.34328	11.14	0.0009
city_urb5	0.21557	0.05360	9.21424	16.18	<.0001

Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	lanes_speed_rte17		2, 55, SR	1	0.3778	0.3778	5874.04	2081.93	<.0001
2	cycle_surface6		Vari, Asph	2	0.1315	0.5093	3909.65	919.03	<.0001
3	rte_class1		INTER	3	0.0590	0.5683	3029.48	468.53	<.0001
4	city_urb1		out, out	4	0.0481	0.6164	2312.43	429.67	<.0001
5	cycle_surface2		Odd, Conc	5	0.0191	0.6355	2029.38	179.19	<.0001
6	rte_class2		US	6	0.0134	0.6489	1831.64	130.30	<.0001
7	lanes_rte11		4, SR	7	0.0177	0.6665	1569.73	181.22	<.0001
8	Factor1		Other	8	0.0126	0.6791	1383.82	134.06	<.0001
9	city_urb4		city, out	9	0.0097	0.6888	1241.12	106.41	<.0001
10	day4		Thu	10	0.0068	0.6956	1141.08	76.69	<.0001
11	lanes_rte4		2, local	11	0.0062	0.7018	1051.13	70.52	<.0001
12	lanes_rte3		2, SR	12	0.0071	0.7089	947.004	83.35	<.0001
13	cycle_surface7		Vari, Conc	13	0.0094	0.7182	808.735	113.80	<.0001
14	lanes_rte1		2, US	14	0.0044	0.7227	744.677	54.43	<.0001
15	cycle_surface1		Ann, Asph	15	0.0044	0.7270	681.400	54.63	<.0001
16	Factor5		Farm	16	0.0052	0.7322	606.067	65.95	<.0001
17	speed_rte6		25, SR	17	0.0038	0.7361	550.527	49.77	<.0001

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
18	lanes_rte2		2, NC	18	0.0040	0.7400	493.332	51.97	<.0001
19		lanes_rte11	4, SR	17	0.0000	0.7400	491.715	0.34	0.5619
20	cycle_surface5		Even, Soil	18	0.0035	0.7435	441.417	46.54	<.0001
21	co_urb4		Durham, 103	19	0.0035	0.7470	390.754	47.50	<.0001
22	cycle_surface4		Even, Conc	20	0.0030	0.7500	348.141	40.71	<.0001
23	lucode4		Trans, Comm, Util	21	0.0028	0.7528	308.660	38.26	<.0001
24	month5		May	22	0.0023	0.7551	275.903	32.36	<.0001
25	lucode1		Residential	23	0.0019	0.7570	248.941	27.17	<.0001
26	Factor2		Single	24	0.0019	0.7590	222.199	27.17	<.0001
27	speed_rte20		55, US	25	0.0016	0.7606	200.054	22.97	<.0001
28	co_urb6		Orange	26	0.0016	0.7621	178.468	22.58	<.0001
29	speed_rte19		55, INTER	27	0.0012	0.7634	162.252	17.52	<.0001
30	city_urb5		city, major	28	0.0011	0.7645	147.262	16.42	<.0001
31	lanes_speed_rte41		4, 55, NC	29	0.0010	0.7655	134.446	14.37	0.0002
32	lanes_speed_rte5		2, 25, SR	30	0.0009	0.7664	123.033	13.06	0.0003
33	month1		Jan	31	0.0008	0.7672	112.553	12.19	0.0005
34	cycle_surface3		Odd, Soil	32	0.0005	0.7677	106.971	7.42	0.0065
35	lucode5		Indust & Comm Complxs	33	0.0005	0.7682	102.203	6.63	0.0100
36	lanes_speed_rte50		6, 45, US	34	0.0005	0.7686	97.2951	6.78	0.0092
37	month6		June	35	0.0005	0.7691	91.9919	7.18	0.0074
38	year4		2000	36	0.0005	0.7696	86.5739	7.31	0.0069
39	lucode10		Other Ag Land	37	0.0004	0.7701	81.9501	6.54	0.0106
40	lanes_speed_rte23		3, 35, SR	38	0.0004	0.7704	78.3802	5.51	0.0190
41	lucode2		Comm & Services	39	0.0004	0.7708	74.4807	5.84	0.0157

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
42	lucode6		Mxd Urban or Built-up	40	0.0004	0.7712	70.6225	5.81	0.0160
43	lanes_speed_rte55		6, 65, INTER	41	0.0004	0.7716	67.1166	5.47	0.0195
44	lanes_speed_rte9			42	0.0004	0.7720	63.7162	5.37	0.0206
45	lucode11		Shrub & Brush Rangeland	43	0.0003	0.7723	60.9301	4.76	0.0292
46	lanes_speed_rte47		5, 45, US	44	0.0003	0.7726	58.8243	4.09	0.0432
47	city_urb3		out, minor	45	0.0003	0.7728	56.9793	3.83	0.0503
48	year2		1998	46	0.0003	0.7731	54.8882	4.08	0.0434
49	year3		1999	47	0.0002	0.7733	53.3546	3.53	0.0604
50		year4	2000	46	0.0000	0.7733	51.8773	0.52	0.4701
51	month11		Nov	47	0.0002	0.7735	50.1430	3.73	0.0535
52	co_urb5		Johnston	48	0.0003	0.7738	48.3987	3.74	0.0531
53	month2		Feb	49	0.0002	0.7740	46.8379	3.56	0.0591
54	lanes_speed_rte33		4, 35, NC	50	0.0002	0.7742	45.6005	3.24	0.0718
55		rte_class2	US	49	0.0001	0.7741	45.2725	1.67	0.1957
56	lanes_rte11		4, SR	50	0.0003	0.7744	43.3298	3.95	0.0469
57	lanes_speed_rte20		3, 30, SR	51	0.0003	0.7747	40.1735	5.17	0.0230
58	a_control2		Medium	52	0.0003	0.7750	37.9455	4.25	0.0394
59	lanes_speed_rte49		6, 35, SR	53	0.0002	0.7752	36.6651	3.30	0.0695
60	lanes_speed_rte21		3, 35, US	54	0.0002	0.7754	35.7580	2.92	0.0874
61	Factor7		Wealthy	55	0.0002	0.7756	35.1133	2.66	0.1029
62	lanes_speed_rte36		4, 45, US	56	0.0002	0.7758	34.4407	2.69	0.1010
63	lanes_speed_rte56		8, 45, US	57	0.0002	0.7760	34.1321	2.32	0.1274
64	Factor3		College	58	0.0002	0.7761	33.8621	2.29	0.1306
65	lanes_speed_rte57		8, 60, INTER	59	0.0001	0.7762	33.7117	2.17	0.1411
66	lanes_rte17		8, INTER	60	0.0003	0.7766	31.1230	4.63	0.0315

Fit final model for aadt^.2 regression from full model (PCA of census + all station data)
The REG Procedure
Model: MODEL1
Dependent Variable: baadt

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	60	6671.70425	111.19507	195.20	<.0001
Error	3370	1919.72432	0.56965		
Corrected Total	3430	8591.42857			

Root MSE	0.75475	R-Square	0.7766
Dependent Mean	5.39260	Adj R-Sq	0.7726
Coeff Var	13.99608		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	6.23009	0.18230	34.17	<.0001
Factor1	Other	1	0.18055	0.01431	12.62	<.0001
Factor2	Single	1	0.06189	0.01426	4.34	<.0001
Factor3	College	1	0.02196	0.01424	1.54	0.1232
Factor5	Farm	1	-0.22093	0.02482	-8.90	<.0001
Factor7	Wealthy	1	0.02437	0.01353	1.80	0.0719
rte_class1	INTER	1	1.58235	0.19193	8.24	<.0001
co_urb4	Durham, 103	1	-0.27598	0.05143	-5.37	<.0001
co_urb5	Johnston	1	-0.10479	0.05705	-1.84	0.0663
co_urb6	Orange	1	0.23604	0.06945	3.40	0.0007
lanes_rte1	2, US	1	-0.72604	0.08761	-8.29	<.0001
lanes_rte2	2, NC	1	-0.68759	0.08106	-8.48	<.0001
lanes_rte3	2, SR	1	-1.40233	0.14900	-9.41	<.0001

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lanes_rte4	2, local	1	-2.14652	0.15592	-13.77	<.0001
lanes_rte11	4, SR	1	-0.54427	0.15121	-3.60	0.0003
lanes_rte17	8, INTER	1	0.67292	0.31274	2.15	0.0315
speed_rte6	25, SR	1	-1.99556	0.22916	-8.71	<.0001
speed_rte19	55, INTER	1	-0.78039	0.22330	-3.49	0.0005
speed_rte20	55, US	1	0.61472	0.08981	6.84	<.0001
lanes_speed_rte5	2, 25, SR	1	1.16671	0.29628	3.94	<.0001
lanes_speed_rte9		1	-0.12178	0.05291	-2.30	0.0214
lanes_speed_rte17	2, 55, SR	1	-0.36403	0.05036	-7.23	<.0001
lanes_speed_rte20	3, 30, SR	1	-1.12314	0.36648	-3.06	0.0022
lanes_speed_rte21	3, 35, US	1	-0.43645	0.27367	-1.59	0.1109
lanes_speed_rte23	3, 35, SR	1	-1.07605	0.26215	-4.10	<.0001
lanes_speed_rte33	4, 35, NC	1	-0.41579	0.16305	-2.55	0.0108
lanes_speed_rte36	4, 45, US	1	0.29614	0.17145	1.73	0.0842
lanes_speed_rte41	4, 55, NC	1	1.21275	0.34416	3.52	0.0004
lanes_speed_rte47	5, 45, US	1	1.15271	0.53774	2.14	0.0321
lanes_speed_rte49	6, 35, SR	1	-0.64290	0.34152	-1.88	0.0599
lanes_speed_rte50	6, 45, US	1	0.55461	0.21203	2.62	0.0089
lanes_speed_rte55	6, 65, INTER	1	0.58097	0.24362	2.38	0.0171
lanes_speed_rte56	8, 45, US	1	1.16814	0.76019	1.54	0.1245
lanes_speed_rte57	8, 60, INTER	1	-1.05003	0.42207	-2.49	0.0129
cycle_surface1	Ann, Asph	1	1.12672	0.12889	8.74	<.0001
cycle_surface2	Odd, Conc	1	0.66526	0.09340	7.12	<.0001
cycle_surface3	Odd, Soil	1	1.90623	0.76120	2.50	0.0123
cycle_surface4	Even, Conc	1	1.09071	0.21282	5.13	<.0001
cycle_surface5	Even, Soil	1	0.78204	0.14952	5.23	<.0001

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
cycle_surface6	Vari, Asph	1	1.51836	0.10935	13.89	<.0001
cycle_surface7	Vari, Conc	1	1.05681	0.08294	12.74	<.0001
a_control2	Medium	1	-0.36254	0.13969	-2.60	0.0095
year2	1998	1	-0.22280	0.05931	-3.76	0.0002
year3	1999	1	-0.15888	0.05608	-2.83	0.0046
month1	Jan	1	-0.28200	0.08586	-3.28	0.0010
month2	Feb	1	-0.19963	0.09564	-2.09	0.0369
month5	May	1	-0.33902	0.04747	-7.14	<.0001
month6	June	1	-0.23031	0.06234	-3.69	0.0002
month11	Nov	1	0.22131	0.10669	2.07	0.0381
day4	Thu	1	-0.41573	0.12556	-3.31	0.0009
lucode1	Residential	1	-0.12207	0.04154	-2.94	0.0033
lucode2	Comm & Services	1	0.16336	0.05495	2.97	0.0030
lucode4	Trans, Comm, Util	1	0.40724	0.06583	6.19	<.0001
lucode5	Indust & Comm Complxs	1	0.45578	0.14543	3.13	0.0017
lucode6	Mxd Urban or Built-up	1	0.13388	0.04741	2.82	0.0048
lucode10	Other Ag Land	1	0.67339	0.25360	2.66	0.0080
lucode11	Shrub & Brush Rangeland	1	-0.97617	0.43900	-2.22	0.0262
city_urb1	out, out	1	-0.69996	0.05688	-12.31	<.0001
city_urb3	out, minor	1	-0.39946	0.22589	-1.77	0.0771
city_urb4	city, out	1	-0.24726	0.07410	-3.34	0.0009
city_urb5	city, major	1	0.21557	0.05360	4.02	<.0001

Fit final model for aadt^.2 regression from full model (PCA of census + all station data)
Normality check on residuals
The UNIVARIATE Procedure

Variable: raadt (Residual)

Moments			
N	3431	Sum Weights	3431
Mean	0	Sum Observations	0
Std Deviation	0.74812191	Variance	0.55968639
Skewness	-0.1071289	Kurtosis	0.35491668
Uncorrected SS	1919.72432	Corrected SS	1919.72432
Coeff Variation	.	Std Error Mean	0.01277209

Basic Statistical Measures			
Location		Variability	
Mean	0.00000	Std Deviation	0.74812
Median	0.00630	Variance	0.55969
Mode	-0.40331	Range	5.51173
		Interquartile Range	0.98066

NOTE: The mode displayed is the smallest of 2 modes with a count of 2.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	14.5	Pr >= |M|	0.6326
Signed Rank	S	32379.5	Pr >= |S|	0.5769

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.018968	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.173174	Pr > W-Sq	0.0124
Anderson-Darling	A-Sq	1.429019	Pr > A-Sq	<0.0050

Quantiles (Definition 5)

Quantile	Estimate
100% Max	2.88057776
99%	1.70085774
95%	1.17946684
90%	0.91402860
75% Q3	0.51021024
50% Median	0.00630117
25% Q1	-0.47044508
10%	-0.94605575
5%	-1.22076473
1%	-1.90908802
0% Min	-2.63115423

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-2.63115	805	2.48006	2675
-2.58413	811	2.52092	2006
-2.48411	3039	2.73624	2607
-2.41656	896	2.78593	2629
-2.41159	561	2.88058	2039

Fit final model for aadt^.2 regression from full model (PCA of census + all station data)
Normality check on residuals
The UNIVARIATE Procedure
Variable: raadt (Residual)

```
                                        Histogram                        #
Boxplot
                    2.75+*                                               4
0
                        .*                                              13
```

```
0
                         .****                                        59
|
                         .***********                                199
|
                         .******************************             606          +--
---+
                         .*******************************************  849          *--
+--*
                         .************************************************ 899          +--
---+
                         .***************************                500
|
                         .***********                                205
|
                         .****                                        70
0
                         .**                                          25
0
                    -2.75+*                                            2
0
                          ----+----+----+----+----+----+----+----+----+---
                          * may represent up to 19 counts
```

```
                                   Normal Probability Plot
                         2.75+                                                *
                             |                                                *
                             |                                          ******
                             |                                     *******
                             |                                 ********
                             |                            ********
                             |                       ********
                             |                  ********
                             |     +******
                             |+******
                             |**
                        -2.75+*
                              +----+----+----+----+----+----+----+----+----+----+
                                  -2        -1         0        +1        +2
```

Fit final model for aadt^.2 regression from full model (PCA of census + all station data) abs(residual)>2

Obs	UNIQ_ID	AADT	baadt	paadt	raadt
1	0181759	6532	5.79441	3.77547	2.01894
2	1030304	699	3.70591	5.73771	-2.03179

Obs	UNIQ_ID	AADT	baadt	paadt	raadt
3	1030658	352	3.23079	5.64238	-2.41159
4	1030657	525	3.49971	5.64238	-2.14267
5	1030053	169	2.78983	4.99464	-2.20481
6	1030644	343	3.21410	5.58444	-2.37035
7	1030561	18302	7.12032	5.05602	2.06430
8	1030544	36	2.04767	4.67883	-2.63115
9	1030705	95	2.48625	5.07038	-2.58413
10	1030434	695	3.70166	5.78271	-2.08105
11	1030654	462	3.41137	5.82793	-2.41656
12	1030235	8009	6.03553	8.04917	-2.01364
13	0501816	160	2.75946	4.78944	-2.02998
14	0501608	9082	6.18923	3.99492	2.19430
15	0501613	8153	6.05708	3.86104	2.19604
16	1030768	70	2.33894	4.48748	-2.14854
17	1080861	27731	7.73737	5.60554	2.13183
18	1080915	24018	7.51809	5.14976	2.36833
19	1080168	85405	9.68939	7.16847	2.52092
20	1080175	228	2.96201	5.23324	-2.27123
21	1080641	66102	9.20541	6.32483	2.88058
22	1080264	483	3.44183	5.47290	-2.03107
23	1080498	1756	4.45559	6.55738	-2.10180
24	0910584	52	2.20394	4.41330	-2.20936
25	1080698	10510	6.37266	4.20687	2.16578
26	1080616	469	3.42164	5.49988	-2.07824
27	0912026	10429	6.36280	4.23721	2.12560
28	1080512	925	3.91948	6.30312	-2.38364
29	1080514	2135	4.63318	6.81135	-2.17817
30	1080450	336	3.20087	5.56312	-2.36225

Obs	UNIQ_ID	AADT	baadt	paadt	raadt
31	1080819	17257	7.03709	5.03433	2.00276
32	1080815	36575	8.17781	5.44157	2.73624
33	1080812	25800	7.62648	5.25713	2.36934
34	0910673	10436	6.36366	3.57772	2.78593
35	1080451	618	3.61574	6.00247	-2.38673
36	1080715	17801	7.08091	4.60084	2.48006
37	0910034	2158	4.64312	6.95505	-2.31192
38	0910842	95	2.48625	4.51263	-2.02638
39	1080473	847	3.85103	6.00840	-2.15737
40	1080714	15188	6.85962	4.70427	2.15535
41	1080111	597	3.59083	6.07494	-2.48411
42	1080879	39077	8.28675	6.08340	2.20335
43	1080174	245	3.00492	5.34986	-2.34494
44	1080511	788	3.79582	6.12820	-2.33238

Fit final model to ptcdemo1
The REG Procedure
Model: MODEL1
Dependent Variable: baadt

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	59	3493.20507	59.20687	102.92	<.0001
Error	1698	976.77720	0.57525		
Corrected Total	1757	4469.98227			

Root MSE	0.75845	R-Square	0.7815
Dependent Mean	5.35009	Adj R-Sq	0.7739
Coeff Var	14.17647		

NOTE: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.

NOTE: The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

lanes_speed_rte56 =	0

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	6.16072	0.26399	23.34	<.0001
Factor1	Other	1	0.19228	0.02052	9.37	<.0001
Factor2	Single	1	0.03861	0.02019	1.91	0.0560
Factor3	College	1	0.03763	0.01973	1.91	0.0566
Factor5	Farm	1	-0.20279	0.03511	-5.78	<.0001
Factor7	Wealthy	1	0.00727	0.01901	0.38	0.7024
rte_class1	INTER	1	1.94563	0.27186	7.16	<.0001
co_urb4	Durham, 103	1	-0.31186	0.07350	-4.24	<.0001
co_urb5	Johnston	1	-0.08849	0.07967	-1.11	0.2669
co_urb6	Orange	1	0.37249	0.09782	3.81	0.0001
lanes_rte1	2, US	1	-0.59980	0.12542	-4.78	<.0001
lanes_rte2	2, NC	1	-0.61260	0.11672	-5.25	<.0001
lanes_rte3	2, SR	1	-1.26288	0.21925	-5.76	<.0001
lanes_rte4	2, local	1	-2.05798	0.22916	-8.98	<.0001
lanes_rte11	4, SR	1	-0.35141	0.21975	-1.60	0.1100
lanes_rte17	8, INTER	1	0.49323	0.35890	1.37	0.1695
speed_rte6	25, SR	1	-2.11764	0.30899	-6.85	<.0001
speed_rte19	55, INTER	1	-0.86655	0.33574	-2.58	0.0099
speed_rte20	55, US	1	0.68487	0.13002	5.27	<.0001
lanes_speed_rte5	2, 25, SR	1	1.58360	0.40749	3.89	0.0001
lanes_speed_rte9		1	-0.09478	0.07421	-1.28	0.2017

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lanes_speed_rte17	2, 55, SR	1	-0.32660	0.06966	-4.69	<.0001
lanes_speed_rte20	3, 30, SR	1	-1.32113	0.78520	-1.68	0.0926
lanes_speed_rte21	3, 35, US	1	-0.02690	0.44790	-0.06	0.9521
lanes_speed_rte23	3, 35, SR	1	-1.07166	0.40153	-2.67	0.0077
lanes_speed_rte33	4, 35, NC	1	-0.37053	0.23483	-1.58	0.1148
lanes_speed_rte36	4, 45, US	1	0.36942	0.25507	1.45	0.1477
lanes_speed_rte41	4, 55, NC	1	-0.29140	0.76954	-0.38	0.7050
lanes_speed_rte47	5, 45, US	1	1.26587	0.54438	2.33	0.0202
lanes_speed_rte49	6, 35, SR	1	-0.16536	0.49193	-0.34	0.7368
lanes_speed_rte50	6, 45, US	1	0.56724	0.28341	2.00	0.0455
lanes_speed_rte55	6, 65, INTER	1	0.53124	0.33653	1.58	0.1146
lanes_speed_rte56	8, 45, US	0	0	.	.	.
lanes_speed_rte57	8, 60, INTER	1	-0.93888	0.49348	-1.90	0.0573
cycle_surface1	Ann, Asph	1	0.96392	0.18179	5.30	<.0001
cycle_surface2	Odd, Conc	1	0.58145	0.13034	4.46	<.0001
cycle_surface3	Odd, Soil	1	1.93991	0.77035	2.52	0.0119
cycle_surface4	Even, Conc	1	1.07360	0.29406	3.65	0.0003
cycle_surface5	Even, Soil	1	0.64034	0.20334	3.15	0.0017
cycle_surface6	Vari, Asph	1	1.35009	0.15553	8.68	<.0001
cycle_surface7	Vari, Conc	1	0.95545	0.11601	8.24	<.0001
a_control2	Medium	1	-0.24606	0.20381	-1.21	0.2275
year2	1998	1	-0.23841	0.08390	-2.84	0.0045
year3	1999	1	-0.13190	0.07830	-1.68	0.0923
month1	Jan	1	-0.45789	0.11934	-3.84	0.0001
month2	Feb	1	-0.15791	0.13100	-1.21	0.2282
month5	May	1	-0.43176	0.06651	-6.49	<.0001

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
month6	June	1	-0.24486	0.08613	-2.84	0.0045
month11	Nov	1	0.21079	0.13935	1.51	0.1305
day4	Thu	1	-0.31799	0.18559	-1.71	0.0868
lucode1	Residential	1	-0.15474	0.05948	-2.60	0.0094
lucode2	Comm & Services	1	0.08627	0.07802	1.11	0.2690
lucode4	Trans, Comm, Util	1	0.38200	0.09673	3.95	<.0001
lucode5	Indust & Comm Complxs	1	0.45097	0.19885	2.27	0.0235
lucode6	Mxd Urban or Built-up	1	0.09417	0.06552	1.44	0.1509
lucode10	Other Ag Land	1	0.86892	0.34366	2.53	0.0115
lucode11	Shrub & Brush Rangeland	1	-1.52756	0.78356	-1.95	0.0514
city_urb1	out, out	1	-0.74138	0.08399	-8.83	<.0001
city_urb3	out, minor	1	-0.24241	0.35675	-0.68	0.4969
city_urb4	city, out	1	-0.26101	0.10922	-2.39	0.0170
city_urb5	city, major	1	0.30430	0.07872	3.87	0.0001

Fit final model to ptcdemo1
Normality check on residuals for ptcdemo1
The UNIVARIATE Procedure
Variable: raadt (Residual)

Moments			
N	1673	Sum Weights	1673
Mean	0.05009181	Sum Observations	83.8036015
Std Deviation	0.76691043	Variance	0.58815161
Skewness	-0.0466899	Kurtosis	0.26080656
Uncorrected SS	987.587365	Corrected SS	983.38949
Coeff Variation	1531.00956	Std Error Mean	0.0187498

Basic Statistical Measures			
Location		Variability	
Mean	0.05009	Std Deviation	0.76691
Median	0.03845	Variance	0.58815
Mode	-0.62193	Range	5.32280
		Interquartile Range	1.01634

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	2.671592	Pr > |t|	0.0076
Sign	M	44.5	Pr >= |M|	0.0314
Signed Rank	S	55665.5	Pr >= |S|	0.0048

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.998732	Pr < W	0.2676
Kolmogorov-Smirnov	D	0.01978	Pr > D	0.1111
Cramer-von Mises	W-Sq	0.06801	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.439355	Pr > A-Sq	>0.2500

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	2.7563194
99%	1.8687248
95%	1.3131357
90%	1.0005458
75% Q3	0.5692602
50% Median	0.0384536
25% Q1	-0.4470749
10%	-0.9312025

Quantiles (Definition 5)	
Quantile	Estimate
5%	-1.2013290
1%	-1.8086955
0% Min	-2.5664786

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-2.56648	399	2.27267	189
-2.42820	1342	2.28342	1441
-2.42307	1327	2.40610	963
-2.36191	325	2.56911	1328
-2.34536	1223	2.75632	984

Fit final model to ptcdemo1
Normality check on residuals for ptcdemo1
The UNIVARIATE Procedure
Variable: raadt (Residual)

```
                                        Histogram                       #
Boxplot
                    2.75+*                                              2               0
                        .*                                              9               0
                        .*****                                         39               |
                        .**************                                118               |
                        .********************************             297             +---
--+
                        .***********************************************  416             *--
+--*
                        .**********************************************   412             +---
--+
                        .***************************                    236               |
                        .************                                   103               |
                        .****                                           31               0
                        .*                                               9               0
                   -2.75+*                                               1               0
                         ----+----+----+----+----+----+----+----+----+--
                         * may represent up to 9 counts
```

```
                          Normal Probability Plot
 2.75+                                                        *
     |                                                        *
     |                                                 *******
     |                                          *******
     |                                   ********
     |                            ********
     |                      *******
     |                ********
     |          *******
     |******
     |*
-2.75+*
      +----+----+----+----+----+----+----+----+----+----+
          -2        -1         0        +1        +2
```

Fit final model to ptcdemo2
The REG Procedure
Model: MODEL1
Dependent Variable: baadt

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	58	3203.97635	55.24097	97.87	<.0001
Error	1614	910.95464	0.56441		
Corrected Total	1672	4114.93099			

Root MSE	0.75127	**R-Square**	0.7786
Dependent Mean	5.43727	**Adj R-Sq**	0.7707
Coeff Var	13.81707		

NOTE: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.

NOTE: The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

lanes_speed_rte47 =	0

cycle_surface3 =	0

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	6.37424	0.25897	24.61	<.0001
Factor1	Other	1	0.17069	0.02024	8.44	<.0001
Factor2	Single	1	0.08844	0.02050	4.31	<.0001
Factor3	College	1	-0.00705	0.02140	-0.33	0.7419
Factor5	Farm	1	-0.24034	0.03567	-6.74	<.0001
Factor7	Wealthy	1	0.03862	0.01954	1.98	0.0482
rte_class1	INTER	1	1.22790	0.27534	4.46	<.0001
co_urb4	Durham, 103	1	-0.25317	0.07314	-3.46	0.0006
co_urb5	Johnston	1	-0.12553	0.08237	-1.52	0.1277
co_urb6	Orange	1	0.09771	0.09975	0.98	0.3274
lanes_rte1	2, US	1	-0.85017	0.12458	-6.82	<.0001
lanes_rte2	2, NC	1	-0.78417	0.11437	-6.86	<.0001
lanes_rte3	2, SR	1	-1.59799	0.20727	-7.71	<.0001
lanes_rte4	2, local	1	-2.29278	0.21693	-10.57	<.0001
lanes_rte11	4, SR	1	-0.79847	0.21262	-3.76	0.0002
lanes_rte17	8, INTER	1	0.99845	0.77861	1.28	0.1999
speed_rte6	25, SR	1	-1.65520	0.34835	-4.75	<.0001
speed_rte19	55, INTER	1	-0.69684	0.30441	-2.29	0.0222
speed_rte20	55, US	1	0.55701	0.12717	4.38	<.0001
lanes_speed_rte5	2, 25, SR	1	0.54071	0.44122	1.23	0.2206
lanes_speed_rte9		1	-0.13894	0.07625	-1.82	0.0686
lanes_speed_rte17	2, 55, SR	1	-0.39618	0.07368	-5.38	<.0001
lanes_speed_rte20	3, 30, SR	1	-1.26560	0.42759	-2.96	0.0031
lanes_speed_rte21	3, 35, US	1	-0.74039	0.34749	-2.13	0.0333
lanes_speed_rte23	3, 35, SR	1	-1.16509	0.34921	-3.34	0.0009

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lanes_speed_rte33	4, 35, NC	1	-0.40803	0.22958	-1.78	0.0757
lanes_speed_rte36	4, 45, US	1	0.20512	0.23271	0.88	0.3782
lanes_speed_rte41	4, 55, NC	1	1.50530	0.38681	3.89	0.0001
lanes_speed_rte47	5, 45, US	0	0	.	.	.
lanes_speed_rte49	6, 35, SR	1	-1.25160	0.48210	-2.60	0.0095
lanes_speed_rte50	6, 45, US	1	0.56126	0.32553	1.72	0.0849
lanes_speed_rte55	6, 65, INTER	1	0.57565	0.35641	1.62	0.1065
lanes_speed_rte56	8, 45, US	1	0.98542	0.76249	1.29	0.1964
lanes_speed_rte57	8, 60, INTER	1	-1.42131	0.93329	-1.52	0.1280
cycle_surface1	Ann, Asph	1	1.26479	0.18839	6.71	<.0001
cycle_surface2	Odd, Conc	1	0.73528	0.13660	5.38	<.0001
cycle_surface3	Odd, Soil	0	0	.	.	.
cycle_surface4	Even, Conc	1	1.07478	0.31220	3.44	0.0006
cycle_surface5	Even, Soil	1	0.92944	0.22650	4.10	<.0001
cycle_surface6	Vari, Asph	1	1.64122	0.15814	10.38	<.0001
cycle_surface7	Vari, Conc	1	1.13441	0.12132	9.35	<.0001
a_control2	Medium	1	-0.49467	0.19526	-2.53	0.0114
year2	1998	1	-0.20037	0.08658	-2.31	0.0208
year3	1999	1	-0.18186	0.08230	-2.21	0.0273
month1	Jan	1	-0.09292	0.12489	-0.74	0.4570
month2	Feb	1	-0.22252	0.14292	-1.56	0.1197
month5	May	1	-0.24003	0.06869	-3.49	0.0005
month6	June	1	-0.21025	0.09134	-2.30	0.0215
month11	Nov	1	0.20819	0.16763	1.24	0.2144
day4	Thu	1	-0.46465	0.17383	-2.67	0.0076
lucode1	Residential	1	-0.09151	0.05860	-1.56	0.1186

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lucode2	Comm & Services	1	0.22269	0.07810	2.85	0.0044
lucode4	Trans, Comm, Util	1	0.39708	0.09239	4.30	<.0001
lucode5	Indust & Comm Complxs	1	0.46206	0.21750	2.12	0.0338
lucode6	Mxd Urban or Built-up	1	0.19013	0.07000	2.72	0.0067
lucode10	Other Ag Land	1	0.43178	0.37845	1.14	0.2541
lucode11	Shrub & Brush Rangeland	1	-0.87487	0.53717	-1.63	0.1036
city_urb1	out, out	1	-0.66967	0.07869	-8.51	<.0001
city_urb3	out, minor	1	-0.50400	0.29992	-1.68	0.0931
city_urb4	city, out	1	-0.25631	0.10237	-2.50	0.0124
city_urb5	city, major	1	0.13085	0.07434	1.76	0.0786

Fit final model to ptcdemo2
Normality check on residuals for ptcdemo2
The UNIVARIATE Procedure
Variable: raadt (Residual)

Moments			
N	1758	Sum Weights	1758
Mean	-0.0462379	Sum Observations	-81.286237
Std Deviation	0.7704909	Variance	0.59365623
Skewness	-0.0898357	Kurtosis	0.43119154
Uncorrected SS	1046.8125	Corrected SS	1043.05399
Coeff Variation	-1666.3621	Std Error Mean	0.0183763

Basic Statistical Measures			
Location		Variability	
Mean	-0.04624	Std Deviation	0.77049
Median	-0.05242	Variance	0.59366

Basic Statistical Measures			
Location		Variability	
Mode	.	Range	5.64615
		Interquartile Range	1.01242

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	-2.51617	Pr > |t|	0.0120
Sign	M	-45	Pr >= |M|	0.0338
Signed Rank	S	-44997.5	Pr >= |S|	0.0345

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.996361	Pr < W	0.0003
Kolmogorov-Smirnov	D	0.018095	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.099761	Pr > W-Sq	0.1162
Anderson-Darling	A-Sq	0.921704	Pr > A-Sq	0.0205

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	3.0118885
99%	1.6413868
95%	1.1608537
90%	0.9095572
75% Q3	0.4832669
50% Median	-0.0524183
25% Q1	-0.5291542
10%	-1.0208215
5%	-1.3074882
1%	-2.0360264

Quantiles (Definition 5)	
Quantile	Estimate
0% Min	-2.6342618

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-2.63426	411	2.38361	1312
-2.54241	1542	2.43035	1168
-2.43225	292	2.73823	1310
-2.37803	1585	2.91592	1043
-2.36114	461	3.01189	1322

Fit final model to ptcdemo2
Normality check on residuals for ptcdemo2
The UNIVARIATE Procedure
Variable: raadt (Residual)

```
                                        Histogram                        #
Boxplot
                     3.25+*                                              1               0
                         .*                                              2               0
                         .*                                              6               0
                         .***                                           26               |
                         .**********                                    96               |
                         .******************************                296               |
                     0.25+*****************************************     407            +---
--+
                         .*********************************************** 460            *--
+--*
                         .****************************                   284            +---
--+
                         .************                                   119               |
                         .*****                                          43               |
                         .**                                             16               0
                    -2.75+*                                              2               0
                          ----+----+----+----+----+----+----+----+----+-
                          * may represent up to 10 counts


                                                Normal Probability Plot
```

```
 3.25+                                                  *
     |                                                  *
     |                                                  *
     |                                              *****
     |                                        *******
     |                                   ********
 0.25+                              *******
     |                        ********
     |              ********
     |        ******
     |++******
     |***
-2.75+*
      +----+----+----+----+----+----+----+----+----+----+
          -2        -1         0        +1        +2
```

APPENDIX 9:

9. Determining the Final Mean Model for the Statewide Area, Version 2

Jacqueline M. Hughes-Oliver
Tae-Young Heo

March 2005

Determining the Final Mean Model for the Statewide Area, Version 2
Jacqueline M. Hughes-Oliver and Tae-Young Heo
March 28, 2005

Based on

Determining the Final Mean Model for the Statewide Area
Jacqueline M. Hughes-Oliver and Tae-Young Heo
Dec. 19, 2003

1. Introduction

This report closely follows the format of the July 2003 report "Determining the Final Mean Model for the 5-County Test Area," which will be referred to here simply as the July Report. In fact, the current report chronicles the extension of activities applied to the test area PTC stations, as outlined in the July Report, to the set of PTC stations of the entire state. Because of the strong similarities between activities leading to the two reports, we will only highlight differences from the July Report.

Appendix A includes the SAS code for all analyses presented here.

2. Data Description

2.1 Station Data

The July Report was based on data modified to reflect findings form the "List of 55" outliers. As such, that station data included the new method (see "New Urban and Municipal Field Value Assignments and Their Relation to Station Locations and the PTC" in Appendix 6) for classifying municipal and urban location of stations. The new method renamed the existing attribute URBAN_FLG to MUNICIPAL_FLG, in order to more accurately indicate its meaning. Station inside a municipal boundary has MUNICIPAL_FLG of one, with value zero otherwise. The new method also recreated an attribute URBAN_FLG to indicate a station's location in urban areas: zero outside urban areas, one inside major urban areas, and two inside minor urban areas. Another edit motivated by the "List of 55" was an update to the classification scheme used to identify route segments as interstate, US routes, NC routes, SR routes, or local routes. Station data used in the July Report contained the new attribute RTE_ADJUST.

In the July Report, the interaction between MUNICIPAL_FLG and URBAN_FLG was a significant component in the final model, as were several variables that depended on attribute RTE_ADJUST. Unfortunately, the statewide data that we received in May 2002 did not include the new version of URBAN_FLG, MUNICIPAL_FLG, or RTE_ADJUST. In fact, while the statewide PTC count data benefited from the first set of extensive edits described in "Procedures and Corrections of PTC Data" and "Data Description for Spatial Study" (Appendices 2 and 3), it did not undergo the second set of extensive edits described in "Data Edits Determined from 'List of 55'" (Appendix 5).

To accommodate the distinctions between available data for the test area and the entire state, we performed a specialized merge. Starting with the October 2002 (most updated) version of PTC count data for the test area, we

- Deleted the newly created URBAN_FLG
- Kept MUNICIPAL_FLG without changes
- Kept RTE_ADJUST without changes

For the May 2002 version of statewide PTC count data, we

- Deleted data for the 3,431 stations within the test area
- Renamed URB_FLAG to MUNICIPAL_FLG
- If RTE_TYPE=7, replace it with RTE_TYPE=4, which is the most prevalent route type
- Renamed RTE_TYPE to RTE_ADJUST
- Inserted the newly modified version of test area PTC count data

This was the data we used for developing the mean model for the state.

2.2 Census Data

Census data was unchanged from the July Report.

3. Category Summaries

3.1 Census Data

Following the July Report, we again split the data into two halves and considered keeping either 21 principal components or a smaller set of seven principal components. While the 21 principal components account for 89% of the variability in the 146 census attributes, the seven principal components only account for approximately 76% of the variability. On the other hand, the r-square value for the overall mean model changed only slightly, dropping from .7431 for the model using all 21 components to .7287 for the model with only seven components. Furthermore, correlations between the sets of seven principal components were always at least 0.98, indicating strong agreement. Even though the number of components retained was determined using the smaller datasets, the entire dataset was used to obtain the principal components to be included in refining the mean model. A summary of the seven principal components is given in Table 1. These differ in substantial ways from the census summaries reported in the July Report. In other words, the 5-county test area has unique features that separate it from the rest of North Carolina. Additional details of the principal components analysis are given in Appendix B.

