`

A Rural Transit Asset Management System

By

Michael D. Anderson and Nathan S. Davenport
Department of Civil and Environmental Engineering
The University of Alabama in Huntsville
Huntsville, Alabama

Prepared by

UTCA

University Transportation Center for Alabama
The University of Alabama, The University of Alabama in Birmingham,
And The University of Alabama in Huntsville

UTCA Report 04401
June 2005

Technical Report Documentation Page

1.Report No FHWA/CA/OR-	**2.Government Accession No.**	**3.Recipient Catalog No.**

4.Title and Subtitle A Rural Transit Asset Management System	**5.Report Date** June 2005
7. Authors Michael David Anderson and Nathan Scott Davenport	**8. Performing Organization Report No.** UTCA Report 04401
9. Performing Organization Name and Address Civil and Environmental Engineering Department The University of Alabama in Huntsville Huntsville, AL 35899	**10.Work Unit No.**
	11. Contract or Grant No. Alabama Department of Transportation Research Project
12. Sponsoring Agency Name and Address Alabama Department of Transportation 1409 Coliseum Blvd. Montgomery, AL 36130-3050	**13.Type of Report and Period Covered** Final Report: October 1, 2003 – May 15, 2005
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Advanced asset management systems have emerged as important tools in the management, maintenance and procurement of vehicles for transit fleet operators. This project presents the research undertaken to create an interactive, geographic information system based asset management system for the Alabama Department of Transportation to manage vehicles purchased and operated through Section 5310 and 5311 federal grant programs. The benefits of the system for the transportation department include the ability to estimate the overall fleet quality, to identify vehicles that need to be replaced each year, to predict future funding and budgetary needs, and to access other agency information.

17.Key Words Asset Management, Fleet Management	**18.Distribution Statement**

19.Security Class (of this report)	**20.Security Class (of this page)**	**21.No. of pages** 45	**22.Price**

Contents

Contents (continued)

List of Tables

List of Figures

Executive Summary

Advanced asset management systems have emerged as important tools in the management, maintenance and procurement of vehicles for transit fleet operators. This project presents the research undertaken to create an interactive, geographic information system (GIS) based asset management system for the Alabama Department of Transportation to manage vehicles purchased and operated through Section 5310 and 5311 federal grant programs.

Using GIS, along with a traditional database technology, enabled simplified access to the data through spatial selections and queries. A system was created to retain vehicle and agency information and predict future vehicle serviceability using a combination of factors. The benefits of the system for the transportation department include the ability to estimate the overall fleet quality, to identify vehicles that need to be replaced each year, to provide a basis for predicting future funding and budgetary needs, and to access other agency information.

Chapter 1
Introduction

Asset management strategies using existing data enable trained individuals to analyze, summarize and convey asset characteristics and aggregate information efficiently. Advances in technology, data collection and storage software have facilitated the development of advanced asset management techniques as the logical next step in the application and use of database information. Combining mathematics, engineering, and statistical analysis techniques with hands-on experience and raw data, asset management systems can be an indispensable tool for managing existing resources and allocating new resources effectively.

Asset management principles have been applied by many public and private sector agencies to improve understanding in a wide variety of applications. State transportation departments have applied asset management systems to improve decision making processes in areas such as allocating funds, bridge maintenance and pavement maintenance (FHWA, 1999). Often, these applications focus on managing the current assets, not for the prediction of future needs (Montgomery et al., 2001).

Another important area within transportation departments where the application of asset management techniques can provide decision support is public transportation. Capital equipment procurement and maintenance, and the prediction of future capital expenditures are important in today's public transportation operations. To address this need, a geographic information system (GIS) based fleet asset management system with statistically valid prediction capabilities was researched and developed to assist department of transportation (DOT) personnel in determining needs, budget requirements and equitable resource allocation.

Background information

Rural public transportation in the United States is a vital service for many citizens, providing access to employment and health care as well as social and recreational activities. The United States Department of Transportation (USDOT) through the Federal Transit Administration, created grant programs to fund agencies working with public transportation in rural areas. These grant programs, identified as Section 5310 (transportation specifically designed for elderly and disabled passengers) and Section 5311 (general public transportation to rural residences), provide funds for capital purchases. The programs meet transportation needs by providing funding for vehicle purchases through an 80:20 federal:local purchase arrangement.

Alabama currently has approximately 26 agencies receiving support under the Section 5311 program and 150 agencies receiving support under the 5310 program. The vehicles operated by these agencies comprise the "statewide fleet" consisting of 1,024 active vehicles. The majority of the vehicles in the statewide fleet are "cutaway" vans modified to seat 12 to 21 passengers, depending on interior configurations and the number of wheelchair tie-down spaces available.

The Alabama Department of Transportation (ALDOT) has oversight responsibility for the activities of the 26 agencies and for the purchase and disposal of all vehicles.

Objective of study

The intent of this project was the development and implementation of an asset management system to enable ALDOT personnel monitoring the Section 5310 and Section 5311 grants to access existing data readily and to improve vehicle procurement decisions. A GIS-based database of vehicles being operated within the state was needed to manage existing records in the ALDOT central office. The asset management database was to include information on vehicles and agencies using Section 5310 and Section 5311 federal grants.

Research tasks

To complete this research, work was divided into five tasks. The tasks include a literature review, data review, system design, statistical analysis, and procurement model.

Literature review

A review of journal publications, online material and other resources was preformed to determine the successful approaches in the design and implementation of asset management systems. Existing transportation related software was examined to determine which packages were best suited to the needs of ALDOT. Reference material for selected software packages was procured in digital or hard copy for later use.

Data review and initial analysis

Existing ALDOT vehicle records were obtained and reviewed for accuracy and possible inclusion in an asset management system. This review included interpreting the database terminology, correcting entries, conducting agency reviews and updating the database as necessary.

Asset management system design

After reviewing the records provided by ALDOT, the asset management infrastructure was created. A link was established between the GIS interface and external database using agency names and other common data values to define the relationships between the data sets.

Linear regression and discriminant analysis model design

A statistically valid model was created to predict vehicle conditions as a function of operational and socioeconomic characteristics. The model was validated using several validation tests. Using the variable analysis performed in the creation of the linear regression analysis model, discriminant analysis was conducted to produce an alternate future prediction model.

Procurement model system integration

After the prediction model was developed and the discriminant analysis was conducted, the prediction model was integrated into the asset management system. Custom macros, basic logic statements and scripts were used to automate the process of predicting future fleet quality.

Document organization

This report includes seven chapters. The first chapter provides a brief overview of asset management, the area of application this project addresses and the tasks involved. The second chapter includes an overview of previous studies, ongoing research and existing applications of asset management. The third chapter covers the underlying infrastructure for the asset management system and the data used in the creation of the system. The fourth chapter describes the creation of a prediction model to forecast the condition of vehicles in the future and the use of the model as a decision making tool. The fifth chapter describes the use of discriminant analysis to supply categorical equations to determine the accuracy of the model and an application in the procurement model. The sixth chapter concludes with an overview of the final system and its potential applications, and the application of the entire system in the decision making process. The final chapter contains a list of references.

Chapter 2
Literature Review

This chapter begins with a brief overview of asset management, including the goals, the structure and some specific applications within the transportation industry. It concludes with a summary of prediction models used to manage vehicle fleets, bridges and pavement sections as well as an analysis of approaches and drawbacks.

Asset management system overview

Advanced asset management systems have become an important tool in the management, maintenance and procurement of vehicles for operators of transportation fleets (FHWA, 1999). As defined by the Federal Highway Administration, asset management systems are "a systematic process of maintaining, upgrading and operating physical assets cost-effectively" (FHWA, 1999). Asset management systems are designed to provide part of the infrastructure for the planning and decision making process (FHWA, 1999). Asset management systems can incorporate geographical information systems (GIS), raw database information, mathematical and statistical analysis, hands-on experience, policies, goals, the Internet and other tools to provide an easily accessible system to analyze and process data/information into a form that is readily usable to individuals or businesses (FHWA, 1999) (Figure 2-1).

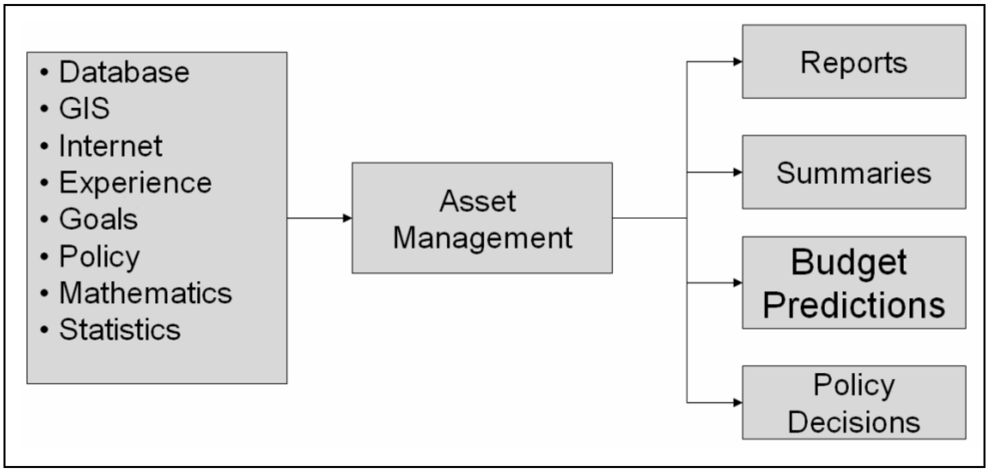


Figure 2-1. Asset management structure flowchart

The New York State Department of Transportation defines the purpose of asset management as a process to "maximize the benefits of a transportation system to its customers and users, based on well-defined goals and with available resources" (FHWA, 1999). Simply stated, an asset management system uses existing data and resources to provide an informed basis for key decisions.

The federal government supported the development of management systems through legislation in all parts of governmental operations. The introduction of the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA) and the 1997 update, the Transportation Efficiency Act of the 21st Century (TEA-21) established stringent rules on the management of assets that where allocated under the USDOT jurisdiction (ISTEA, 1991 and TEA21, 2001). These rules brought about a need to improve existing management systems and to develop new systems throughout state DOT's across the country.

Another driving force behind asset management in government agencies is the Government Accounting Standards Board (GASB) statement 34 (GASB34, 1999). This statement requires more financial accountability for state and local governments (Kurt et al., 2003). GASB34 expanded reporting requirements to all capital and long-term assets, and suggested that policies be established to require reporting of these in financial statements. Part of these needs can be addressed with a well-designed asset management system. Reporting, information summary and well-defined policy based decisions can all be a function in an asset management system.