Table 1: Summary of Rotated Principal Components Retained

Principal Component	% variability explained	Attributes in Factor (- means negative impact on component, 0 means negligible impact on all components). Attribute names are explained in Appendix D of Appendix 8.
Factor1: With Kids	44.68	BORNWEST, OTH_STAT, HISPANIC, ARMDFORC, AGE21_24, COM_LT15, ABLEENGL, BORNMIDW, MILIQUAR, BORNSOUT, WORKERS, MALE, WRK_HOME, WRKINCTY, AGEIS_20, AGELT_05, CARPOOL, SINGLE_, AGE25_34, AGE05_09, NATIVE, AGEGE_03, AGEGE_16, BORNNORT, ONLYENGL, ASIAN, ATTACHED, MARRWICH, AGE18_19, SOMECOLG, NATURAL_, INC_1525, MARRIED, WHITE, BLACK, ELEMSCND, PREPRIMA, PERINFAM, AGE10_14, PRUNIT34, SEPARATE, RNT_2550, DRVALONE, HIGHSCHL, COM_1529, VEHICL_1, ROOM_4_6, ELECTRIC, WAGE_SAL, VEHICL_2, AGE35_44, UNEMPLOY, PUBSEWER, COLLEGE, BLTBFR70, INC_2535, PUBWATER, CHILDPOV, AGE15_17, DUPLEX, PUBLWORK, APARTMNT, OTH_HEAT, PUBTRANS Note that this component includes middle income families with children.
Factor2: No Kids	13.41	PRIVWORK, EMPLOYED, OWNR_OCC, AGE45_54, TECHSALE, TRADE, MARRNOCH, PRUNIT12, INC_5075,

		INC_3550, EXECPROF, UTILITY, SELF_INC, VAL_501C, SERVICES, FIRE, SERVICE, DETACHED, BLTAFT84, AGE55_64, INTE_INC, BLT_8084, SAMEHOUS, BORN_INS, PROFSERV, ONEPERHH, BLT_7079, SELFWORK, COM_3044, MANUFACT, SAMECNTY, PRIMARY_, VAL_1C2C, COLGGRAD, RETI_INC, WRKEXCTY, ROOM_1_3 Note that this component includes middle income families with no children.
Factor3: Poor	6.98	PUBL_INC, NOVEHICL, INC_LT15, RNT_LT25, INPOVRTY, SINGWICH, VAL_LT25, NODIPLOM, AMIND, RNT_MEDI(-), VAL_MEDI(-), INCPRCAP(-), INC_MEDN(-)
Factor4: City Life	5.29	FAMILIES, HOUSHOLD, HOUSEDEN, POPDEN, PUBL_GAS SHELTERS, F_MIGRNT(-), BOTL_GAS(-), FARM_INC(-), COALWOOD(-), LAND_KM(-)
Factor5: Elderly	2.26	AGE75_84, AGE65_74, WIDOWED, SOCS_INC, FUELKERO, NURSHOME, MEDYRBLT, O_INSTIT
Factor6: Wealthy	1.88	RNT_751K, VAL_GT3C, VAL_2C3C, RNT_5075, RNT_GT1K, SEASONAL, WATER_KM, INSTREET(0), VAL_2550(-)
Factor7: Unstable	1.64	DORMITOR, SAMESTAT, MENTAL, CORRINST, O_NOINST

3.2 Statewide Station Data

All interaction variables for the entire state were created as in the July Report, with one exception. The July Report used attributes MUNICIPAL_FLG and URBAN_FLG to create CITY_URB. Because the statewide data does not have attribute URBAN_FLG, the definition of CITY_URB changed to

CITY_URB=**100***(MUNICPAL_FLG=**1).**

4. Model Selection

Combining the seven principal components (called FACTOR1, FACTOR2, …, FACTOR7) from the census data with the attributes from the station data, a full regression model for AADT yields an r-square value of 0.7296 and well-behaved residuals. We are, however, in search of a simpler model that uses fewer attributes. We accomplish this using a stepwise selection procedure with significance-level-to-enter of 0.50 and significance-level-to-stay of 0.10. The selected model is based on 178 attributes, with r-square value of 0.7287 and Mallow's Cp of 174.2. A summary of the model is given in Table 2 below and some details are given in Appendix C. Because of linear dependencies, the INTERCEPT term actually also corresponds to the following attributes:
RTE_TYPE 5 (Local), CO_URB128 (Yancey county), LANES_RTE 23 (12 lanes, INTER), SPEED_RTE 34 (70 mph, INTER), LANES_SPEED_RTE 58 (12 lanes, 55 mph, INTER), CYCLE_SURFACE 10 (variable cycle, Soil surface), A_CONTROL 3 (interstate), YEAR 5 (2001), MONTH 0 (missing month), DAY 0 (missing day), LUCODE 0 (transitional areas), CITY_URB 2 (Urban area).

Table 2: Summary of the Selected Model

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	2.89975	0.13146	22.06	<.0001
Factor1	With Kids	1	0.04044	0.00224	18.02	<.0001
Factor2	No Kids	1	0.06372	0.00235	27.15	<.0001
Factor3	Poor	1	-0.04177	0.00238	-17.58	<.0001
Factor4	City Life	1	0.11557	0.00321	36.01	<.0001
Factor6	Wealthy	1	0.01611	0.00236	6.83	<.0001
Rte_class1	INTER	1	0.73955	0.12976	5.70	<.0001
Rte_class2	US	1	0.46149	0.09600	4.81	<.0001
Rte_class4	SR	1	0.25370	0.13692	1.85	0.0639
Co_urb2	Alamance,101	1	-0.05037	0.01970	-2.56	0.0106
Co_urb5	Alleghary	1	-0.31888	0.03209	-9.94	<.0001
Co_urb6	Anson	1	-0.24327	0.02209	-11.01	<.0001
Co_urb7	Ashe	1	-0.09501	0.02772	-3.43	0.0006
Co_urb8	Averry	1	-0.06778	0.03523	-1.92	0.0544
Co_urb9	Beanfort	1	-0.04624	0.02293	-2.02	0.0437
Co_urb10	Bertie	1	-0.09002	0.02709	-3.32	0.0009
Co_urb11	Bladen	1	-0.06975	0.02416	-2.89	0.0039
Co_urb12	Brunswick	1	0.22186	0.02592	8.56	<.0001
Co_urb13	Brunswick,109	1	0.81044	0.38151	2.12	0.0337
Co_urb14	Buncombe	1	0.04653	0.02461	1.89	0.0587
Co_urb16	Burke	1	0.09365	0.02274	4.12	<.0001

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Co_urb17	Burke,114	1	0.09461	0.04802	1.97	0.0488
Co_urb18	Cabamus	1	0.09471	0.02633	3.60	0.0003
Co_urb19	Cabamus,111	1	0.14474	0.02630	5.50	<.0001
Co_urb20	Caldwell	1	0.11415	0.02481	4.60	<.0001
Co_urb21	Caldwell,114	1	0.15614	0.08619	1.81	0.0701
Co_urb22	Camden	1	-0.10877	0.03996	-2.72	0.0065
Co_urb23	Carteret	1	0.10947	0.03385	3.23	0.0012
Co_urb24	Caswell	1	-0.09026	0.02820	-3.20	0.0014
Co_urb25	Catawba	1	0.18278	0.03117	5.86	<.0001
Co_urb26	Catawba,114	1	0.06809	0.02308	2.95	0.0032
Co_urb27	Chatham	1	0.03865	0.02274	1.70	0.0892
Co_urb29	Cherokee	1	-0.08635	0.03294	-2.62	0.0088
Co_urb30	Chowan	1	-0.07768	0.03671	-2.12	0.0343
Co_urb31	Clay	1	-0.15403	0.03721	-4.14	<.0001
Co_urb34	Craven	1	0.05920	0.02503	2.36	0.0180
Co_urb35	Cumberland,104	1	0.08844	0.01691	5.23	<.0001
Co_urb36	Cumituck	1	-0.06633	0.04207	-1.58	0.1149
Co_urb37	Dare	1	-0.17554	0.04087	-4.29	<.0001
Co_urb38	Davison	1	0.09107	0.01771	5.14	<.0001
Co_urb39	Davison,107	1	0.05746	0.03265	1.76	0.0784
Co_urb43	Durham,103	1	0.08753	0.01858	4.71	<.0001

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Co_urb44	Edgecombe	1	-0.12207	0.02641	-4.62	<.0001
Co_urb45	Edgecombe,116	1	-0.19169	0.04913	-3.90	<.0001
Co_urb46	Forsyth	1	0.10441	0.02051	5.09	<.0001
Co_urb48	Franklin	1	0.24077	0.01920	12.54	<.0001
Co_urb50	Gaston,105	1	0.06179	0.03419	1.81	0.0707
Co_urb52	Graham	1	-0.20030	0.03103	-6.45	<.0001
Co_urb53	Granville	1	-0.20094	0.04424	-4.54	<.0001
Co_urb55	Guilford	1	-0.07316	0.02579	-2.84	0.0046
Co_urb57	Guilford,106	1	-0.40715	0.05637	-7.22	<.0001
Co_urb58	Guilford,107,	1	0.24018	0.02110	11.38	<.0001
Co_urb60	Harnett	1	-0.11157	0.02193	-5.09	<.0001
Co_urb61	Harnett,104,	1	0.13411	0.02058	6.52	<.0001
co_urb62	Haywood	1	0.23887	0.09910	2.41	0.0159
co_urb63	Henderson	1	0.04047	0.02502	1.62	0.1058
co_urb64	Hertford	1	0.18482	0.02320	7.97	<.0001
co_urb65	Hoke	1	-0.11413	0.02714	-4.21	<.0001
co_urb66	Hyde	1	0.07272	0.02938	2.48	0.0133
co_urb67	Iredell	1	-0.19964	0.04101	-4.87	<.0001
co_urb69	Johnston	1	0.07050	0.03009	2.34	0.0191
co_urb70	Jones	1	0.07419	0.01800	4.12	<.0001
co_urb71	Lee	1	-0.11026	0.03701	-2.98	0.0029

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
co_urb72	Lenoir	1	0.14829	0.02652	5.59	<.0001
co_urb73	Lincoln	1	0.13403	0.02143	6.25	<.0001
co_urb74	Macon	1	0.11116	0.02088	5.32	<.0001
co_urb75	Madison	1	-0.11504	0.02577	-4.46	<.0001
co_urb76	Martin	1	-0.21342	0.03077	-6.94	<.0001
co_urb77	McDowell	1	-0.13394	0.02500	-5.36	<.0001
co_urb80	Mitchell	1	0.26488	0.01612	16.43	<.0001
co_urb81	Montgomery	1	-0.14516	0.03632	-4.00	<.0001
co_urb82	Moore	1	-0.20998	0.02323	-9.04	<.0001
co_urb84	Nash,116	1	-0.07992	0.01979	-4.04	<.0001
co_urb86	Northampton	1	0.15816	0.02408	6.57	<.0001
co_urb87	Onslow	1	-0.16820	0.02697	-6.24	<.0001
co_urb88	Onslow,112	1	0.13216	0.02591	5.10	<.0001
co_urb89	Orange	1	0.29597	0.03730	7.93	<.0001
co_urb90	Orange,101	1	0.12297	0.02820	4.36	<.0001
co_urb92	Pamlico	1	0.10736	0.03273	3.28	0.0010
co_urb93	Pasquotank	1	-0.12698	0.04288	-2.96	0.0031
co_urb97	Pitt	1	0.04977	0.02468	2.02	0.0438
co_urb98	Pitt, 115	1	-0.05650	0.02208	-2.56	0.0105
co_urb99	Pork	1	0.08170	0.02202	3.71	0.0002
co_urb100	Randolph	1	-0.17441	0.03064	-5.69	<.0001

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
co_urb101	Randolph,107	1	0.06395	0.01806	3.54	0.0004
co_urb102	Richmond	1	0.13783	0.04904	2.81	0.0049
co_urb103	Robeson	1	-0.04805	0.02300	-2.09	0.0367
co_urb104	Rockingham	1	0.06789	0.01611	4.21	<.0001
co_urb106	Rowan,111	1	0.09791	0.01978	4.95	<.0001
co_urb107	Rutherford	1	-0.05430	0.03367	-1.61	0.1068
co_urb109	Scotland	1	-0.04576	0.01786	-2.56	0.0104
co_urb111	Stokes	1	-0.06280	0.02087	-3.01	0.0026
co_urb113	Swain	1	0.08120	0.01988	4.08	<.0001
co_urb114	Transylvania	1	-0.07980	0.03615	-2.21	0.0273
co_urb115	Tyrrell	1	-0.06324	0.03143	-2.01	0.0443
co_urb116	Union	1	-0.19310	0.05011	-3.85	0.0001
co_urb117	Vance	1	0.08530	0.01777	4.80	<.0001
co_urb118	Wake	1	0.16520	0.02635	6.27	<.0001
co_urb119	Wake,108	1	0.13638	0.01544	8.83	<.0001
co_urb120	Warren	1	-0.08077	0.02655	-3.04	0.0024
co_urb121	Washington	1	-0.08843	0.03686	-2.40	0.0164
co_urb123	Wayne	1	-0.04677	0.02278	-2.05	0.0401
co_urb125	Wilkes	1	0.12113	0.02101	5.77	<.0001
co_urb127	Yadkin	1	0.04972	0.02628	1.89	0.0585
lanes_rte2	1, SR	1	-0.43949	0.17699	-2.48	0.0130

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lanes_rte4	2, US	1	-0.33799	0.02236	-15.12	<.0001
lanes_rte5	2, NC	1	-0.14325	0.09285	-1.54	0.1229
lanes_rte6	2, SR	1	-0.39262	0.04728	-8.30	<.0001
lanes_rte7	2, LOCAL	1	-0.40653	0.12809	-3.17	0.0015
lanes_rte8	3, US	1	-0.13747	0.02950	-4.66	<.0001
lanes_rte13	4, NC	1	0.15421	0.09479	1.63	0.1038
lanes_rte14	4, SR	1	0.13166	0.04879	2.70	0.0070
lanes_rte16	5, SR	1	0.34247	0.15157	2.26	0.0239
lanes_rte19	6, NC	1	0.37013	0.12497	2.96	0.0031
lanes_rte21	8, INTER	1	0.47534	0.07464	6.37	<.0001
speed_rte1	20, US	1	-0.40483	0.08283	-4.89	<.0001
speed_rte3	20, SR	1	-0.30530	0.06190	-4.93	<.0001
speed_rte5	25, NC	1	-0.29746	0.06659	-4.47	<.0001
speed_rte7	25, LOCAL	1	-0.94873	0.10460	-9.07	<.0001
speed_rte12	35, NC	1	-0.18336	0.02175	-8.43	<.0001
speed_rte13	35, SR	1	0.03342	0.01639	2.04	0.0415
speed_rte14	40, US	1	-0.13636	0.01641	-8.31	<.0001
speed_rte20	45, SR	1	0.13384	0.01957	6.84	<.0001
speed_rte21	50, INTER	1	-0.02835	0.01825	-1.55	0.1205
speed_rte27	55, NC	1	-0.04880	0.02184	-2.23	0.0254
speed_rte30	60, US	1	-0.25867	0.06963	-3.71	0.0002

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lanes_speed_rte5	2, 20, US	1	0.20744	0.09448	2.20	0.0281
lanes_speed_rte8	2, 25, US	1	-0.57891	0.38151	-1.52	0.1292
lanes_speed_rte11	2, 25, LOCAL	1	0.73829	0.10807	6.83	<.0001
lanes_speed_rte26	2, 50, SR	1	0.23929	0.05629	4.25	<.0001
lanes_speed_rte31	2, 60, US	1	-0.30309	0.01674	-18.11	<.0001
lanes_speed_rte40	3, 35, US	1	-0.42754	0.17746	-2.41	0.0160
lanes_speed_rte45	3, 45, US	1	0.16677	0.09718	1.72	0.0862
lanes_speed_rte49	3, 55, NC	1	0.20937	0.03578	5.85	<.0001
lanes_speed_rte65	4, 45, NC	1	-0.08422	0.03314	-2.54	0.0110
lanes_speed_rte73	4, 55, NC	1	0.09696	0.03196	3.03	0.0024
lanes_speed_rte74	4, 55, SR	1	0.26884	0.04733	5.68	<.0001
lanes_speed_rte76	4, 60, NC	1	0.24306	0.14413	1.69	0.0917
lanes_speed_rte83	5, 45, SR	1	0.28616	0.19217	1.49	0.1365
lanes_speed_rte93	6, 50, US	1	0.34502	0.11612	2.97	0.0030
lanes_speed_rte95	6, 55, US	1	0.16556	0.06475	2.56	0.0106
lanes_speed_rte99	6, 65, US	1	0.30016	0.08284	3.62	0.0003
lanes_speed_rte102	8, 45, US	1	-0.70542	0.28005	-2.52	0.0118
lanes_speed_rte104	8, 60, INTER	1	-0.81319	0.10325	-7.88	<.0001
cycle_surface1	Ann, Asph	1	1.18513	0.03160	37.50	<.0001
cycle_surface2	Ann, Conc	1	0.87198	0.05551	15.71	<.0001
cycle_surface4	Odd, Conc	1	0.50895	0.02889	17.62	<.0001

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
cycle_surface5	Odd, Soil	1	0.47010	0.13023	3.61	0.0003
cycle_surface6	Even, Conc	1	0.63321	0.04048	15.64	<.0001
cycle_surface7	Even, Soil	1	0.58909	0.04604	12.80	<.0001
cycle_surface8	Vari, Asph	1	1.20750	0.02990	40.38	<.0001
cycle_surface9	Vari, Conc	1	0.71481	0.02685	26.62	<.0001
a_control2	Medium	1	-0.21018	0.08707	-2.41	0.0158
year4	2000	1	0.07120	0.01071	6.65	<.0001
month1	Jan	1	-0.04404	0.00913	-4.82	<.0001
month5	May	1	-0.05882	0.00870	-6.76	<.0001
month6	June	1	-0.04412	0.00965	-4.57	<.0001
month7	July	1	-0.05511	0.01097	-5.03	<.0001
month8	Aug	1	-0.08073	0.01104	-7.31	<.0001
month9	Sept	1	-0.06269	0.00987	-6.35	<.0001
month10	Oct	1	-0.04291	0.00967	-4.44	<.0001
month11	Nov	1	-0.04713	0.00979	-4.82	<.0001
day1	Mon	1	-0.01254	0.00430	-2.91	0.0036
day4	Thu	1	-0.06818	0.02201	-3.10	0.0019
lucode2	12,Comm & Services	1	0.07571	0.00934	8.10	<.0001
lucode3	13,Industrial	1	0.05594	0.01755	3.19	0.0014
lucode4	14,Trans,Comm,Util	1	0.15557	0.01546	10.06	<.0001
lucode5	15,INDUST & COMMERC CMPLXS	1	0.10757	0.02381	4.52	<.0001

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
lucode7	17,OTHER URBAN OR BUILT-U	1	0.06503	0.02453	2.65	0.0080
lucode8	21,CROPLAND AND PASTURE	1	-0.05711	0.00612	-9.33	<.0001
lucode10	23,CONFINED FEEDING OPS	1	-0.11308	0.07109	-1.59	0.1117
lucode11	24,OTHER AGRICULTURAL LAND	1	0.20524	0.07643	2.69	0.0072
lucode13	32,SHRUB & BRUSH RANGELAND	1	-0.25025	0.14424	-1.73	0.0828
lucode14	41,DECIDUOUS FOREST LAND	1	-0.08770	0.00863	-10.16	<.0001
lucode15	42,EVERGREEN FOREST LAND	1	-0.04123	0.01194	-3.45	0.0006
lucode16	43, MIXED FOREST LAND	1	-0.05632	0.00805	-7.00	<.0001
lucode21	61,FORESTED WETLAND	1	-0.03831	0.02012	-1.90	0.0569
lucode24	73,SANDY AREA (NON-BEACH)	1	0.37643	0.22364	1.68	0.0923
city_urb1	rural	1	-0.17076	0.00755	-22.63	<.0001

Stations having absolute residuals larger than 2 are listed in Table 3 below.

Table 3: Outlier Stations From the Selected Model

Obs	uniq_id	AADT	baadt	paadt	raadt
1	1020008	7705	3.82839	6.29663	-2.46824
2	1020452	17	1.52956	3.57594	-2.04638
3	1040388	60	1.84809	4.14013	-2.29205

To ensure validity of the selected attributes, we ran the regression on one part of the dataset and used the results to predict for the other dataset. Residuals were all less than three in absolute value and more than 95% had absolute values less than two. An assumption of normality is also well supported.

6. Conclusions

The selected model is effective in capturing more than 72% of the variability in AADT counts (after a power transformation) in the statewide area. It is based on only 178 attributes instead of approximately 550 possible attributes. The model has been validated by splitting the statewide area data into approximately two equal parts.

APPENDIX A: SAS Program

```
LIBNAME dot "d:\ncdot\sasdata";
LIBNAME dotm "m:\ncdot";


ods trace on;
*OPTIONS PAGENO=1 NODATE NONUMBER PS=44 LS=120;      *landscape printing;
OPTIONS PAGENO=1 LS=120;    *landscape printing;
data ptc;set dot.ptc;run;
proc sort data=ptc;by county90 group90 tract90;run;
data demo;set dot.de;run;
data demo;set demo;if (county90=31 and tract90=9708 and group90=1 and land_km=0) then delete; run;
data demo;set demo;if (county90=49 and tract90=9613 and group90=7 and land_km=0) then delete; run;
proc sort data=demo;by county90 group90 tract90;run;
data ptcdemo;merge ptc demo;by county90 group90 tract90;run;
data t3;set ptcdemo;if (persons=0 and lucode=11);run;
data ptcdemo;set ptcdemo;if (uniq_id=.) then delete;run;
data ptcdemo_r; set ptcdemo;if (uniq_id=1080369|uniq_id=1080574|uniq_id=1080599) then delete;run;
data aaa;set ptcdemo;if rte_type=7;run;
proc sort data=ptcdemo;by co_num;run;
proc univariate data=ptcdemo noprint;
var aadt;
by co_num;
output out=hr mean=m sum=total n=n;
run;

proc freq data=dot.ptc;
table co_name;
run;

/*proc freq data=dotm.out1;
table co_name;
run;
*/
/*
data aaa;
set dotm.out1;
if (co_name='Cherokee'or co_name='Clay' or co_name='Graham' or co_name='Macon' or co_name='Swain'
or co_name='Jackson'or co_name='Haywood' or co_name='Transylv' or co_name='Madison' or
co_name='Buncombe'
or co_name='Henderso'or co_name='Yancey' or co_name='McDowell' or co_name='Polk' or co_name='Rutherfo'or
co_name='Mitchell');
run;

proc freq data=aaa;
table co_name;
run;

data bbb;
set dot.ptc;
if (co_name='Cherokee'or co_name='Clay' or co_name='Graham' or co_name='Macon' or co_name='Swain'
or co_name='Jackson'or co_name='Haywood' or co_name='Transylvania' or co_name='Madison' or
co_name='Buncombe'
```

```
or co_name='Henderson'or co_name='Yancey' or co_name='McDowell' or co_name='Polk' or
co_name='Rutherford'or co_name='Mitchell');
run;

proc freq data=bbb;
table co_name;
run;
*/
data testarea;
set dot.ptcdemo;
length uniq_idn 8;
length uniq_id8n 8;
length routecoden 8;
uniq_idn=uniq_id;
uniq_id8n=uniq_id8;
routecoden=routecode;
drop uniq_id uniq_id8 routecode comments rte_type;
RUN;
data t4;set testarea;if (persons=0 and lucode=11);run;
data testarea(rename=(rte_adjust=rte_type urban_flg=urbanflg));
set testarea;
label rte_adjust=rte_type urban_flg=urbanflg;
*drop rte_adjust;
run;

/*
data ddd;
set testarea;keep rte_type urbanflg;
run;
*/
data testarea;
set testarea;
uniq_id=uniq_idn;
uniq_id8=uniq_id8n;
routecode=routecoden;
drop uniq_idn uniq_id8n routecoden;
run;

data ptcdemo_test(rename=(urbanflg=municpal_f));set ptcdemo;
if (co_num=18|co_num=31|co_num=50|co_num=67|co_num=91) then delete;
run;

data ptcdemo;
set testarea ptcdemo_test;
run;


data fromid(rename=(fromst=uniq_id));
set dot.fromid;
*label fromst=uniq_id;
*drop fromst;
run;
proc sort data=ptcdemo;by uniq_id;run;

data ptcdemo;
```

```
merge fromid ptcdemo;
by uniq_id;
if n=. then delete;
run;

proc sort data=ptc;by uniq_id;run;

DATA PTCdemo;set ptcdemo;id=_n_; *proc print;
*The following variables did not come from DOT/Shannon, but were created by Tae-Young. Omit them.;
        *drop urb_no aadt_up lanes_new rte_new;
*The most recently created routecode variable is rte_adjust;
        if rte_type=7 then rte_type=4;
        rte_class=rte_type;
*Combine co_num & urb_num to create "replacement" for sips;
        co_urb=co_num*10000 + urb_num*(urb_num>0);
*As of 7/14/03, it's not clear what I should use for sips, so for now I'll just take the first 3 digits of uniq_id
        *--this assumes the correct id is uniq_id and not uniq_id8 ...;
        sips_new=floor(uniq_id/10000);
/*Leave the following data out of the model. Use model to
        predict at these sites, thus validating the model ...
*TSU updated some stations according to 2001 counts;
        if year_new>0 then do;
                year=year_new;
                aadt=aadt_new;
                month=month_01;
                day=day_01;
                count1=count1_01;
                count2=count2_01;
                end;
        *AADT_UP=AADT+(AADT_NEW-AADT)*(AADT_NEW>0);
*/
*Double LRS-based lanes for interstates;
        lanes=lanes + lanes*(rte_class=1);
        run;
/*
data bbb;set ptcdemo;if co_urb=0;run;
data ccc;set m;if (co_num=18|co_num=31|co_num=50|co_num=67|co_num=91);run;
data ddd;set ptcdemo;if
(co_name="Wake"|co_name="Johnston"|co_name="Orange"|co_name="Durham"|co_name="Chatham");run;
*/
data eee;set ptcdemo;if lanes=8 and rte_type=2;run;

proc freq data=ptcdemo; table lucode;run;
***************************** BEGIN: Summaries for report ****************************************;
ods html file='d:\NCDOT\OUTPUT\summaries.html';
ods select Freq.CO_URB_by_SIPS.CrossTabFreqs2 Freq.SIPS_NEW_by_SIPS.CrossTabFreqs4
   Freq.sips_by_CO_URB.CrossTabFreqs7
        Freq.SIPS_NEW_by_CO_URB.CrossTabFreqs8;
proc freq data=ptcdemo; tables (sips co_urb sips_new)*(sips co_urb sips_new);
title 'Frequency Tables for "Surrogate" Attributes'; run;
****************************** END: Summaries for report ****************************************;
ods html close;

***************************** BEGIN: PCA of 1990 Census Data
****************************************;
```

```
ods html file='d:\NCDOT\OUTPUT\pca_census.html';
*Prepare data for principal components analysis of census information ...;
data ptcdemo; set ptcdemo;
        *drop miliquar;                         *This only takes 0 values;
*Need population density (persons per area), but don't really know what census variables to use.
*Here are my "guesses" of variable meanings as of 7/15/03:
*       persons=total/actual number of persons in region      **will use this for numerator**
*       agege_03 < persons2 < persons  !!NOT TRUE!! PERSONS2 ACTUALLY SUMS TO REPORTED
VALUE! But I don't have things like male & female to sum to persons2, only for summing to persons.
*       samp_pop=persons
*       per_samp ??? people in region/people in county ??? -- takes values like 12, 14, 9, ...
*       land_km=area of region, in sq km (?), obtained from census     BUT IS SOMETIMES 0!!
*       shape_area=area of region, in sq m (?), obtained from ArcView          THIS IS NEVER 0;
        /*if popden=0 then delete; *Won't have census values; NOT TRUE!! */
        if persons=0 then persons=.0001;    *if persons=0 then delete;
        if housing=0 then housing=.0001;    *if housing=0 then delete;
        popden=persons/(shape_area/1000000); *per sq km;
        houseden=housing/(shape_area/1000000);
*Adjust some other census variables to be relative to "shape_area";
        families=families/(shape_area/1000000);
        houshold=houshold/(shape_area/1000000);
*Ignore census variables that cause singularities;
        drop female o_ethnic age15_19 age20_24 agegt_84 agege_05 agege_25 divorced singnoch
        juvenile occupied vacant rent_occ room_gt6 prunitg4 agg_valu persons2--per_samp cantengl
        noncitiz born_for frm_abrd com_gt44 gradschl notinwrk public_ manual famiwork inc_gt75
        inc_aggr prvwater prvsewer vehiclg2 aggvehcl;
        run;


*Create random numbers that will allow record selection for training & validating models;
data ptcdemo; set ptcdemo; myran=rannor(id); run;

OPTIONS PAGENO=1 NODATE NONUMBER PS=44 LS=120;
ods select eigenvalues screeplot orthrotfactpat;
/*prin comp anal for full data--demo*/ PROC FACTOR DATA=ptcdemo
SIMPLE METHOD=PRIN priors=one score
nfactor=7 SCREE ROTATE=VARIMAX outstat=spc7 reorder;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
title1 'Principal Components Analysis of PTCDEMO';
RUN;
proc score data=ptcdemo score=spc7 out=pc7;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
run;

data ptcdemo; set pc7; /*This adds new output from PCA to previous PCA output*/
data ptcdemo1; set ptcdemo; if myran<0 then delete; *if id>1715 then delete; *if (id/2=floor(id/2)) then delete; run;
ods select eigenvalues screeplot orthrotfactpat;
/*prin comp anal for 1st part of data--demo*/ PROC FACTOR DATA=ptcdemo1 METHOD=PRIN priors=one
score
nfactor=7 SCREE ROTATE=VARIMAX outstat=spc7 reorder;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
title1 'Principal Components Analysis of PTCDEMO1';
RUN;
proc score data=ptcdemo score=spc7 out=pc71;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
```

```
run;

data ptcdemo1; set pc71; /*This adds new output from PCA to previous PCA output*/
data ptcdemo2; set ptcdemo; if myran ge 0 then delete; *if id le 1715 then delete; *if (id/2 ne floor(id/2)) then delete;
run;
ods select eigenvalues screeplot orthrotfactpat;
/*prin comp anal for 2nd part of data--demo*/ PROC FACTOR DATA=ptcdemo2 METHOD=PRIN priors=one
score
nfactor=7 SCREE ROTATE=VARIMAX outstat=spc72 reorder;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
title1 'Principal Components Analysis of PTCDEMO2';
RUN;
proc score data=ptcdemo score=spc72 out=pc72;
VAR popden houseden LAND_KM water_km families--o_noinst ownr_occ--vehicl_2;
run;

data ptcdemo2; set pc72; run;
***************************** END: PCA of 1990 Census Data
****************************************;
ods html close; run;

***************************** BEGIN: Model selection ****************************************;
*OPTIONS PAGENO=1 NODATE NONUMBER PS=40 LS=120;
ods html file='d:\NCDOT\OUTPUT\selection.html';
data ptcdemo; set ptcdemo;          /* Create new variables, not class variables */
*Create interactions between lanes, route classification, and speed;
        lanes_rte=lanes*100 + rte_class;
        speed_rte=speed*100 + rte_class;
        lanes_speed_rte=lanes*100000 + speed*100 + rte_class;
*Create interactions between cycle year and route surface;
        cycle_surface=100*(cycle_yr="A") + 200*(cycle_yr="O") +
                                300*(cycle_yr="E") + 400*(cycle_yr="V") +
                                rd_surface;
*Create interactions between municipality and urban flags;
        city_urb=100*(municpal_f=1);
*Change missing values for month and day to zeros;
        if month=. then month=0;
        if day=. then day=0;
* DROP VARIABLES;
drop countyid comments urbanflg countid stationid urb_num1 case revised blocknum state00 county00 tract00
group00
urb_no aadt_up lanes_new rte_new co_num_urb_no_1 l_n_rte_new_1 speed_rte_new_1 lanes_new_speed n m
total;
        run;
proc freq data=ptcdemo;   tables level2 rte_class sips_new co_urb lanes_rte speed_rte lanes_speed_rte
cycle_surface a_control year month day lucode city_urb;
title1 'Frequency tables for all class variables of station data'; run;

data ptcdemo;set ptcdemo;if speed=22 then speed=25;run;

/*Class variables created using IML*/ proc iml;
        use ptcdemo;
        read all var{rte_class} into rte_class; rte_class=design(rte_class)[,1:4];
        varnames='rte_class1':'rte_class99';
        create temp1 from rte_class[colname=varnames]; append from rte_class;
```

```
use ptcdemo;
read all var{co_urb} into co_urb; co_urb=design(co_urb)[,1:127];
varnames='co_urb1':'co_urb199';
create temp2 from co_urb[colname=varnames]; append from co_urb;
use ptcdemo;
read all var{lanes_rte} into lanes_rte; lanes_rte=design(lanes_rte)[,1:22];
varnames='lanes_rte1':'lanes_rte99';
create temp3 from lanes_rte[colname=varnames]; append from lanes_rte;
use ptcdemo;
read all var{speed_rte} into speed_rte; speed_rte=design(speed_rte)[,1:33];
varnames='speed_rte1':'speed_rte99';
create temp4 from speed_rte[colname=varnames]; append from speed_rte;
use ptcdemo;
read all var{lanes_speed_rte} into lanes_speed_rte; lanes_speed_rte=design(lanes_speed_rte)[,1:105];
varnames='lanes_speed_rte1':'lanes_speed_rte199';
create temp5 from lanes_speed_rte[colname=varnames]; append from lanes_speed_rte;
use ptcdemo;
read all var{cycle_surface} into cycle_surface; cycle_surface=design(cycle_surface)[,1:9];
varnames='cycle_surface1':'cycle_surface99';
create temp6 from cycle_surface[colname=varnames]; append from cycle_surface;
use ptcdemo;
read all var{a_control} into a_control; a_control=design(a_control)[,1:2];
varnames='a_control1':'a_control99';
create temp7 from a_control[colname=varnames]; append from a_control;
use ptcdemo;
read all var{year} into year; year=design(year)[,1:4];
varnames='year1':'year99';
create temp8 from year[colname=varnames]; append from year;
use ptcdemo;
read all var{month} into month; month=design(month)[,2:13];
varnames='month1':'month99';
create temp9 from month[colname=varnames]; append from month;
use ptcdemo;
read all var{day} into day; day=design(day)[,2:5];
varnames='day1':'day99';
create temp10 from day[colname=varnames]; append from day;
use ptcdemo;
read all var{lucode} into lucode; lucode=design(lucode)[,2:27];
varnames='lucode1':'lucode99';
create temp11 from lucode[colname=varnames]; append from lucode;
use ptcdemo;
read all var{city_urb} into city_urb; city_urb=design(city_urb)[,1];
varnames='city_urb1':'city_urb99';
create temp12 from city_urb[colname=varnames]; append from city_urb;
/*Merge class variables*/data ptcdemo;
merge ptcdemo temp1 temp2 temp3 temp4 temp5 temp6 temp7 temp8 temp9 temp10 temp11 temp12;
label
factor1='Other' factor2='Single' factor3='College' factor4='Poor' factor5='Farm' factor6='Elderly' factor7='Wealthy'

rte_class1='INTER' rte_class2='US' rte_class3='NC' rte_class4='SR' rte_class5='LOCAL'

co_urb1='Alamance'co_urb2='Alamance,101'co_urb3='Alexander'co_urb4          ='Alexander,114'
co_urb5 ='Alleghary'co_urb6          ='Anson'co_urb7  ='Ashe'co_urb8    ='Averry'co_urb9 ='Beanfort'
co_urb10='Bertie'co_urb11='Bladen'co_urb12='Brunswick'co_urb13='Brunswick,109'co_urb14='Buncombe'
co_urb15='Buncombe,100'co_urb16='Burke'co_urb17='Burke,114'co_urb18='Cabamus'co_urb19='Cabamus,111'
```

co_urb20='Caldwell'co_urb21='Caldwell,114'co_urb22='Camden'co_urb23='Carteret'co_urb24='Caswell'
co_urb25='Catawba'co_urb26='Catawba,114'co_urb27='Chatham'co_urb28='Chatham,103'co_urb29='Cherokee'
co_urb30='Chowan'co_urb31='Clay'co_urb32='Cleveland'co_urb33='Columbus'co_urb34='Craven'
co_urb35='Cumberland,104'co_urb36='Cumituck'co_urb37='Dare'co_urb38='Davison'
co_urb39='Davison,107'co_urb40='Daive'co_urb41='Duplin'co_urb42='Durham'co_urb43='Durham,103'
co_urb44='Edgecombe'co_urb45='Edgecombe,116'co_urb46='Forsyth' co_urb47='Forsyth,107',co_urb48='Franklin'
co_urb49='Gaston'
co_urb50='Gaston,105'co_urb51='Gates'co_urb52='Graham'co_urb53='Granville'co_urb54='Greene'
co_urb55='Guilford'co_urb56='Guilford,101' co_urb57='Guilford,106' co_urb58='Guilford,107',
co_urb59='Halifax'co_urb60='Harnett'co_urb61='Harnett,104',co_urb62='Haywood'
co_urb63='Henderson'co_urb64='Hertford'co_urb65='Hoke'co_urb66='Hyde'co_urb67='Iredell'
co_urb68='Jackson'co_urb69='Johnston'co_urb70='Jones'co_urb71='Lee'co_urb72='Lenoir'
co_urb73='Lincoln'co_urb74='Macon'co_urb75='Madison'co_urb76='Martin'co_urb77='McDowell'
co_urb78='Mecklenburg'co_urb79='Mecklenburg,102'co_urb80='Mitchell'co_urb81='Montgomery'co_urb82='Moor
e'
co_urb83='Nash'co_urb84='Nash,116'co_urb85='New Hanover,109'co_urb86='Northampton'co_urb87='Onslow'
co_urb88='Onslow,112'co_urb89='Orange'co_urb90='Orange,101'co_urb91='Orange,103'co_urb92='Pamlico'
co_urb93='Pasquotank'co_urb94='Pender'co_urb95='Perquimans'co_urb96='Person'co_urb97='Pitt'
co_urb98='Pitt, 115'co_urb99='Pork'co_urb100='Randolph'co_urb101='Randolph,107'co_urb102='Richmond'
co_urb103='Robeson'co_urb104='Rockingham'co_urb105='Rowan'co_urb106='Rowan,111'co_urb107='Rutherford'
co_urb108='Sampson'co_urb109='Scotland'co_urb110='Stanly'co_urb111='Stokes'co_urb112='Surry'
co_urb113='Swain'co_urb114='Transylvania'co_urb115='Tyrrell'co_urb116='Union'co_urb117='Vance'
co_urb118='Wake'co_urb119='Wake,108'co_urb120='Warren'co_urb121='Washington'co_urb122='Watauga'
co_urb123='Wayne'co_urb124='Wayne,113'co_urb125='Wilkes'co_urb126='Wilson'co_urb127='Yadkin'co_urb128
='Yancey'

lanes_rte1='1, US' lanes_rte2='1, SR' lanes_rte3='2, INTER' lanes_rte4='2, US'
lanes_rte5='2, NC' lanes_rte6='2, SR' lanes_rte7='2, LOCAL' lanes_rte8='3, US'
lanes_rte9='3, NC' lanes_rte10='3, SR'lanes_rte11='4, INTER' lanes_rte12='4, US'
lanes_rte13='4, NC' lanes_rte14='4, SR' lanes_rte15='5, US'lanes_rte16='5, SR'
lanes_rte17='6, INTER'lanes_rte18='6, US' lanes_rte19='6, NC' lanes_rte20='6, SR'
lanes_rte21='8, INTER' lanes_rte22='8, US'lanes_rte23='12, INTER'

speed_rte1='20, US'speed_rte2='20, NC' speed_rte3='20, SR'
speed_rte4='25, US'speed_rte5='25, NC' speed_rte6='25, SR' speed_rte7='25, LOCAL'
speed_rte8='30, US'speed_rte9='30, NC' speed_rte10='30, SR'
speed_rte11='35, US' speed_rte12='35, NC' speed_rte13='35, SR'
speed_rte14='40, US' speed_rte15='40, NC' speed_rte16='40, SR'
speed_rte17='45, INTER' speed_rte18='45, US' speed_rte19='45, NC'speed_rte20='45, SR'
speed_rte21='50, INTER' speed_rte22='50, US' speed_rte23='50, NC'speed_rte24='50, SR'
speed_rte25='55, INTER' speed_rte26='55, US' speed_rte27='55, NC'speed_rte28='55, SR'
speed_rte29='60, INTER' speed_rte30='60, US' speed_rte31='60, NC'
speed_rte32='65, INTER' speed_rte33='65, US'
speed_rte34='70, INTER'

lanes_speed_rte1='1, 20, US' lanes_speed_rte2='1, 35, US' lanes_speed_rte3='1, 35, SR'
lanes_speed_rte4='1, 55, SR' lanes_speed_rte5='2, 20, US' lanes_speed_rte6='2, 20, NC'
lanes_speed_rte7='2, 20, SR' lanes_speed_rte8='2, 25, US' lanes_speed_rte9='2, 25, NC'
lanes_speed_rte10='2, 25, SR' lanes_speed_rte11='2, 25, LOCAL' lanes_speed_rte12='2, 30, US'
lanes_speed_rte13='2, 30, NC' lanes_speed_rte14='2, 30, SR' lanes_speed_rte15='2, 35, US'
lanes_speed_rte16='2, 35, NC' lanes_speed_rte17='2, 35, SR' lanes_speed_rte18='2, 40, US'
lanes_speed_rte19='2, 40, NC' lanes_speed_rte20='2, 40, SR' lanes_speed_rte21='2, 45, US'
lanes_speed_rte22='2, 45, NC' lanes_speed_rte23='2, 45, SR' lanes_speed_rte24='2, 50, US'
lanes_speed_rte25='2, 50, NC' lanes_speed_rte26='2, 50, SR' lanes_speed_rte27='2, 55, INTER'
lanes_speed_rte28='2, 55, US' lanes_speed_rte29='2, 55, NC' lanes_speed_rte30='2, 55, SR'

lanes_speed_rte31='2, 60, US' lanes_speed_rte32='2, 65, INTER' lanes_speed_rte33='2, 65, US'
lanes_speed_rte34='2, 70, INTER'
lanes_speed_rte35='3, 20, US' lanes_speed_rte36='3, 20, SR' lanes_speed_rte37='3, 25, US'
lanes_speed_rte38='3, 25, SR' lanes_speed_rte39='3, 30, SR' lanes_speed_rte40='3, 35, US'
lanes_speed_rte41='3, 35, NC' lanes_speed_rte42='3, 35, SR' lanes_speed_rte43='3, 40, US'
lanes_speed_rte44='3, 40, NC' lanes_speed_rte45='3, 45, US' lanes_speed_rte46='3, 45, NC'
lanes_speed_rte47='3, 45, SR'
lanes_speed_rte48='3, 55, US' lanes_speed_rte49='3, 55, NC' lanes_speed_rte50='3, 55, SR'
lanes_speed_rte51='4, 20, US' lanes_speed_rte52='4, 20, NC' lanes_speed_rte53='4, 20, SR'
lanes_speed_rte54='4, 25, US' lanes_speed_rte55='4, 25, NC' lanes_speed_rte56='4, 25, SR'
lanes_speed_rte57='4, 30, US' lanes_speed_rte58='4, 35, US' lanes_speed_rte59='4, 35, NC'
lanes_speed_rte60='4, 35, SR' lanes_speed_rte61='4, 40, US' lanes_speed_rte62='4, 40, NC'
lanes_speed_rte63='4, 40, SR' lanes_speed_rte64='4, 45, US'
lanes_speed_rte65='4, 45, NC' lanes_speed_rte66='4, 45, SR' lanes_speed_rte67='4, 50, INTER'
lanes_speed_rte68='4, 50, US' lanes_speed_rte69='4, 50, NC' lanes_speed_rte70='4, 50, SR'
lanes_speed_rte71='4, 55, INTER' lanes_speed_rte72='4, 55, US' lanes_speed_rte73='4, 55, NC'
lanes_speed_rte74='4, 55, SR' lanes_speed_rte75='4, 60, INTER' lanes_speed_rte76='4, 60, NC'
lanes_speed_rte77='4, 65, INTER' lanes_speed_rte78='4, 65, US'lanes_speed_rte79='4, 70, INTER'
lanes_speed_rte80='5, 35, US' lanes_speed_rte81='5, 35, SR' lanes_speed_rte82='5, 45, US'
lanes_speed_rte83='5, 45, SR' lanes_speed_rte84='5, 50, SR' lanes_speed_rte85='5, 55, SR'
lanes_speed_rte86='6, 35, US' lanes_speed_rte87='6, 35, NC' lanes_speed_rte88='6, 35, SR'
lanes_speed_rte89='6, 40, US' lanes_speed_rte90='6, 45, US' lanes_speed_rte91='6, 45, NC'
lanes_speed_rte92='6, 45, SR' lanes_speed_rte93='6, 50, US'
lanes_speed_rte94='6, 55, INTER' lanes_speed_rte95='6, 55, US' lanes_speed_rte96='6, 55, SR'
lanes_speed_rte97='6, 60, INTER' lanes_speed_rte98='6, 65, INTER' lanes_speed_rte99='6, 65, US'
lanes_speed_rte100='6, 70, INTER' lanes_speed_rte101='8, 45, INTER' lanes_speed_rte102='8, 45, US'
lanes_speed_rte103='8, 55, INTER' lanes_speed_rte104='8, 60, INTER' lanes_speed_rte105='8, 65, INTER'
lanes_speed_rte106='12, 55, INTER'

cycle_surface1='Ann, Asph' cycle_surface2='Ann, Conc' cycle_surface3='Ann, Soil'
cycle_surface4='Odd, Conc' cycle_surface5='Odd, Soil'
cycle_surface6='Even, Conc' cycle_surface7='Even, Soil'
cycle_surface8='Vari, Asph' cycle_surface9='Vari, Conc' cycle_surface10='Vari, Soil'

a_control1='Low' a_control2='Medium'

year1='1997' year2='1998' year3='1999' year4='2000'year5='2001'
month1='Jan' month2='Feb' month3='Mar' month4='Apr' month5='May' month6='June' month7='July' month8='Aug'
month9='Sept' month10='Oct' month11='Nov' month12='Dec'
day1='Mon' day2='Tue' day3='Wed' day4='Thu'

lucode1='11,Residential' lucode2='12,Comm & Services' lucode3='13,Industrial'
lucode4='14,Trans,Comm,Util' lucode5='15,INDUST & COMMERC CMPLXS '
lucode6='16,MXD URBAN OR BUILT-UP' lucode7='17,OTHER URBAN OR BUILT-U'
lucode8='21,CROPLAND AND PASTURE' lucode9='22,ORCHGROVVNYRDNURSORN'
lucode10='23,CONFINED FEEDING OPS'
lucode11='24,OTHER AGRICULTURAL LAND' lucode12='31,HERBACEOUS RANGELAND'
lucode13='32,SHRUB & BRUSH RANGELAND' lucode14='41,DECIDUOUS FOREST LAND'
lucode15='42,EVERGREEN FOREST LAND' lucode16='43, MIXED FOREST LAND' lucode17='51,STREAMS AND CANALS'
lucode18='52,LAKES' lucode19='53,RESERVOIRS'
lucode20='54,BAYS AND ESTUARIES' lucode21='61,FORESTED WETLAND'
lucode22='62,NONFORESTED WETLAND' lucode23='72,BEACHES' lucode24='73,SANDY AREA (NON-BEACH)'
lucode25='75,STRIP MINES' lucode26='76,TRANSITIONAL AREAS'

```
city_urb1='rural' city_urb2='urban';
run;

data tem;set ptcdemo;if factor1--factor7=. ;run;

ods select boxcox;

/*Run full model and determine Box-Cox transformation*/
/*PROC TRANSREG DATA=ptcdemo SS2 DETAILS OUTTEST=B;
MODEL BOXCOX(AADT/LAMBDA=-2 -1 -0.5 TO 0.5 BY 0.05 1 2 CONVENIENT ALPHA=0.00001)
=IDENTITY(
        factor1--factor7 rte_class1--city_urb1);
OUTPUT OUT=OUT5 RDPREFIX=R PPREFIX=P;
title 'Box-Cox regression using full model (PCA of census + all station data)';
RUN;
*/

/*Best Box-Cox power is 0.15*/ data ptcdemo; set ptcdemo; baadt=aadt**(.15); run;
/*Run full model using aadt^0.15*/proc reg data=ptcdemo;
model baadt=factor1--factor7 rte_class1--city_urb1;
output out=out1 p=pred1 r=raadt;
title1 "aadt^.15 regression using full model (PCA of census + all station data)"; run;
/* NORMALITY check*/ PROC UNIVARIATE DATA=OUT1 NORMAL PLOT; VAR RAADT;
title2 'Normality check on residuals'; RUN;
data mycopy; set ptcdemo; run;
data ptcdemo; set mycopy; run;
ods select step192.anova step192.selparmest selectionsummary;
/*Do model selection using aadt^.15*/proc reg data=ptcdemo;
model baadt=factor1--factor7 rte_class1--city_urb1 / selection=stepwise details=steps;
output out=out1 p=pred1 r=raadt;
title1 "Stepwise Selection for aadt^.15 regression from full model (PCA of census + all station data)"; run; quit;


/*Fit final model using aadt^.15*/proc reg data=ptcdemo outest=est;
model baadt=Factor1 Factor2 Factor3 Factor4 Factor6
rte_class1 rte_class2 rte_class4
co_urb2 co_urb5 co_urb6 co_urb7 co_urb8
co_urb9 co_urb10 co_urb11 co_urb12 co_urb13
co_urb14 co_urb16 co_urb17 co_urb18
co_urb19 co_urb20 co_urb21 co_urb22
co_urb23 co_urb24 co_urb25 co_urb26
co_urb27 co_urb29 co_urb30 co_urb31
co_urb34 co_urb35 co_urb36 co_urb37 co_urb38
co_urb39 co_urb43 co_urb44 co_urb45 co_urb46 co_urb48
co_urb50 co_urb52 co_urb53 co_urb55 co_urb57 co_urb58
co_urb60 co_urb61 co_urb62 co_urb63 co_urb64 co_urb65 co_urb66
co_urb67 co_urb69 co_urb70 co_urb71 co_urb72 co_urb73
co_urb74 co_urb75 co_urb76 co_urb77 co_urb80 co_urb81 co_urb82 co_urb84
co_urb86 co_urb87 co_urb88 co_urb89
co_urb90 co_urb92 co_urb93 co_urb97 co_urb98 co_urb99 co_urb100 co_urb101 co_urb102
co_urb103 co_urb104 co_urb106 co_urb107
co_urb109 co_urb111 co_urb113 co_urb114 co_urb115 co_urb116 co_urb117 co_urb118 co_urb119
co_urb120 co_urb121 co_urb123 co_urb125 co_urb127
lanes_rte2 lanes_rte4 lanes_rte5 lanes_rte6 lanes_rte7
```

```
lanes_rte8 lanes_rte13 lanes_rte14 lanes_rte16 lanes_rte19 lanes_rte21
speed_rte1 speed_rte3 speed_rte5 speed_rte7
speed_rte12 speed_rte13 speed_rte14 speed_rte20 speed_rte21 speed_rte27 speed_rte30
lanes_speed_rte5 lanes_speed_rte8 lanes_speed_rte11 lanes_speed_rte26 lanes_speed_rte31 lanes_speed_rte40
lanes_speed_rte45 lanes_speed_rte49 lanes_speed_rte65
lanes_speed_rte73 lanes_speed_rte74 lanes_speed_rte76 lanes_speed_rte83
lanes_speed_rte93 lanes_speed_rte95 lanes_speed_rte99 lanes_speed_rte102 lanes_speed_rte104
cycle_surface1 cycle_surface2 cycle_surface4
cycle_surface5 cycle_surface6 cycle_surface7 cycle_surface8 cycle_surface9
a_control2 year4 month1 month5 month6 month7 month8 month9
month10 month11 day1 day4 lucode2 lucode3 lucode4 lucode5 lucode7 lucode8 lucode10 lucode11 lucode13
lucode14 lucode15 lucode16 lucode21 lucode24 city_urb1;
output out=dotm.out1 p=paadt r=raadt;
title1 "Fit final model for aadt^.15 regression from full model (PCA of census + all station data)"; run; quit;

data dotm.out2;
set dotm.out1;
*if (co_name='Guilford'or co_name='Forsyth' or co_name='Davidson' or co_name='Alamance' or
co_name='Randolph');
keep baadt Factor1 Factor2 Factor3 Factor4 Factor6
rte_class1 rte_class2 rte_class4
co_urb2 co_urb5 co_urb6 co_urb7 co_urb8
co_urb9 co_urb10 co_urb11 co_urb12 co_urb13
co_urb14 co_urb16 co_urb17 co_urb18
co_urb19 co_urb20 co_urb21 co_urb22
co_urb23 co_urb24 co_urb25 co_urb26
co_urb27 co_urb29 co_urb30 co_urb31
co_urb34 co_urb35 co_urb36 co_urb37 co_urb38
co_urb39 co_urb43 co_urb44 co_urb45 co_urb46 co_urb48
co_urb50 co_urb52 co_urb53 co_urb55 co_urb57 co_urb58
co_urb60 co_urb61 co_urb62 co_urb63 co_urb64 co_urb65 co_urb66
co_urb67 co_urb69 co_urb70 co_urb71 co_urb72 co_urb73
co_urb74 co_urb75 co_urb76 co_urb77 co_urb80 co_urb81 co_urb82 co_urb84
co_urb86 co_urb87 co_urb88 co_urb89
co_urb90 co_urb92 co_urb93 co_urb97 co_urb98 co_urb99 co_urb100 co_urb101 co_urb102
co_urb103 co_urb104 co_urb106 co_urb107
co_urb109 co_urb111 co_urb113 co_urb114 co_urb115 co_urb116 co_urb117 co_urb118 co_urb119
co_urb120 co_urb121 co_urb123 co_urb125 co_urb127
lanes_rte2 lanes_rte4 lanes_rte5 lanes_rte6 lanes_rte7
lanes_rte8 lanes_rte13 lanes_rte14 lanes_rte16 lanes_rte19 lanes_rte21
speed_rte1 speed_rte3 speed_rte5 speed_rte7
speed_rte12 speed_rte13 speed_rte14 speed_rte20 speed_rte21 speed_rte27 speed_rte30
lanes_speed_rte5 lanes_speed_rte8 lanes_speed_rte11 lanes_speed_rte26 lanes_speed_rte31 lanes_speed_rte40
lanes_speed_rte45 lanes_speed_rte49 lanes_speed_rte65
lanes_speed_rte73 lanes_speed_rte74 lanes_speed_rte76 lanes_speed_rte83
lanes_speed_rte93 lanes_speed_rte95 lanes_speed_rte99 lanes_speed_rte102 lanes_speed_rte104
cycle_surface1 cycle_surface2 cycle_surface4
cycle_surface5 cycle_surface6 cycle_surface7 cycle_surface8 cycle_surface9
a_control2 year4 month1 month5 month6 month7 month8 month9
month10 month11 day1 day4 lucode2 lucode3 lucode4 lucode5 lucode7 lucode8 lucode10 lucode11 lucode13
lucode14 lucode15 lucode16 lucode21 lucode24 city_urb1;
;
run;

data dotm.residualarea;
```

```
set dotm.out1;
keep raadt uniq_id rte_type co_name;
run;


data dotm.residualareaup;
set dotm.residualarea;
if (co_name='Guilford'or co_name='Forsyth' or co_name='Davidson' or co_name='Alamance' or
co_name='Randolph');
run;

proc freq data=dotm.residualarea;
table co_name;
run;

proc freq data=dotm.residualareaup;
table co_name;
run;

****************************** END: Model selection ****************************************;
ods html close;

libname ncdot "d:\Spatial";

/*
proc sql;
create table dotm.dist as
select distinct fromst, tost, dist, cost
from ncdot.dist;
quit;
*/

data dotm.residualareaup(rename=(uniq_id=fromst));
set dotm.residualareaup;
id=_n_;
RUN;
proc sort data=dotm.residualareaup;by fromst;run;

data dotm.area(rename=(raadt=r1 rte_type=rte_type1));
merge dotm.residualareaup(in=inr) ncdot.dist(in=ind); by fromst;if inr;
keep rte_type fromst raadt tost dist cost;
run;

proc sort data=dotm.area nodupkey;
by fromst tost dist cost;
run;

data dotm.residualareaup1(rename=(fromst=tost));
set dotm.residualareaup;
id=_n_;
RUN;

proc sort data=dotm.residualareaup1;
by tost;
run;
```

```
proc sort data=dotm.area;
by tost;
run;

data dotm.area1(rename=(raadt=r1 rte_type=rte_type1));
merge dotm.residualareaup1(in=inr) dotm.area(in=ind); by tost;if inr;
keep rte_type fromst raadt tost dist cost;
run;

data dotm.area1;
set dotm.area1;
if dist=. then delete;
run;


/*
data bb;
set dotm.area1;
if tost=310009;
run;

data cc;
set dotm.area1;
if fromst=310009;
run;
*/
/*
proc sql;
create table dotm.id as
select distinct fromst
from dotm.area;
quit;
*/
/*
data aa;
merge dotm.residualareaup dotm.id;
by fromst;
run;
data aaa;set aa; if fromst=190001;run;
*/
/*
data bbb;
set ncdot.dist;
if fromst=180003 and tost=180069;
run;
*/


proc iml;reset nolog;;
distance=j(4096,4096,160000);
*distance=diag(0); *check;
use dotm.residualareaup(keep=fromst);
read all into x;
close dotm.residualareaup;
```

```
use dotm.area1;
read all into raw;
close dotm.area1;
do i=1 to nrow(raw);
distance[loc(x=raw[i,4]),loc(x=raw[i,1])]=raw[i,5];
end;
do i=1 to nrow(distance)-1;
do j=i+1 to nrow(distance);
distance[j,i]=distance[i,j];
end;
distance[i,i]=0;
end;
distance[nrow(distance),nrow(distance)]=0;
nugget=0.0946;sill=0.0503;range=676.1;
cov=sill*exp(-(distance#distance)/(range*range));
*cov=diag(nugget+sill); *check;
do i=1 to 4096; *Do with Diagonal function;
cov[i,i]=1+sill/nugget;
end;

icov=inv(cov);
p=root(icov);
create dotm.pp from p;
append from p;
create dotm.distance from distance;
append from distance;
create dotm.cov from cov;
append from cov;
quit;

proc iml;
use dotm.out2;
read all into xx;
close dotm.out2;
y=xx[,179]; *check;
ex=xx[,1:178];
use dotm.pp;
read all into p;
y_star=p*y; *element multiplication;
x_star=p*ex;
create dotm.ystar from y_star;
append from y_star;
create dotm.xstar from x_star;
append from x_star;
quit;

data dotm.ystar(rename=(col1=y));
set dotm.ystar;
run;

data dotm.ols;
merge dotm.ystar dotm.xstar;
run;

proc reg data=dotm.ols;
```

```
model y=col1--col178/cli clm;
output out=dotm.outreg predicted=yhat l95=li95 u95=ui95;
run;

data dotm.flag;
set dotm.outreg;
if (y<=li95 or y>=ui95) then flagi=0;
else flagi=1;
keep flagi li95 ui95;
run;

proc freq data=dotm.flag;
tables flagi;
run;


proc reg data=dotm.ols;
model y=col1--col178/cli alpha=0.01;
** Create output data set for predictions **
        UCL and LCL are the limits for the confidence intervals for an individual
        STDP is the standard error that will be used to create confidence intervals
        for the means.;
output out=dotm.outreg99 p=pred LCL=Li99 UCL=ui99 STDP=STDP;
run;

*proc print data=dotm.outreg99; *run;

data dotm.flag99;
set dotm.outreg99;
if (y<=li99 or y>=ui99) then flagi99=0;
else flagi99=1;
*if (y<=lm or y>=um) then flagm=0;
*else flagm=1;
keep flagi99 li99 ui99 y;
run;

data dotm.both;
merge dotm.flag dotm.flag99 dotm.residualareaup;
run;
data dotm.both5;
set dotm.both;
if (flagi=0 and flagi99=1) then flag=1;
else if (flagi=0 and flagi99=0) then flag=2;
else flag=0;
run;

proc freq data=dotm.both5;
tables flag;
run;

proc freq data=dotm.flag99;
tables flagi99;
run;
```

```
data dotm.pre(rename=(fromst=uniq_id));
set dotm.both5;
run;

proc iml;
use dotm.pp;
read all into p;
close dotm.pp;
use dotm.pre;
read all into pre;
close dotm.pre;
ip=inv(p);
y_lpred95=(ip*pre[,1]);
y_upred95=(ip*pre[,2]);
y_lpred99=(ip*pre[,5]);
y_upred99=(ip*pre[,6]);
aadt=(ip*pre[,4]);
temp=y_lpred95||y_upred95||y_lpred99||y_upred99||pre[,8]||aadt;
create dotm.prediction from temp[colname={lpred95 upred95 lpred99 upred99 uniq_id aadt}];
append from temp;
*create dotm.ylpred from y_lpred95;
*append from y_lpred95;
quit;


data dotm.final;
set dotm.prediction;
lpred95=lpred95**(1/0.15);
upred95=upred95**(1/0.15);
lpred99=lpred99**(1/0.15);
upred99=upred99**(1/0.15);
aadt_actual=aadt**(1/0.15);
run;


data dotm.final1;
set dotm.final;
if (aadt_actual<=lpred95 or aadt_actual>=upred95) then flagi95=0;
else flagi95=1;
if (aadt_actual<=lpred99 or aadt_actual>=upred99) then flagi99=0;
else flagi99=1;
if (flagi95=0 and flagi99=1) then flag=1;
else if (flagi95=0 and flagi99=0) then flag=2;
else flag=0;
drop flagi95 flagi99;
run;
proc freq data=dotm.final1;
tables flag;
run;

/*
proc iml;
use dotm.cov;
read all into so;
close dotm.cov;
```

```
use dotm.out2;
read all into xxx;
y=xxx[,179];
ex=xxx[,1:178];
soXstar=solve(so,ex);
soYstar=solve(so,y);
create dotm.soXstar from soXstar;
append from soXstar;
create dotm.soYstar from soYstar;
append from soYstar;
quit;


proc iml;
use dotm.cov;
read all into sp;
close dotm.cov;
call eigen(val, vec, sp);
sqeiv=diag(1/sqrt(val[,1]));
sq=vec*sqeiv*t(vec);
use dotm.out2;
read all into xxx;
y=xxx[,179];
ex=xxx[,1:178];
spXstar=sq*ex;
spYstar=sq*y;
create dotm.spXstar from spXstar;
append from spXstar;
create dotm.spYstar from spYstar;
append from spYstar;
quit;




*************************************************************************;
*options obs=20000;
/*
data ncdot.dist;
infile "d:\Spatial\sorted.txt" delimiter="," dsd missover;
input fromst tost dist cost;
run;
*/
/*
proc univariate data=ncdot.dist noprint;
by fromst;
var dist;
output out=ncdot.fromst sum=total mean=m n=n;
run;
*/
/*data temp; set ncdot.fromst; if n<1000; run;*/
/*
proc  sort data=ncdot.dist;by tost;run;
proc univariate data=ncdot.dist noprint;
```

```
by tost;
var dist;
output out=ncdot.tost sum=total mean=m n=n;
run;

proc sql;
create table froms as
select count(distinct fromst)
from ncdot.dist;
quit;

proc sql;
create table tos as
select count(distinct tost)
from ncdot.dist;
quit;
*/
/*
%macro adding(last);
  %do i= 1 %to &last;
    data nodot&i;
      infile "d:\\Spatial\sorted.txt" delimiter=",";
      input fromst tost dist cost; id=_n_;
      %if (_n_ >= (&i-1)*10+1) & (_n_ < &i*10)  %then stop;
    run;
            %if &i=1 %then %do;
              data total; set nodot&i; run;
            %end;
    %else %do;
              data total; set total nodot&i; run;
            %end;
  %end;
%mend;

%adding(2);
*/
/*
proc freq data=ncdot.dist;table fromst/list missing out=out;run;
proc freq data=ncdot.dist;table tost/list missing out=out1;run;
*
proc means data=ncdot.dist;
var fromst tost dist cost;
OUTPUT OUT=summary MIN=FMA TMA MAX=FMI TMI;
RUN;
*/

/*proc print data=residual(obs=10);run;*/
/*proc freq data=residual; table rte_type;run;

proc sort data=residual;
by uniq_id;
run;
/*
proc univariate data=new;
var resid5;
```

```
by rte_new;
run;
*/

/*
proc sort data=ncdot.dist out=ncdot.dist1;
by fromst;
run;
*/



/*
data dotm.temp(rename=(raadt=r1 rte_type=rte_type1));
set dotm.temp;
run;
*/
/*
data dotm.dist(rename=(uniq_id=fromst));
merge ncdot.residual(rename=(raadt=r1 rte_type=rte_type1)) ncdot.dist(rename=(fromst=uniq_id));
by uniq_id;
run;
*/
/*
data _null_;
set dotm.temp nobs=na;
call symput("na",na);
run;
%put &na;
*/
/*proc options option=work;run;*/
/*
proc sort data=dotm.temp sortsize=Max out=dotm.temp2;
by tost;
run;

data dotm.residual(rename=(fromst=uniq_id));
set dotm.residual;
*drop id SelectionProb SamplingWeight;
RUN;

data dotm.temp2(rename=(uniq_id=tost));
merge dotm.residual(rename=(raadt=r2 rte_type=rte_type2)) dotm.temp2(rename=(tost=uniq_id));
by uniq_id;
run;
* until here;

data dotm.temp2;
set dotm.temp2;
if tost=. then delete;
if fromst=. then delete;
if rte_type2=. then delete;
keep tost fromst r1 r2 rte_type1 rte_type2 dist cost;
run;
```

```
*/
/*
proc sort data=dotm.residualarea;
by rte_type;
run;
proc surveyselect data=dotm.residualarea
        method=sys samprate=15
        out=dotm.residuala;
   strata rte_type;
   run;

proc sort data=dotm.residuala;
by uniq_id;
run;
*/

proc sort data=dotm.residualarea;
by uniq_id; run;
data dotm.temp(rename=(uniq_id=fromst));
merge dotm.residualarea(rename=(raadt=r1 rte_type=rte_type1)) ncdot.dist(rename=(fromst=uniq_id));
by uniq_id;
run;

/*
data _null_;
set dotm.dist nobs=na;
call symput("na",na);
run;
%put &na;
*/
/*proc options option=work;run;*/

proc sort data=dotm.temp sortsize=Max out=dotm.temp2;
by tost;
run;

data dotm.residualarea(rename=(fromst=uniq_id));
set dotm.residualarea;
*drop id SelectionProb SamplingWeight;
RUN;

data dotm.temp2(rename=(uniq_id=tost));
merge dotm.residualarea(rename=(raadt=r2 rte_type=rte_type2)) dotm.temp2(rename=(tost=uniq_id));
by uniq_id;
run;



data dotm.temp2;
set dotm.temp2;
if tost=. then delete;
if fromst=. then delete;
if rte_type2=. then delete;
keep co_name tost fromst r1 r2 rte_type1 rte_type2 dist cost;
run;
```

```
/*data a; set dotm.temp2; if dist>165000;run;*/

data dotm.rte1 dotm.rte2 dotm.rte3 dotm.rte4 dotm.rte5;
set dotm.temp2;
label cost='Time to traverse path-unit:second' dist='Actual road distance along path-unit:meters';
if (rte_type1=1)& (rte_type2=1) then output dotm.rte1;
else if (rte_type1=2)& (rte_type2=2) then output dotm.rte2;
else if (rte_type1=3)& (rte_type2=3) then output dotm.rte3;
else if (rte_type1=4)& (rte_type2=4) then output dotm.rte4;
else if (rte_type1=5)& (rte_type2=5) then output dotm.rte5;
run;

proc means data=dotm.rte1;
var dist cost;
OUTPUT OUT=rte1 MIN=FMA TMA MAX=FMI TMI;
RUN;

proc means data=dotm.rte2;
var dist cost;
OUTPUT OUT=rte2 MIN=FMA TMA MAX=FMI TMI;
RUN;

proc means data=dotm.rte3;
var dist cost;
OUTPUT OUT=rte3 MIN=FMA TMA MAX=FMI TMI;
RUN;

proc means data=dotm.rte4;
var dist cost;
OUTPUT OUT=rte4 MIN=FMA TMA MAX=FMI TMI;
RUN;

proc means data=dotm.rte5;
var dist cost;
OUTPUT OUT=rte5 MIN=FMA TMA MAX=FMI TMI;
RUN;

/*
data aa;
set dotm.temp2;
if dist<675;run;
*/
/*
symbol v=dot height=0.1 i=none;
proc gplot data=final.five3;
plot distance*cost/ haxis=axis1 vaxis=axis2;
axis1 order=0 to 7200 by 1000 major=none;
axis2 order=0 to 170000 by 10000 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Actual distance vs time traverse distance';
run ;
*/
/*
```

```
%macro lag(data,fdist,tdist);
proc iml;
use &data;
read all into e;
n=nrow(e[,1]);
q=choose((e[,7]>&fdist)&&(e[,7]<&tdist),1,0);
nh=sum(q);
d=&tdist;
ssq=0;
do i=1 to n;
if(q[i]=1) then do;
diff=e[i,6]-e[i,3];
ssq=ssq+diff*diff;
end;
end;
gammahat=(1/(2#nh))#ssq;
print d nh gammahat;
%mend lag;
/* for distance, all has lag 100 meters*/
/* for distance, interstate has lag 2000 meters */
/* for distance, US has lag 300 meters */
/* for distance, NC has lag 600 meters */
/* for distance, SR has lag 100 meters */
/* for distance, Local has lag 300 meters */
/*
%macro a(b,c,d);
%do b=&b %to &c %by &d;
%lag(data=final.rte1,fdist=&b,tdist=&b+&d)
create out var{d nh gammahat};
append from out;
%end;
%mend a;
%a(b=0,c=6000,d=3000);
*/

data dotm.temp3;
set dotm.temp2;
if cost=6000 then delete;
run;

%macro lag(type,data, fcost,tcost,f);
proc iml;
use &data;
read all into e;
n=nrow(e[,1]);
ctemp=j(int((&tcost-&fcost)/&f)+1,3,0);
do b=&fcost to &tcost by &f;
d=(b+&f)/2;
ssq=0;
q=choose((e[,9]>b)&&(e[,9]<(b+&f)),1,0);
nh=sum(q);
  do i=1 to n;
    if(q[i]=1) then do;
diff=e[i,7]-e[i,4];
ssq=ssq+diff*diff;
```

```
    end;
  end;
gammahat=(1/(2#nh))#ssq;
ctemp[int((b-&fcost)/&f)+1,]=d||nh||gammahat;
end;
*print temp;
create dotm.clag&type from ctemp[colname={d nh gammahat}];
append from ctemp;
*close dotm.clag;
quit;
%mend lag;

*%lag(type=inter,data=dotm.rte1, fcost=0,tcost=20400,f=400);
*%lag(type=us,data=dotm.rte2, fcost=0,tcost=35750,f=200);
*%lag(type=nc,data=dotm.rte3, fcost=0,tcost=8800,f=100);
*%lag(type=sr,data=dotm.rte4, fcost=0,tcost=9000,f=20);
*%lag(type=local,data=dotm.rte5, fcost=0,tcost=8600,f=100);
%lag(type=all,data=dotm.temp3, fcost=0,tcost=6000,f=20);

****************************************** INTERSTATE ******************************************;
/* proc nlin for measurelength*/
data laginter;
set dotm.claginter;
label d='Cost' nh='number of points' gammahat='gammahat';
if d=. then delete;
*if nh<30 then delete;
if d>8000 then delete;
run;


/* rte_interstate*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
proc gplot data=laginter;
plot gammahat*d/haxis=axis1
                vaxis=axis2;
axis1 order=0 to 8000 by 100 major=none;
axis2 order=0 to 0.15 by 0.05 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Sample Variogram with Cost';
title2 'Interstate Road';
run ;

*Exponential;
proc nlin data=laginter method=gauss nohalve;
parms c0=0.03 a0=1000 nugg=0.02;
gamma=nugg+c0*(1-exp(-d/a0));
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(OLS)- with nugget';
run;

proc nlin data=laginter method=gauss nohalve;
parms c0=0.03 a0=1000 nugg=0.02;
gamma=nugg+c0*(1-exp(-d/a0));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
```

```
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(WLS)- with nugget';
run;

*Gaussian;
proc nlin data=laginter method=gauss nohalve;
parms c0=0.03 a0=1000 nugg=0.02;
gamma=nugg+c0*(1-exp(-d*d/(a0*a0)));
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(OLS)- with nugget';
run;

proc nlin data=laginter method=gauss nohalve;
parms c0=0.03 a0=1000 nugg=0.02;
gamma=nugg+c0*(1-exp(-d*d/(a0*a0)));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(WLS)- with nugget';
run;

*Spherical;

proc nlin data=laginter nohalve;
parms c0=0.03 a0=1000 nugg=0.02;
if (d < 1000) then gamma=nugg+c0*(3*d/(2*a0)-((b/a0)**3)/2);
if (d > 1000) then gamma=nugg+c0;
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(OLS)- with nugget';
run;

proc nlin data=laginter nohalve;
parms c0=0.03 a0=1000 nugg=0.02;
if (d < 1000) then gamma=nugg+c0*(3*d/(2*a0)-((d/a0)**3)/2);
if (d > 1000) then gamma=nugg+c0;
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(WLS)- with nugget';
run;

data route1_1;
  set nlinoutsn;
  vari = fittedsn ; type = 'W Spherical'; output ;
  vari = gammahat  ; type = 'regular'    ; output ;
  label vari='variogram';
run ;



 /*- Plot all the Variograma   ---------------------------------*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
symbol2 i=join l=1 v='square' c=black height=0.3;
```

```
proc gplot data=route1_1;
  plot vari*d=type/haxis=axis1
                   vaxis=axis2;
axis1 order=0 to 120000 by 3000 major=none;
axis2 order=0 to 1 by 0.1 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Theoretical vs Sample Variogram:';
title2 'Interstate';
run ;



************************************ US ********************************************;

/* proc nlin for measurelength*/
data lagus;
set dotm.clagus;
label d='Distance' nh='number of points' gammahat='gammahat';
if d=. then delete;
*if nh<30 then delete;
if d>10000 then delete;
run;


/* rte_interstate*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
proc gplot data=lagus;
plot gammahat*d/haxis=axis1
                   vaxis=axis2;
axis1 order=0 to 10000 by 100 major=none;
axis2 order=0 to 0.20 by 0.1 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Sample Variogram with Cost';
title2 'US Road';
run ;

*Exponential;
proc nlin data=lagus method=gauss nohalve;
parms c0=0.7 a0=800 nugg=0.5;
gamma=nugg+c0*(1-exp(-d/a0));
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(OLS)- with nugget';
run;

proc nlin data=lagus method=gauss nohalve;
parms c0=0.7 a0=800 nugg=0.5;
gamma=nugg+c0*(1-exp(-d/a0));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(WLS)- with nugget';
run;

*Gaussian;
proc nlin data=lagus method=gauss nohalve;
parms c0=0.7 a0=800 nugg=0.5;
```

```
gamma=nugg+c0*(1-exp(-d*d/(a0*a0)));
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(OLS)- with nugget';
run;

proc nlin data=lagus method=gauss nohalve;
parms c0=0.7 a0=800 nugg=0.5;
gamma=nugg+c0*(1-exp(-d*d/(a0*a0)));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(WLS)- with nugget';
run;

*Spherical;

proc nlin data=lagus nohalve;
parms c0=0.7 a0=800 nugg=0.5;
if (d < 800) then gamma=nugg+c0*(3*d/(2*a0)-((b/a0)**3)/2);
if (d > 800) then gamma=nugg+c0;
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(OLS)- with nugget';
run;

proc nlin data=lagus nohalve;
parms c0=0.7 a0=800 nugg=0.5;
if (d < 800) then gamma=nugg+c0*(3*d/(2*a0)-((d/a0)**3)/2);
if (d > 800) then gamma=nugg+c0;
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(WLS)- with nugget';
run;

data route1_1;
  set nlinoutsn;
  vari = fittedsn ; type = 'W Spherical'; output ;
  vari = gammahat  ; type = 'regular'    ; output ;
  label vari='variogram';
run ;



 /*- Plot all the Variograma   ----------------------------------*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
symbol2 i=join l=1 v='square' c=black height=0.3;
proc gplot data=route1_1;
  plot vari*d=type/haxis=axis1
                   vaxis=axis2;
axis1 order=0 to 120000 by 3000 major=none;
axis2 order=0 to 1 by 0.1 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Theoretical vs Sample Variogram:';
title2 'US Road';
```

```
run ;

************************************** NC ************************************;

/* proc nlin for measurelength*/
data lagnc;
set dotm.clagnc;
label d='Distance' nh='number of points' gammahat='gammahat';
if d=. then delete;
*if nh<30 then delete;
if d>3500 then delete;
*if d=250 or d=500 then delete;
run;


/* rte_interstate*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
proc gplot data=lagnc;
plot gammahat*d/haxis=axis1
                vaxis=axis2;
axis1 order=0 to 3500 by 100 major=none;
axis2 order=0 to 0.30 by 0.1 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Sample Variogram with Distance';
title2 'All route_NC Road';
run ;

*Exponential;
proc nlin data=lagnc method=gauss nohalve;
parms c0=0.08 a0=500 nugg=0.02;
gamma=nugg+c0*(1-exp(-3*d/a0));
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(OLS)- with nugget';
run;

proc nlin data=lagnc method=gauss nohalve;
parms c0=0.08 a0=500 nugg=0.02;
gamma=nugg+c0*(1-exp(-3*d/a0));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(WLS)- with nugget';
run;

*Gaussian;
proc nlin data=lagnc method=gauss nohalve;
parms c0=0.08 a0=500 nugg=0.01;
gamma=nugg+c0*(1-exp(-3*d*d/(a0*a0)));
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(OLS)- with nugget';
run;

proc nlin data=lagnc method=gauss nohalve;
parms c0=0.08 a0=500 nugg=0.02;
```

```
gamma=nugg+c0*(1-exp(-3*d*d/(a0*a0)));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(WLS)- with nugget';
run;

*Spherical;

proc nlin data=lagnc nohalve;
parms c0=0.08 a0=500 nugg=0.02;
if (d < 500) then gamma=nugg+c0*(3*d/(2*a0)-((b/a0)**3)/2);
if (d > 500) then gamma=nugg+c0;
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(OLS)- with nugget';
run;

proc nlin data=lagnc nohalve;
parms c0=0.08 a0=500 nugg=0.02;
if (d < 500) then gamma=nugg+c0*(3*d/(2*a0)-((d/a0)**3)/2);
if (d > 500) then gamma=nugg+c0;
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(WLS)- with nugget';
run;

data route1_1;
  set nlinoutsn;
  vari = fittedsn ; type = 'W Spherical'; output ;
  vari = gammahat  ; type = 'regular'    ; output ;
  label vari='variogram';
run ;



 /*- Plot all the Variograma   ---------------------------------*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
symbol2 i=join l=1 v='square' c=black height=0.3;
proc gplot data=route1_1;
  plot vari*d=type/haxis=axis1
                   vaxis=axis2;
axis1 order=0 to 120000 by 3000 major=none;
axis2 order=0 to 1 by 0.1 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Theoretical vs Sample Variogram:';
title2 'NC';
run ;


****************************************** SR ******************************************;
/* proc nlin for measurelength*/
data lagsr;
set dotm.clagsr;
label d='Cost' nh='number of points' gammahat='gammahat';
```

```
if d=. then delete;
if nh<30 then delete;
if d>3000 then delete;
run;


/* rte_interstate*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
proc gplot data=lagsr;
plot gammahat*d/haxis=axis1
                vaxis=axis2;
axis1 order=0 to 3000 by 100 major=none;
axis2 order=0 to 0.30 by 0.05 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Sample Variogram with Distance';
title2 'All route_interstate Road';
run ;

*Exponential;
proc nlin data=lagsr method=gauss nohalve;
parms c0=0.15 a0=300 nugg=0;
gamma=nugg+c0*(1-exp(-d/a0));
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(OLS)- with nugget';
run;

proc nlin data=lagsr method=gauss nohalve;
parms c0=0.15 a0=300 nugg=0;
gamma=nugg+c0*(1-exp(-d/a0));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(WLS)- with nugget';
run;

*Gaussian;
proc nlin data=lagsr method=gauss nohalve;
parms c0=0.15 a0=300 nugg=0;
gamma=nugg+c0*(1-exp(-d*d/(a0*a0)));
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(OLS)- with nugget';
run;

proc nlin data=lagsr method=gauss nohalve;
parms c0=0.15 a0=300 nugg=0;
gamma=nugg+c0*(1-exp(-d*d/(a0*a0)));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(WLS)- with nugget';
run;

*Spherical;
```

```
proc nlin data=lagsr nohalve;
parms c0=0.15 a0=300 nugg=0;
if (d < 300) then gamma=nugg+c0*(3*d/(2*a0)-((b/a0)**3)/2);
if (d > 300) then gamma=nugg+c0;
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(OLS)- with nugget';
run;

proc nlin data=lagsr nohalve;
parms c0=0.15 a0=300 nugg=0;
if (d < 300) then gamma=nugg+c0*(3*d/(2*a0)-((d/a0)**3)/2);
if (d > 300) then gamma=nugg+c0;
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(WLS)- with nugget';
run;

data route1_1;
  set nlinoutsn;
  vari = fittedsn ; type = 'W Spherical'; output ;
  vari = gammahat  ; type = 'regular'    ; output ;
  label vari='variogram';
run ;



 /*- Plot all the Variograma   ----------------------------------*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
symbol2 i=join l=1 v='square' c=black height=0.3;
proc gplot data=route1_1;
  plot vari*d=type/haxis=axis1
                   vaxis=axis2;
axis1 order=0 to 120000 by 3000 major=none;
axis2 order=0 to 1 by 0.1 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Theoretical vs Sample Variogram:';
title2 'SR Road';
run ;

************************************** Local ****************************************;

/* proc nlin for measurelength*/
data laglocal;
set dotm.claglocal;
label d='Distance' nh='number of points' gammahat='gammahat';
if d=. then delete;
*if nh<30 then delete;
if d>1000 then delete;
run;


/* rte_interstate*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
proc gplot data=laglocal;
```

```
plot gammahat*d/haxis=axis1
                vaxis=axis2;
axis1 order=0 to 1000 by 50 major=none;
axis2 order=0 to 0.50 by 0.1. offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Sample Variogram with Distance';
title2 'All route_Local Road';
run ;

*Exponential;
proc nlin data=laglocal method=gauss nohalve;
parms c0=0.35 a0=200 nugg=0.05;
gamma=nugg+c0*(1-exp(-d/a0));
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(OLS)- with nugget';
run;

proc nlin data=laglocal method=gauss nohalve;
parms c0=0.35 a0=200 nugg=0.05;
gamma=nugg+c0*(1-exp(-d/a0));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(WLS)- with nugget';
run;

*Gaussian;
proc nlin data=laglocal method=gauss nohalve;
parms c0=0.35 a0=200 nugg=0.05;
gamma=nugg+c0*(1-exp(-d*d/(a0*a0)));
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(OLS)- with nugget';
run;

proc nlin data=laglocal method=gauss nohalve;
parms c0=0.35 a0=200 nugg=0.05;
gamma=nugg+c0*(1-exp(-d*d/(a0*a0)));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(WLS)- with nugget';
run;

*Spherical;

proc nlin data=laglocal nohalve;
parms c0=0.35 a0=200 nugg=0.05;
if (d < 200) then gamma=nugg+c0*(3*d/(2*a0)-((b/a0)**3)/2);
if (d > 200) then gamma=nugg+c0;
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(OLS)- with nugget';
run;
```

```
proc nlin data=laglocal nohalve;
parms c0=0.35 a0=200 nugg=0.05;
if (d < 200) then gamma=nugg+c0*(3*d/(2*a0)-((d/a0)**3)/2);
if (d > 200) then gamma=nugg+c0;
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(WLS)- with nugget';
run;

data route1_1;
  set nlinoutsn;
  vari = fittedsn ; type = 'W Spherical'; output ;
  vari = gammahat  ; type = 'regular'    ; output ;
  label vari='variogram';
run ;



 /*- Plot all the Variograma   ----------------------------------*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
symbol2 i=join l=1 v='square' c=black height=0.3;
proc gplot data=route1_1;
  plot vari*d=type/haxis=axis1
                   vaxis=axis2;
axis1 order=0 to 120000 by 3000 major=none;
axis2 order=0 to 1 by 0.1 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Theoretical vs Sample Variogram:';
title2 'Local Road';
run ;

************************************** ALL ******************************************;
/* proc nlin for measurelength*/
data lagall;
set dotm.clagall;
label d='Distance' nh='number of points' gammahat='gammahat';
if d=. then delete;
*if nh<30 then delete;
if d>1000 then delete;
run;


/* rte_interstate*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
proc gplot data=lagall;
plot gammahat*d/haxis=axis1
                   vaxis=axis2;
axis1 order=0 to 3500 by 500 major=none;
axis2 order=0 to 0.30 by 0.1 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Sample Variogram with Distance';
title2 'All route_all Road';
run ;

*Exponential;
proc nlin data=lagall method=gauss nohalve;
```

```
parms c0=0.15 a0=300 nugg=0;
gamma=nugg+c0*(1-exp(-d/a0));
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(OLS)- with nugget';
run;

proc nlin data=lagall method=gauss nohalve;
parms c0=0.15 a0=300 nugg=0;
gamma=nugg+c0*(1-exp(-d/a0));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=route1 p=vhat r=resid;
title ' NLS for exponential model(WLS)- with nugget';
run;

*Gaussian;
proc nlin data=lagall method=gauss nohalve;
parms c0=0.13 a0=200 nugg=0;
gamma=nugg+c0*(1-exp(-d*d/(a0*a0)));
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(OLS)- with nugget';
run;

proc nlin data=lagall method=gauss nohalve;
parms c0=0.15 a0=300 nugg=0;
gamma=nugg+c0*(1-exp(-d*d/(a0*a0)));
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutgn p=fittedgn;
title ' NLS for gaussina model(WLS)- with nugget';
run;

*Spherical;

proc nlin data=lagall nohalve;
parms c0=0.15 a0=300 nugg=0;
if (d < 300) then gamma=nugg+c0*(3*d/(2*a0)-((b/a0)**3)/2);
if (d > 300) then gamma=nugg+c0;
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(OLS)- with nugget';
run;

proc nlin data=lagall nohalve;
parms c0=0.15 a0=1000 nugg=0;
if (d < 300) then gamma=nugg+c0*(3*d/(2*a0)-((d/a0)**3)/2);
if (d > 300) then gamma=nugg+c0;
_weight_= nh/(2*gamma**2);
model gammahat=gamma;
output out=nlinoutsn p=fittedsn;
title ' NLS for spherical model(WLS)- with nugget';
run;
```

```
data route1_1;
  set nlinoutsn;
  vari = fittedsn ; type = 'W Spherical'; output ;
  vari = gammahat  ; type = 'regular'    ; output ;
  label vari='variogram';
run ;



 /*- Plot all the Variograma   ----------------------------------*/
symbol1 i=join l=1 v=v='circle' c=black height=0.3 ;
symbol2 i=join l=1 v='square' c=black height=0.3;
proc gplot data=route1_1;
  plot vari*d=type/haxis=axis1
                   vaxis=axis2;
axis1 order=0 to 120000 by 3000 major=none;
axis2 order=0 to 1 by 0.1 offset=(0,0) minor=none label=(a=-90 r=90);
title1 'Theoretical vs Sample Variogram:';
title2 'Local Road';
run ;
```