Asset management system structure

The basic structure of any asset management system requires an underlying information database, a performance rating and a goal for the area that the system covers (FHWA, 1999) (Figure 2-2).

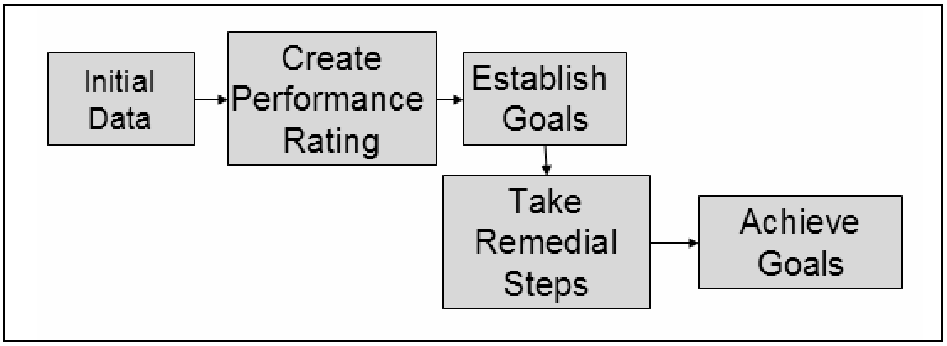


Figure 2-2. Asset management process flowchart

The information database includes individualized data identifying each asset and its characteristics. The integrity of the data is imperative in determining the quality of the system, with almost all analysis being based on the initial database. Data entry errors and incorrect asset information can result in erroneous output, skewing data summaries and analysis and possibly leading to incorrect decisions.

Analysis of the asset characteristics enables the system to create a baseline to determine if that asset is performing above or below standards. The performance rating system can be applied across other data to determine if the assets are falling short of the expected performance criteria.

Agency goals can then be established to evaluate the system's overall performance and remedial steps can be taken if necessary. The steps needed to reach these goals can often be gleaned from the asset management system using the variables that determine the performance ratings.

Applications of prediction models in asset management systems

The use of prediction models in an asset management system adds another tool for the decision making and management process. An asset management system with a well-designed prediction model can estimate system changes, not merely quantify the existing system. Prediction of asset performance or condition ratings enables the user to address needs and budgetary requirements early on, reducing the need for frequent and costly physical inspections and providing insight into factors that affect the overall goal of the asset management system (FHWA, 1999). Prediction models can also be used to test alternatives to determine their overall affect on the performance of the system. Some of the ongoing applications of future prediction models in the transportation sector include equipment and fleet maintenance and procurement, bridge maintenance and pavement maintenance (FHWA, 1999).

Bridge maintenance systems are one of the more commonly found applications of asset management in the transportation sector. Commercial asset management programs available for bridge maintenance such as PONTIS are capable of developing future maintenance cost estimates, based on existing condition data and past maintenance history (Cambridge, 2003). Pavement maintenance systems have also become popular applications of asset management and future prediction models. Applied systems include PASER and custom systems developed for some state and county operations. These systems most often employ a linear regression statistical approach to future prediction (Kurt et al., 2003). Asset management systems for fleet and equipment maintenance and procurement have been created by several states including New Jersey, Indiana and Iowa. These models perform life-cycle analysis, analyze the benefits of different maintenance practices, determine procurement needs and estimate future budgetary needs by predicting overall fleet quality and individual vehicle conditions.

The asset management system developed for the New Jersey Transit Public Transportation Facility applied a rating system to vehicles and equipment that analyzed multiple aspects of the assets, such as electrical and mechanical systems, taking into account the difference in deterioration of each system (Ludwig, 1997). The system used deterioration curves to determine the transition between rating categories for assets. Designed to apply to different forms of equipment besides traditional rolling stock, New Jersey's system took a broader approach to fleet management by creating individual prediction curves for a diversified inventory.

The Indiana Department of Transportation (INDOT) developed an asset management system that predicted vehicle conditions based on a set of variables including weather and maintenance data (Karlaftis and Sinha, 1997). This system applied an Ordered Probit model to predict deterioration in the overall fleet and discriminant analysis to predict individual vehicle conditions (Kurt et al., 2003). Distinctions in the model were made between vehicle size because of the relative maintenance and procurement expense differences. Some the variables involved in this system were relatively hard to quantify, such as weather. Due to the complexity of the predictors, the system included input of a large amount of data from outside the existing department database.

In Iowa, Kurt, Weaver and Kroeger took the INDOT model a step further and created a system that predicted the condition rating based on average maintenance cost, mileage and age (Kurt et al., 2003). Surveys were sent out to participating agencies to obtain the data necessary to create a prediction model, which covered different types of vehicles. After data corrections, 63 observations were used to generate the prediction curves that determined the future condition rating of the fleet and each vehicle. This system proved a statistically valid model to predict average fleet condition ratings using the surveyed database. The main disadvantage of the model was that surveys were needed to obtain the data, requiring extra cost and time for data on a small portion of the fleet.

Another approach was a simplified model based on linear regression analysis with variables including age, mileage and socio-economic data (Anderson and Sandlin, 2001). This approach, though not thoroughly validated, suggested a simplified model could be created from on-hand information to predict future values with accuracy comparable to other models. The major advantage of this prediction model was that the data was readily available from ALDOT and the US census. This removed the need for additional surveys and data collection, reducing the error, expense and limitations incurred during these processes. A second advantage to this model was that it used a large number of vehicles (over 400) from the existing ALDOT database.