APPENDIX B: Some details of Principal Components Analysis

Principal Components Analysis of PTCDEMO
The FACTOR Procedure
Initial Factor Method: Principal Components

Eigenvalues of the Correlation Matrix: Total = 147 Average = 1				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	65.5242808	45.8362184	0.4457	0.4457
2	19.6880624	9.3945541	0.1339	0.5797
3	10.2935084	2.4396778	0.0700	0.6497
4	7.8538306	4.5670192	0.0534	0.7031
5	3.2868113	0.5242448	0.0224	0.7255
6	2.7625665	0.3433602	0.0188	0.7443
7	2.4192063	0.1752181	0.0165	0.7607
8	2.2439882	0.4798951	0.0153	0.7760
9	1.7640931	0.2960274	0.0120	0.7880
10	1.4680657	0.1103746	0.0100	0.7980
11	1.3576911	0.0264856	0.0092	0.8072
12	1.3312055	0.0895052	0.0091	0.8163
13	1.2417003	0.0481652	0.0084	0.8247
14	1.1935351	0.0359607	0.0081	0.8328
15	1.1575744	0.0907978	0.0079	0.8407
16	1.0667766	0.0443502	0.0073	0.8480
17	1.0224264	0.0130121	0.0070	0.8549
18	1.0094143	0.0480343	0.0069	0.8618
19	0.9613800	0.0157187	0.0065	0.8683
20	0.9456613	0.0953490	0.0064	0.8748
21	0.8503123	0.0109939	0.0058	0.8806
22	0.8393185	0.0236260	0.0057	0.8863
23	0.8156925	0.0110502	0.0055	0.8918

Eigenvalues of the Correlation Matrix: Total = 147 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
24	0.8046423	0.0133938	0.0055	0.8973
25	0.7912486	0.0828527	0.0054	0.9027
26	0.7083959	0.0232973	0.0048	0.9075
27	0.6850986	0.0492905	0.0047	0.9122
28	0.6358081	0.0190993	0.0043	0.9165
29	0.6167089	0.0060298	0.0042	0.9207
30	0.6106791	0.0912260	0.0042	0.9248
31	0.5194531	0.0319292	0.0035	0.9284
32	0.4875239	0.0102553	0.0033	0.9317
33	0.4772687	0.0288518	0.0032	0.9349
34	0.4484168	0.0183971	0.0031	0.9380
35	0.4300197	0.0143493	0.0029	0.9409
36	0.4156704	0.0383005	0.0028	0.9437
37	0.3773699	0.0198132	0.0026	0.9463
38	0.3575567	0.0392452	0.0024	0.9487
39	0.3183115	0.0043319	0.0022	0.9509
40	0.3139796	0.0106780	0.0021	0.9530
41	0.3033015	0.0194651	0.0021	0.9551
42	0.2838365	0.0026555	0.0019	0.9570
43	0.2811809	0.0232949	0.0019	0.9589
44	0.2578860	0.0070817	0.0018	0.9607
45	0.2508044	0.0101089	0.0017	0.9624
46	0.2406954	0.0088696	0.0016	0.9640
47	0.2318258	0.0141698	0.0016	0.9656
48	0.2176560	0.0042463	0.0015	0.9671
49	0.2134097	0.0036212	0.0015	0.9685

Eigenvalues of the Correlation Matrix: Total = 147 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
50	0.2097886	0.0064924	0.0014	0.9700
51	0.2032962	0.0060723	0.0014	0.9714
52	0.1972239	0.0100789	0.0013	0.9727
53	0.1871449	0.0090652	0.0013	0.9740
54	0.1780797	0.0115471	0.0012	0.9752
55	0.1665326	0.0050247	0.0011	0.9763
56	0.1615079	0.0026619	0.0011	0.9774
57	0.1588460	0.0124983	0.0011	0.9785
58	0.1463477	0.0026918	0.0010	0.9795
59	0.1436559	0.0089644	0.0010	0.9805
60	0.1346915	0.0026825	0.0009	0.9814
61	0.1320091	0.0026298	0.0009	0.9823
62	0.1293793	0.0060494	0.0009	0.9832
63	0.1233299	0.0095147	0.0008	0.9840
64	0.1138152	0.0052639	0.0008	0.9848
65	0.1085513	0.0015809	0.0007	0.9855
66	0.1069705	0.0062460	0.0007	0.9862
67	0.1007245	0.0069634	0.0007	0.9869
68	0.0937611	0.0022264	0.0006	0.9876
69	0.0915347	0.0016925	0.0006	0.9882
70	0.0898422	0.0034335	0.0006	0.9888
71	0.0864088	0.0037319	0.0006	0.9894
72	0.0826769	0.0026722	0.0006	0.9899
73	0.0800047	0.0039222	0.0005	0.9905
74	0.0760825	0.0039106	0.0005	0.9910
75	0.0721719	0.0050629	0.0005	0.9915

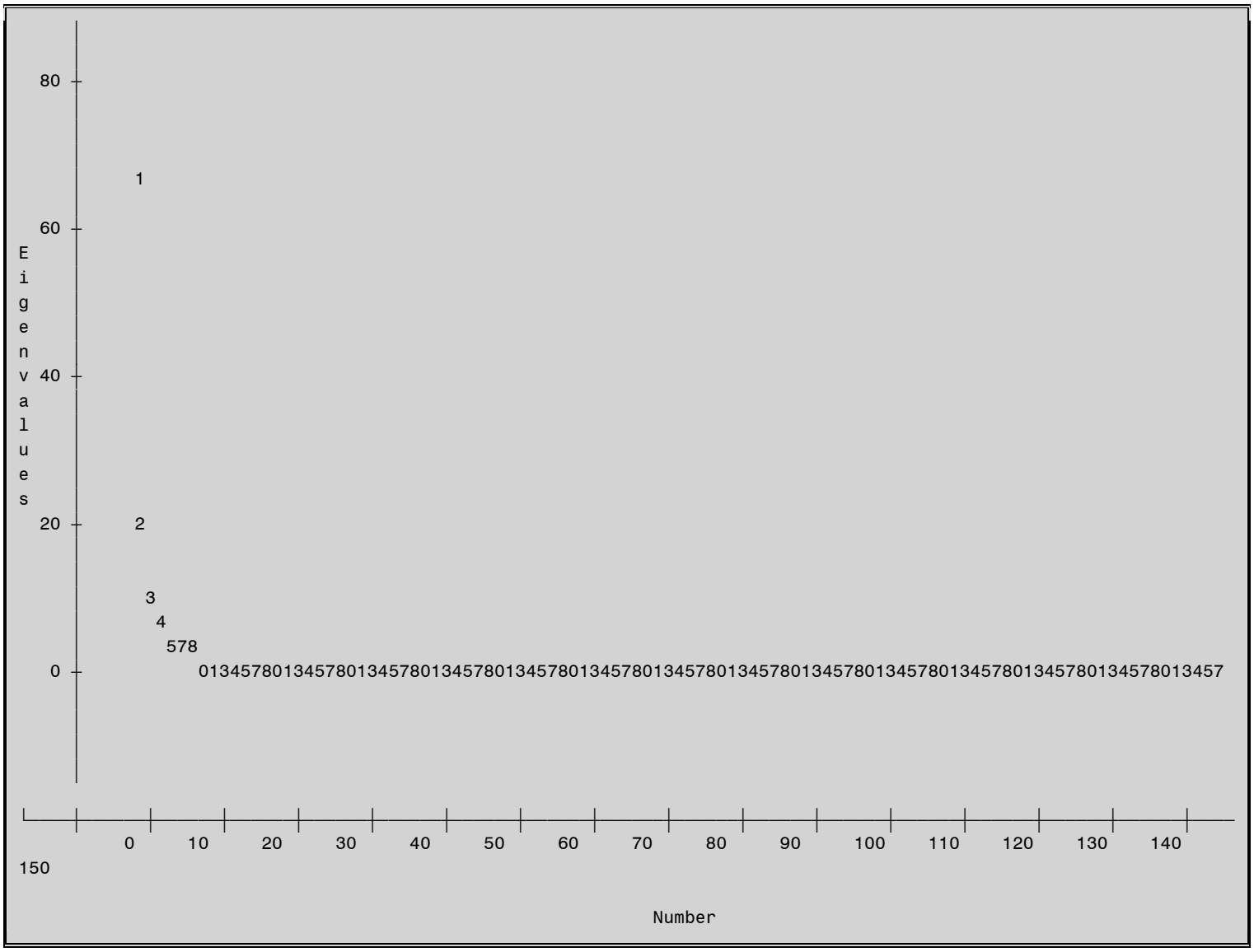
Eigenvalues of the Correlation Matrix: Total = 147 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
76	0.0671090	0.0020830	0.0005	0.9920
77	0.0650260	0.0027196	0.0004	0.9924
78	0.0623064	0.0025366	0.0004	0.9928
79	0.0597698	0.0041361	0.0004	0.9932
80	0.0556337	0.0012277	0.0004	0.9936
81	0.0544060	0.0037237	0.0004	0.9940
82	0.0506823	0.0011943	0.0003	0.9943
83	0.0494880	0.0019132	0.0003	0.9947
84	0.0475748	0.0027812	0.0003	0.9950
85	0.0447936	0.0027243	0.0003	0.9953
86	0.0420692	0.0011397	0.0003	0.9956
87	0.0409295	0.0014526	0.0003	0.9959
88	0.0394769	0.0023723	0.0003	0.9961
89	0.0371047	0.0012107	0.0003	0.9964
90	0.0358939	0.0013297	0.0002	0.9966
91	0.0345642	0.0018422	0.0002	0.9969
92	0.0327220	0.0013770	0.0002	0.9971
93	0.0313450	0.0020298	0.0002	0.9973
94	0.0293152	0.0006135	0.0002	0.9975
95	0.0287017	0.0033714	0.0002	0.9977
96	0.0253303	0.0005271	0.0002	0.9979
97	0.0248031	0.0005141	0.0002	0.9980
98	0.0242891	0.0011142	0.0002	0.9982
99	0.0231749	0.0010510	0.0002	0.9983
100	0.0221239	0.0015987	0.0002	0.9985
101	0.0205252	0.0017508	0.0001	0.9986

Eigenvalues of the Correlation Matrix: Total = 147 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
102	0.0187744	0.0020795	0.0001	0.9988
103	0.0166949	0.0022210	0.0001	0.9989
104	0.0144739	0.0003360	0.0001	0.9990
105	0.0141379	0.0021479	0.0001	0.9991
106	0.0119900	0.0004581	0.0001	0.9992
107	0.0115320	0.0005215	0.0001	0.9992
108	0.0110105	0.0014822	0.0001	0.9993
109	0.0095283	0.0002605	0.0001	0.9994
110	0.0092678	0.0004463	0.0001	0.9994
111	0.0088215	0.0003868	0.0001	0.9995
112	0.0084347	0.0005311	0.0001	0.9996
113	0.0079036	0.0011762	0.0001	0.9996
114	0.0067273	0.0002658	0.0000	0.9997
115	0.0064616	0.0009410	0.0000	0.9997
116	0.0055206	0.0007439	0.0000	0.9997
117	0.0047767	0.0007675	0.0000	0.9998
118	0.0040093	0.0004736	0.0000	0.9998
119	0.0035357	0.0002642	0.0000	0.9998
120	0.0032715	0.0003091	0.0000	0.9998
121	0.0029624	0.0001435	0.0000	0.9999
122	0.0028189	0.0001211	0.0000	0.9999
123	0.0026978	0.0003806	0.0000	0.9999
124	0.0023172	0.0002747	0.0000	0.9999
125	0.0020425	0.0002275	0.0000	0.9999
126	0.0018150	0.0003904	0.0000	0.9999
127	0.0014246	0.0001326	0.0000	0.9999

Eigenvalues of the Correlation Matrix: Total = 147 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
128	0.0012920	0.0001734	0.0000	1.0000
129	0.0011187	0.0000995	0.0000	1.0000
130	0.0010192	0.0001054	0.0000	1.0000
131	0.0009138	0.0001193	0.0000	1.0000
132	0.0007945	0.0001178	0.0000	1.0000
133	0.0006767	0.0001341	0.0000	1.0000
134	0.0005426	0.0001740	0.0000	1.0000
135	0.0003686	0.0001089	0.0000	1.0000
136	0.0002597	0.0000910	0.0000	1.0000
137	0.0001687	0.0000343	0.0000	1.0000
138	0.0001343	0.0000433	0.0000	1.0000
139	0.0000911	0.0000099	0.0000	1.0000
140	0.0000812	0.0000145	0.0000	1.0000
141	0.0000667	0.0000097	0.0000	1.0000
142	0.0000570	0.0000312	0.0000	1.0000
143	0.0000258	0.0000046	0.0000	1.0000
144	0.0000211	0.0000041	0.0000	1.0000
145	0.0000170	0.0000072	0.0000	1.0000
146	0.0000099	0.0000040	0.0000	1.0000
147	0.0000058		0.0000	1.0000

Principal Components Analysis of PTCDEMO
The FACTOR Procedure
Initial Factor Method: Principal Components

```
Scree Plot of Eigenvalues
      |
      |
```



Principal Components Analysis of PTCDEMO
The FACTOR Procedure
Rotation Method: Varimax

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
BORNWEST	BORNWEST	0.99361	-0.02502	-0.02134	0.00077	-0.03173	0.03542	-0.00014
OTH_STAT	OTH_STAT	0.99338	0.00133	-0.02370	0.01368	-0.03451	0.05632	0.01892
HISPANIC	HISPANIC	0.99320	-0.04593	0.00047	-0.00596	-0.03565	0.01609	-0.01287

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
ARMDFORC	ARMDFORC	0.98997	-0.09370	-0.01708	-0.01482	-0.02879	0.00495	-0.02029
AGE21_24	AGE21_24	0.98922	0.02333	0.02813	0.01357	-0.03631	0.02213	0.07787
COM_LT15	COM_LT15	0.98776	0.03686	0.00404	0.03393	0.03545	0.01064	0.02704
ABLEENGL	ABLEENGL	0.98703	0.04295	0.00279	0.02881	-0.03655	0.06136	0.04035
BORNMIDW	BORNMIDW	0.98653	0.05501	-0.05075	0.02288	-0.02588	0.08444	0.01273
MILIQUAR	MILIQUAR	0.98452	-0.10247	-0.01992	-0.01753	-0.02399	0.00430	-0.01529
BORNSOUT	BORNSOUT	0.98260	0.10905	-0.02617	0.03399	-0.00836	0.04341	0.02318
WORKERS	WORKERS	0.98214	0.15196	0.00792	-0.02075	0.02955	-0.00188	0.04082
MALE	MALE	0.98140	0.17206	0.02034	-0.02552	0.03063	-0.00763	0.02453
WRK_HOME	WRK_HOME	0.97596	-0.00503	-0.02838	-0.05217	0.00343	0.02761	0.01795
WRKINCTY	WRKINCTY	0.97298	0.18106	-0.00140	0.04106	0.00563	0.00708	0.00065
AGEIS_20	AGEIS_20	0.97270	-0.03316	-0.00545	-0.00114	-0.02341	-0.00778	0.20023
AGELT_05	AGELT_05	0.97055	0.18217	0.06236	-0.01425	0.00102	-0.01554	-0.04163
CARPOOL	CARPOOL	0.96804	0.12487	0.08465	-0.06283	0.01549	-0.05414	-0.00951
SINGLE_	SINGLE_	0.96765	0.08309	0.05539	0.03071	-0.01506	0.03303	0.19611
AGE25_34	AGE25_34	0.96025	0.24851	0.05303	0.02116	-0.03740	0.02248	-0.01168
AGE05_09	AGE05_09	0.95159	0.24354	0.06482	-0.04908	0.02890	-0.02774	-0.05459
NATIVE	NATIVE	0.95114	0.28705	0.05034	-0.02289	0.07026	-0.01727	0.04656
AGEGE_03	AGEGE_03	0.95056	0.28984	0.04531	-0.01663	0.07091	-0.00757	0.05455
AGEGE_16	AGEGE_16	0.94888	0.28715	0.03705	-0.00629	0.07603	0.00058	0.08025
BORNNORT	BORNNORT	0.94405	0.16485	-0.05546	0.05776	-0.02834	0.17757	0.06665
ONLYENGL	ONLYENGL	0.93408	0.32996	0.05074	-0.02495	0.08765	-0.01833	0.06234
ASIAN	ASIAN	0.93278	0.13834	-0.01721	0.09553	-0.05992	0.10704	0.14305
ATTACHED	ATTACHED	0.93190	0.06954	0.00536	0.07652	-0.04554	0.15553	-0.01454
MARRWICH	MARRWICH	0.92841	0.32258	-0.02711	-0.09138	0.03490	-0.05246	-0.05453
AGE18_19	AGE18_19	0.91597	-0.00060	-0.01408	-0.01068	-0.01119	-0.04395	0.36194
SOMECOLG	SOMECOLG	0.90445	0.37211	-0.04172	0.05182	0.02942	0.05239	-0.05017

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
NATURAL_	NATURAL_	0.89805	0.18794	-0.04900	0.09091	-0.03981	0.18073	0.03378
INC_1525	INC_1525	0.88959	0.31583	0.17602	0.02380	0.07258	-0.01701	-0.01795
MARRIED	MARRIED	0.87451	0.44224	-0.04844	-0.08232	0.11936	-0.04150	-0.04791
WHITE	WHITE	0.87103	0.41998	-0.09748	-0.03927	0.11700	-0.03145	0.06338
BLACK	BLACK	0.86619	0.00743	0.30654	0.05427	-0.01385	0.02702	0.02126
ELEMSCND	ELEMSCND	0.85930	0.39432	0.12288	-0.09234	0.07488	-0.06808	-0.07046
PREPRIMA	PREPRIMA	0.85568	0.30769	-0.04101	0.03384	0.00865	0.11143	-0.03797
PERINFAM	PERINFAM	0.85007	0.48178	0.05855	-0.07155	0.12286	-0.06225	-0.06432
AGE10_14	AGE10_14	0.84991	0.41775	0.11498	-0.10154	0.08606	-0.07639	-0.07968
PRUNIT34	PRUNIT34	0.84619	0.49400	0.03791	-0.06283	0.08572	-0.07623	-0.05325
SEPARATE	SEPARATE	0.81136	0.31419	0.37686	0.10066	0.10057	-0.02971	-0.03248
RNT_2550	RNT_2550	0.79962	0.30274	0.13858	0.29452	-0.10484	0.20141	0.04419
DRVALONE	DRVALONE	0.79541	0.58950	-0.02806	0.02571	0.04149	-0.00810	-0.02521
HIGHSCHL	HIGHSCHL	0.78952	0.45634	0.08683	-0.12608	0.15951	-0.20908	-0.06869
COM_1529	COM_1529	0.76523	0.58713	-0.01193	0.02349	-0.00694	-0.04345	-0.01437
VEHICL_1	VEHICL_1	0.74143	0.47894	0.25297	0.20068	0.10008	0.15151	0.00721
ROOM_4_6	ROOM_4_6	0.72871	0.58166	0.21627	-0.02127	0.14177	-0.05801	-0.03922
ELECTRIC	ELECTRIC	0.72563	0.59716	0.05084	0.08357	-0.03902	0.13370	-0.02434
WAGE_SAL	WAGE_SAL	0.71978	0.67319	0.08364	0.05610	0.06552	0.04130	-0.02404
VEHICL_2	VEHICL_2	0.69697	0.67790	-0.02364	0.00515	0.09294	0.05943	-0.04488
AGE35_44	AGE35_44	0.69485	0.67656	-0.00148	-0.04626	0.09321	0.00685	-0.05574
UNEMPLOY	UNEMPLOY	0.68866	0.28328	0.34382	0.04824	0.05257	0.00629	0.10575
PUBSEWER	PUBSEWER	0.68054	0.35098	0.16717	0.45644	0.05709	0.29651	0.00998
COLLEGE	COLLEGE	0.67570	0.17362	-0.01106	0.07716	-0.04901	0.05443	0.67338
BLTBFR70	BLTBFR70	0.67216	0.33428	0.20359	0.08696	0.36455	-0.08637	0.04189
INC_2535	INC_2535	0.65997	0.62566	0.12261	0.03696	0.08797	-0.08612	-0.03645
PUBWATER	PUBWATER	0.65784	0.45735	0.15307	0.37851	0.10938	0.22513	-0.02253

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
CHILDPOV	CHILDPOV	0.65298	-0.00052	0.55798	-0.02987	0.05928	0.02149	-0.04509
AGE15_17	AGE15_17	0.61662	0.61326	0.16490	-0.14019	0.16718	-0.13483	-0.08907
DUPLEX	DUPLEX	0.60750	0.07399	0.24276	0.39217	0.16377	0.08583	0.05017
PUBLWORK	PUBLWORK	0.60122	0.54882	0.06177	0.03306	-0.00254	0.17738	0.26633
APARTMNT	APARTMNT	0.59935	0.34703	0.22261	0.37522	-0.13150	0.38201	0.07235
OTH_HEAT	OTH_HEAT	0.56267	0.12734	0.01832	-0.01710	0.08405	-0.05179	0.15023
PUBTRANS	PUBTRANS	0.41065	0.14296	0.28436	0.34047	-0.09768	0.13707	0.26464
PRIVWORK	PRIVWORK	0.19923	0.93651	0.04491	0.05390	0.11784	-0.09652	0.04210
EMPLOYED	EMPLOYED	0.30861	0.92811	0.04810	0.02684	0.10863	-0.03694	0.09515
OWNR_OCC	OWNR_OCC	0.01542	0.88902	-0.05710	-0.19143	0.31355	-0.14023	-0.08629
AGE45_54	AGE45_54	0.15757	0.88192	-0.02458	-0.11397	0.29535	-0.09006	-0.06775
TECHSALE	TECHSALE	0.36881	0.86426	-0.06151	0.14237	-0.01041	0.08507	0.16191
TRADE	TRADE	0.34028	0.85983	0.01569	0.12547	0.02265	0.02051	0.07995
MARRNOCH	MARRNOCH	0.26882	0.85765	-0.11050	-0.07125	0.34669	-0.02667	-0.04554
PRUNIT12	PRUNIT12	0.21776	0.84094	0.19138	0.20647	0.23057	0.20503	0.02063
INC_5075	INC_5075	0.22089	0.83476	-0.26083	0.09398	0.04947	0.18048	-0.05389
INC_3550	INC_3550	0.45977	0.80803	-0.05029	0.04314	0.07643	-0.05935	-0.05271
EXECPROF	EXECPROF	0.22750	0.80769	-0.17859	0.18093	0.03937	0.36424	0.08681
UTILITY	UTILITY	0.10914	0.80564	-0.06460	0.08123	-0.05372	-0.04899	-0.02728
SELF_INC	SELF_INC	0.23362	0.79605	-0.10528	-0.07272	0.18013	0.13540	0.00725
VAL_501C	VAL_501C	0.04555	0.79456	-0.12915	0.13339	0.14834	-0.22929	-0.09201
SERVICES	SERVICES	0.27300	0.79377	0.03526	0.17439	-0.03596	0.09337	0.13421
FIRE	FIRE	0.15703	0.79107	-0.13410	0.21553	-0.04346	0.25177	-0.02023
SERVICE	SERVICE	0.29302	0.78794	0.23493	-0.17136	0.11596	-0.24903	0.07685
DETACHED	DETACHED	0.26758	0.78062	-0.05866	-0.08582	0.39873	-0.08915	-0.10384
BLTAFT84	BLTAFT84	0.04049	0.76799	-0.01998	-0.01515	-0.15542	0.30414	-0.03922
AGE55_64	AGE55_64	0.00397	0.76623	0.05165	-0.06564	0.52475	-0.09278	-0.06295

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
INTE_INC	INTE_INC	0.38564	0.76163	-0.21091	0.14091	0.19292	0.31273	0.00843
BLT_8084	BLT_8084	0.08389	0.75776	0.10874	-0.01476	-0.06145	0.28837	-0.05393
SAMEHOUS	SAMEHOUS	0.17889	0.74723	0.16432	-0.23039	0.40085	-0.27811	-0.04559
BORN_INS	BORN_INS	0.39357	0.73070	0.26041	-0.15222	0.28847	-0.25478	0.09243
PROFSERV	PROFSERV	0.33953	0.70876	-0.03615	0.15952	0.08074	0.27193	0.30660
ONEPERHH	ONEPERHH	0.12541	0.70278	0.32643	0.32850	0.18567	0.30648	0.04338
BLT_7079	BLT_7079	0.44207	0.70256	0.15837	-0.05989	0.07013	0.03406	-0.04713
SELFWORK	SELFWORK	0.17687	0.68065	-0.02498	-0.28725	0.18787	0.00482	0.01472
COM_3044	COM_3044	0.27600	0.67971	0.00276	-0.20551	-0.09529	-0.05420	-0.00946
MANUFACT	MANUFACT	0.08823	0.63683	0.13509	-0.12304	0.28648	-0.42345	-0.06205
SAMECNTY	SAMECNTY	0.61041	0.63200	0.17858	0.15496	0.07638	0.04505	-0.06441
PRIMARY_	PRIMARY_	0.03868	0.61418	0.17889	-0.44680	0.01162	-0.07438	0.02033
VAL_1C2C	VAL_1C2C	0.03058	0.60963	-0.39890	0.07712	0.12053	0.39873	-0.04400
COLGGRAD	COLGGRAD	0.57031	0.58386	-0.18391	0.19736	0.03456	0.42681	0.01346
RETI_INC	RETI_INC	0.05126	0.57968	-0.01873	0.13718	0.50057	0.15453	-0.06353
WRKEXCTY	WRKEXCTY	0.14056	0.52638	0.02068	-0.29753	0.05706	-0.10734	0.10371
ROOM_1_3	ROOM_1_3	0.19093	0.48073	0.39200	0.34511	-0.05426	0.40560	0.09805
PUBL_INC	PUBL_INC	0.07157	0.08015	0.75870	-0.05388	0.23190	0.01957	-0.04039
NOVEHICL	NOVEHICL	0.00540	0.09548	0.73917	0.25383	0.32044	0.08523	0.02273
INC_LT15	INC_LT15	0.44027	0.31358	0.69016	0.02337	0.31968	0.01439	0.04892
RNT_LT25	RNT_LT25	-0.00598	0.06875	0.68853	0.24430	0.33733	-0.06675	0.01995
INPOVRTY	INPOVRTY	0.54860	0.11211	0.68807	-0.04193	0.10086	0.09174	0.04217
SINGWICH	SINGWICH	0.40078	0.43551	0.61999	0.17812	0.13151	-0.05754	-0.06830
VAL_LT25	VAL_LT25	-0.04425	-0.06172	0.52044	-0.30452	0.33618	-0.16453	-0.05179
NODIPLOM	NODIPLOM	0.15070	0.42208	0.47141	-0.24823	0.43940	-0.41580	-0.03413
AMIND	AMIND	0.13207	0.01958	0.22783	-0.13176	-0.05571	0.01466	-0.04793
RNT_MEDI	RNT_MEDI	0.13614	0.36175	-0.50559	0.39183	0.04074	0.26009	-0.02053

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
VAL_MEDI	VAL_MEDI	0.03156	0.34867	-0.58099	0.13571	0.10069	0.41605	0.09697
INCPRCAP	INCPRCAP	-0.01401	0.31309	-0.62998	0.19610	0.17327	0.31781	-0.10803
INC_MEDN	INC_MEDN	0.01228	0.42337	-0.71740	0.03678	0.05399	0.10929	-0.08101
FAMILIES	FAMILIES	0.00736	0.07652	0.10412	0.86709	0.06361	0.00897	0.02251
HOUSHOLD	HOUSHOLD	-0.00755	0.05776	0.14994	0.86335	0.02816	0.07056	0.06249
houseden		-0.00943	0.05077	0.15476	0.85963	0.02062	0.07892	0.06228
popden		0.02370	0.04955	0.13145	0.84704	0.02102	0.01813	0.24396
PUBL_GAS	PUBL_GAS	0.30789	0.39182	0.03251	0.56269	0.13003	0.26109	-0.00719
SHELTERS	SHELTERS	-0.00817	-0.04143	0.06756	0.09366	-0.05214	0.01074	0.02789
F_MIGRNT	F_MIGRNT	-0.03255	-0.00913	0.11592	-0.27460	-0.06469	0.06895	0.06745
BOTL_GAS	BOTL_GAS	-0.03817	0.19787	0.33722	-0.52097	0.00062	0.04123	0.00299
FARM_INC	FARM_INC	-0.00242	0.14059	0.14475	-0.55741	-0.01397	0.06183	0.07368
COALWOOD	COALWOOD	-0.04985	0.21546	0.13488	-0.56835	0.18451	-0.26955	0.00198
LAND_KM	LAND_KM	0.12456	-0.03798	0.21401	-0.58524	-0.02859	0.08569	0.01226
AGE75_84	AGE75_84	-0.02639	0.30448	0.20888	0.06357	0.79739	0.11109	0.03535
AGE65_74	AGE65_74	-0.02073	0.52879	0.14685	0.00351	0.72956	0.02171	-0.03351
WIDOWED	WIDOWED	0.01868	0.38460	0.40284	0.07994	0.71685	0.04949	0.01084
SOCS_INC	SOCS_INC	0.03832	0.48616	0.26559	0.01185	0.70629	0.01275	-0.03929
FUELKERO	FUELKERO	0.21939	0.35472	0.11695	-0.17026	0.51727	-0.40668	-0.03517
NURSHOME	NURSHOME	0.00350	0.01939	0.05215	0.07066	0.39049	0.15329	0.06691
MEDYRBLT	MEDYRBLT	0.01715	0.12526	-0.00620	-0.03700	0.17079	0.01752	0.00964
O_INSTIT	O_INSTIT	0.00133	-0.01706	-0.02682	0.02542	0.07749	0.01549	0.05675
RNT_751K	RNT_751K	0.22733	0.23734	-0.07515	0.11890	0.07959	0.52303	0.06829
VAL_GT3C	VAL_GT3C	0.01794	0.18026	-0.31554	0.02834	0.27573	0.47936	-0.00173
VAL_2C3C	VAL_2C3C	0.02463	0.30647	-0.36306	-0.00547	0.21666	0.47676	-0.00999
RNT_5075	RNT_5075	0.38749	0.46295	0.03498	0.24855	-0.21649	0.46406	0.04053
RNT_GT1K	RNT_GT1K	0.01648	0.06952	-0.03650	0.01618	0.20192	0.41760	0.04528

Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
SEASONAL	SEASONAL	-0.02986	-0.05372	-0.02714	-0.13890	0.00892	0.21035	-0.04414
WATER_KM	WATER_KM	-0.00696	-0.03411	0.04216	-0.12191	-0.01318	0.13312	-0.01344
INSTREET	INSTREET	-0.00933	-0.04584	0.04613	0.06392	-0.02439	0.06709	0.02734
VAL_2550	VAL_2550	-0.01926	0.26184	0.38037	-0.02628	0.48347	-0.49302	-0.08195
DORMITOR	DORMITOR	0.02599	0.00402	-0.06383	0.03146	-0.01570	-0.06876	0.92399
SAMESTAT	SAMESTAT	0.38963	0.45026	0.00391	0.01503	-0.00509	0.06619	0.68261
MENTAL	MENTAL	0.00198	-0.00424	-0.01487	0.01561	0.01800	-0.00729	0.12710
CORRINST	CORRINST	0.00073	-0.04085	0.04698	-0.00558	0.02664	0.01970	0.07803
O_NOINST	O_NOINST	0.01384	-0.01023	0.03344	0.00765	-0.01390	0.05503	0.06516

APPENDIX C: Some details of Model Selection

Box-Cox Transformation

Box-Cox regression using full model (PCA of census + all station data)

The TRANSREG Procedure

Transformation Information for BoxCox(AADT)				
Lambda		**R-Square**	**Log Like**	
-2.000		0.04	-443925	
-1.000		0.24	-317613	
-0.500		0.49	-278189	
-0.450		0.51	-275295	
-0.400		0.54	-272600	
-0.350		0.56	-270108	
-0.300		0.58	-267825	
-0.250		0.60	-265758	
-0.200		0.63	-263912	
-0.150		0.64	-262297	
-0.100		0.66	-260919	
-0.050		0.68	-259788	
0.000		0.69	-258911	
0.050		0.71	-258297	
0.100		0.72	-257954	
0.150		0.73	-257886	<
0.200		0.74	-258100	
0.250		0.75	-258598	
0.300		0.75	-259383	
0.350		0.76	-260453	
0.400		0.76	-261809	
0.450		0.77	-263446	
0.500		0.77	-265361	

Transformation Information for BoxCox(AADT)				
Lambda		R-Square	Log Like	
1.000		0.75	-298167	
2.000		0.63	-410732	
< - Best Lambda * - Confidence Interval + - Convenient Lambda				

Summary of Stepwise Selection

Stepwise Selection for aadt^.15 regression from full model (PCA of census + all station data)
The REG Procedure
Model: MODEL1
Dependent Variable: baadt
Stepwise Selection: Step 192
Variable lanes_speed_rte83 Entered: R-Square = 0.7287 and C(p) = 174.2006

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	178	13547	76.10582	525.28	<.0001
Error	34815	5044.24393	0.14489		
Corrected Total	34993	18591			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.89975	0.13146	70.50091	486.59	<.0001
Factor1	0.04044	0.00224	47.07215	324.89	<.0001
Factor2	0.06372	0.00235	106.76403	736.88	<.0001
Factor3	-0.04177	0.00238	44.78081	309.07	<.0001
Factor4	0.11557	0.00321	187.91122	1296.95	<.0001
Factor6	0.01611	0.00236	6.76856	46.72	<.0001
rte_class1	0.73955	0.12976	4.70609	32.48	<.0001
rte_class2	0.46149	0.09600	3.34800	23.11	<.0001
rte_class4	0.25370	0.13692	0.49744	3.43	0.0639

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
co_urb2	-0.05037	0.01970	0.94753	6.54	0.0106
co_urb5	-0.31888	0.03209	14.30451	98.73	<.0001
co_urb6	-0.24327	0.02209	17.56780	121.25	<.0001
co_urb7	-0.09501	0.02772	1.70177	11.75	0.0006
co_urb8	-0.06778	0.03523	0.53619	3.70	0.0544
co_urb9	-0.04624	0.02293	0.58921	4.07	0.0437
co_urb10	-0.09002	0.02709	1.59936	11.04	0.0009
co_urb11	-0.06975	0.02416	1.20794	8.34	0.0039
co_urb12	0.22186	0.02592	10.61660	73.27	<.0001
co_urb13	0.81044	0.38151	0.65382	4.51	0.0337
co_urb14	0.04653	0.02461	0.51774	3.57	0.0587
co_urb16	0.09365	0.02274	2.45737	16.96	<.0001
co_urb17	0.09461	0.04802	0.56246	3.88	0.0488
co_urb18	0.09471	0.02633	1.87458	12.94	0.0003
co_urb19	0.14474	0.02630	4.38678	30.28	<.0001
co_urb20	0.11415	0.02481	3.06795	21.17	<.0001
co_urb21	0.15614	0.08619	0.47545	3.28	0.0701
co_urb22	-0.10877	0.03996	1.07356	7.41	0.0065
co_urb23	0.10947	0.03385	1.51575	10.46	0.0012
co_urb24	-0.09026	0.02820	1.48404	10.24	0.0014
co_urb25	0.18278	0.03117	4.98236	34.39	<.0001
co_urb26	0.06809	0.02308	1.26090	8.70	0.0032
co_urb27	0.03865	0.02274	0.41863	2.89	0.0892
co_urb29	-0.08635	0.03294	0.99553	6.87	0.0088
co_urb30	-0.07768	0.03671	0.64887	4.48	0.0343
co_urb31	-0.15403	0.03721	2.48317	17.14	<.0001
co_urb34	0.05920	0.02503	0.81032	5.59	0.0180

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
co_urb35	0.08844	0.01691	3.96489	27.37	<.0001
co_urb36	-0.06633	0.04207	0.36013	2.49	0.1149
co_urb37	-0.17554	0.04087	2.67235	18.44	<.0001
co_urb38	0.09107	0.01771	3.82959	26.43	<.0001
co_urb39	0.05746	0.03265	0.44873	3.10	0.0784
co_urb43	0.08753	0.01858	3.21424	22.18	<.0001
co_urb44	-0.12207	0.02641	3.09621	21.37	<.0001
co_urb45	-0.19169	0.04913	2.20532	15.22	<.0001
co_urb46	0.10441	0.02051	3.75543	25.92	<.0001
co_urb48	0.24077	0.01920	22.79552	157.33	<.0001
co_urb50	0.06179	0.03419	0.47328	3.27	0.0707
co_urb52	-0.20030	0.03103	6.03582	41.66	<.0001
co_urb53	-0.20094	0.04424	2.98892	20.63	<.0001
co_urb55	-0.07316	0.02579	1.16570	8.05	0.0046
co_urb57	-0.40715	0.05637	7.55918	52.17	<.0001
co_urb58	0.24018	0.02110	18.77133	129.56	<.0001
co_urb60	-0.11157	0.02193	3.75161	25.89	<.0001
co_urb61	0.13411	0.02058	6.15263	42.46	<.0001
co_urb62	0.23887	0.09910	0.84172	5.81	0.0159
co_urb63	0.04047	0.02502	0.37910	2.62	0.1058
co_urb64	0.18482	0.02320	9.19523	63.46	<.0001
co_urb65	-0.11413	0.02714	2.56224	17.68	<.0001
co_urb66	0.07272	0.02938	0.88784	6.13	0.0133
co_urb67	-0.19964	0.04101	3.43371	23.70	<.0001
co_urb69	0.07050	0.03009	0.79563	5.49	0.0191
co_urb70	0.07419	0.01800	2.46060	16.98	<.0001
co_urb71	-0.11026	0.03701	1.28592	8.88	0.0029

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
co_urb72	0.14829	0.02652	4.52955	31.26	<.0001
co_urb73	0.13403	0.02143	5.66729	39.12	<.0001
co_urb74	0.11116	0.02088	4.10597	28.34	<.0001
co_urb75	-0.11504	0.02577	2.88633	19.92	<.0001
co_urb76	-0.21342	0.03077	6.96947	48.10	<.0001
co_urb77	-0.13394	0.02500	4.15879	28.70	<.0001
co_urb80	0.26488	0.01612	39.10641	269.91	<.0001
co_urb81	-0.14516	0.03632	2.31468	15.98	<.0001
co_urb82	-0.20998	0.02323	11.83488	81.68	<.0001
co_urb84	-0.07992	0.01979	2.36206	16.30	<.0001
co_urb86	0.15816	0.02408	6.25115	43.14	<.0001
co_urb87	-0.16820	0.02697	5.63613	38.90	<.0001
co_urb88	0.13216	0.02591	3.76910	26.01	<.0001
co_urb89	0.29597	0.03730	9.12061	62.95	<.0001
co_urb90	0.12297	0.02820	2.75542	19.02	<.0001
co_urb92	0.10736	0.03273	1.55898	10.76	0.0010
co_urb93	-0.12698	0.04288	1.27079	8.77	0.0031
co_urb97	0.04977	0.02468	0.58913	4.07	0.0438
co_urb98	-0.05650	0.02208	0.94830	6.55	0.0105
co_urb99	0.08170	0.02202	1.99389	13.76	0.0002
co_urb100	-0.17441	0.03064	4.69572	32.41	<.0001
co_urb101	0.06395	0.01806	1.81716	12.54	0.0004
co_urb102	0.13783	0.04904	1.14448	7.90	0.0049
co_urb103	-0.04805	0.02300	0.63216	4.36	0.0367
co_urb104	0.06789	0.01611	2.57228	17.75	<.0001
co_urb106	0.09791	0.01978	3.55160	24.51	<.0001
co_urb107	-0.05430	0.03367	0.37685	2.60	0.1068

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
co_urb109	-0.04576	0.01786	0.95097	6.56	0.0104
co_urb111	-0.06280	0.02087	1.31182	9.05	0.0026
co_urb113	0.08120	0.01988	2.41642	16.68	<.0001
co_urb114	-0.07980	0.03615	0.70614	4.87	0.0273
co_urb115	-0.06324	0.03143	0.58637	4.05	0.0443
co_urb116	-0.19310	0.05011	2.15150	14.85	0.0001
co_urb117	0.08530	0.01777	3.33793	23.04	<.0001
co_urb118	0.16520	0.02635	5.69482	39.31	<.0001
co_urb119	0.13638	0.01544	11.30417	78.02	<.0001
co_urb120	-0.08077	0.02655	1.34065	9.25	0.0024
co_urb121	-0.08843	0.03686	0.83399	5.76	0.0164
co_urb123	-0.04677	0.02278	0.61083	4.22	0.0401
co_urb125	0.12113	0.02101	4.81565	33.24	<.0001
co_urb127	0.04972	0.02628	0.51872	3.58	0.0585
lanes_rte2	-0.43949	0.17699	0.89341	6.17	0.0130
lanes_rte4	-0.33799	0.02236	33.11742	228.57	<.0001
lanes_rte5	-0.14325	0.09285	0.34486	2.38	0.1229
lanes_rte6	-0.39262	0.04728	9.99239	68.97	<.0001
lanes_rte7	-0.40653	0.12809	1.45936	10.07	0.0015
lanes_rte8	-0.13747	0.02950	3.14576	21.71	<.0001
lanes_rte13	0.15421	0.09479	0.38346	2.65	0.1038
lanes_rte14	0.13166	0.04879	1.05497	7.28	0.0070
lanes_rte16	0.34247	0.15157	0.73972	5.11	0.0239
lanes_rte19	0.37013	0.12497	1.27102	8.77	0.0031
lanes_rte21	0.47534	0.07464	5.87640	40.56	<.0001
speed_rte1	-0.40483	0.08283	3.46105	23.89	<.0001
speed_rte3	-0.30530	0.06190	3.52421	24.32	<.0001

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
speed_rte5	-0.29746	0.06659	2.89111	19.95	<.0001
speed_rte7	-0.94873	0.10460	11.91903	82.26	<.0001
speed_rte12	-0.18336	0.02175	10.29858	71.08	<.0001
speed_rte13	0.03342	0.01639	0.60198	4.15	0.0415
speed_rte14	-0.13636	0.01641	10.00869	69.08	<.0001
speed_rte20	0.13384	0.01957	6.77676	46.77	<.0001
speed_rte21	-0.02835	0.01825	0.34940	2.41	0.1205
speed_rte27	-0.04880	0.02184	0.72359	4.99	0.0254
speed_rte30	-0.25867	0.06963	1.99926	13.80	0.0002
lanes_speed_rte5	0.20744	0.09448	0.69838	4.82	0.0281
lanes_speed_rte8	-0.57891	0.38151	0.33360	2.30	0.1292
lanes_speed_rte11	0.73829	0.10807	6.76215	46.67	<.0001
lanes_speed_rte26	0.23929	0.05629	2.61798	18.07	<.0001
lanes_speed_rte31	-0.30309	0.01674	47.51763	327.96	<.0001
lanes_speed_rte40	-0.42754	0.17746	0.84095	5.80	0.0160
lanes_speed_rte45	0.16677	0.09718	0.42669	2.95	0.0862
lanes_speed_rte49	0.20937	0.03578	4.96245	34.25	<.0001
lanes_speed_rte65	-0.08422	0.03314	0.93588	6.46	0.0110
lanes_speed_rte73	0.09696	0.03196	1.33336	9.20	0.0024
lanes_speed_rte74	0.26884	0.04733	4.67484	32.27	<.0001
lanes_speed_rte76	0.24306	0.14413	0.41205	2.84	0.0917
lanes_speed_rte83	0.28616	0.19217	0.32128	2.22	0.1365
lanes_speed_rte93	0.34502	0.11612	1.27897	8.83	0.0030
lanes_speed_rte95	0.16556	0.06475	0.94715	6.54	0.0106
lanes_speed_rte99	0.30016	0.08284	1.90218	13.13	0.0003
lanes_speed_rte102	-0.70542	0.28005	0.91927	6.34	0.0118
lanes_speed_rte104	-0.81319	0.10325	8.98731	62.03	<.0001

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
cycle_surface1	1.18513	0.03160	203.73563	1406.17	<.0001
cycle_surface2	0.87198	0.05551	35.74881	246.74	<.0001
cycle_surface4	0.50895	0.02889	44.96950	310.38	<.0001
cycle_surface5	0.47010	0.13023	1.88781	13.03	0.0003
cycle_surface6	0.63321	0.04048	35.45017	244.67	<.0001
cycle_surface7	0.58909	0.04604	23.72094	163.72	<.0001
cycle_surface8	1.20750	0.02990	236.27730	1630.77	<.0001
cycle_surface9	0.71481	0.02685	102.66617	708.59	<.0001
a_control2	-0.21018	0.08707	0.84426	5.83	0.0158
year4	0.07120	0.01071	6.39884	44.16	<.0001
month1	-0.04404	0.00913	3.36915	23.25	<.0001
month5	-0.05882	0.00870	6.62903	45.75	<.0001
month6	-0.04412	0.00965	3.03065	20.92	<.0001
month7	-0.05511	0.01097	3.66049	25.26	<.0001
month8	-0.08073	0.01104	7.74617	53.46	<.0001
month9	-0.06269	0.00987	5.84751	40.36	<.0001
month10	-0.04291	0.00967	2.85329	19.69	<.0001
month11	-0.04713	0.00979	3.35911	23.18	<.0001
day1	-0.01254	0.00430	1.22987	8.49	0.0036
day4	-0.06818	0.02201	1.39066	9.60	0.0019
lucode2	0.07571	0.00934	9.51706	65.69	<.0001
lucode3	0.05594	0.01755	1.47211	10.16	0.0014
lucode4	0.15557	0.01546	14.66576	101.22	<.0001
lucode5	0.10757	0.02381	2.95712	20.41	<.0001
lucode7	0.06503	0.02453	1.01794	7.03	0.0080
lucode8	-0.05711	0.00612	12.61769	87.09	<.0001
lucode10	-0.11308	0.07109	0.36659	2.53	0.1117

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
lucode11	0.20524	0.07643	1.04492	7.21	0.0072
lucode13	-0.25025	0.14424	0.43611	3.01	0.0828
lucode14	-0.08770	0.00863	14.96641	103.30	<.0001
lucode15	-0.04123	0.01194	1.72768	11.92	0.0006
lucode16	-0.05632	0.00805	7.09037	48.94	<.0001
lucode21	-0.03831	0.02012	0.52517	3.62	0.0569
lucode24	0.37643	0.22364	0.41051	2.83	0.0923
city_urb1	-0.17076	0.00755	74.19592	512.09	<.0001

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	cycle_surface4		Odd, Conc	1	0.4259	0.4259	38662.2	25961.4	<.0001
2	lanes_speed_rte31		2, 60, US	2	0.0833	0.5093	27972.6	5942.01	<.0001
3	cycle_surface8		Vari, Asph	3	0.0512	0.5605	21401.0	4079.14	<.0001
4	rte_class1		INTER	4	0.0307	0.5912	17459.1	2631.28	<.0001
5	Factor4		Poor	5	0.0272	0.6184	13974.5	2491.81	<.0001
6	rte_class2		US	6	0.0181	0.6365	11659.3	1738.23	<.0001
7	lanes_rte14		4, SR	7	0.0117	0.6482	10161.1	1162.79	<.0001
8	Factor2		No Kids	8	0.0082	0.6564	9105.38	839.46	<.0001
9	cycle_surface1		Ann, Asph	9	0.0076	0.6640	8130.24	793.06	<.0001
10	lanes_rte4		2, US	10	0.0078	0.6719	7125.79	836.36	<.0001
11	Factor3		Poor	11	0.0058	0.6776	6389.24	624.67	<.0001
12	city_urb1		rural	12	0.0061	0.6837	5607.86	675.36	<.0001
13	Factor1		With Kids	13	0.0044	0.6882	5041.16	497.24	<.0001
14	lanes_rte13		4, NC	14	0.0034	0.6915	4611.79	381.27	<.0001
15	cycle_surface9		Vari, Conc	15	0.0024	0.6940	4301.49	278.21	<.0001

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
16		cycle_surface4	Odd, Conc	14	0.0000	0.6940	4299.92	0.38	0.5363
17	lucode4		14,Trans,Comm,Util	15	0.0017	0.6956	4087.28	192.26	<.0001
18	rte_class4		SR	16	0.0015	0.6972	3894.60	175.24	<.0001
19	lanes_speed_rte45		3, 45, US	17	0.0013	0.6985	3724.86	155.28	<.0001
20	co_urb80		Mitchell	18	0.0012	0.6997	3570.73	141.74	<.0001
21	co_urb6		Anson	19	0.0012	0.7009	3414.97	143.80	<.0001
22	lanes_speed_rte18		2, 40, US	20	0.0011	0.7020	3275.52	129.41	<.0001
23	cycle_surface2		Ann, Conc	21	0.0009	0.7029	3167.17	101.24	<.0001
24	co_urb5		Alleghary	22	0.0007	0.7036	3077.01	84.76	<.0001
25	lanes_rte6		2, SR	23	0.0007	0.7043	2987.45	84.41	<.0001
26	co_urb64		Hertford	24	0.0007	0.7050	2902.57	80.28	<.0001
27	co_urb12		Brunswick	25	0.0007	0.7057	2820.07	78.24	<.0001
28	co_urb89		Orange	26	0.0006	0.7063	2744.22	72.25	<.0001
29	co_urb82		Moore	27	0.0006	0.7069	2670.51	70.38	<.0001
30	co_urb57		Guilford,106	28	0.0006	0.7074	2601.46	66.19	<.0001
31	speed_rte7		25, LOCAL	29	0.0005	0.7079	2536.99	62.02	<.0001
32	co_urb61		Harnett,104,	30	0.0005	0.7084	2473.73	61.00	<.0001
33	lucode2		12,Comm & Services	31	0.0005	0.7090	2410.64	60.95	<.0001
34	lanes_rte19		6, NC	32	0.0005	0.7094	2349.47	59.25	<.0001
35	speed_rte12		35, NC	33	0.0005	0.7099	2290.77	57.02	<.0001
36	lanes_speed_rte23		2, 45, SR	34	0.0004	0.7104	2235.97	53.44	<.0001
37	co_urb104		Rockingham	35	0.0004	0.7107	2188.38	46.71	<.0001
38	co_urb58		Guilford,107,	36	0.0004	0.7111	2144.61	43.17	<.0001
39	co_urb48		Franklin	37	0.0004	0.7115	2095.78	48.01	<.0001
40	co_urb73		Lincoln	38	0.0003	0.7118	2054.21	41.20	<.0001

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
41	cycle_surface6		Even, Conc	39	0.0003	0.7122	2013.85	40.09	<.0001
42	cycle_surface4		Odd, Conc	40	0.0007	0.7129	1922.63	88.46	<.0001
43	cycle_surface7		Even, Soil	41	0.0013	0.7142	1752.94	163.68	<.0001
44	co_urb76		Martin	42	0.0003	0.7146	1712.73	40.29	<.0001
45	co_urb52		Graham	43	0.0003	0.7148	1678.57	34.54	<.0001
46	lucode14		41,DECIDUOUS FOREST LAND	44	0.0003	0.7151	1643.72	35.24	<.0001
47	lucode8		21,CROPLAND AND PASTURE	45	0.0004	0.7156	1589.57	53.78	<.0001
48	co_urb107		Rutherford	46	0.0003	0.7158	1557.49	32.66	<.0001
49	lanes_speed_rte104		8, 60, INTER	47	0.0003	0.7161	1525.71	32.41	<.0001
50	co_urb100		Randolph	48	0.0003	0.7164	1494.23	32.15	<.0001
51	co_urb45		Edgecombe,116	49	0.0003	0.7166	1463.04	31.91	<.0001
52	lanes_speed_rte11		2, 25, LOCAL	50	0.0003	0.7169	1432.46	31.34	<.0001
53	lucode16		43, MIXED FOREST LAND	51	0.0002	0.7171	1402.83	30.45	<.0001
54	co_urb118		Wake	52	0.0002	0.7174	1372.87	30.80	<.0001
55	co_urb125		Wilkes	53	0.0002	0.7176	1345.62	28.21	<.0001
56	co_urb74		Macon	54	0.0002	0.7178	1319.96	26.69	<.0001
57	co_urb25		Catawba	55	0.0002	0.7180	1295.92	25.15	<.0001
58	lanes_speed_rte73		4, 55, NC	56	0.0002	0.7182	1271.98	25.07	<.0001
59	Factor6		Wealthy	57	0.0002	0.7184	1248.57	24.57	<.0001
60	co_urb35		Cumberland,104	58	0.0002	0.7186	1225.30	24.46	<.0001
61	co_urb20		Caldwell	59	0.0002	0.7188	1202.69	23.83	<.0001
62	co_urb70		Jones	60	0.0002	0.7190	1180.31	23.62	<.0001
63	co_urb88		Onslow,112	61	0.0002	0.7192	1158.79	22.80	<.0001
64	co_urb106		Rowan,111	62	0.0002	0.7194	1137.12	22.97	<.0001

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
65	co_urb38		Davison	63	0.0002	0.7196	1114.18	24.21	<.0001
66	co_urb117		Vance	64	0.0002	0.7198	1091.07	24.40	<.0001
67	co_urb72		Lenoir	65	0.0002	0.7199	1069.76	22.66	<.0001
68	speed_rte1		20, US	66	0.0002	0.7201	1049.73	21.43	<.0001
69	co_urb16		Burke	67	0.0002	0.7203	1030.58	20.58	<.0001
70	lanes_rte21		8, INTER	68	0.0002	0.7204	1011.15	20.87	<.0001
71	lanes_speed_rte26		2, 50, SR	69	0.0002	0.7206	993.082	19.55	<.0001
72	co_urb90		Orange,101	70	0.0001	0.7207	975.873	18.72	<.0001
73	co_urb119		Wake,108	71	0.0001	0.7209	958.734	18.66	<.0001
74	year4		2000	72	0.0002	0.7210	941.258	19.00	<.0001
75	co_urb86		Northampton	73	0.0002	0.7213	916.492	26.13	<.0001
76	co_urb2		Alamance,101	74	0.0002	0.7214	895.764	22.21	<.0001
77	co_urb51		Gates	75	0.0002	0.7216	869.682	27.46	<.0001
78	lucode5		15,INDUST & COMMERC CMPLXS	76	0.0002	0.7218	850.879	20.35	<.0001
79	lanes_speed_rte49		3, 55, NC	77	0.0001	0.7220	833.669	18.80	<.0001
80	speed_rte3		20, SR	78	0.0001	0.7221	817.146	18.14	<.0001
81	speed_rte14		40, US	79	0.0002	0.7223	799.157	19.59	<.0001
82		lanes_speed_rte18	2, 40, US	78	0.0000	0.7223	797.263	0.10	0.7469
83	co_urb113		Swain	79	0.0001	0.7224	780.792	18.11	<.0001
84	co_urb101		Randolph,107	80	0.0001	0.7225	764.351	18.09	<.0001
85	co_urb46		Forsyth	81	0.0002	0.7227	746.700	19.28	<.0001
86	month8		Aug	82	0.0001	0.7228	730.411	17.96	<.0001
87	month5		May	83	0.0001	0.7230	713.589	18.49	<.0001
88	co_urb87		Onslow	84	0.0001	0.7231	696.411	18.85	<.0001
89	co_urb77		McDowell	85	0.0002	0.7233	678.737	19.35	<.0001

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
90	speed_rte30		60, US	86	0.0001	0.7234	664.559	15.92	<.0001
91	co_urb97		Pitt	87	0.0001	0.7235	650.398	15.90	<.0001
92	lanes_speed_rte74		4, 55, SR	88	0.0001	0.7237	636.592	15.56	<.0001
93	speed_rte5		25, NC	89	0.0001	0.7238	622.874	15.48	<.0001
94	co_urb67		Iredell	90	0.0001	0.7239	609.508	15.14	<.0001
95	co_urb116		Union	91	0.0001	0.7240	597.965	13.35	0.0003
96	co_urb37		Dare	92	0.0001	0.7241	586.602	13.18	0.0003
97	lanes_speed_rte99		6, 65, US	93	0.0001	0.7242	575.429	12.99	0.0003
98	lucode3		13,Industrial	94	0.0001	0.7243	564.773	12.49	0.0004
99	lanes_rte7		2, LOCAL	95	0.0001	0.7244	553.716	12.89	0.0003
100	co_urb81		Montgomery	96	0.0001	0.7245	543.087	12.47	0.0004
101	co_urb44		Edgecombe	97	0.0001	0.7246	532.332	12.60	0.0004
102	co_urb84		Nash,116	98	0.0001	0.7247	521.622	12.56	0.0004
103	lanes_rte8		3, US	99	0.0001	0.7248	511.116	12.36	0.0004
104	co_urb53		Granville	100	0.0001	0.7249	500.978	12.00	0.0005
105	co_urb34		Craven	101	0.0001	0.7250	491.055	11.79	0.0006
106	lucode15		42,EVERGREEN FOREST LAND	102	0.0001	0.7251	480.167	12.75	0.0004
107	co_urb18		Cabamus	103	0.0001	0.7252	470.283	11.76	0.0006
108	co_urb23		Carteret	104	0.0001	0.7253	460.507	11.66	0.0006
109	co_urb43		Durham,103	105	0.0001	0.7254	451.092	11.30	0.0008
110	co_urb60		Harnett	106	0.0001	0.7255	442.622	10.37	0.0013
111	lanes_speed_rte93		6, 50, US	107	0.0001	0.7255	434.913	9.62	0.0019
112	co_urb66		Hyde	108	0.0001	0.7256	427.414	9.41	0.0022
113	cycle_surface5		Odd, Soil	109	0.0001	0.7257	420.019	9.31	0.0023
114	co_urb92		Pamlico	110	0.0001	0.7257	413.339	8.61	0.0034

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
115	co_urb19		Cabamus,111	111	0.0001	0.7258	405.730	9.53	0.0020
116	co_urb69		Johnston	112	0.0001	0.7259	399.210	8.45	0.0036
117	co_urb99		Pork	113	0.0001	0.7260	392.814	8.33	0.0039
118	co_urb102		Richmond	114	0.0001	0.7260	386.237	8.51	0.0035
119	lucode11		24,OTHER AGRICULTURAL LAND	115	0.0001	0.7261	379.873	8.30	0.0040
120	co_urb14		Buncombe	116	0.0001	0.7262	373.579	8.23	0.0041
121	co_urb63		Henderson	117	0.0001	0.7262	367.499	8.02	0.0046
122	month1		Jan	118	0.0001	0.7263	359.929	9.50	0.0021
123	co_urb50		Gaston,105	119	0.0001	0.7264	353.787	8.09	0.0045
124	co_urb31		Clay	120	0.0001	0.7264	347.780	7.96	0.0048
125	day4		Thu	121	0.0001	0.7265	342.328	7.41	0.0065
126	month9		Sept	122	0.0001	0.7265	336.476	7.80	0.0052
127	day1		Mon	123	0.0001	0.7266	329.730	8.69	0.0032
128	co_urb75		Madison	124	0.0001	0.7267	323.372	8.31	0.0039
129	co_urb65		Hoke	125	0.0001	0.7267	317.198	8.13	0.0044
130	lanes_rte16		5, SR	126	0.0001	0.7268	312.055	7.11	0.0077
131	lucode7		17,OTHER URBAN OR BUILT-U	127	0.0001	0.7268	306.921	7.10	0.0077
132	lanes_speed_rte102		8, 45, US	128	0.0001	0.7269	302.044	6.84	0.0089
133	co_urb93		Pasquotank	129	0.0001	0.7269	297.394	6.62	0.0101
134	lanes_speed_rte95		6, 55, US	130	0.0001	0.7270	292.792	6.57	0.0104
135	co_urb7		Ashe	131	0.0000	0.7270	288.719	6.05	0.0139
136	co_urb62		Haywood	132	0.0000	0.7271	284.785	5.91	0.0151
137	co_urb22		Camden	133	0.0000	0.7271	280.866	5.89	0.0152
138	a_control2		Medium	134	0.0000	0.7272	277.023	5.82	0.0159

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
139	co_urb24		Caswell	135	0.0000	0.7272	273.528	5.47	0.0193
140	lanes_rte2		1, SR	136	0.0000	0.7273	270.485	5.02	0.0250
141	lanes_speed_rte40		3, 35, US	137	0.0000	0.7273	267.375	5.09	0.0240
142	month11		Nov	138	0.0000	0.7273	264.289	5.07	0.0244
143	month10		Oct	139	0.0000	0.7274	261.164	5.11	0.0238
144	month6		June	140	0.0001	0.7274	256.693	6.45	0.0111
145	month7		July	141	0.0001	0.7275	247.413	11.25	0.0008
146	co_urb127		Yadkin	142	0.0000	0.7276	243.221	6.17	0.0130
147	co_urb27		Chatham	143	0.0000	0.7276	238.992	6.21	0.0127
148	co_urb33		Columbus	144	0.0000	0.7277	235.167	5.81	0.0159
149	co_urb10		Bertie	145	0.0000	0.7277	232.163	4.99	0.0255
150	co_urb71		Lee	146	0.0000	0.7277	229.188	4.96	0.0259
151	lucode21		61,FORESTED WETLAND	147	0.0000	0.7278	226.391	4.79	0.0287
152	co_urb111		Stokes	148	0.0000	0.7278	223.798	4.58	0.0323
153	lanes_speed_rte5		2, 20, US	149	0.0000	0.7279	221.299	4.49	0.0341
154	lanes_speed_rte65		4, 45, NC	150	0.0000	0.7279	218.433	4.86	0.0276
155	speed_rte27		55, NC	151	0.0000	0.7279	215.981	4.44	0.0350
156	co_urb13		Brunswick,109	152	0.0000	0.7280	213.720	4.25	0.0392
157	co_urb26		Catawba,114	153	0.0000	0.7280	211.535	4.18	0.0410
158		co_urb51	Gates	152	0.0000	0.7280	211.164	1.63	0.2021
159	co_urb120		Warren	153	0.0000	0.7280	208.727	4.43	0.0353
160	month12		Dec	154	0.0000	0.7281	206.056	4.66	0.0308
161	co_urb29		Cherokee	155	0.0000	0.7281	203.395	4.66	0.0310
162	co_urb122		Watauga	156	0.0000	0.7281	200.867	4.52	0.0335
163	speed_rte20		45, SR	157	0.0000	0.7282	198.969	3.89	0.0485

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
164	co_urb121		Washington	158	0.0000	0.7282	197.193	3.77	0.0521
165	co_urb17		Burke,114	159	0.0000	0.7282	195.426	3.76	0.0524
166		lanes_speed_rte23	2, 45, SR	158	0.0000	0.7282	195.494	2.07	0.1506
167	speed_rte13		35, SR	159	0.0000	0.7282	192.757	4.73	0.0296
168	co_urb11		Bladen	160	0.0000	0.7283	190.982	3.77	0.0521
169	co_urb21		Caldwell,114	161	0.0000	0.7283	189.680	3.30	0.0693
170	co_urb114		Transylvania	162	0.0000	0.7283	188.550	3.13	0.0770
171	co_urb39		Davison,107	163	0.0000	0.7283	187.684	2.86	0.0906
172	lanes_speed_rte76		4, 60, NC	164	0.0000	0.7284	186.817	2.87	0.0905
173	lucode13		32,SHRUB & BRUSH RANGELAND	165	0.0000	0.7284	185.963	2.85	0.0913
174	co_urb55		Guilford	166	0.0000	0.7284	185.123	2.84	0.0920
175	co_urb30		Chowan	167	0.0000	0.7284	184.245	2.88	0.0898
176	co_urb98		Pitt, 115	168	0.0000	0.7284	183.280	2.96	0.0852
177	co_urb9		Beanfort	169	0.0000	0.7285	182.510	2.77	0.0961
178	lucode24		73,SANDY AREA (NON-BEACH)	170	0.0000	0.7285	181.757	2.75	0.0971
179	co_urb115		Tyrrell	171	0.0000	0.7285	181.042	2.71	0.0995
180	co_urb109		Scotland	172	0.0000	0.7285	180.151	2.89	0.0891
181	co_urb123		Wayne	173	0.0000	0.7286	179.073	3.08	0.0794
182	co_urb103		Robeson	174	0.0000	0.7286	177.801	3.27	0.0705
183		co_urb33	Columbus	173	0.0000	0.7286	177.855	2.05	0.1519
184	co_urb8		Averry	174	0.0000	0.7286	176.709	3.15	0.0762
185		co_urb122	Watauga	173	0.0000	0.7286	176.769	2.06	0.1512
186		month12	Dec	172	0.0000	0.7286	176.527	1.76	0.1849
187	lucode10		23,CONFINED FEEDING OPS	173	0.0000	0.7286	175.980	2.55	0.1105

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
188	co_urb36		Cumituck	174	0.0000	0.7286	175.495	2.49	0.1149
189	speed_rte21		50, INTER	175	0.0000	0.7286	175.113	2.38	0.1227
190	lanes_rte5		2, NC	176	0.0000	0.7286	174.732	2.38	0.1229
191	lanes_speed_rte8		2, 25, US	177	0.0000	0.7287	174.418	2.31	0.1281
192	lanes_speed_rte83		5, 45, SR	178	0.0000	0.7287	174.201	2.22	0.1365

APPENDIX 10:

10. Preparations for November 19, 2004 Meeting with DOT Research Team

Jacqueline M. Hughes-Oliver

November 2004

Meeting 19Nov04.doc

Preparations for November 19, 2004 Meeting with DOT Research Team
Authored by JH-O on November 18, 2004

JH-O and T-YH will be primarily responsible for informing Shannon of what we did to convert the data originally given to us by DOT into the format that was input to SAS's PROC REG to obtain the mean model.

The JH-O & T-YH Process to Determine the Mean Model:

1. Receive "original" data from NCDOT
    a. Station data, i.e., PTC station locations and counts
    b. Census data from 1990
2. Edit data—received replacement files from DOT
3. Create new variables (more details below)
4. Perform principal components analysis on census data
5. Perform model selection to arrive at "final mean model"

DOT current project needs, as JH-O understands it:

1. How can DOT **run the "final mean model" obtained by JH-O and T-YH** on newly obtained data? E.g., this data may be from stations originally included in the modeling by JH-O and T-YH, but have somehow changed since then (more lanes, for instance?). Or it could be from newly digitized stations that were not included in the modeling by JH-O and T-YH. **This is what I recommend.**
2. How can DOT/Shannon **create their own mean model**, in a fashion following JH-O and T-YH, using newly obtained data? E.g., this data may be from newly digitized stations that were not included in the modeling by JH-O and T-YH. The thinking here is that the new stations will be "different enough" from stations used to build the JH-O&T-YH model that this model will not adequately capture their behavior. **This is NOT recommended.**

How to run the JH-O and T-YH "final mean model"?

- Input:
    - MSE=0.14489. This is the mean squared error from the model
    - $\hat{\beta}$ is the vector of estimated regression coefficients. It is available from report "Determining the Final Mean Model for the Statewide Area," dated December 19, 2003.
    - Attribute matrix X. This matrix contains columns like "factor1," "factor2," "factor3," "factor4," "factor6," "rte_class1," "rte_class2," etc. It was delivered earlier as file cov_X.csv. **Similar matrices will need to be calculated for new or modified PTC stations.**
- Output:
    - 95% prediction interval for aadt^.15
    - 99% prediction interval for aadt^.15
- Formulas:
    - X : attribute matrix used to develop model

- X_0: attribute matrix for "new" stations
- $\hat{Y}_0 = X_0\hat{\beta}$: prediction of aadt^.15 for "new" stations. This is a vector.
- Var($\hat{Y}_0 - Y_0$) $= MSE \cdot \left[I + X_0(X^T X)^{-1} X_0^T\right]$: mean squared prediction error for "new" stations. This is a matrix. We will only use the diagonal elements of this matrix, which will be denoted as $diag(Var(\hat{Y}_0 - Y_0))$.
- $\hat{Y}_0 \pm z_{\alpha/2}\sqrt{diag(Var(\hat{Y}_0 - Y_0))}$: endpoints of (1-α)100% prediction interval. For 95% interval, use 1.96 for $z_{\alpha/2}$. For 99% interval, use 2.58 for $z_{\alpha/2}$.