Another approach proposed by Khasnabis suggested optimization equations along with a prediction model (Montgomery et al., 2001). This system was applied to large transit buses, such as school buses, and allowed three maintenance types: replacement, rebuild and remanufacture. The maintenance types and their respective affects on vehicle quality were used in the optimization model to determine the best application of the various maintenance types. This UTCA project examined the feasibility of a remanufacture process for ALDOT's 5311 and 5310 fleet. It was found that rebuilding applications for cutaways were not cost effective due to the large amount of fiberglass/plastic body work used in the construction of the vehicle.

Analysis of data for procurement model

Two forms of statistical analysis were incorporated in the asset management system, the first was regression analysis and the second was discriminant analysis. Regression analysis has become a staple in many fields of research as a method of predicting and determining population characteristics based on the relationships of variables. To create a statically sound model, many aspects of the model must be investigated including the adequacy of the model and correct use of regressors. The following sections outline the basics of multiple linear regression analysis and the process of discriminant analysis.

Fundamentals of linear regression

Regression analysis relates one population, designated *Y*, to another population or populations, designated *x*, based on observations in one of the populations (equation 2-1) (Montgomery et al., 2001).

$$Y = (\mathrm{f}(x)) \qquad (2\text{-}1)$$

In simple linear regression, *Y* is expressed as a function of two-regression coefficients $ß_0$ and $ß_{1,}$ and the independent variable, *X* (equation 2-2).

$$Y = ß_0 + ß_1 X + \mathrm{e} \qquad (2\text{-}2)$$

The coefficient $ß_0$ is known as the intercept and indicates the point at which the model intersects the y-axis. The $ß_1$ coefficient, known as the slope, defines the slope of the prediction line. The magnitude of the predictor determines the placement of the point along the slope and thus the value of the response, *Y.* The statistical error e, inherent in all models, is usually shown in the general model. Montgomery, Peck and Vining define the e as “a random (undefined) variable that accounts for the failure of a model to fit the data exactly” (Montgomery et al., 2001). The relative size of the e plays an important role in determining the quality and applicability of the model.

Three assumptions are made when conducting linear regression analysis on a dataset. These assumptions must hold true for the regression to be considered acceptable. The first is that model errors, e, are assumed to be normally distributed. The second is the assumption that the sum of the error, e, is zero and the variance of the error is constant. The third is that the errors are independent (Montgomery et al., 2001). These assumptions are checked for a given model by evaluating a series of data plots of the model errors (i.e., residual plots).

Multiple linear regression analysis enables the response to be a function of multiple predictors, allowing the model to take into account multiple factors that could be left out of a simpler model. As with simple linear regression, the three assumptions must be checked. The form of the equation follows the basic format of the simple linear regression adding additional slopes, $ß_1$ through $ß_{n,}$ where *n* is the total number of regressors, and the regressor variables, X_1 through X_n. As with the previous model, the statistical error is represented by e and retains the same definition (equation 2-3).

$$Y = ß_0 + ß_1 X_1 + ß_2 X_2 + ß_3 X_3 \ldots + ß_n X_n + \mathrm{e} \qquad (2\text{-}3)$$

The use of multiple regressors introduces a new concern, known as multicollinearity, during the model creation process. Multicollinearity occurs when two regressors are linearly related, making it hard to distinguish the effects of each variable on the model (Montgomery et al., 2001). The simplest way to check for multicollinearity is to plot regressors against each other and look for linear trends in the data.

In the first stage of model creation in this project, multiple linear regression analysis was applied to determine the potential variables for use in discriminant analysis. Then discriminant analysis was used to create an alternative model for predicting condition ratings.

Discriminant Analysis

Discriminant analysis takes a different approach to data analysis; instead of predicting values based on existing data, it categorizes the data into 2 or more groups (Johnson and Wichern, 2002). The process of data categorization used in discriminant analysis is referred to as classification (Rice, 1995). In discriminant analysis, a set of equations, called the linear discriminant function, calculates the probabilities that each data point is within a group. Discriminant analysis begins by locating the centroid of each category. For each individual data point, the respective distances are calculated from the centroid of each category. These distances are used to determine the probability that a point is from a specific category. The distance, Z_i, is determined based on the population characteristics X_n, where n is the number of characteristics; C_i is a constant and a series of weights (W_i) (equation 2-4) (Statsoft, 2003).

$$Z_i = C_i + W_{i1}X_1 + W_{i2}X_2 + ... + W_{in}X_n \qquad (2\text{-}4)$$

After calculating the probability that an individual point belongs to each group, the probabilities are then compared and the point is placed in the group with the highest probability (Table 2-1).

Table 2-1. Example discriminant analysis single point probabilities

Group	Probability	
	Pred	X-val
1	0.6	0.62
2	0.3	0.28
3	0.1	0.1
4	0	0
5	0	0

The example point (in Table 2-1) will be placed in group 1 according to the probabilities. This process is repeated for every point in the data set until all points are categorized. To evaluate the statistical quality of the model, the factor called the apparent error rate (APER) is calculated (Johnson and Wichern, 2002). The APER is an estimation of the error rate of a discriminant analysis classification system and is calculated by dividing the sum of miscalculated points by the sum of the correctly classified points. As the APER becomes greater, the unexplained variability in the model also increases.