- More on building the X matrix:
  - Factor1, factor2, factor3, factor4, factor6: these are the principal components obtained from 1990 census data. Because census data is available only to the tract level, all stations within the same census tract will have the same values for these variables. To obtain these variables for new stations, simply identify other stations in the same census tract then duplicate their values for the principal components.
  - Rte_class1, rte_class2, rte_class4: these are binary (indicator) variables that are one when the station lies on a route of type indicated. We actually went through a series of redefinitions of route types. Here's what was used
    ```
    if rte_type=7 then rte_type=4;
    rte_class=rte_type;
    ```
  - Co_urb2, co_urb5, etc: these attempt to resolve the conflicting messages created by variables SIPS, CO_NUM, URB_NUM, and COUNTYID, which are similar but all slightly different. CO_URB is actually an interaction between CO_NUM and URB_NUM as follows below. Once again, binary (indicator) variables are created to indicate each of the 127 possibilities.
    ```
    co_urb=co_num*10000 + urb_num*(urb_num>0);
    ```
  - Lanes_rte2, lanes_rte4, etc: these are interaction terms between LANES and RTE_CLASS. First, LANES was modified using
    ```
    lanes=lanes + lanes*(rte_class=1);
    ```
    then LANES_RTE was calculated using
    ```
    lanes_rte=lanes*100 + rte_class;
    ```
    and finally indicator variables are obtained.
  - Speed_rte1, speed_rte3, etc.: `speed_rte=speed*100 + rte_class;`
  - Lanes_speed_rte5, lanes_speed_rte8, etc.: `lanes_speed_rte=lanes*100000 + speed*100 + rte_class;`
  - Cycle_surface1, cycle_surface2, etc.:
    ```
    cycle_surface=100*(cycle_yr="A") + 200*(cycle_yr="O") +
                  300*(cycle_yr="E") + 400*(cycle_yr="V") +
                  rd_surface;
    ```
  - A_control2:
  - Year4:
  - Month1, month5, etc.: `if month=. then month=0;`
  - Day1, day4: `if day=. then day=0;`
  - Lucode2, lucode3:
  - City_urb1: `city_urb=100*(municpal_f=1);`

- **Variable labels from SAS program:**

```
factor1='Other' factor2='Single' factor3='College' factor4='Poor'
factor5='Farm' factor6='Elderly' factor7='Wealthy'

rte_class1='INTER' rte_class2='US' rte_class3='NC' rte_class4='SR'
rte_class5='LOCAL'

co_urb1='Alamance'co_urb2='Alamance,101'co_urb3 ='Alexander'co_urb4
      ='Alexander,114'
co_urb5     ='Alleghary'co_urb6     ='Anson'co_urb7   ='Ashe'co_urb8
      ='Averry'co_urb9  ='Beanfort'
co_urb10='Bertie'co_urb11='Bladen'co_urb12='Brunswick'co_urb13='Brunswick,109
'co_urb14='Buncombe'
co_urb15='Buncombe,100'co_urb16='Burke'co_urb17='Burke,114'co_urb18='Cabamus'
co_urb19='Cabamus,111'
co_urb20='Caldwell'co_urb21='Caldwell,114'co_urb22='Camden'co_urb23='Carteret
'co_urb24='Caswell'
co_urb25='Catawba'co_urb26='Catawba,114'co_urb27='Chatham'co_urb28='Chatham,1
03'co_urb29='Cherokee'
co_urb30='Chowan'co_urb31='Clay'co_urb32='Cleveland'co_urb33='Columbus'co_urb
34='Craven'
co_urb35='Cumberland,104'co_urb36='Cumituck'co_urb37='Dare'co_urb38='Davison'
co_urb39='Davison,107'co_urb40='Daive'co_urb41='Duplin'co_urb42='Durham'co_ur
b43='Durham,103'
co_urb44='Edgecombe'co_urb45='Edgecombe,116'co_urb46='Forsyth'
co_urb47='Forsyth,107',co_urb48='Franklin'
co_urb49='Gaston'
co_urb50='Gaston,105'co_urb51='Gates'co_urb52='Graham'co_urb53='Granville'co_
urb54='Greene'
co_urb55='Guilford'co_urb56='Guilford,101' co_urb57='Guilford,106'
co_urb58='Guilford,107',
co_urb59='Halifax'co_urb60='Harnett'co_urb61='Harnett,104',co_urb62='Haywood'
co_urb63='Henderson'co_urb64='Hertford'co_urb65='Hoke'co_urb66='Hyde'co_urb67
='Iredell'
co_urb68='Jackson'co_urb69='Johnston'co_urb70='Jones'co_urb71='Lee'co_urb72='
Lenoir'
co_urb73='Lincoln'co_urb74='Macon'co_urb75='Madison'co_urb76='Martin'co_urb77
='McDowell'
co_urb78='Mecklenburg'co_urb79='Mecklenburg,102'co_urb80='Mitchell'co_urb81='
Montgomery'co_urb82='Moore'
co_urb83='Nash'co_urb84='Nash,116'co_urb85='New
Hanover,109'co_urb86='Northampton'co_urb87='Onslow'
co_urb88='Onslow,112'co_urb89='Orange'co_urb90='Orange,101'co_urb91='Orange,1
03'co_urb92='Pamlico'
co_urb93='Pasquotank'co_urb94='Pender'co_urb95='Perquimans'co_urb96='Person'c
o_urb97='Pitt'
co_urb98='Pitt,
115'co_urb99='Pork'co_urb100='Randolph'co_urb101='Randolph,107'co_urb102='Ric
hmond'
co_urb103='Robeson'co_urb104='Rockingham'co_urb105='Rowan'co_urb106='Rowan,11
1'co_urb107='Rutherford'
co_urb108='Sampson'co_urb109='Scotland'co_urb110='Stanly'co_urb111='Stokes'co
_urb112='Surry'
co_urb113='Swain'co_urb114='Transylvania'co_urb115='Tyrrell'co_urb116='Union'
co_urb117='Vance'
```

```
co_urb118='Wake'co_urb119='Wake,108'co_urb120='Warren'co_urb121='Washington'c
o_urb122='Watauga'
co_urb123='Wayne'co_urb124='Wayne,113'co_urb125='Wilkes'co_urb126='Wilson'co_
urb127='Yadkin'co_urb128='Yancey'

lanes_rte1='1, US' lanes_rte2='1, SR' lanes_rte3='2, INTER' lanes_rte4='2,
US'
lanes_rte5='2, NC' lanes_rte6='2, SR' lanes_rte7='2, LOCAL' lanes_rte8='3,
US'
lanes_rte9='3, NC' lanes_rte10='3, SR'lanes_rte11='4, INTER' lanes_rte12='4,
US'
lanes_rte13='4, NC' lanes_rte14='4, SR' lanes_rte15='5, US'lanes_rte16='5,
SR'
lanes_rte17='6, INTER'lanes_rte18='6, US' lanes_rte19='6, NC' lanes_rte20='6,
SR'
lanes_rte21='8, INTER' lanes_rte22='8, US'lanes_rte23='12, INTER'

speed_rte1='20, US'speed_rte2='20, NC' speed_rte3='20, SR'
speed_rte4='25, US'speed_rte5='25, NC' speed_rte6='25, SR' speed_rte7='25,
LOCAL'
speed_rte8='30, US'speed_rte9='30, NC' speed_rte10='30, SR'
speed_rte11='35, US' speed_rte12='35, NC' speed_rte13='35, SR'
speed_rte14='40, US' speed_rte15='40, NC' speed_rte16='40, SR'
speed_rte17='45, INTER' speed_rte18='45, US' speed_rte19='45,
NC'speed_rte20='45, SR'
speed_rte21='50, INTER' speed_rte22='50, US' speed_rte23='50,
NC'speed_rte24='50, SR'
speed_rte25='55, INTER' speed_rte26='55, US' speed_rte27='55,
NC'speed_rte28='55, SR'
speed_rte29='60, INTER' speed_rte30='60, US' speed_rte31='60, NC'
speed_rte32='65, INTER' speed_rte33='65, US'
speed_rte34='70, INTER'

lanes_speed_rte1='1, 20, US' lanes_speed_rte2='1, 35, US'
lanes_speed_rte3='1, 35, SR'
lanes_speed_rte4='1, 55, SR' lanes_speed_rte5='2, 20, US'
lanes_speed_rte6='2, 20, NC'
lanes_speed_rte7='2, 20, SR' lanes_speed_rte8='2, 25, US'
lanes_speed_rte9='2, 25, NC'
lanes_speed_rte10='2, 25, SR' lanes_speed_rte11='2, 25, LOCAL'
lanes_speed_rte12='2, 30, US'
lanes_speed_rte13='2, 30, NC' lanes_speed_rte14='2, 30, SR'
lanes_speed_rte15='2, 35, US'
lanes_speed_rte16='2, 35, NC' lanes_speed_rte17='2, 35, SR'
lanes_speed_rte18='2, 40, US'
lanes_speed_rte19='2, 40, NC' lanes_speed_rte20='2, 40, SR'
lanes_speed_rte21='2, 45, US'
lanes_speed_rte22='2, 45, NC' lanes_speed_rte23='2, 45, SR'
lanes_speed_rte24='2, 50, US'
lanes_speed_rte25='2, 50, NC' lanes_speed_rte26='2, 50, SR'
lanes_speed_rte27='2, 55, INTER'
lanes_speed_rte28='2, 55, US' lanes_speed_rte29='2, 55, NC'
lanes_speed_rte30='2, 55, SR'
lanes_speed_rte31='2, 60, US' lanes_speed_rte32='2, 65, INTER'
lanes_speed_rte33='2, 65, US'
lanes_speed_rte34='2, 70, INTER'
```

```
lanes_speed_rte35='3, 20, US' lanes_speed_rte36='3, 20, SR'
lanes_speed_rte37='3, 25, US'
lanes_speed_rte38='3, 25, SR' lanes_speed_rte39='3, 30, SR'
lanes_speed_rte40='3, 35, US'
lanes_speed_rte41='3, 35, NC' lanes_speed_rte42='3, 35, SR'
lanes_speed_rte43='3, 40, US'
lanes_speed_rte44='3, 40, NC' lanes_speed_rte45='3, 45, US'
lanes_speed_rte46='3, 45, NC'
lanes_speed_rte47='3, 45, SR'
lanes_speed_rte48='3, 55, US' lanes_speed_rte49='3, 55, NC'
lanes_speed_rte50='3, 55, SR'
lanes_speed_rte51='4, 20, US' lanes_speed_rte52='4, 20, NC'
lanes_speed_rte53='4, 20, SR'
lanes_speed_rte54='4, 25, US' lanes_speed_rte55='4, 25, NC'
lanes_speed_rte56='4, 25, SR'
lanes_speed_rte57='4, 30, US' lanes_speed_rte58='4, 35, US'
lanes_speed_rte59='4, 35, NC'
lanes_speed_rte60='4, 35, SR' lanes_speed_rte61='4, 40, US'
lanes_speed_rte62='4, 40, NC'
lanes_speed_rte63='4, 40, SR' lanes_speed_rte64='4, 45, US'
lanes_speed_rte65='4, 45, NC' lanes_speed_rte66='4, 45, SR'
lanes_speed_rte67='4, 50, INTER'
lanes_speed_rte68='4, 50, US' lanes_speed_rte69='4, 50, NC'
lanes_speed_rte70='4, 50, SR'
lanes_speed_rte71='4, 55, INTER' lanes_speed_rte72='4, 55, US'
lanes_speed_rte73='4, 55, NC'
lanes_speed_rte74='4, 55, SR' lanes_speed_rte75='4, 60, INTER'
lanes_speed_rte76='4, 60, NC'
lanes_speed_rte77='4, 65, INTER' lanes_speed_rte78='4, 65,
US'lanes_speed_rte79='4, 70, INTER'
lanes_speed_rte80='5, 35, US' lanes_speed_rte81='5, 35, SR'
lanes_speed_rte82='5, 45, US'
lanes_speed_rte83='5, 45, SR' lanes_speed_rte84='5, 50, SR'
lanes_speed_rte85='5, 55, SR'
lanes_speed_rte86='6, 35, US' lanes_speed_rte87='6, 35, NC'
lanes_speed_rte88='6, 35, SR'
lanes_speed_rte89='6, 40, US' lanes_speed_rte90='6, 45, US'
lanes_speed_rte91='6, 45, NC'
lanes_speed_rte92='6, 45, SR' lanes_speed_rte93='6, 50, US'
lanes_speed_rte94='6, 55, INTER' lanes_speed_rte95='6, 55, US'
lanes_speed_rte96='6, 55, SR'
lanes_speed_rte97='6, 60, INTER' lanes_speed_rte98='6, 65, INTER'
lanes_speed_rte99='6, 65, US'
lanes_speed_rte100='6, 70, INTER' lanes_speed_rte101='8, 45, INTER'
lanes_speed_rte102='8, 45, US'
lanes_speed_rte103='8, 55, INTER' lanes_speed_rte104='8, 60, INTER'
lanes_speed_rte105='8, 65, INTER'
lanes_speed_rte106='12, 55, INTER'

cycle_surface1='Ann, Asph' cycle_surface2='Ann, Conc' cycle_surface3='Ann,
Soil'
cycle_surface4='Odd, Conc' cycle_surface5='Odd, Soil'
cycle_surface6='Even, Conc' cycle_surface7='Even, Soil'
cycle_surface8='Vari, Asph' cycle_surface9='Vari, Conc'
cycle_surface10='Vari, Soil'
```

```
a_control1='Low' a_control2='Medium'

year1='1997' year2='1998' year3='1999' year4='2000'year5='2001'
month1='Jan' month2='Feb' month3='Mar' month4='Apr' month5='May'
month6='June' month7='July' month8='Aug'
month9='Sept' month10='Oct' month11='Nov' month12='Dec'
day1='Mon' day2='Tue' day3='Wed' day4='Thu'

lucode1='11,Residential' lucode2='12,Comm & Services' lucode3='13,Industrial'
lucode4='14,Trans,Comm,Util' lucode5='15,INDUST & COMMERC CMPLXS '
lucode6='16,MXD URBAN OR BUILT-UP' lucode7='17,OTHER URBAN OR BUILT-U'
lucode8='21,CROPLAND AND PASTURE' lucode9='22,ORCHGROVVNYRDNURSORN'
lucode10='23,CONFINED FEEDING OPS'
lucode11='24,OTHER AGRICULTURAL LAND' lucode12='31,HERBACEOUS RANGELAND'
lucode13='32,SHRUB & BRUSH RANGELAND' lucode14='41,DECIDUOUS FOREST LAND'
lucode15='42,EVERGREEN FOREST LAND' lucode16='43, MIXED FOREST LAND'
lucode17='51,STREAMS AND CANALS'
lucode18='52,LAKES' lucode19='53,RESERVOIRS'
lucode20='54,BAYS AND ESTUARIES' lucode21='61,FORESTED WETLAND'
lucode22='62,NONFORESTED WETLAND' lucode23='72,BEACHES' lucode24='73,SANDY
AREA (NON-BEACH)'
lucode25='75,STRIP MINES' lucode26='76,TRANSITIONAL AREAS'

   city_urb1='rural' city_urb2='urban';
```

APPENDIX 11:

11. Primary Road Statewide Traffic Continuity Map

Shannon McDonald

March 29, 2005

PRIMARY ROAD STATEWIDE TRAFFIC CONTINUITY MAP

Data Preparation

This section is designed as a reference to the procedures and assumptions used to create the statewide traffic volume map, as deemed by project task 6. Problems and deviations are also included in this report.

Exercise: To create a statewide traffic continuity map for primary highways. The two primary data sources are: the ptc station data, represented by a point type shapefile and associated database; and the road network currently being used within this project, derived from the NCDOT road network linework and Universe database.

Terminology

Thiessen (Voronoi) Polygons: polygon whose boundaries define the area that is closest to each point relative to all other points (defined mathematically as the perpendicular bisectors of the lines between all points).

Problems: Divided road assignments in route1 field contain a digit in the third column depicting it as a paired set. This is inconsistent in the PTC station data. The solution was to create a route field (ROUTESEL) that would be calculated for each dataset using existing values from their respected databases. Discrepancies in the actual input values between the PTC and road network files are retained and thus may still occur in ROUTESEL.

The relationship between the road network and PTC stations is irregular. That is, several segments have no station within its arc, while others have one or more. This is similar to a spatial many-to-many type relationship. To resolve this, an assumption has to be made as to where an individual station's range lies. Our solution was to create a thiessen polygon coverage for each route from the PTC station x,y coordinates. The thiessen, or proximal polygons (Voronoi cells) can be used in conjunction with spatial analysis to assign all segments values from the source dataset (PTC stations).

Software used: ArcGIS 8.3 (ArcMap, ArcEditor, ArcInfo, ARCEDIT, ARCPLOT, INFO, Microsoft Access 2000, AML)

Step 1: Creating Primary Roads dataset

Assumptions: The primary roads dataset is derived from the mergeroads shapefile. This file is a union of the NDOT Universal database and the statewide road linework (produced currently from 2001 data). The primary roads is an extracted set consisting of only features within mergeroads where values of field RTE_TYPE are less than 4 (i.e. 1 = Interstate, 2 = US, and 3 = NC). State routes, local roads, ramps and federal roads are not included. The intended use of the output is as a visual display and only attributes used for processing and displaying the data are to retain usable values. This includes RTE_TYPE, ROUTE1 and ROUTESEL.

Procedures:

1. create primary road coverage from mergeroads.shp

```
Arc: shapearc mergeroads.shp rds
Arc: copy rds prds
Arc: ae
Arcedit: ec prds arc
Arcedit: select rte_type > 3
Arcedit: delete
```

2. create and populate a new route field for consistency when selecting roads and ptc stations for the station assignment process.

```
Arcedit: additem route_sel 10 10 I
Arcedit: q
Arc: ap
Arcplot: &r calcrouteprds.aml
Arcplot: q
```

3. remove psuedonodes from segments within same route using ROUTESEL

```
Arc: copy prds uprds; ae
Arcedit: ec uprds arc
Arcedit: sel all
Arcedit: unsplit routesel
```

Notes: the AML calcrouteprds.aml copies the values from RTE to ROUTESEL. For those segments without an RTE value, a simple cursor routine extracts the first two and last three numbers from the field ROUTE1 to create a five-digit value depicting each segments route type and route number. This format is congruent with that of the RTE field in the road layer. Though the fourth digit is also used to depict route numbers, it is negligent for use with the primary routes as described above. That is, no Interstate, US or NC routes has more than three numbers in its route number designation. The second digit disassociates different routes with the same route number designation (e.g. business vs. bypass routes).

Using the unsplit method to join the arcs created a smaller file size for the road network. The downside is the usability of the fields (see assumptions), but this is offset by the need to create a more usable dataset for the PTC data assignment routine and the fact that no original road data other than the route values from ROUTESEL are needed in the output. Of 6865 arcs in uprds, 226 have a ROUTESEL value of 0.

CALCROUTEPRDS.AML

```
&if [show program] ne ARCPLOT &then
  &return &inform Run from ArcPlot

&if ^ [exists prds -cover] &then
  &return &inform Cover PRDS does not exist!

&if [iteminfo prds.aat -info routesel -exists] &then
  &do
    arc dropitem prds.aat prds.aat routesel
    arc additem prds.aat prds.aat routesel 10 10 i
  &end
&else
  arc additem prds.aat prds.aat routesel 10 10 i

calc prds.aat info routesel = rte
resel prds.aat info routesel = 0

cursor c1 declare prds.aat info
cursor c1 open

&do &while %:c1.aml$next%
  &s a [substr %:c1.route1% 1 2]
  &s b [substr %:c1.route1% 6 3]
  calc prds.aat info routesel = %a%%b%
  cursor c1 next
&end

cursor c1 close
cursor c1 remove

clearsel
q

&return
```

Step 2: Creating Primary PTC dataset

Assumptions: The primary stations is an extracted set consisting of only features within ptc_final5.shp where values of field RTE_TYPE are less than 4 (i.e. 1 = Interstate, 2 = US, and 3 = NC).

Procedures:

1. Create primary ptc coverage from ptc_final5.shp

```
Arc: shapearc ptc_final5.shp ptc
Arc: copy ptc pptc
Arc: ae
Arcedit: ec pptc point
Arcedit: sel rte_type > 3
Arcedit: delete
```

2. Create and populate a new route field for consistency when selecting roads and ptc stations for the station assignment process.

```
Arcedit: additem route_sel 10 10 I
Arcedit: q
Arc: ap
Arcplot: &r calcroutepptc.aml
Arcplot: q
```

Notes: the AML calcroutepptc.aml contains a simple cursor routine that extracts the first two and last three numbers from the field ROUTE1 to create a five-digit value depicting each segments route type and route number. This procedure populates all but 45 stations. A second routine concatenates values from the ROUTENUM and ROUTECODE fields to create the ROUTESEL values for the all but 7 of the remaining stations (see below). These last seven stations were manually assigned.

CALCROUTEPPTC.AML

```
&if [show program] ne ARCPLOT &then
  &return &inform Run from ArcPlot

&if ^ [exists pptc -cover] &then
  &return &inform Cover PPTC does not exist!

&if [iteminfo pptc.pat -info routesel -exists] &then
  &do
    arc dropitem pptc.pat pptc.pat routesel
    arc additem pptc.pat pptc.pat routesel 10 10 i
  &end
&else
  arc additem pptc.pat pptc.pat routesel 10 10 i

clearsel

cursor c1 declare pptc.pat info
cursor c1 open

&severity &error &routine bailout

&do &while %:c1.aml$next%
  &s a [substr %:c1.route1% 1 2]
  &s b [substr %:c1.route1% 6 3]
  calc pptc.pat info routesel = %a%%b%
  cursor c1 next
&end

cursor c1 close
cursor c1 remove

&severity &error &fail

clearsel
resel pptc.pat info routesel = 0

cursor c1 declare pptc.pat info
cursor c1 open

&severity &error &routine bailout

&do &while %:c1.aml$next%
  &s l [length %:c1.routenum%]
  &if %l% le 3 &then
    &do
      &if %l% = 1 &then
        &s z 000
      &if %l% = 2 &then
        &s z 00
      &if %l% = 3 &then
        &s z 0
      calc pptc.pat info routesel = %:c1.routecode%%z%%:c1.routenum%
    &end
```

```
  cursor c1 next
&end

cursor c1 close
cursor c1 remove

&severity &error &fail

&return

&routine bailout

&severity &error &fail
cursor c1 close
cursor c1 remove
```

The following adjustments were made to the seven unassigned stations:

UNIQ_ID	ROUTESEL
0550053	20023
0100090	20025
1000244	20019
1000117	20019
0800022	29074
0230025	99999999*
1090025	20017

* Station was left out of routine and later manually entered for accuracy due to its location along the route in respect to another station and intersection.

Step 3: Assigning the PTC station values to primary road segments

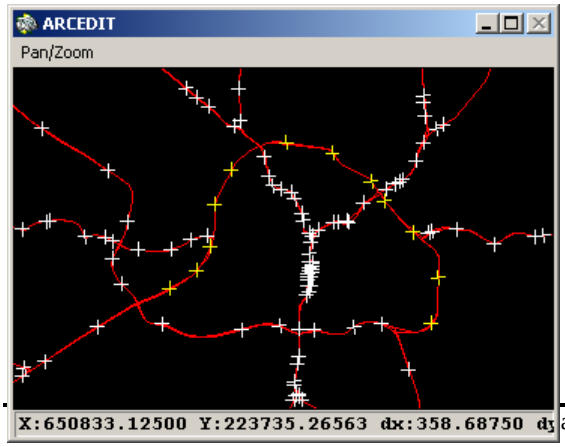
Assumptions: The range of each PTC station is exactly the distance along a give route closest to that station in relation to all others.

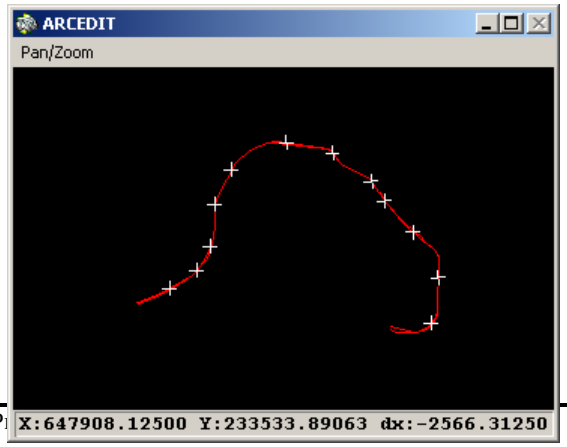
Procedures: This procedure was completed by executing tsplit.aml. The following outlines the AML:

1. A temporary file is created listing all unique values from the ROUTESEL field of cover PPTC. The file is used to select the associated records from both data sources; PPTC and UPRDS.

2. A routine opens the file, reads the first line (routesel value), and selects the PPTC and UPRDS features accordingly. These are exported into temporary covers:

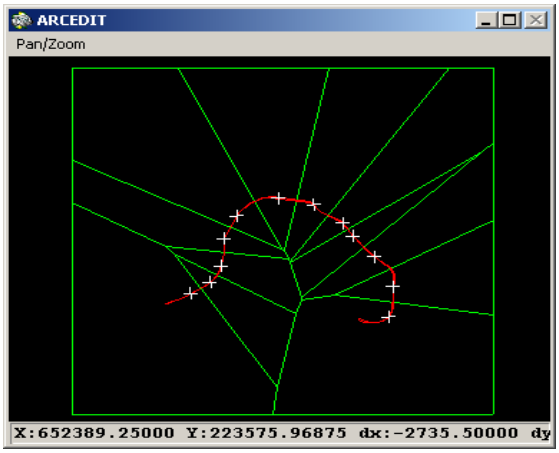
```
Arcedit: ec uprds arc
Arcedit: sel routesel = 10440
Arcedit: put temparc
Arcedit: ec pptc point
Arcedit: sel routesel = 10440
Arcedit: put temppoint
```





3. The temporary point file is used as an x,y input for creating a theissen polygon coverage. The values from the point file are retained in the polygon cover.

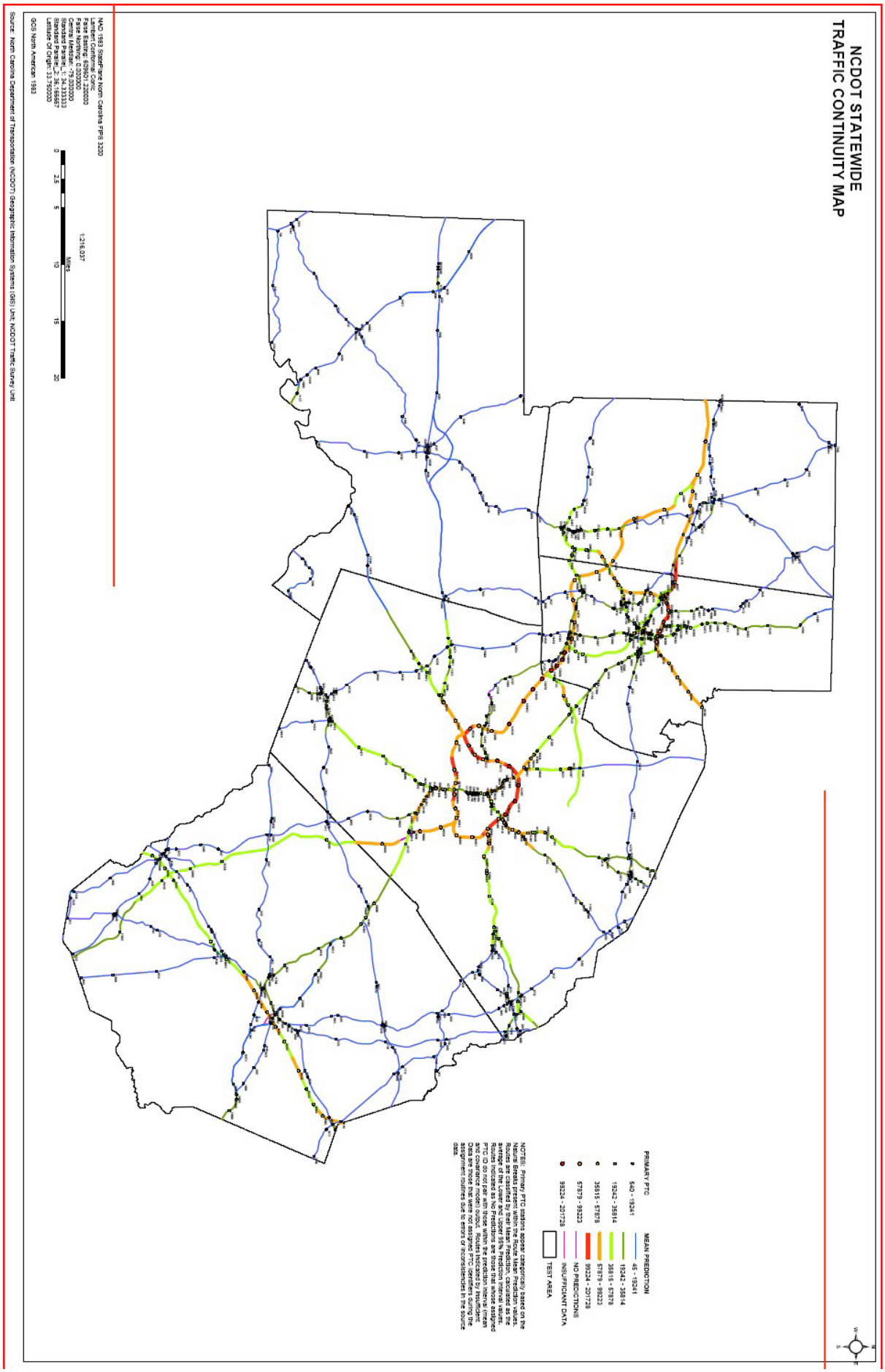
```
Arc: Thiessen temppoint temppoly
```



4. The polygon cover is then overlayed using the spatial analysis tool IDENTITY to split segments at the polygon/arc intersections and assign the PTC data to each segment within the corresponding polygon.

The final output, named IPRDS, will be used to create a statewide thematic map depicting the primary road segments as they relate to traffic volume data for their associated PTC stations.

The example map on the next page illustrates the mean prediction continuity for the five-county test area primary routes.


NCDOT STATEWIDE
TRAFFIC CONTINUITY MAP
PRIMARY PTC
MEAN PREDICTION
540 - 19241
45 - 19241
19242 - 35814
19242 - 35814
35815 - 57978
35815 - 57978
57979 - 99223
57979 - 99223
99224 - 201728
NO PREDICTIONS
INSUFFICIANT DATA
TEST AREA
NOTES: Primary PTC sections appear categorically based on the Natural Breaks present within the Route Mean Prediction values. Routes are classified by their Mean Prediction, calculated as the average of the 24 Predictions for those routes assigned PTC ID do not pair with those within the prediction intervals (mean and covariance model) output. Routes indicated by insufficient data have no predictions assigned to them. Gaps in the assigned routes are a result of inconsistencies in the source data.
NAD 1983 StatePlane North Carolina FIPS 3200
Lambert Conformal Conic
False Easting: 609601.220000
False Northing: 0.000000
Central Meridian: -79.000000
Standard Parallel_1: 34.333333
Standard Parallel_2: 36.166667
Latitude Of Origin: 33.750000
GCS North American 1983
1:216,037
Miles
0 2.5 5 10 15 20
Source: North Carolina Department of Transportation (NCDOT) Geographic Information Systems (GIS) Unit, NCDOT Traffic Survey Unit

APPENDIX 12:

12. Automation With GIS Application Procedures

Shannon McDonald

June 7, 2006

AUTOMATION WITH GIS APPLICATION PROCEDURES

The following report is designed as a reference to the design procedures for the automation process for obtaining station count data prediction intervals. The application is a function of Task 3: *Automation and GIS Implementation*.

Exercise: To create an application that programmatically obtains predictions and prediction intervals. To do this, a series of input data, both spatial and non-spatial is provided and modeled through a series of program procedures. To better organize design and documentation, the application is divided into four phases. The first three phases are ordered as prerequisite building blocks for creating the predictions. The output will be used within a custom analytical GIS interface for the display of the model output. The following outlines these phases:

Application Phase I – Build a station data table that includes information from PTC, AADT, land use, urban/municipal, x-y coordinate, and census data. This table is used as an input for Phase II. This phase reflects the initial database creation as was used as input for the statistical models.

Application Phase II – Creation of data for use in statistical process to build prediction intervals (Phase III). This phase reflects the process of interpreting the table(s) from Phase I for use with the statistical models.

Application Phase III – Applying the statistical mean model to produce MEAN model output predictions and prediction intervals (originally performed in SAS).

Custom GIS Interface – An application will be built based on the out-of-box ArcGIS GUI with customized tools and interface. This phase will utilize some of the same practices and output characteristics as used with making the primary roads continuity maps. This application will be dynamic and change over the course of user interaction to best suit the needs of the analysts.

The spatial input data is in the form of shapefiles, represented by geographic entities that are referenced to a station by spatial proximity. A station identifier links non-spatial data representing count information as a table-join. It is of principal note that the data fields and files outlined in the *Program Procedures* section of this report be present and unchanged from the format described in order for the application to execute properly.

Statistical modeling provides other data such as that for the matrix calculation algorithm. This data is used during the matrix calculations algorithm, or Phase III.

The application is written using ESRI® ArcObjects™ 8.3 in Microsoft® Visual Basic Editor (VBA) version 6.0.

Problems: In addition to the large amount of data and long processes, the primary concern regards precision. The process failed to duplicate the numbers provided from the SAS output (created on a UNIX system), and the cause is believed to be with the range abilities provided for the PC environment.

Program Procedures

The following section describes in detail each of the application procedures. The application is again divided into four phases as described above. Each phase has a series of steps, or *procedures* performed

during the application execution. The insertion of code snippets—indicated by a change in font (Courier New)—is intended as a reading aid and may not reflect the final code.

Supportive procedures—those that do not directly build or model the datasets but rather act as administrative or maintenance routines—may not be included in this section. An exception is the file-checking procedures. Knowledge of data structure is vital to the application's successful execution.

Application Phase I

The first application phase builds the input table for Phase II. The input data is built through a series of spatial and non-spatial joins and calculations.

1. The initial procedure allows the user to select a county. The countywide extent is acceptable as the data is analyzed on a county-by-county basis. In addition, the smaller extent provides for faster processing. The procedure prompts the user to select a county. The county name *object* is returned and stored as a variable.

2. An output table for Phase I, named *Export_Table.dbf* is created. This table is blank initially, and will be populated with output fields during a later procedure.

3. The presences of the input shapefiles are checked.

```
...

'++ Create an array of the required shapefile names
Dim a
a = Array("mergedroads", "mb", "landuse", "majurb", & _
    "minurb", "demographics")

'++ Loop through array elements to check presence
'++ of shapefiles
...

'++ Check for presence of PTC shapefile in data folder
Set pFeatureWorkspace = & _
    pWorkspaceFactory.OpenFromFile(strDataFolder, 0)
Set pFeatureLayer = New FeatureLayer
Set pFeatureLayer.FeatureClass = & _
    pFeatureWorkspace.OpenFeatureClass("ptc.shp")
...
```

4. A new map frame is created within the ArcMap application. The data processing occurs within this new map, which is deleted near the script's termination. This organization prevents conflict with data that may exist within the table of contents (TOC) during the scripts' execution. The map, named *Build Table* will appear in the ArcMap TOC during the execution. At finalization, *Build Table* is removed.

5. The shapefiles (input layers and PTC station file) are added to the *Build Table* map frame.

6. The stations are selected from the PTC shapefile based on the county name variable (step 1). These are put into a temporary shapefile and added to the TOC. It is with this table that the number of records in the export table will be defined.

```
...

'++ Select ptc by county
Dim pActiveView As IActiveView
Dim pQueryFilter As IQueryFilter

Set pActiveView = pMap

'++ Create the query filter
Set pQueryFilter = New QueryFilter
pQueryFilter.WhereClause = "CO_NAME = '" & m_CountyName & "'"

'++ Invalidate only the selection cache
'++ Flag the original selection
pActiveView.PartialRefresh esriViewGeoSelection, Nothing, Nothing
'++ Perform the selection
pFSel.SelectFeatures pQueryFilter, esriSelectionResultNew, False
'++ Flag the new selection
pActiveView.PartialRefresh esriViewGeoSelection, Nothing, Nothing

'++ Get the selection set
Dim pSelSet As ISelectionSet
Set pSelSet = pFSel.SelectionSet
'++ If none were selected, prompt user and select all features in '++ PTC
If pSelSet.Count = 0 Then
  MsgBox "There are no PTC stations in " & m_CountyName & " County"
  DeleteDataFrame
  Exit Sub
End If

'Export
Dim pExportOp As IExportOperation
Dim pFCNew As IFeatureClass
Set pExportOp = New ExportOperation
pExportOp.ExportFeatureClass pInDsName, Nothing, pSelSet, & _
  Nothing, pName, 0
...
```

Note the CO_NAME field must be present in the PTC shapefile. If no records are returned, an error message is thrown and the bailout routine is run.

7. Fields are added to the temporary shapefile for the calculations of the municipal flag, urban flag, x-coordinate and y-coordinate values. These fields are named MUNI_FLG, URB_FLG, X-COORD and Y-COORD, respectfully.

```
...

'++ Add fields to output layer of temp shapefile
Dim pMuniFieldEdit As IFieldEdit
Dim pUrbFieldEdit As IFieldEdit
```

```
Dim pXFieldEdit As IFieldEdit
Dim pYFieldEdit As IFieldEdit

Set pMuniFieldEdit = New Field
With pMuniFieldEdit
  .name = "MUNI_FLG"
  .Type = esriFieldTypeSmallInteger
End With
pPntLayer.FeatureClass.AddField pMuniFieldEdit

Set pUrbFieldEdit = New Field
With pUrbFieldEdit
  .name = "URB_FLG"
  .Type = esriFieldTypeSmallInteger
End With
pPntLayer.FeatureClass.AddField pUrbFieldEdit

Set pXFieldEdit = New Field
With pXFieldEdit
  .name = "X-COORD"
  .Type = esriFieldTypeDouble
End With
pPntLayer.FeatureClass.AddField pXFieldEdit

Set pYFieldEdit = New Field
With pYFieldEdit
  .name = "Y-COORD"
  .Type = esriFieldTypeDouble
End With
pPntLayer.FeatureClass.AddField pYFieldEdit
…
```

Note the field types for each field indicated in the code above. The X- and Y-COORD fields were omitted from the final output of this application (step 12). However, processing time is not effected by the inclusion of these procedure statements.

8. The MUNI_FLG and URB_FLG fields are calculated using a select-by-location method with the temporary PTC shapefile and the municipal boundary and urban boundary shapefiles. The values are classifications, indicating whether a station falls within or outside of an urban or municipal polygon. Note in the code below, the ‘l’ variable represents a polygon shapefile (municipal boundary, major urban boundary, or minor urban boundary).

```
…

'++ Create a case scenerio for each layer

Select Case l
  Case "mb"
    f = pFields.FindField("MUNI_FLG")

    Set pFCursor = pFc.Update(pSpFilter, False)
    Set pFeature = pFCursor.NextFeature

    Do Until pFeature Is Nothing
```

```
      pFeature.Value(f) = "1"
      pFCursor.UpdateFeature pFeature
      Set pFeature = pFCursor.NextFeature
    Loop
    Set pFCursor = Nothing
    Case "majurb"
      f = pFields.FindField("URB_FLG")
      Set pFCursor = pFc.Update(pSpFilter, False)
      Set pFeature = pFCursor.NextFeature
      Do Until pFeature Is Nothing
        pFeature.Value(f) = "1"
        pFCursor.UpdateFeature pFeature
        Set pFeature = pFCursor.NextFeature
      Loop
      Set pFCursor = Nothing
    Case "minurb"
      f = pFields.FindField("URB_FLG")
      Set pFCursor = pFc.Update(pSpFilter, False)
      Set pFeature = pFCursor.NextFeature
      Do Until pFeature Is Nothing
        pFeature.Value(f) = "2"
        pFCursor.UpdateFeature pFeature
        Set pFeature = pFCursor.NextFeature
      Loop
      Set pFCursor = Nothing
  End Select

Next l
...
```

9. Next, the X- and Y-coordinates are calculated for the X-COORD and Y-COORD fields. The *shape* property of the PTC point featureclass is referenced for the calculation. Thus, the output values will reflect the shapefile geometry (from the *IGeometry* interface). In this case the coordinates reflect those of the NAD 1983 State Plane North Carolina 3200 (meters).

```
...
pFeature.Value(pFClass.FindField("X-COORD")) = pPoint.X
pFeature.Value(pFClass.FindField("Y-COORD")) = pPoint.Y
...
```

   Note that *pFeature* represents a feature within a cursor routine.

10. A spatial join, or join-by-location is performed to relate the data from the *demographics* (census) and *landuse* shapefiles to the temporary PTC file. The output from this routine is a new temporary point feature shapefile that contains the fields of all three datasets. The output data is added as a layer to the TOC.