Summary

The models reviewed in this chapter provide statistically based predictions of future maintenance, procurement and budgetary needs. The system created in this research provided a statistically sound model that can predict the replacement needs of vehicles over a five-year cycle, based on existing ALDOT data. Data surveys and advanced methodologies were considered but found to provide little if any advantage over traditional linear regression methods.

Chapter 3
Database Creation

All asset management systems rely on an underlying database structure to analyze, maintain and update information. The database determines the amount of time needed to update the system, the interface and the abilities of the overall system. The quality of the original data, the way the data is handled, and the maintenance of the database determine the quality and applicability of the system in the decision making process. The following chapter explains the steps taken to ensure the creation of a quality database, input/output processes and query methods.

Original data

The foundation of an asset management system is data. Part of the process of creating any asset management system is the determination of data quality. Quality data is imperative to the construction of an asset management system because the entire system uses data to perform the analysis, establish goals and identify the solutions to obtain the goals. When applying a future prediction model, the importance of data quality becomes even more pronounced as the quality of the prediction is based on the data and analysis.

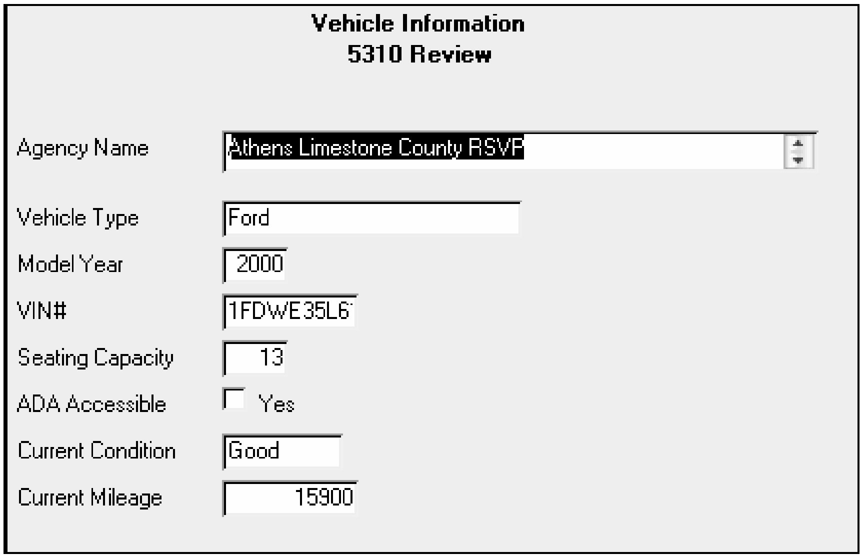
The current ALDOT database was built in Microsoft Access and contains transit vehicle information. Vehicles from Alabama's 5310 and 3211 programs are listed in the database with information dating back to 1987. A checkbox designation is used to distinguish between active and inactive vehicles. The database includes 40 attributes for each vehicle, ranging from make and model to purchase and delivery dates. Three DOT personnel manage the database, with updates being sent from computer to computer upon completion.

The DOT database required cleaning and updating before being used in the asset management system. Double entries were a problem as personnel entered vehicles multiple times, due to incorrect vehicle identification numbers (VIN), agency names, abbreviations and model years. The first step in correcting the database, was a standardization of the agency names, model type and vehicle nomenclature. Next, all VIN's were manually inspected and corrected if needed using the alphanumeric replacement and title information. The last step was the creation of a script to eliminate double entries. A temporary centralized version of the database was created to reduce any possible error introduced by having multiple personnel updating the database at one time.

Access design inputs and outputs

To manage and simplify the process, the researchers used the Microsoft Access database management software for data entry and output. This program was selected because ALDOT personnel were familiar with the database software, which decreased their learning curve for the new system. The types of data the system handles include vehicle acquisition and disposal, agency review data, agency contact data and annual and monthly agency reporting data.

The input sections of the asset management system were designed as forms in Access and are displayed through a web browser on the user's computer. Forms include drop down menus whenever feasible to reduce the problem of abbreviation and misspelling of entries such as agency names and vehicle manufacturers. Active/inactive vehicle forms edit the active/inactive checkbox in the main vehicle database to indicate whether a vehicle is retired from active duty. Forms and custom update macros were created in Access to streamline the data input process (Figure 3-1). These forms were designed for agency information modification and addition, vehicle procurement and agency review summaries.



Vehicle Information
5310 Review

Agency Name	Athens Limestone County RSVP
Vehicle Type	Ford
Model Year	2000
VIN#	1FDWE35L6
Seating Capacity	13
ADA Accessible	☐ Yes
Current Condition	Good
Current Mileage	15900

Figure 3-1. Example of vehicle input form

Using MS Access created reports and MS Word macros, agency specific review summaries, vehicle inventories and prediction scenarios were produced to aid in information retrieval and analysis. Full 5310 reviews are included in the system and are available for print or digital review. Output from the GIS interface is controlled via System Query Language (SQL) coding. The user has the option of paper copies and digital versions for each output.

GIS advantage

Integration of GIS into the asset management systems enables users to access and display information more effectively. Users can analyze the data through graphical, spatial, tabular, and query based selection methods. These methods improve data analysis and make the database more user friendly, compared to traditional text databases. Arcview software was chosen as the GIS tool for incorporation into the asset management system.