11. A second spatial join procedure joins the line-type shapefile *mergedroads* file to the temporary output file from step 10. A new temporary point-type shapefile is created that contains the fields of both datasets. The new shapefile is added to the TOC.

12. A final output table is created. This table is exported as a dBase format file and is derived from the output file from step 11.

```
...

'++ Define the output feature class name
Dim pFeatureClassName As IFeatureClassName
Dim pOutDatasetName As IDatasetName
Dim pName As IName
Set pFeatureClassName = New FeatureClassName
Set pOutDatasetName = pFeatureClassName
With pOutDatasetName
  .name = "Export_Table"
  Set .WorkspaceName = pWkSpName
End With
Set pName = pOutDatasetName

'++ Select fields to include in output
Dim pQFilter As IQueryFilter
Set pQFilter = New QueryFilter
pQFilter.SubFields = "FACTOR1, FACTOR2, FACTOR3, FACTOR4, " & _
                     "FACTOR6, " & _
                     "UNIQ_ID, SIPS, CO_NUM, CYCLE_YR, " & _
                     "YEAR, MONTH, DAY, RTE_TYPE, LANES, " & _
                     "SPEED, MUNI_FLG, URB_FLG, " & _
                     "LUCODE, RD_SURFACE, A_CONTROL"

'++ Export
Dim pExportOp As IExportOperation
Dim pFCNew As IFeatureClass
Set pExportOp = New ExportOperation
pExportOp.ExportTable pInDsName, pQFilter, Nothing, pName, 0

...
```

Note the fields that are exported. These fields were determined to be significant factors and are used within the statistical mean model.

The remaining procedures for *Application Phase I* remove the data frame from the ArcMap TOC and delete the temporary files.

The export file remains in the *Input* data folder until the application is rerun.

The data dictionary for the output file *Export_Table.dbf* is as follows:

NAME	TYPE	LENGTH	PRECISION	SCALE	DATA SOURCE
FACTOR1*	DOUBLE	19	18	11	demographics.dbf
FACTOR2*	DOUBLE	19	18	11	demographics.dbf
FACTOR3*	DOUBLE	19	18	11	demographics.dbf
FACTOR4*	DOUBLE	19	18	11	demographics.dbf
FACTOR6*	DOUBLE	19	18	11	demographics.dbf
UNIQ_ID+	STRING	7	0	0	ptc.dbf
SIPS	LONG	9	9	0	ptc.dbf
CO_NUM	LONG	9	9	0	ptc.dbf
CYCLE_YR++	STRING	1	0	0	ptc.dbf
YEAR++	DOUBLE	19	18	11	ptc.dbf
MONTH++	DOUBLE	19	18	11	ptc.dbf
DAY++	DOUBLE	19	18	11	ptc.dbf
RTE_TYPE	SHORT	2	2	0	mergedroads.dbf
LANES	DOUBLE	11	11	0	mergedroads.dbf
SPEED	DOUBLE	11	11	0	mergedroads.dbf
MUNI_FLG	SHORT	4	4	0	application procedure
URB_FLG	SHORT	4	4	0	application procedure
LUCODE	SHORT	4	4	0	landuse.dbf
RD_SURFACE	DOUBLE	19	18	5	mergedroads.dbf
A_CONTROL	DOUBLE	19	18	5	mergedroads.dbf

Table 1: *Export_Table.dbf* data dictionary

The fields and their values are derived from the statistical methodology and were joined to the original demographics table prior to and for this application. The factor values derived from the PTC prediction interval table and was joined by spatial process. Because not all demographic polygons contain a PTC point, they do not all have a value for these factors. Such instances are flagged with an error code in a later application procedure.

+ A number-type field has replaced the unique identifier. A crossover table has been created to enable a conversion between these two identifiers.

++ This data present in the ptc.dbf table is derived from the TSU database and is described in the report *Data Description for Spatial Study* submitted by Larry Wikoff on May 17th, 2002.

Application Phase II

The second application phase builds the input table for Phase III. The table is interpreted for its use in the statistical mean model application. The events parallel those endured for the data preparation for the SAS modeling. The output table is the attribute matrix for *new* stations (those without prediction intervals).

The procedure is simply a series of case scenarios and line-writes ultimately creating a comma-delimited ASCII file (*XO.txt*) populated with the X-Matrix table values for the target stations.

The fields are representative of the model *variables*. The FactorX values are calculated directly from the *demographics* data table. The remaining variables are calculated with binary values by *Select Case* statements.

The following snippet shows the *Select Case* code for assigning the data based on a station's route type (intRTE field value). Each station has a value representing its route type. For the analysis, we are

interested in three of these types: ‘1’, ‘2’ and ‘4’, representing interstate, US highway, and local routes, respectfully. All other routes (SR and miscellaneous) are grouped into a fourth set (Case Else). Each case can be represented and stored using three binary digits, as the code indicates.

…

```
'++ Capture rte_classX variables
    intRTE = pRow.Value(pTable.FindField("RTE_TYPE"))
    Select Case intRTE
      Case 1
        strRTE = "1,0,0"
      Case 2
        strRTE = "0,1,0"
      Case 4
        strRTE = "0,0,1"
      Case Else
        strRTE = "0,0,0"
    End Select
```

…

The table excerpts illustrates the data population taking for six stations*:

Export_Table.dbf (from Phase I)

…	Uniq_ID	RTE_TYPE	…
	0580001	2	
	0580002	4	
	0580003	1	
	0580004	3	
	0580005	1	
	0580006	5	

XO.txt populated fields for route class

…	rte_class1	rte_class2	rte_class4	…
	0	1	0	
	0	0	1	
	1	0	0	
	0	0	0	
	1	0	0	
	0	0	0	

* this data is not representative of the actual stations

Table 2: Cross Reference for route class field population

This process is continued for each of the 179 fields, and in turn for each station until all the selected stations are represented. Refer to Appendix 9 Table 2 for the complete field listing.

Application Phase III

The product of the third phase of the application is the station endpoints. These data represent the lower and upper 95 and 99 percentile prediction intervals for each station. Using the given data produced during the statistical evaluation of the known stations, together with the matrix produced in phase II of this application, the endpoints were derived from the mean model formulae (refer to section 3.2 of the report).

Given:

Comma-delimited file *Beta.txt* – Parameter Estimates as a vector derived from the diagonal of the 179 x 179 matrix.

Comma-delimited file *XTXinv.txt* – 179 x 179 matrix provided by Tae-Young Hoe.

Comma-delimited file *XO.txt* derived from phase II of the application.

The result is a comma-delimited file *PI_Results.txt*. This file contains the unique station identifier followed by the lower 99%, lower 95%, upper 95% and upper 99% values, respectively. At this point, the data is available to be joined with the PTC station shapefile(s) or any data set that contains the unique identifier.

UNIQ_ID	LPRED99	LPRED95	UPRED95	UPRED99
0910628	46.6902624633054	110.655934753657	4934.12764619288	7683.16728401482
0911510	76.6830045926381	172.850750026247	6567.28193447143	10088.3058794311
0910119	16.1286811408223	51.425266272942	5389.69042812855	8972.51552582362

Table 3: Excerpt of *PI_Results.txt* (may not represent actual data)

VBA Code (ArcObjects)

The following pages document the code used during the development of each of these phases. They are shown separated and in sequence (order-of-execution). This is the *beta* version of the application. The working application combines the scripts and calls each sub from *Main*. The workspace and subfolders are declared first and once (as global) for maintenance simplicity.

Application Code Phase I (beta)

```
Option Explicit
Public Const strDataFolder As String = "C:\NCSU\GIS\Application\Data"
Public Const strAnalysisFolder As String = "C:\NCSU\GIS\Application\Data\Analysis"
Public Const strTempFolder As String = "C:\NCSU\GIS\Application\Data\aaTemp"
Public Const strInputFolder As String = "C:\NCSU\GIS\Application\Data\Input"
Public m_CountyName As String

'++ 0) Populate county variable using frmCountyPick

Public Sub LaunchForm()
  m_CountyName = ""
  load frmCountyPick
  frmCountyPick.Show
  Do While m_CountyName = ""
    MsgBox "Please pick a county!"
    frmCountyPick.Show
  Loop

  ExitRoutine

End Sub

'++ 1) Exit routine runs when premature exit occurs
'++    and at startup to ensure the TOC map frame
'++    Build Table and temp directory aaTemp and
'++    the export table do not
'++    exist or cause error during runtime

Public Sub ExitRoutine()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Checking environment..."

  '++ Run DeleteDataFrame to expunge Build Table from TOC
  '++ (DeleteDataFrame consequently runs DeleteTempFolder

  DeleteDataFrame

  '++ check to see if export file exists
  Dim pTargetFileName As String
  pTargetFileName = strInputFolder & "\Export_Table.dbf"
  Dim fs
  Set fs = CreateObject("Scripting.FileSystemObject")
  If fs.fileexists(pTargetFileName) Then
    '++ present the user with an option to exit the routine if
    '++ the export file exists
    If (MsgBox("The file " & pTargetFileName & " exists." & vbNewLine & "Do you want" & _
        " to overwrite the file?", vbYesNo, "File Not Found") = vbYes) Then
      fs.deletefile pTargetFileName
    Else
      Exit Sub
    End If
  End If

  '++ Run next procedure
  CheckForFiles

End Sub

'++ 2) Check for required files and
'++    create the temp workspace

Public Sub CheckForFiles()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
```

```
  psbar.ProgressBar.Message = "Checking for workspace and input files..."

  Dim pWorkspaceFactory As IWorkspaceFactory
  Dim pFeatureWorkspace As IFeatureWorkspace
  Dim pFeatureLayer As IFeatureLayer

  '++ Create a new ShapefileWorkspaceFactory object and open a shapefile folder
  Set pWorkspaceFactory = New ShapefileWorkspaceFactory
  On Error Resume Next
  Set pFeatureWorkspace = pWorkspaceFactory.OpenFromFile(strInputFolder, 0)

  '++ Create an array of the required shapefile names
  Dim a
  a = Array("mergedroads", "mb", "landuse", "majurb", "minurb", "demographics")

  '++ Loop through array elements to check presence
  '++ of shapefiles
  Dim i
  For Each i In a
    Set pFeatureLayer = New FeatureLayer
    Set pFeatureLayer.FeatureClass = pFeatureWorkspace.OpenFeatureClass(i)
    If pFeatureLayer.FeatureClass Is Nothing Then
      If (MsgBox("The " & i & " shapefile can not be found.  Exiting Procedure." & _
        , vbExclamation, "File Not Found") = vbOK) Then
        Exit Sub
      End If
    End If
  Next i

  '++ Check for presence of PTC shapefile in data folder
  Set pFeatureWorkspace = pWorkspaceFactory.OpenFromFile(strDataFolder, 0)
  Set pFeatureLayer = New FeatureLayer
  Set pFeatureLayer.FeatureClass = pFeatureWorkspace.OpenFeatureClass("ptc.shp")
  If pFeatureLayer.FeatureClass Is Nothing Then
    If (MsgBox("The PTC shapefile can not be found.  Exiting Procedure.", & _
        vbExclamation, "File Not Found") = vbOK) Then
      Exit Sub
    End If
  End If

  psbar.ProgressBar.Message = "Creating temp file..."

  '++ Create the temp shapefile workspace
  Dim pworkspacename As IWorkspaceName
  Set pWorkspaceFactory = New ShapefileWorkspaceFactory
  Set pworkspacename = pWorkspaceFactory.Create(strDataFolder, "aaTemp", Nothing, 0)

  '++ Run next procedure
  CreateAndAddNewMap

End Sub

'++ 3) Creates a new Map named Build Table for process

Public Sub CreateAndAddNewMap()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Creating Build Table data frame from TOC..."

  Dim pMxDoc As IMxDocument
  Set pMxDoc = ThisDocument

  'Create a new map
  Dim pMap As IMap
  Set pMap = New Map
  pMap.name = "Build Table"

  'Create a new MapFrame and associate map with it
  Dim pMapFrame As IMapFrame
```

```
  Set pMapFrame = New MapFrame
  Set pMapFrame.Map = pMap

  'Set the position of the new map frame
  Dim pElement As IElement
  Dim pEnv As IEnvelope
  Set pElement = pMapFrame
  Set pEnv = New Envelope
  pEnv.PutCoords 0, 0, 5, 5
  pElement.Geometry = pEnv

  'Add mapframe to the layout
  Dim pGraphicsContainer As IGraphicsContainer
  Set pGraphicsContainer = pMxDoc.PageLayout
  pGraphicsContainer.AddElement pMapFrame, 0

  'Make the newly added map the focus map
  Dim pActiveView As IActiveView
  Set pActiveView = pMxDoc.ActiveView
  If TypeOf pActiveView Is IPageLayout Then
    Set pActiveView.FocusMap = pMap
  Else
    Set pMxDoc.ActiveView = pMap
  End If

  'Refresh ActiveView and TOC
  pActiveView.Refresh
  pMxDoc.CurrentContentsView.Refresh 0

  '++ Run next routine
  AddShapefiles

End Sub

'++ 4) Add shapefiles to the new map in the TOC

Public Sub AddShapefiles()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Adding input data to TOC..."

  Dim pMxDoc As IMxDocument
  Dim pMap As IMap
  Dim pWorkspaceFactory As IWorkspaceFactory
  Dim pFeatureWorkspace As IFeatureWorkspace
  Dim pFeatureLayer As IFeatureLayer
  Set pMxDoc = ThisDocument
  Set pMap = pMxDoc.FocusMap

  '++ Create a new ShapefileWorkspaceFactory object and open a shapefile folder
  Set pWorkspaceFactory = New ShapefileWorkspaceFactory
  On Error Resume Next
  Set pFeatureWorkspace = pWorkspaceFactory.OpenFromFile(strInputFolder, 0)

  '++ Create an array of the required shapefile names
  Dim a
  a = Array("mergedroads", "mb", "landuse", "majurb", "minurb", "demographics")

  '++ Add shapefiles to TOC
  Dim i
  For Each i In a
    Set pFeatureLayer = New FeatureLayer
    Set pFeatureLayer.FeatureClass = pFeatureWorkspace.OpenFeatureClass(i)
    pFeatureLayer.name = pFeatureLayer.FeatureClass.AliasName
    pMap.AddLayer pFeatureLayer
  Next i

  '++ Add PTC shapefile to TOC
  Set pFeatureWorkspace = pWorkspaceFactory.OpenFromFile(strDataFolder, 0)
```

```
  Set pFeatureLayer = New FeatureLayer
  Set pFeatureLayer.FeatureClass = pFeatureWorkspace.OpenFeatureClass("ptc")
  pFeatureLayer.name = pFeatureLayer.FeatureClass.AliasName
  pMap.AddLayer pFeatureLayer

  '++ Run next routine
  ExportPTC

End Sub

'++ 5) Export PTC points to ptctemp shapefile for use
'++    throughout this module

Public Sub ExportPTC()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Exporting PTC data..."

  Dim pMxDoc As IMxDocument
  Dim pMap As IMap
  Set pMxDoc = ThisDocument
  Set pMap = pMxDoc.FocusMap

  '++ Set output location
  Dim pWkSpName As IWorkspaceName
  Set pWkSpName = New WorkspaceName
  pWkSpName.WorkspaceFactoryProgID = "esricore.ShapeFileWorkspaceFactory.1"
  pWkSpName.PathName = strTempFolder

  '++ Set featureclass name
  Dim pFLayer As IFeatureLayer
  Dim pFc As IFeatureClass
  Dim i As Integer
  Dim sName As String
  sName = "ptc"
  For i = 0 To pMap.LayerCount - 1
    If pMap.Layer(i).name = sName Then
      Set pFLayer = pMap.Layer(i)
      Exit For
    End If
  Next i
  Set pFc = pFLayer.FeatureClass

  'Get the FcName from the featureclass
  Dim pINFeatureClassName As IFeatureClassName
  Dim pDataset As IDataset
  Dim pInDsName As IDatasetName
  Set pDataset = pFc
  Set pINFeatureClassName = pDataset.FullName
  Set pInDsName = pINFeatureClassName

  'Define the output feature class name
  Dim pFSel As IFeatureSelection
  Dim pFeatureClassName As IFeatureClassName
  Dim pOutDatasetName As IDatasetName
  Dim pName As IName
  Set pFSel = pFLayer
  Set pFeatureClassName = New FeatureClassName
  With pFeatureClassName
    .FeatureType = esriFTSimple
    .ShapeFieldName = "Shape"
    .ShapeType = esriGeometryAny
  End With
  Set pOutDatasetName = pFeatureClassName
  With pOutDatasetName
    .name = "ptc_export"
    Set .WorkspaceName = pWkSpName
  End With
  Set pName = pOutDatasetName
```

```
  '++ Select ptc by county
  Dim pActiveView As IActiveView
  Dim pQueryFilter As IQueryFilter

  Set pActiveView = pMap

  '++ Create the query filter
  Set pQueryFilter = New QueryFilter
  pQueryFilter.WhereClause = "CO_NAME = '" & m_CountyName & "'"

  '++ Invalidate only the selection cache
  '++ Flag the original selection
  pActiveView.PartialRefresh esriViewGeoSelection, Nothing, Nothing
  '++ Perform the selection
  pFSel.SelectFeatures pQueryFilter, esriSelectionResultNew, False
  '++ Flag the new selection
  pActiveView.PartialRefresh esriViewGeoSelection, Nothing, Nothing

  'Get the selection set
  Dim pSelSet As ISelectionSet
  Set pSelSet = pFSel.SelectionSet

  '++ If none were selected, prompt user and select all features in PTC
  If pSelSet.Count = 0 Then
    MsgBox "There are no PTC stations in " & m_CountyName & " County"
    DeleteDataFrame
    Exit Sub
  End If

  'Export
  Dim pExportOp As IExportOperation
  Dim pFCNew As IFeatureClass
  Set pExportOp = New ExportOperation
  pExportOp.ExportFeatureClass pInDsName, Nothing, pSelSet, Nothing, pName, 0

  '++ Clear selections
  pFSel.SelectFeatures Nothing, esriSelectionResultSubtract, False

  '++ Add shapefile to the the TOC
  Dim pWorkspaceFactory As IWorkspaceFactory
  Dim pFeatureWorkspace As IFeatureWorkspace
  Dim pFeatureLayer As IFeatureLayer

  '++ Create a new ShapefileWorkspaceFactory object and open a shapefile folder
  Set pWorkspaceFactory = New ShapefileWorkspaceFactory
  On Error Resume Next
  Set pFeatureWorkspace = pWorkspaceFactory.OpenFromFile(strTempFolder, 0)

  '++ Add shapefile to TOC
  Set pFeatureLayer = New FeatureLayer
  Set pFeatureLayer.FeatureClass = pFeatureWorkspace.OpenFeatureClass("ptc_export")
  pFeatureLayer.name = pFeatureLayer.FeatureClass.AliasName
  pMap.AddLayer pFeatureLayer

  '++ Run next routine
  AddFields

End Sub

'++ 6) Add Fields to be populated for output table: MUNI_FLG,
'++    URB_FLG, X-COORD, Y-COORD

Public Sub AddFields()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Adding fields to ptc export file..."

  Dim pMxDoc As IMxDocument
```

```
  Dim pMap As IMap
  Set pMxDoc = ThisDocument
  Set pMap = pMxDoc.FocusMap

  '++ Get final spatial join output layer
  Dim pPntLayer As IFeatureLayer
  Dim i As Integer
  Dim sName As String
  sName = "ptc_export"
  For i = 0 To pMap.LayerCount - 1
    If pMap.Layer(i).name = sName Then
      Set pPntLayer = pMap.Layer(i)
      Exit For
    End If
  Next i

  '++ Add fields to output layer of spatial join process
  Dim pMuniFieldEdit As IFieldEdit
  Dim pUrbFieldEdit As IFieldEdit
  Dim pXFieldEdit As IFieldEdit
  Dim pYFieldEdit As IFieldEdit

  Set pMuniFieldEdit = New Field
  With pMuniFieldEdit
    .name = "MUNI_FLG"
    .Type = esriFieldTypeSmallInteger
  End With
  pPntLayer.FeatureClass.AddField pMuniFieldEdit

  Set pUrbFieldEdit = New Field
  With pUrbFieldEdit
    .name = "URB_FLG"
    .Type = esriFieldTypeSmallInteger
  End With
  pPntLayer.FeatureClass.AddField pUrbFieldEdit

  Set pXFieldEdit = New Field
  With pXFieldEdit
    .name = "X-COORD"
    .Type = esriFieldTypeDouble
  End With
  pPntLayer.FeatureClass.AddField pXFieldEdit

  Set pYFieldEdit = New Field
  With pYFieldEdit
    .name = "Y-COORD"
    .Type = esriFieldTypeDouble
  End With
  pPntLayer.FeatureClass.AddField pYFieldEdit

  '++ Run next procedure
  SpatialQuery

End Sub


'++ 7)  Calculate MUNI_FLG and URB_FLG fields
'++     using select by location

Sub SpatialQuery()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Calculating Municipal and Urban Flags..."

  Dim pMap As IMap
  Dim pDoc As IMxDocument
  Dim pFLayer As IFeatureLayer
  Dim pPntLayer As IFeatureLayer
  Dim pFc As IFeatureClass
```

```
Set pDoc = ThisDocument
Set pMap = pDoc.FocusMap

'++ This calculation is to be done outside of an edit session.
Dim pEditor As IEditor
Dim pID As New UID
pID = "esriCore.Editor"
Set pEditor = Application.FindExtensionByCLSID(pID)
If pEditor.EditState = esriStateEditing Then
  MsgBox "This sample requires that ArcMap is not in edit mode"
  Exit Sub
End If

'++ Specify the point layer
Dim i As Integer
Dim sName As String
sName = "ptc_export"
For i = 0 To pMap.LayerCount - 1
  If pMap.Layer(i).name = sName Then
    Set pPntLayer = pMap.Layer(i)
    Exit For
  End If
Next i

'++ Create array of poly layers to use as selectors
Dim a
a = Array("mb", "majurb", "minurb")

Dim l
For Each l In a

  '++ Specify the polygon layer
  sName = l
  For i = 0 To pMap.LayerCount - 1
    If pMap.Layer(i).name = sName Then
      Set pFLayer = pMap.Layer(i)
      Exit For
    End If
  Next i
  Dim pFeatSel As IFeatureSelection
  Set pFeatSel = pFLayer

  '++ Select all features in poly layer
  pFeatSel.SelectFeatures Nothing, esriSelectionResultNew, False

  Dim pSelSet As ISelectionSet
  Set pSelSet = pFeatSel.SelectionSet

  Dim pEnumGeom As IEnumGeometry
  Dim pEnumGeomBind As IEnumGeometryBind
  Set pEnumGeom = New EnumFeatureGeometry
  Set pEnumGeomBind = pEnumGeom
  pEnumGeomBind.BindGeometrySource Nothing, pSelSet

  Dim pGeomFactory As IGeometryFactory
  Dim pGeom As IGeometry
  Set pGeomFactory = New GeometryEnvironment
  Set pGeom = pGeomFactory.CreateGeometryFromEnumerator(pEnumGeom)

  Dim pSpFilter As ISpatialFilter
  Set pSpFilter = New SpatialFilter
  With pSpFilter
    Set .Geometry = pGeom
    .GeometryField = "SHAPE"
    .SpatialRel = esriSpatialRelContains
  End With

  '++ Create update cursor and calc selected set flag (field) values
  Dim pFCursor As IFeatureCursor
  Dim pFields As IFields
```

```
    Dim pFeature As IFeature
    Dim f As Long

    Set pFc = pPntLayer.FeatureClass
    Set pFields = pFc.Fields

    '++ Create a case scenerio for each layer

    Select Case l
      Case "mb"
        f = pFields.FindField("MUNI_FLG")

        Set pFCursor = pFc.Update(pSpFilter, False)
        Set pFeature = pFCursor.NextFeature

        Do Until pFeature Is Nothing
          pFeature.Value(f) = "1"
          pFCursor.UpdateFeature pFeature
          Set pFeature = pFCursor.NextFeature
        Loop
        Set pFCursor = Nothing
      Case "majurb"
        f = pFields.FindField("URB_FLG")
        Set pFCursor = pFc.Update(pSpFilter, False)
        Set pFeature = pFCursor.NextFeature
        Do Until pFeature Is Nothing
          pFeature.Value(f) = "1"
          pFCursor.UpdateFeature pFeature
          Set pFeature = pFCursor.NextFeature
        Loop
        Set pFCursor = Nothing
      Case "minurb"
        f = pFields.FindField("URB_FLG")
        Set pFCursor = pFc.Update(pSpFilter, False)
        Set pFeature = pFCursor.NextFeature
        Do Until pFeature Is Nothing
          pFeature.Value(f) = "2"
          pFCursor.UpdateFeature pFeature
          Set pFeature = pFCursor.NextFeature
        Loop
        Set pFCursor = Nothing
    End Select

  Next l

  '++ Run next procedure
  GetXY

End Sub

'++ 8)  Calculate X and Y Coordinates


Public Sub GetXY()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Calculating X,Y values..."

  Dim pMxDoc As IMxDocument
  Dim pMap As IMap
  Dim pFLayer As IFeatureLayer
  Dim pFClass As IFeatureClass
  Dim pFCursor As IFeatureCursor
  Dim pFeature As IFeature
  Dim pPoint As IPoint
  Set pMxDoc = ThisDocument
  Set pMap = pMxDoc.FocusMap
  Set pFLayer = pMap.Layer(0)
  Set pFClass = pFLayer.FeatureClass
```

```
  Set pFCursor = pFClass.Update(Nothing, False)
  Set pFeature = pFCursor.NextFeature

  Do While Not pFeature Is Nothing
    Set pPoint = pFeature.ShapeCopy
    If Not pPoint.IsEmpty Then
      pFeature.Value(pFClass.FindField("X-COORD")) = pPoint.X
      pFeature.Value(pFClass.FindField("Y-COORD")) = pPoint.Y
    End If
    pFCursor.UpdateFeature pFeature
    Set pFeature = pFCursor.NextFeature
  Loop

  Set pMxDoc = Nothing

  '++ Run next routine
  SpatialJoinPolys

  Exit Sub

EH:
  MsgBox Err.Description

End Sub

'++ 9) Join by location the demographics and landuse
'++    area shapefiles to the PTC shapefile to create
'++    new shapefile

Public Sub SpatialJoinPolys()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Spatial overlay: Areas to point (ptc)..."

  Dim pMxDoc As IMxDocument
  Dim pMap As IMap
  Set pMxDoc = ThisDocument
  Set pMap = pMxDoc.FocusMap

  '++ This section adds the joins by proximity the demographics and landuse
  '++ poly layers to the ptc point layer

  '++ Set output location and feature class name
  Dim pWkSpName As IWorkspaceName
  Set pWkSpName = New WorkspaceName
  pWkSpName.WorkspaceFactoryProgID = "esriCore.ShapeFileWorkspaceFactory.1"
  pWkSpName.PathName = strTempFolder

  '++ set source layer as point layer
  Dim pPntLayer As IFeatureLayer
  Dim i As Integer
  Dim sName As String
  sName = "ptc"
  'loop through layers
  For i = 0 To pMap.LayerCount - 1
    If pMap.Layer(i).name = sName Then
      Set pPntLayer = pMap.Layer(i)
      Exit For
    End If
  Next i

  '++ set join layers as area layers
  Dim pAreaLayer1 As IFeatureLayer
  sName = "landuse"
  'loop through layers
  For i = 0 To pMap.LayerCount - 1
    If pMap.Layer(i).name = sName Then
      Set pAreaLayer1 = pMap.Layer(i)
      Exit For
```

```
  End If
Next i

Dim pAreaLayer2 As IFeatureLayer
sName = "demographics"
'loop through layers
For i = 0 To pMap.LayerCount - 1
  If pMap.Layer(i).name = sName Then
    Set pAreaLayer2 = pMap.Layer(i)
    Exit For
  End If
Next i


'++ loop through the spatial join routine twice.  First to join demographics
'++ to ptc to create sp_join1, next to join landuse to sp_join1 to create
'++ sp_join2.  The latter output contains fields from all three attribute
'++ tables.

Dim l As Long
For l = 1 To 2

  '++ create the name object for the output join by location shapefile
  Dim strOutName As String
  Dim pFCName As IFeatureClassName
  Dim pOutDSName As IDatasetName
  Dim pName As IName
  strOutName = "sp_join" & l
  Set pFCName = New FeatureClassName
  With pFCName
    .FeatureType = esriFTSimple
    .ShapeFieldName = "Shape"
    .ShapeType = esriGeometryPoint
  End With
  Set pOutDSName = pFCName
  With pOutDSName
    .name = strOutName
    Set .WorkspaceName = pWkSpName
  End With
  Set pName = pOutDSName

  '++ Set the first layer as the source (point) layer
  Set pPntLayer = pMap.Layer(0)

  '++ Do a join by location that joins the attributes of the
  '++ first point cantained within each polygon.
  Dim pSpJoin As ISpatialJoin
  Dim pFCNew As IFeatureClass
  Set pSpJoin = New SpatialJoin

  With pSpJoin
    If l = 1 Then
      Set .JoinTable = pAreaLayer1.FeatureClass
    ElseIf l = 2 Then
      Set .JoinTable = pAreaLayer2.FeatureClass
    End If
    Set .SourceTable = pPntLayer.FeatureClass
    .LeftOuterJoin = True
  End With

  '++ setting maxMapDist to 0 means that only points within
  '++ each each polygon will be considered
  Set pFCNew = pSpJoin.JoinNearest(pName, 0)

  '++ Create a new layer and add it to the Map
  If Not pFCNew Is Nothing Then
    Dim pNewFLayer As IFeatureLayer
    Set pNewFLayer = New FeatureLayer
    Set pNewFLayer.FeatureClass = pFCNew
    pNewFLayer.name = "Join by Location" & l
```

```
      pMap.AddLayer pNewFLayer
      pMxDoc.UpdateContents
    End If
    Set pPntLayer = Nothing

  Next l

  '++ Run next procedure
  SpatialJoinLine

End Sub

'++ 10) Join by location the mergedroads polyline
'++    shapefile to the joined shapefile (9) to create
'++    new shapefile

Public Sub SpatialJoinLine()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Spatial overlay: Line (mergedroads) to point (ptc)..."

  Dim pMxDoc As IMxDocument
  Dim pMap As IMap
  Set pMxDoc = ThisDocument
  Set pMap = pMxDoc.FocusMap

  '++ This section joins by proximity the mergedroads
  '++ line layer to the final output point layer from the
  '++ SpatialJoinPolys procedure

  '++ Set output location and feature class name
  Dim pWkSpName As IWorkspaceName
  Set pWkSpName = New WorkspaceName
  pWkSpName.WorkspaceFactoryProgID = "esriCore.ShapeFileWorkspaceFactory.1"
  pWkSpName.PathName = strTempFolder

  '++ set source layer as point layer
  Dim pPntLayer As IFeatureLayer
  Dim i As Integer
  Dim sName As String
  sName = "Join by Location2"
  'loop through layers
  For i = 0 To pMap.LayerCount - 1
    If pMap.Layer(i).name = sName Then
      Set pPntLayer = pMap.Layer(i)
      Exit For
    End If
  Next i

  '++ set join layers as area layers
  Dim pLineLayer As IFeatureLayer
  sName = "mergedroads"
  For i = 0 To pMap.LayerCount - 1
    If pMap.Layer(i).name = sName Then
      Set pLineLayer = pMap.Layer(i)
      Exit For
    End If
  Next i

  '++ create the name object for the output join by location shapefile
  Dim strOutName As String
  Dim pFCName As IFeatureClassName
  Dim pOutDSName As IDatasetName
  Dim pName As IName
  strOutName = "sp_join3"
  Set pFCName = New FeatureClassName
  With pFCName
    .FeatureType = esriFTSimple
    .ShapeFieldName = "Shape"
```

```
    .ShapeType = esriGeometryPoint
  End With
  Set pOutDSName = pFCName
  With pOutDSName
    .name = strOutName
    Set .WorkspaceName = pWkSpName
  End With
  Set pName = pOutDSName

  '++ Do a join by location that joins the attributes of the
  '++ first point cantained within each polygon.
  Dim pSpJoin As ISpatialJoin
  Dim pFCNew As IFeatureClass
  Set pSpJoin = New SpatialJoin

  With pSpJoin
    Set .JoinTable = pLineLayer.FeatureClass
    Set .SourceTable = pPntLayer.FeatureClass
    .LeftOuterJoin = True
  End With

  '++ Set maxMapDist to 20 allows points within 20
  '++ meters to join to linework
  Set pFCNew = pSpJoin.JoinNearest(pName, 20)

  '++ Create a new layer and add it to the Map
  If Not pFCNew Is Nothing Then
    Dim pNewFLayer As IFeatureLayer
    Set pNewFLayer = New FeatureLayer
    Set pNewFLayer.FeatureClass = pFCNew
    pNewFLayer.name = "Join by Location3"
    pMap.AddLayer pNewFLayer
    pMxDoc.UpdateContents
  End If
  Set pPntLayer = Nothing

  '++ Run next routine
  ExportTable

End Sub

'++ 11) Export output dbf

Public Sub ExportTable()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Exporting table..."

  Dim pMxDoc As IMxDocument
  Dim pMap As IMap
  Set pMxDoc = ThisDocument
  Set pMap = pMxDoc.FocusMap

  '++ Set output location
  Dim pWkSpName As IWorkspaceName
  Set pWkSpName = New WorkspaceName
  pWkSpName.WorkspaceFactoryProgID = "esricore.ShapeFileWorkspaceFactory.1"
  pWkSpName.PathName = strInputFolder

  '++ Set featureclass name
  Dim pFLayer As IFeatureLayer
  Dim pFc As IFeatureClass
  Dim i As Integer
  Dim sName As String
  sName = "Join by Location2"
  For i = 0 To pMap.LayerCount - 1
    If pMap.Layer(i).name = sName Then
      Set pFLayer = pMap.Layer(i)
      Exit For
```

```
    End If
  Next i
  Set pFLayer = pMap.Layer(0)
  Set pFc = pFLayer.FeatureClass

  'Get the FcName from the featureclass
  Dim pINFeatureClassName As IFeatureClassName
  Dim pDataset As IDataset
  Dim pInDsName As IDatasetName
  Set pDataset = pFc
  Set pINFeatureClassName = pDataset.FullName
  Set pInDsName = pINFeatureClassName

  'Define the output feature class name
  Dim pFeatureClassName As IFeatureClassName
  Dim pOutDatasetName As IDatasetName
  Dim pName As IName
  Set pFeatureClassName = New FeatureClassName
  Set pOutDatasetName = pFeatureClassName
  With pOutDatasetName
    .name = "Export_Table"
    Set .WorkspaceName = pWkSpName
  End With
  Set pName = pOutDatasetName

  '++ Select fields to include in output
  Dim pQFilter As IQueryFilter
  Set pQFilter = New QueryFilter
  pQFilter.SubFields = "FACTOR1, FACTOR2, FACTOR3, FACTOR4, FACTOR6, " & _
                       "UNIQ_ID, SIPS, CO_NUM, CYCLE_YR, " & _
                       "YEAR, MONTH, DAY, RTE_TYPE, LANES, " & _
                       "SPEED, MUNI_FLG, URB_FLG, " & _
                       "LUCODE, RD_SURFACE, A_CONTROL"

  'Export
  Dim pExportOp As IExportOperation
  Dim pFCNew As IFeatureClass
  Set pExportOp = New ExportOperation
  pExportOp.ExportTable pInDsName, pQFilter, Nothing, pName, 0

  '++ Run next procedure
  DeleteDataFrame

End Sub

'++ 12) Clean up by first deleting data frame...

Public Sub DeleteDataFrame()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Deleting Build Table data frame from TOC..."

  Dim pMxDoc As IMxDocument
  Dim pActiveView As IActiveView
  Set pMxDoc = ThisDocument
  Set pActiveView = pMxDoc.ActiveView

  '++ Create array of maps in TOC
  Dim pMaps() As IMap
  ReDim pMaps(pMxDoc.Maps.Count - 1)

  Dim l As Long
  For l = 0 To pMxDoc.Maps.Count - 1
    Set pMaps(l) = pMxDoc.Maps.Item(l)
  Next l

  '++ Remove "Build Table" frame(s)
  Dim pGC As IGraphicsContainer
  Set pGC = pMxDoc.PageLayout
```

```
  For l = 0 To UBound(pMaps)
    If pMaps(l).name <> "Build Table" Then
      Debug.Print "skip " & l
    Else
      pGC.DeleteElement pGC.FindFrame(pMaps(l))
    End If
  Next l

  '++ Activate first data frame
  For l = 0 To 0
    If TypeOf pActiveView Is IPageLayout Then
      Set pMxDoc.ActiveView.FocusMap = pMaps(l)
    Else
      Set pMxDoc.ActiveView = pMaps(l)
    End If
  Next l

  '++ Refresh ActiveView and TOC
  pActiveView.Refresh
  pMxDoc.CurrentContentsView.Refresh Nothing

  '++ Run next procedure
  DeleteTempFolder

End Sub

'++ 13) Continue cleanup by deleting temp folder

Public Sub DeleteTempFolder()

  Dim psbar As IStatusBar
  Set psbar = StatusBar
  psbar.ProgressBar.Message = "Deleting temp folder..."

  Dim fs
  Set fs = CreateObject("Scripting.FileSystemObject")
  If fs.folderexists(strTempFolder) Then
    fs.deletefolder strTempFolder
  End If

End Sub

Public Sub ExportSelectedPTC()

End Sub
```