GIS data relationships

Because GIS combines many forms of data, such as maps, census data and tabular data, relationships are needed to provide the links between the data sets. These relationships are dynamic, allowing real-time selection of one entry, which in turn selects all related entries and displays. The first relationship was established between the map data and 5310 agency data using a spatially defined zip code file and a digital Alabama state map. Using GIS format allowed zip code and Alabama state maps to be related, so that both the state and zip codes could be shown in one display. Thought a *select by theme* query, which selects features based on their spatial characteristics, all zip codes lying outside of Alabama were selected and hidden from view. All of the 5310 funded agencies were linked to the map by their corresponding zip codes.

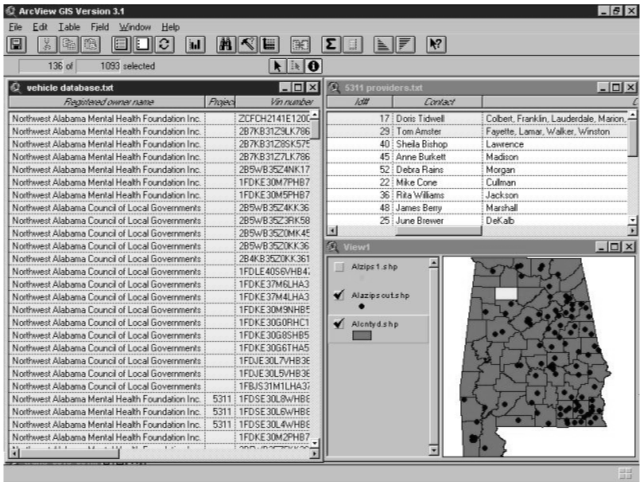
Since some of the 5311 agencies are operating in multiple counties, a new table was created to link the grant providers with the map. This table consisted of two fields: one for the ALDOT designated agency ID numbers and the other for one of the counties in which the agency operated. Agencies that covered multiple counties would have multiple lines in the new table providing an accurate visual representation of the counties and their respective agencies.

ALDOT's vehicle database combined vehicles purchased by both grant types in one large database. It included a large amount of data that was superfluous to the user, such as vehicle title numbers, which was hidden in the initial revision. The names of the agencies were chosen as the linking value between the vehicles and the maps for both the 5310 and 5311 funded agencies. After standardizing the agency names between the provider list and ALDOT's vehicle data, the vehicle table was linked to the 5310 and 5311 provider lists. The relationships established between the vehicle, agency and map data provide the basis for advanced analysis and queries as well as the SQL linking between Arcview and Access.

Analysis inside the GIS environment

Because GIS combines maps with data, the graphical selection method is one of the main features of GIS. It enables the user to select certain areas or points on a map, which in turn selects all related data that corresponds to that location. For example, selecting Winston County in the GIS interface selects all corresponding data, highlighting all specific data linked associated with the county of interest (Figure 3-2).

By opening the 5311 provider table and clicking the promote button, one can see the addresses and locations of 5311 agencies in Winston County as well as the vehicles that operate for the respective agencies in that county (Figure 3-3).

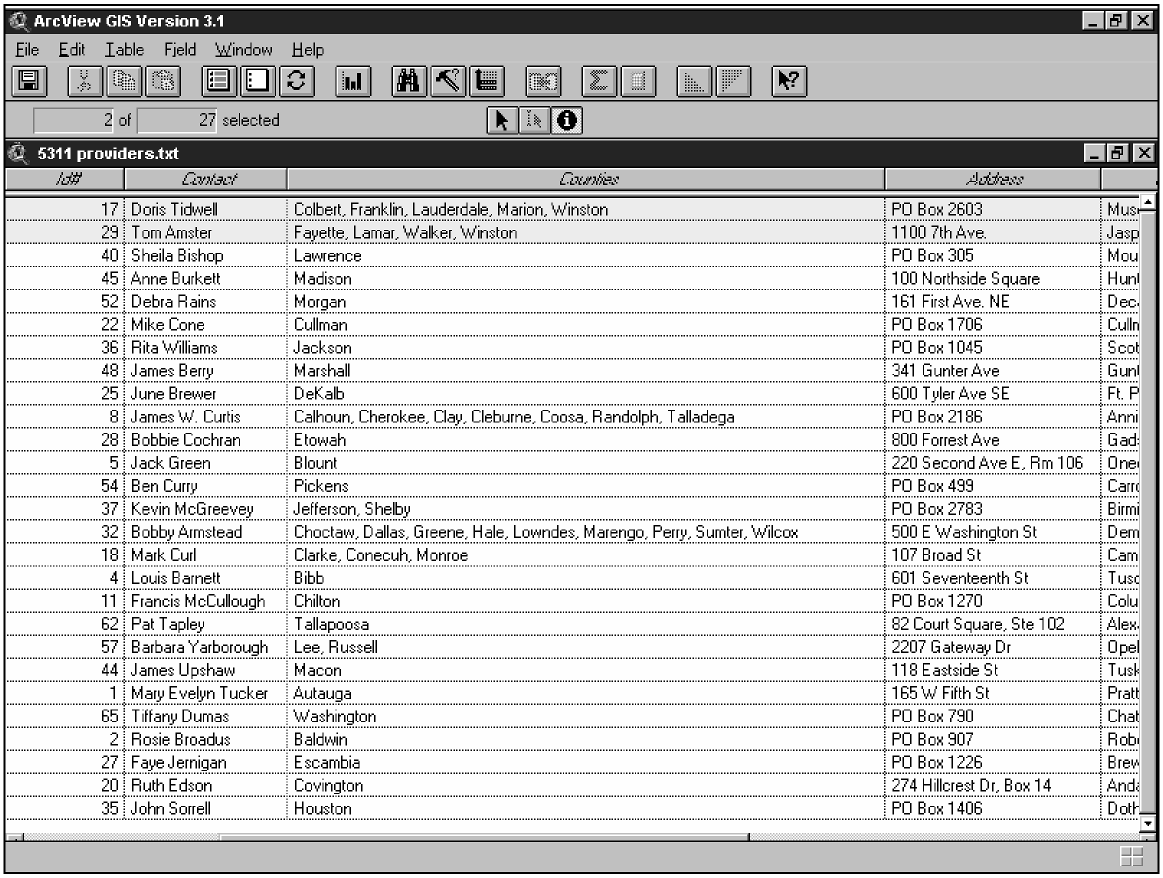


136 of 1093 selected

Registered owner name	Projec	Vin number
Northwest Alabama Mental Health Foundation Inc.		ZCFCH2141E120C
Northwest Alabama Mental Health Foundation Inc.		2B7KB31Z9LK786
Northwest Alabama Mental Health Foundation Inc.		2B7KB31Z8SK575
Northwest Alabama Mental Health Foundation Inc.		2B7KB31Z7LK786
Northwest Alabama Mental Health Foundation Inc.		2B5WB35Z4NK17
Northwest Alabama Mental Health Foundation Inc.		1FDKE30M7PHB7
Northwest Alabama Mental Health Foundation Inc.		1FDKE30M5PHB7
Northwest Alabama Council of Local Governments		2B5WB35Z4KK36
Northwest Alabama Council of Local Governments		2B5WB35Z3RK58
Northwest Alabama Council of Local Governments		2B5WB35Z0MK45
Northwest Alabama Council of Local Governments		2B5WB35Z0KK36
Northwest Alabama Council of Local Governments		2B4KB35Z0KK361
Northwest Alabama Council of Local Governments		1FDLE40S6VHB4
Northwest Alabama Council of Local Governments		1FDKE37M6LHA3
Northwest Alabama Council of Local Governments		1FDKE37M4LHA3
Northwest Alabama Council of Local Governments		1FDKE30M9NHB5
Northwest Alabama Council of Local Governments		1FDKE30G0RHC1
Northwest Alabama Council of Local Governments		1FDKE30G8SHB5
Northwest Alabama Council of Local Governments		1FDKE30G6THA5
Northwest Alabama Council of Local Governments		1FDJE30L7VHB3E
Northwest Alabama Council of Local Governments		1FDJE30L5VHB3E
Northwest Alabama Council of Local Governments		1FBJS31M1LHA3
Northwest Alabama Mental Health Foundation Inc.	5311	1FDSE30L8WHBE
Northwest Alabama Mental Health Foundation Inc.	5311	1FDSE30L6WHBE
Northwest Alabama Mental Health Foundation Inc.	5311	1FDSE30L4WHBE
Northwest Alabama Mental Health Foundation Inc.		1FDKE30M2PHB7

Id#	Contact	
17	Doris Tidwell	Colbert, Franklin, Lauderdale, Marion,
29	Tom Amster	Fayette, Lamar, Walker, Winston
40	Sheila Bishop	Lawrence
45	Anne Burkett	Madison
52	Debra Rains	Morgan
22	Mike Cone	Cullman
36	Rita Williams	Jackson
48	James Berry	Marshall
25	June Brewer	DeKalb

Figure 3-2. GIS initial spatial selection



2 of 27 selected

Id#	Contact	Counties	Address	
17	Doris Tidwell	Colbert, Franklin, Lauderdale, Marion, Winston	PO Box 2603	Mus
29	Tom Amster	Fayette, Lamar, Walker, Winston	1100 7th Ave.	Jasp
40	Sheila Bishop	Lawrence	PO Box 305	Mou
45	Anne Burkett	Madison	100 Northside Square	Hunt
52	Debra Rains	Morgan	161 First Ave. NE	Dec
22	Mike Cone	Cullman	PO Box 1706	Culln
36	Rita Williams	Jackson	PO Box 1045	Scot
48	James Berry	Marshall	341 Gunter Ave	Gunt
25	June Brewer	DeKalb	600 Tyler Ave SE	Ft. P
8	James W. Curtis	Calhoun, Cherokee, Clay, Cleburne, Coosa, Randolph, Talladega	PO Box 2186	Anni
28	Bobbie Cochran	Etowah	800 Forrest Ave	Gad
5	Jack Green	Blount	220 Second Ave E, Rm 106	One
54	Ben Curry	Pickens	PO Box 499	Carr
37	Kevin McGreevey	Jefferson, Shelby	PO Box 2783	Birmi
32	Bobby Armstead	Choctaw, Dallas, Greene, Hale, Lowndes, Marengo, Perry, Sumter, Wilcox	500 E Washington St	Dem
18	Mark Curl	Clarke, Conecuh, Monroe	107 Broad St	Cam
4	Louis Barnett	Bibb	601 Seventeenth St	Tusc
11	Francis McCullough	Chilton	PO Box 1270	Colu
62	Pat Tapley	Tallapoosa	82 Court Square, Ste 102	Alex
57	Barbara Yarborough	Lee, Russell	2207 Gateway Dr	Opel
44	James Upshaw	Macon	118 Eastside St	Tusk
1	Mary Evelyn Tucker	Autauga	165 W Fifth St	Pratt
65	Tiffany Dumas	Washington	PO Box 790	Chat
2	Rosie Broadus	Baldwin	PO Box 907	Rob
27	Faye Jernigan	Escambia	PO Box 1226	Brew
20	Ruth Edson	Covington	274 Hillcrest Dr, Box 14	Anda
35	John Sorrell	Houston	PO Box 1406	Doth

Figure 3-3. Results of spatial selection

Spatial selection includes functions oriented around the spatial relationships of data within the system. For example, special functions allow users to specify a point and distance, and ArcView will select the points within or outside that buffer distance. In Figure 3-4, a "contained within" spatial query is conducted to show all the 5310 agencies that operate within 50 miles of Birmingham.