Application Code Phase II (beta)

```
Option Explicit
Public Const strDataFolder As String = "C:\NCSU\GIS\ApplicationII\Data"
Public Const strInputFolder As String = "C:\NCSU\GIS\ApplicationII\Data\Input"

Public Sub CheckData()

  '++ check to see if XTable file exists
  Dim pTargetFileName As String
  pTargetFileName = strDataFolder & "\XO.txt"
  Dim fs
  Set fs = CreateObject("Scripting.FileSystemObject")
  If fs.fileexists(pTargetFileName) Then
    If (MsgBox("The file " & pTargetFileName & " exists." & vbNewLine & "Do you want" & _
        " to overwrite the file?", vbYesNo, "File Not Found") = vbNo) Then
      If MsgBox("Exiting procedure", vbExclamation, "Exit Procedure") = vbOK Then
        Exit Sub
      End If
    End If
  End If

  '++ check to see if export file exists
  pTargetFileName = strInputFolder & "\Export_Table.dbf"
  Dim fs2
  Set fs2 = CreateObject("Scripting.FileSystemObject")
  If Not fs2.fileexists(pTargetFileName) Then
    If (MsgBox("The file " & pTargetFileName & " does not exist." & vbNewLine & " & _
        " Exiting Procedure.", vbExclamation, "File Not Found") = vbOK) Then
      Exit Sub
    End If
  End If

  CursorRoutine

End Sub

Public Sub CursorRoutine()

  '++ cursor through export table, write lines to X delimited text
  '++ file.

  Dim pFact As IWorkspaceFactory
  Dim pWorkspace As IWorkspace
  Dim pFeatws As IFeatureWorkspace
  Dim pTable As ITable
  Set pFact = New ShapefileWorkspaceFactory
  Set pWorkspace = pFact.OpenFromFile(strInputFolder, 0)
  Set pFeatws = pWorkspace
  Set pTable = pFeatws.OpenTable("/Export_Table.dbf")

  '++ create X delimited text file
  Dim fs
  Dim a
  Set fs = CreateObject("Scripting.FileSystemObject")
  Set a = fs.CreateTextFile(strDataFolder & "\XO.txt", True)

  '++ Populate field values in X table
  a.writeline & _
  ("Intercept,Factor1,Factor2,Factor3,Factor4,Factor6,rte_class1,rte_class2," + _
  "rte_class4,co_urb2,co_urb5,co_urb6,co_urb7,co_urb8,co_urb9,co_urb10,co_urb11," + _
  "co_urb12,co_urb13," + _
  "co_urb14,co_urb16,co_urb17,co_urb18,co_urb19,co_urb20,co_urb21,co_urb22,co_urb23," + _
  "co_urb24,co_urb25,co_urb26,co_urb27,co_urb29,co_urb30,co_urb31,co_urb34,co_urb35," + _
  "co_urb36,co_urb37,co_urb38,co_urb39,co_urb43,co_urb44,co_urb45,co_urb46,co_urb48," + _
  "co_urb50,co_urb52,co_urb53,co_urb55,co_urb57,co_urb58,co_urb60,co_urb61,co_urb62," + _
  "co_urb63,co_urb64,co_urb65,co_urb66,co_urb67,co_urb69,co_urb70,co_urb71,co_urb72," + _
  "co_urb73,co_urb74,co_urb75,co_urb76,co_urb77,co_urb80,co_urb81,co_urb82,co_urb84," + _
  "co_urb86,co_urb87,co_urb88,co_urb89,co_urb90,co_urb92,co_urb93,co_urb97,co_urb98," + _
```

```
  "co_urb99,co_urb100,co_urb101,co_urb102,co_urb103,co_urb104,co_urb106,co_urb107," + _
  "co_urb109,co_urb111,co_urb113,co_urb114,co_urb115,co_urb116,co_urb117,co_urb118," + _
  "co_urb119,co_urb120,co_urb121,co_urb123,co_urb125,co_urb127," + _
  "lanes_rte2,lanes_rte4," + _
  "lanes_rte5,lanes_rte6,lanes_rte7,lanes_rte8,lanes_rte13,lanes_rte14,lanes_rte16, " + _
  "lanes_rte19,lanes_rte21," + _
  "speed_rte1,speed_rte3,speed_rte5,speed_rte7,speed_rte12,speed_rte13,speed_rte14," + _
  "speed_rte20,speed_rte21,speed_rte27,speed_rte30,lanes_speed_rte5," + _
  "lanes_speed_rte8,lanes_speed_rte11,lanes_speed_rte26,lanes_speed_rte31," + _
  "lanes_speed_rte40," + _
  "lanes_speed_rte45," + _
  "lanes_speed_rte49,lanes_speed_rte65,lanes_speed_rte73,lanes_speed_rte74," + _
  "lanes_speed_rte76," + _
  "lanes_speed_rte83,lanes_speed_rte93,lanes_speed_rte95,lanes_speed_rte99," + _
  "lanes_speed_rte102," + _
  "lanes_speed_rte104,cycle_surface1,cycle_surface2,cycle_surface4,cycle_surface5," + _
  "cycle_surface6," + _
  "cycle_surface7,cycle_surface8,cycle_surface9,a_control2,year4," + _
  "month1,month5,month6," + _
  "month7,month8,month9,month10,month11,day1,day4,lucode2,lucode3,lucode4,lucode5," + _
  "lucode7,lucode8,lucode10,lucode11,lucode13,lucode14,lucode15,lucode16," + _
  "lucode21,lucode24,city_urb1")

  Dim pRow As IRow
  Dim pCursor As ICursor
  Set pCursor = pTable.Update(Nothing, True)

'++ Factor Variables
  Dim dblF1 As Double
  Dim dblF2 As Double
  Dim dblF3 As Double
  Dim dblF4 As Double
  Dim dblF6 As Double
  Dim intRTE As Integer
  Dim strRTE As String
  Dim intCo As Integer
  Dim intSIPS As Integer
  Dim strCoUrb As String
  Dim intLanes As Integer
  Dim strLRTE As String
  Dim strLanesRTE As String
  Dim strSRTE As String
  Dim strSpeedRTE As String
  Dim strLSRTE As String
  Dim strLanesSpeedRTE As String
  Dim intSpeed As Integer
  Dim strCycle As String
  Dim intSurf As Integer
  Dim strSC As String
  Dim strSurfCycle As String
  Dim intAControl As Integer
  Dim strAccessControl As String
  Dim intYear As Integer
  Dim strYear As String
  Dim intMonth As Integer
  Dim strMonth As String
  Dim intDay As Integer
  Dim strDay As String
  Dim intLU As Integer
  Dim strLU As String
  Dim intUrbFlg As Integer
  Dim intMuniFlg As Integer
  Dim strMuniFlg As String

  Set pRow = pCursor.NextRow

  Do Until pRow Is Nothing
    '++ First capture Factor variables
    dblF1 = pRow.Value(pTable.FindField("FACTOR1"))
    dblF2 = pRow.Value(pTable.FindField("FACTOR2"))
```

```
dblF3 = pRow.Value(pTable.FindField("FACTOR3"))
dblF4 = pRow.Value(pTable.FindField("FACTOR4"))
dblF6 = pRow.Value(pTable.FindField("FACTOR6"))

'++ Capture rte_classX variables
intRTE = pRow.Value(pTable.FindField("RTE_TYPE"))
Select Case intRTE
  Case 1
    strRTE = "1,0,0"
  Case 2
    strRTE = "0,1,0"
  Case 4
    strRTE = "0,0,1"
  Case Else
    strRTE = "0,0,0"
End Select

'++ capture co_urbX variables
intCo = pRow.Value(pTable.FindField("CO_NUM"))
intSIPS = pRow.Value(pTable.FindField("SIPS"))
Select Case intCo
  Case 0
    If intSIPS > 99 Then
      strCoUrb = & _
           "1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
    End If
  Case 2
    strCoUrb = & _
           "0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 3
    strCoUrb = & _
           "0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 4
    strCoUrb = & _
           "0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 5
    strCoUrb = & _
           "0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 6
    strCoUrb = & _
           "0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 7
    strCoUrb = & _
           "0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 8
    strCoUrb = & _
           "0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 9
    If intSIPS > 99 Then
      strCoUrb = & _
           "0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
           "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
    Else
```

```
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 10
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 11
  If intSIPS > 99 Then
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Else
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 12
  If intSIPS > 99 Then
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Else
    strCoUrb =
         "0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 13
  If intSIPS > 99 Then
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Else
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 14
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 15
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 16
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 17
  If intSIPS > 99 Then
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Else
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0," + _
```

```
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 18
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 19
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 20
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 21
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 24
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 25
  If intSIPS > 99 Then
    strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 26
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 27
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 28
  If intSIPS > 99 Then
    strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Else
    strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 31
  If intSIPS > 99 Then
    strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 32
  If intSIPS > 99 Then
    strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
```

```
      Else
        strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
      End If
    Case 33
      strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
    Case 34
      strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
    Case 35
      If intSIPS > 99 Then
        strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
      End If
    Case 37
      strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
    Case 38
      strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
    Case 40
      If intSIPS > 99 Then
        If intSIPS = 106 Then
          strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
        ElseIf intSIPS = 107 Then
          strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
        End If
      Else
        strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
      End If
    Case 42
      If intSIPS > 99 Then
        strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
      Else
        strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
      End If
    Case 43
      strCoUrb = & _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
             "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
    Case 44
```

```
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 45
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 46
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 47
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 48
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 50
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 51
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 52
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 53
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 54
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 55
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 56
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 57
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 58
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 60
```

```
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 61
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 62
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 63
  If intSIPS > 99 Then
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 65
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 66
  If intSIPS > 99 Then
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Else
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 67
  If intSIPS > 99 Then
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Else
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 68
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 69
  strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 73
  If intSIPS > 99 Then
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Else
    strCoUrb = & _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
         "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
```

```
        "0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 74
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 75
  If intSIPS > 99 Then
    strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Else
    strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 76
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 77
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 78
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 79
  If intSIPS > 99 Then
    strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End If
Case 80
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 82
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 84
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0"
Case 86
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0"
Case 87
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0"
Case 88
  strCoUrb = & _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
        "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
```

```
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0"
    Case 89
      strCoUrb = & _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0"
    Case 90
      strCoUrb = & _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0"
    Case 91
      If intSIPS > 99 Then
        strCoUrb = & _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0"
      Else
        strCoUrb = & _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0"
      End If
    Case 92
      strCoUrb = & _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0"
    Case 93
      strCoUrb = & _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0"
    Case 95
      strCoUrb = & _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0"
    Case 96
      strCoUrb = & _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0"
    Case 98
      strCoUrb = & _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1"
    Case Else
      strCoUrb = & _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0," + _
            "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End Select

  '++ capture lane_rteX variables
  intLanes = pRow.Value(pTable.FindField("LANES"))
  strLRTE = intLanes & intRTE
  Select Case strLRTE
    Case 14
      strLanesRTE = "1,0,0,0,0,0,0,0,0,0,0"
    Case 22
      strLanesRTE = "0,1,0,0,0,0,0,0,0,0,0"
    Case 23
      strLanesRTE = "0,0,1,0,0,0,0,0,0,0,0"
    Case 24
      strLanesRTE = "0,0,0,1,0,0,0,0,0,0,0"
    Case 25
      strLanesRTE = "0,0,0,0,1,0,0,0,0,0,0"
    Case 32
```

```
    strLanesRTE = "0,0,0,0,0,1,0,0,0,0,0"
  Case 43
    strLanesRTE = "0,0,0,0,0,0,1,0,0,0,0"
  Case 44
    strLanesRTE = "0,0,0,0,0,0,0,1,0,0,0"
  Case 54
    strLanesRTE = "0,0,0,0,0,0,0,0,1,0,0"
  Case 63
    strLanesRTE = "0,0,0,0,0,0,0,0,0,1,0"
  Case 81
    strLanesRTE = "0,0,0,0,0,0,0,0,0,0,1"
  Case Else
    strLanesRTE = "0,0,0,0,0,0,0,0,0,0,0"
End Select

'++ capture speed_rteX variables
intSpeed = pRow.Value(pTable.FindField("SPEED"))
strSRTE = intSpeed & intRTE
Select Case strSRTE
  Case 202
    strSpeedRTE = "1,0,0,0,0,0,0,0,0,0,0"
  Case 204
    strSpeedRTE = "0,1,0,0,0,0,0,0,0,0,0"
  Case 253
    strSpeedRTE = "0,0,1,0,0,0,0,0,0,0,0"
  Case 255
    strSpeedRTE = "0,0,0,1,0,0,0,0,0,0,0"
  Case 353
    strSpeedRTE = "0,0,0,0,1,0,0,0,0,0,0"
  Case 354
    strSpeedRTE = "0,0,0,0,0,1,0,0,0,0,0"
  Case 402
    strSpeedRTE = "0,0,0,0,0,0,1,0,0,0,0"
  Case 454
    strSpeedRTE = "0,0,0,0,0,0,0,1,0,0,0"
  Case 501
    strSpeedRTE = "0,0,0,0,0,0,0,0,1,0,0"
  Case 553
    strSpeedRTE = "0,0,0,0,0,0,0,0,0,1,0"
  Case 602
    strSpeedRTE = "0,0,0,0,0,0,0,0,0,0,1"
  Case Else
    strSpeedRTE = "0,0,0,0,0,0,0,0,0,0,0"
End Select

'++ capture lanes_speed_rteX variables
strLSRTE = intLanes & intSpeed & intRTE
Select Case strLSRTE
  Case 2202
    strLanesSpeedRTE = "1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 2252
    strLanesSpeedRTE = "0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 2255
    strLanesSpeedRTE = "0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 2504
    strLanesSpeedRTE = "0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 2602
    strLanesSpeedRTE = "0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 3352
    strLanesSpeedRTE = "0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 3452
    strLanesSpeedRTE = "0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0"
  Case 3553
    strLanesSpeedRTE = "0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0"
  Case 4453
    strLanesSpeedRTE = "0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0"
  Case 4553
    strLanesSpeedRTE = "0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0"
  Case 4554
    strLanesSpeedRTE = "0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0"
```

```
    Case 4603
      strLanesSpeedRTE = "0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0"
    Case 5454
      strLanesSpeedRTE = "0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0"
    Case 6502
      strLanesSpeedRTE = "0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0"
    Case 6552
      strLanesSpeedRTE = "0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0"
    Case 6652
      strLanesSpeedRTE = "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0"
    Case 8452
      strLanesSpeedRTE = "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0"
    Case 8601
      strLanesSpeedRTE = "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1"
    Case Else
      strLanesSpeedRTE = "0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0"
  End Select


  '++ capture cycle_surfaceX variables
  strCycle = pRow.Value(pTable.FindField("CYCLE_YR"))
  intSurf = pRow.Value(pTable.FindField("RD_SURFACE"))
  strSC = strCycle & intSurf
  If strCycle = "A" Then
    Select Case strSC
      Case 1
        strSurfCycle = "1,0,0,0,0,0,0,0"
      Case 2
        strSurfCycle = "0,1,0,0,0,0,0,0"
      Case Else
        strSurfCycle = "0,0,0,0,0,0,0,0"
    End Select
  End If
  If strCycle = "O" Then
    Select Case strSC
      Case 2
        strSurfCycle = "0,0,1,0,0,0,0,0"
      Case 3
        strSurfCycle = "0,0,0,1,0,0,0,0"
      Case Else
        strSurfCycle = "0,0,0,0,0,0,0,0"
    End Select
  End If
  If strCycle = "E" Then
    Select Case strSC
      Case 2
        strSurfCycle = "0,0,0,0,1,0,0,0"
      Case 3
        strSurfCycle = "0,0,0,0,0,1,0,0"
      Case Else
        strSurfCycle = "0,0,0,0,0,0,0,0"
    End Select
  End If
  If strCycle = "V" Then
    Select Case strSC
      Case 1
        strSurfCycle = "0,0,0,0,0,0,1,0"
      Case 2
        strSurfCycle = "0,0,0,0,0,0,0,1"
      Case Else
        strSurfCycle = "0,0,0,0,0,0,0,0"
    End Select
  End If

  '++ capture a_controlX variables
  Select Case intAControl
    Case 2
      strAccessControl = "1"
    Case Else
      strAccessControl = "0"
```

```
End Select

'++ capture yearX variables
intYear = pRow.Value(pTable.FindField("YEAR"))
If intYear = 2000 Then
  strYear = "1"
Else
  strYear = "0"
End If

'++ capture monthX variables
intMonth = pRow.Value(pTable.FindField("MONTH"))
Select Case intMonth
  Case 1
    strMonth = "1,0,0,0,0,0,0,0"
  Case 5
    strMonth = "0,1,0,0,0,0,0,0"
  Case 6
    strMonth = "0,0,1,0,0,0,0,0"
  Case 7
    strMonth = "0,0,0,1,0,0,0,0"
  Case 8
    strMonth = "0,0,0,0,1,0,0,0"
  Case 9
    strMonth = "0,0,0,0,0,1,0,0"
  Case 10
    strMonth = "0,0,0,0,0,0,1,0"
  Case 11
    strMonth = "0,0,0,0,0,0,0,1"
  Case Else
    strMonth = "0,0,0,0,0,0,0,0"
End Select

'++ capture dayX variables
intDay = pRow.Value(pTable.FindField("DAY"))
Select Case intDay
  Case 1
    strDay = "1,0"
  Case 4
    strDay = "0,1"
  Case Else
    strDay = "0,0"
End Select

'++ capture landuse variable
intLU = pRow.Value(pTable.FindField("LUCODE"))
Select Case intLU
  Case 12
    strLU = "1,0,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 13
    strLU = "0,1,0,0,0,0,0,0,0,0,0,0,0,0"
  Case 14
    strLU = "0,0,1,0,0,0,0,0,0,0,0,0,0,0"
  Case 15
    strLU = "0,0,0,1,0,0,0,0,0,0,0,0,0,0"
  Case 17
    strLU = "0,0,0,0,1,0,0,0,0,0,0,0,0,0"
  Case 21
    strLU = "0,0,0,0,0,1,0,0,0,0,0,0,0,0"
  Case 23
    strLU = "0,0,0,0,0,0,1,0,0,0,0,0,0,0"
  Case 24
    strLU = "0,0,0,0,0,0,0,1,0,0,0,0,0,0"
  Case 32
    strLU = "0,0,0,0,0,0,0,0,1,0,0,0,0,0"
  Case 41
    strLU = "0,0,0,0,0,0,0,0,0,1,0,0,0,0"
  Case 42
    strLU = "0,0,0,0,0,0,0,0,0,0,1,0,0,0"
  Case 43
```

```
        strLU = "0,0,0,0,0,0,0,0,0,0,0,1,0,0"
      Case 61
        strLU = "0,0,0,0,0,0,0,0,0,0,0,0,1,0"
      Case 73
        strLU = "0,0,0,0,0,0,0,0,0,0,0,0,0,1"
      Case Else
        strLU = "0,0,0,0,0,0,0,0,0,0,0,0,0,0"
    End Select

    '++ capture city_urb1 code variable
    intUrbFlg = pRow.Value(pTable.FindField("URB_FLG"))
    intMuniFlg = pRow.Value(pTable.FindField("MUNI_FLG"))
    If intMuniFlg = 1 Then
      strMuniFlg = "1"
    Else
      strMuniFlg = "0"
    End If

    '++ Write X table line for record
    a.writeline ("????" & "," & dblF1 & "," & dblF2 & "," & dblF3 & "," & _
                 dblF4 & "," & dblF6 & "," & strRTE & "," & _
                 strCoUrb & "," & strLanesRTE & "," & _
                 strSpeedRTE & "," & strLanesSpeedRTE & "," & _
                 strSurfCycle & "," & strAccessControl & "," & _
                 strYear & "," & strMonth & "," & strDay & "," & _
                 strLU & "," & strMuniFlg)

    Set pRow = pCursor.NextRow
  Loop

End Sub
```

Application Code Phase III (beta)

```
Option Explicit
Public Const strDataFolder As String = "C:\NCSU\GIS\ApplicationIII\Data"
Public Const strInputFolder As String = "C:\NCSU\GIS\ApplicationIII\Data\Input"
Public Const strMatrixFolder As String = "C:\NCSU\GIS\ApplicationIII\Data\MTables"
Public Const strTempMatrixFolder As String =
"C:\NCSU\GIS\ApplicationIII\Data\MTables\xx_temp"
Public Const pLogFileName As String = strDataFolder & "\MatrixFunction.log"

Public Sub Log()

  Open pLogFileName For Output As #1
  Print #1, Now & ":  BEGIN LOG OPPERATION"
  Close #1

  MakeTempFolder

End Sub

Public Sub MakeTempFolder()

  Open pLogFileName For Append As #1
  Print #1, Now & ":  START make temp folder"
  Close #1

  Dim fs
  Set fs = CreateObject("Scripting.FileSystemObject")
  If fs.folderexists(strTempMatrixFolder) Then
    fs.deletefolder strTempMatrixFolder
  End If
  fs.createfolder strTempMatrixFolder

  Open pLogFileName For Append As #1
  Print #1, Now & ":  STOP make temp folder"
  Close #1

  CreateXoB

End Sub

Public Sub CreateXoB()

  Open pLogFileName For Append As #1
  Print #1, Now & ":  START create XoB"
  Close #1

  Dim pXMatrixFileName As String
  Dim pBetaVectorFileName As String
  Dim pTargetFileName As String
  Dim Xa As Variant
  Dim Ba As Variant
  Dim s
  Dim fs
  Dim fs2
  Dim fsOut
  Dim t
  Dim t2
  Dim a
  Dim dblProduct As Double
  Dim dblX As Double
  Dim dblB As Double
  Dim dblSum As Double
  Dim dblProductO As Double

  pXMatrixFileName = strMatrixFolder & "\XO.txt"
  Set fs = CreateObject("Scripting.FileSystemObject")
  If Not fs.fileexists(pXMatrixFileName) Then
    MsgBox "File XO.txt does not exist!"
```

```
      End
    End If

    pBetaVectorFileName = strMatrixFolder & "\Beta.txt"
    Set fs2 = CreateObject("Scripting.FileSystemObject")
    If Not fs2.fileexists(pBetaVectorFileName) Then
      MsgBox "File Beta.txt does not exist!"
      End
    End If

    pTargetFileName = strTempMatrixFolder & "\xx_XoB.txt"
    Set fsOut = CreateObject("Scripting.FileSystemObject")

    '++ create XoB delimited text file
    Set a = fsOut.CreateTextFile(strTempMatrixFolder & "\XoB.txt", True)

    '++ open the X-Matrix table for reading
    Set t = fs.opentextfile(pXMatrixFileName)

    '++ loop through X-Matrix for each line
    Do While t.atendofstream <> True
      Xa = t.readline
      '++ split the array into individual components using "," delimiter
      s = Split(Xa, ",")
      Dim i
      '++ for each component in the X-Matrix line, open Beta-hat Vector,
      '++ and pull the number associated with X-Matrix value to compute
      '++ the product
      Set t2 = fs.opentextfile(pBetaVectorFileName)
      dblProductO = 0
      For Each i In s
        dblX = CDbl(i)
        Ba = t2.readline
        dblB = CDbl(Ba)
        dblProduct = dblX * dblB
        dblSum = dblProduct + dblProductO
        '++ sum the products to solve for each element of the product
        '++ vector
        dblProductO = dblSum
      Next
      '++ write the product to the XoB.txt file
      a.writeline dblSum

    Loop

    Open pLogFileName For Append As #1
    Print #1, Now & ":  STOP create XoB"
    Close #1

    TransposeXo

  End Sub

  Public Sub TransposeXo()

    Open pLogFileName For Append As #1
    Print #1, Now & ":  START Transpose Xo"
    Close #1

    Dim pXMatrixFileName As String
    Dim pTargetFileName As String
    Dim Xa As Variant
    Dim s
    Dim fs
    Dim fsOut
    Dim t
    Dim a

    pXMatrixFileName = strMatrixFolder & "\XO.txt"
    Set fs = CreateObject("Scripting.FileSystemObject")
```

```
  If Not fs.fileexists(pXMatrixFileName) Then
    MsgBox "File XO.txt does not exist!"
    End
  End If

  pTargetFileName = strTempMatrixFolder & "\XoT.txt"
  Set fsOut = CreateObject("Scripting.FileSystemObject")

  '++ create XoT delimited text file
  Set a = fsOut.CreateTextFile(strTempMatrixFolder & "\XoT.txt", True)

  '++ open the X-Matrix copy table for reading
  Set t = fs.opentextfile(pXMatrixFileName)

  '++ get a column count for X-Matrix
  Dim XaCount As Variant
  Dim u
  XaCount = Split(t.readline, ",")
  u = UBound(XaCount)
  t.Close

  '++ loop through X-Matrix for each line
  Dim i As Integer
  i = 0
  Do While i <= u
    '++ open the X-Matrix table for reading
    Set t = fs.opentextfile(pXMatrixFileName)
    Do While t.atendofstream <> True
      Xa = Split(t.readline, ",")
      Dim XaWrite
      XaWrite = Xa(i)
      a.Write Xa(i)
      If t.atendofstream <> True Then
        a.Write ","
      End If
    Loop
    i = i + 1
    t.Close
    a.writeline
  Loop

  Open pLogFileName For Append As #1
  Print #1, Now & ":  STOP transpose Xo"
  Close #1


  TransposeXTXinv

End Sub

Public Sub TransposeXTXinv()

  Open pLogFileName For Append As #1
  Print #1, Now & ":  START transpose XTXinv"
  Close #1

  Dim pXMatrixFileName As String
  Dim pTargetFileName As String
  Dim Xa As Variant
  Dim s
  Dim fs
  Dim fsOut
  Dim t
  Dim a

  pXMatrixFileName = strMatrixFolder & "\XTXinv.txt"
  Set fs = CreateObject("Scripting.FileSystemObject")
  If Not fs.fileexists(pXMatrixFileName) Then
    MsgBox "File XO.txt does not exist!"
    End
```

```
  End If

  pTargetFileName = strTempMatrixFolder & "\XTXinvT.txt"
  Set fsOut = CreateObject("Scripting.FileSystemObject")

  '++ create XTXinvT delimited text file
  Set a = fsOut.CreateTextFile(strTempMatrixFolder & "\XTXinvT.txt", True)

  '++ open the XTXinv-Matrix table for reading
  Set t = fs.opentextfile(pXMatrixFileName)

  '++ get a column count for X-Matrix
  Dim XaCount As Variant
  Dim u
  XaCount = Split(t.readline, ",")
  u = UBound(XaCount)
  t.Close

  '++ loop through X-Matrix for each line
  Dim i As Integer
  i = 0
  Do While i <= u
    '++ open the X-Matrix table for reading
    Set t = fs.opentextfile(pXMatrixFileName)
    Do While t.atendofstream <> True
      Xa = Split(t.readline, ",")
      Dim XaWrite
      XaWrite = Xa(i)
      a.Write Xa(i)
      If t.atendofstream <> True Then
        a.Write ","
      End If
    Loop
    i = i + 1
    t.Close
    a.writeline
  Loop

  Open pLogFileName For Append As #1
  Print #1, Now & ":  STOP transpose XTXinv"
  Close #1

  MultiplyMatrix_XO_XTXinv

End Sub

Public Sub MultiplyMatrix_XO_XTXinv()

  Open pLogFileName For Append As #1
  Print #1, Now & ":  START Xo (XtX)inverse matrix multiplication"
  Close #1

  Dim pXMat1_FN As String
  Dim pXMat2_FN As String
  Dim pTargetFileName As String
  Dim fs1
  Dim fs2
  Dim fsOut
  Dim aMat1 As Variant
  Dim aMat2 As Variant
  Dim aMaster As Variant
  Dim pXMat1
  Dim pXMat2
  Dim pXMatOut
  Dim d
  Dim n
  Dim dblSum As Double
  Dim dblProduct As Double
  Dim dblSumFinal As Double
  Dim u
```

```
  Dim i

  pXMat1_FN = strMatrixFolder & "\XO.txt"
  Set fs1 = CreateObject("Scripting.FileSystemObject")
  If Not fs1.fileexists(pXMat1_FN) Then
    MsgBox "File XO.txt does not exist!"
    End
  End If

  '++ open the inverted matrix in lieu of XtXinv
  '++file for the loop routine
  pXMat2_FN = strTempMatrixFolder & "\XTXinvT.txt"
  Set fs2 = CreateObject("Scripting.FileSystemObject")
  If Not fs2.fileexists(pXMat2_FN) Then
    MsgBox "File XTXinv.txt does not exist!"
    End
  End If

  pTargetFileName = strTempMatrixFolder & "\XoXTXinv.txt"
  Set fsOut = CreateObject("Scripting.FileSystemObject")

  '++ create and open XoXTXinv.txt delimited text file
  Set pXMatOut = fsOut.CreateTextFile(strTempMatrixFolder & "\XoXTXinv.txt", True)


  '++ open the Xo-Matrix table for reading
  Set pXMat1 = fs1.opentextfile(pXMat1_FN)

  '++ multiply the matrices
  Do While pXMat1.atendofstream <> True
    '++ open the XTXinv-Matrix table for reading
    Set pXMat2 = fs2.opentextfile(pXMat2_FN)
    aMat1 = Split(pXMat1.readline, ",")
    aMat2 = Split(pXMat2.readline, ",")
    dblSum = 0
    dblProduct = 0
    u = UBound(aMat1)
    i = 0
    Do While i <= u
      n = 0
      dblProduct = 0
      For Each d In aMat1
        dblSum = aMat1(n) * aMat2(n)
        dblProduct = dblProduct + dblSum
        n = n + 1
      Next
      '++ write product to XoXTXinv
      i = i + 1
      pXMatOut.Write dblProduct
      If i <= u Then
        pXMatOut.Write ","
      End If
      If pXMat2.atendofstream <> True Then
        aMat2 = Split(pXMat2.readline, ",")
      End If
    Loop
    '++ write a new line to XoXTXinv
    pXMatOut.writeline
    Set u = Nothing
    Set i = Nothing
    pXMat2.Close
  Loop

  Open pLogFileName For Append As #1
  Print #1, Now & ":  STOP Xo (XtX)inverse matrix multiplication"
  Close #1

  MultiplyMatrix_XoXTXinv_XoT

End Sub
```

```
Public Sub MultiplyMatrix_XoXTXinv_XoT()

  Open pLogFileName For Append As #1
  Print #1, Now & ":  START Xo(XtX)inverse (XoT) matrix multiplication"
  Close #1

  Dim pXMat1_FN As String
  Dim pXMat2_FN As String
  Dim pXMatXoT_FN As String
  Dim pTargetFileName As String
  Dim fs1
  Dim fs2
  Dim fsXoT
  Dim fsOut
  Dim aMat1 As Variant
  Dim aMat2 As Variant
  Dim aMatXoT As Variant
  Dim pXMat1
  Dim pXMat2
  Dim pXMatXoT
  Dim pXMatOut
  Dim d
  Dim n
  Dim u
  Dim dblSum As Double
  Dim dblProduct As Double
  Dim dblSumFinal As Double
  Dim l
  Dim i

  pXMat1_FN = strTempMatrixFolder & "\XoXTXinv.txt"
  Set fs1 = CreateObject("Scripting.FileSystemObject")
  If Not fs1.fileexists(pXMat1_FN) Then
    MsgBox "File XoXTXinv.txt does not exist!"
    End
  End If

  '++ to use with the loop process below, I would transpose XoT
  '++ but we already have a transposed XoT => Xo
  pXMat2_FN = strMatrixFolder & "\XO.txt"
  Set fs2 = CreateObject("Scripting.FileSystemObject")
  If Not fs2.fileexists(pXMat2_FN) Then
    MsgBox "File XO.txt does not exist!"
    End
  End If

  '++ but I still have to use the XoT dimensions (UBound)
  '++ in the multiplication routine sense the
  '++ multiplier matrix is not square in this case
  pXMatXoT_FN = strTempMatrixFolder & "\XoT.txt"
  Set fsXoT = CreateObject("Scripting.FileSystemObject")
  If Not fsXoT.fileexists(pXMatXoT_FN) Then
    MsgBox "File XoT.txt does not exist!"
    End
  End If

  pTargetFileName = strTempMatrixFolder & "\XoXTXiXoT.txt"
  Set fsOut = CreateObject("Scripting.FileSystemObject")

  '++ create and open XoXTXiXoT.txt delimited text file
  Set pXMatOut = fsOut.CreateTextFile(strTempMatrixFolder & "\XoXTXiXoT.txt", True)

  '++ open the Xo Matrix table for reading
  Set pXMat1 = fs1.opentextfile(pXMat1_FN)

  '++ open the XoT table
  Set pXMatXoT = fsXoT.opentextfile(pXMatXoT_FN)
  '++ grab an array of the first line to set the u variable
  '++ for the counter in the multiplication routine
```

```
  aMatXoT = Split(pXMatXoT.readline, ",")
  u = UBound(aMatXoT)

  '++ multiply the matrices
  Do While pXMat1.atendofstream <> True
    '++ open the XoXTXinv-Matrix table for reading
    Set pXMat2 = fs2.opentextfile(pXMat2_FN)
    aMat1 = Split(pXMat1.readline, ",")
    aMat2 = Split(pXMat2.readline, ",")
    dblSum = 0
    dblProduct = 0
    i = 0
    Do While i <= u
      n = 0
      dblProduct = 0
      For Each d In aMat1
        dblSum = aMat1(n) * aMat2(n)
        dblProduct = dblProduct + dblSum
        n = n + 1
      Next
      '++ write product to XoXTXinv
      i = i + 1
      pXMatOut.Write dblProduct
      If i <= u Then
        pXMatOut.Write ","
      End If
      If pXMat2.atendofstream <> True Then
        aMat2 = Split(pXMat2.readline, ",")
      End If
    Loop
    '++ write a new line to XoXTXinv
    pXMatOut.writeline
    Set i = Nothing
    pXMat2.Close
  Loop

  Open pLogFileName For Append As #1
  Print #1, Now & ":  STOP Xo(XtX)inverse (XoT) matrix multiplication"
  Close #1

  SR_Diag_XoXTXiXoT

End Sub

Public Sub SR_Diag_XoXTXiXoT()

  Open pLogFileName For Append As #1
  Print #1, Now & ":  START sigma-hat^2 + [diagonal of Xo(XtX)i(XoT)]^1/2"
  Close #1

  Dim pXMat1_FN As String
  Dim pTargetFileName As String
  Dim fs
  Dim fsOut
  Dim pXMat1
  Dim pVOut
  Dim aMat1 As Variant
  Dim u
  Dim dblRowVal As Double
  Dim dblSigma As Double

  '++ sigma given by JHO
  dblSigma = 0.14489


  pXMat1_FN = strTempMatrixFolder & "\XoXTXiXoT.txt"
  Set fs = CreateObject("Scripting.FileSystemObject")
  If Not fs.fileexists(pXMat1_FN) Then
    MsgBox "File XoXTXiXoT.txt does not exist!"
    End
```

```
  End If

  pTargetFileName = strTempMatrixFolder & "\Diag.txt"
  Set fsOut = CreateObject("Scripting.FileSystemObject")

  '++ create and open XoXTXiXoT.txt delimited text file
  Set pVOut = fsOut.CreateTextFile(strTempMatrixFolder & "\Diag.txt", True)

  '++ open the Xo Matrix table for reading
  Set pXMat1 = fs.opentextfile(pXMat1_FN)

  '++ create an array of values from first row
  aMat1 = Split(pXMat1.readline, ",")

  '++ use the array to get a column count
  '++ this will be the counter for our loop
  u = UBound(aMat1)

  '++ loop routine uses counter and readline to
  '++ set the column/row intersection at diagonal
  '++ values
  Dim d

  For d = 0 To u
    '++ retrieve the diagonal values and add sigma-hat squared
    dblRowVal = aMat1(d) + dblSigma
    '++ while we have it, let's get its square root value
    dblRowVal = dblRowVal ^ (1 / 2)
    pVOut.writeline dblRowVal
    If pXMat1.atendofstream <> True Then
      aMat1 = Split(pXMat1.readline, ",")
    End If
  Next

  Open pLogFileName For Append As #1
  Print #1, Now & ":  STOP sigma-hat^2 + [diagonal of Xo(XtX)i(XoT)]^1/2"
  Close #1

  EndPoints

End Sub

Public Sub EndPoints()

  Open pLogFileName For Append As #1
  Print #1, Now & ":  START endpoints calculations"
  Close #1

  Dim pXVector_FN As String
  Dim pXoB_FN As String
  Dim pTargetFileName As String
  Dim fs
  Dim fs2
  Dim fsOut
  Dim pXVector
  Dim pXoB
  Dim pVOut
  Dim dblZ95 As Double
  Dim dblZ99 As Double
  Dim dblVLine As Double
  Dim dblXoBLine As Double
  Dim dbl95LE As Double
  Dim dbl95UE As Double
  Dim dbl99LE As Double
  Dim dbl99UE As Double

  '++ z sigma/2 for 95% and 99% CI's given by JHO
  dblZ95 = 1.96
  dblZ99 = 2.58
```

```
  pXVector_FN = strTempMatrixFolder & "\Diag.txt"
  Set fs = CreateObject("Scripting.FileSystemObject")
  If Not fs.fileexists(pXVector_FN) Then
    MsgBox "File Diag.txt does not exist!"
    End
  End If

  pXoB_FN = strTempMatrixFolder & "\XoB.txt"
  Set fs = CreateObject("Scripting.FileSystemObject")
  If Not fs.fileexists(pXoB_FN) Then
    MsgBox "File Diag.txt does not exist!"
    End
  End If

  pTargetFileName = strTempMatrixFolder & "\xx_EndPoints.txt"
  Set fsOut = CreateObject("Scripting.FileSystemObject")

  '++ create and open EP95 and EP99 text file
  Set pVOut = fsOut.CreateTextFile(strTempMatrixFolder & "\xx_EndPoints.txt", True)

  '++ open the vector and XoB for reading
  Set pXVector = fs.opentextfile(pXVector_FN)
  Set pXoB = fs.opentextfile(pXoB_FN)

  '++ loop through vector and multiply by z sigma/2
  '++ to find 99 and 95 CI, then add/subtract these
  '++ to find Lower and Upper Endpoints
  '++ write results to EndPoints.txt as comma-delimited
  '++ file in order LE 99, LE 95, UE 95, UE 99
  '++ Note the vectors used have the
  '++ same dimension, thus only one Do While loop
  '++ using AtEndOfStream is required
  Do While pXVector.atendofstream <> True
    dblVLine = pXVector.readline
    dblXoBLine = pXoB.readline
    dbl95LE = (dblXoBLine - (dblVLine * dblZ95)) ^ (1 / 0.15)
    dbl95UE = (dblXoBLine + (dblVLine * dblZ95)) ^ (1 / 0.15)
'    dbl99LE = (dblXoBLine - (dblVLine * dblZ99)) ^ (1 / 0.15)
    dbl99UE = (dblXoBLine + (dblVLine * dblZ99)) ^ (1 / 0.15)
    pVOut.writeline dbl99LE & "," & dbl95LE & "," & dbl95UE & "," & dbl99UE
  Loop

  Open pLogFileName For Append As #1
  Print #1, Now & ":  STOP endpoints calculations"
  Close #1

  OutputPI

End Sub

Public Sub OutputPI()

  '++ open up the end points table for reading
  Dim pFileName As String
  Dim fsEP
  Dim t
  pFileName = strTempMatrixFolder & "\xx_EndPoints.txt"
  Set fsEP = CreateObject("Scripting.FileSystemObject")
  Set t = fsEP.opentextfile(pFileName)

  '++ cursor through export table, write lines to X delimited text
  '++ file.

  Dim pFact As IWorkspaceFactory
  Dim pWorkspace As IWorkspace
  Dim pFeatws As IFeatureWorkspace
  Dim pTable As ITable
  Set pFact = New ShapefileWorkspaceFactory
  Set pWorkspace = pFact.OpenFromFile(strInputFolder, 0)
  Set pFeatws = pWorkspace
```

```
  Set pTable = pFeatws.OpenTable("/Export_Table.dbf")

  '++ create X delimited text file
  Dim fs
  Dim a
  Set fs = CreateObject("Scripting.FileSystemObject")
  Set a = fs.CreateTextFile(strMatrixFolder & "\PI_Results.txt", True)

  Dim pRow As IRow
  Dim pCursor As ICursor
  Set pCursor = pTable.Update(Nothing, True)

  '++ Factor Variables
  Dim strUID As String

  Set pRow = pCursor.NextRow

  '++ While cursoring through the export table to get unique id's,
  '++ write the PI's to the same output file

  Do Until pRow Is Nothing
    strUID = pRow.Value(pTable.FindField("UNIQ_ID"))
    Set pRow = pCursor.NextRow
    a.writeline strUID & "," & t.readline

  Loop

  CleanUpTempFiles

End Sub

Public Sub CleanUpTempFiles()

  Dim fs
  Set fs = CreateObject("Scripting.FileSystemObject")
  If fs.folderexists(strTempMatrixFolder) Then
    fs.deletefolder strTempMatrixFolder
  End If


End Sub
```