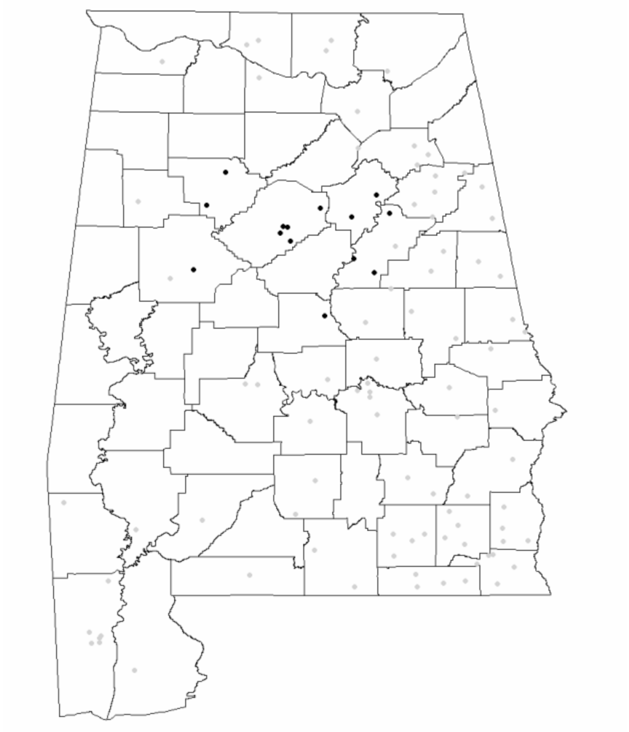
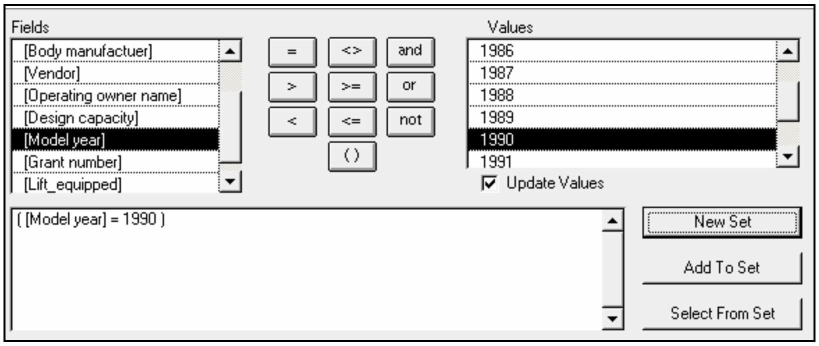


Figure 3-4. Spatial query results

Some of the other methods of spatial selection include intersection, include and exclude queries. The traditional database access method of tabular or direct data selection allows the user to select multiple or individual row entries in the asset management system's tables, again selecting all related entries.

Query based selection enables the user to select entries using arithmetic and logic based operators to include or exclude specific data. For example, setting the model year equal to 1990 selects all vehicles that are specified as 1990 models in the system (Figure 3-5). This selection process can use any alphanumeric column within the database and can be expanded to contain multiple operations.



☑ Update Values

([Model year] = 1990)

Figure 3-5. Query for only 1990 year models

To simplify the use of the system, the more common queries were converted to macros to provide direct access to data without having to rebuild queries. Macros were also developed for custom queries of information not directly stored in the GIS portion of the database, such as agency report summaries. These queries are run in the GIS portion of the database using the integrated System Query Language (SQL) dynamic linking program.

SQL interconnect between programs

The SQL language enables the user to link information between databases for querying and searching needs. Because a large amount of data is stored in the Access portion of the asset management system, the GIS portion was not able to dynamically link and query the data. The solution to this problem is custom scripts that link the two database files. Using the embedded SQL coder in Arcview, all common queries of the database were created and stored in the GIS portion.

Summary

Using Access and GIS a database, and asset management infrastructure were created to maintain, update and query the asset management system. Expanding the system to GIS enabled spatial queries as well as direct selection of points on the map. Interlinks between database sections were established using SQL code.

Chapter 4
Regression Analysis and Model Development

Regression analysis has become an integral tool for developing and modeling data. Regression analysis of characteristics of transit fleets and vehicles can predict a myriad of information including ridership, connection times, and degradation curves for individual and overall vehicle applications. The application of regression analysis in this chapter is concentrated on predicting future vehicle quality based on vehicle and socioeconomic characteristics. The developed model will be used to predict the level of vehicle procurement needed to maintain a specific fleet quality over time. From this analysis, annual budgets can be developed, compared, and tested to determine the effects on the overall fleet.

Initial database

Part of the review process for agencies operating vehicles obtained under 5311 federal grants is a tri-annual review including a full physical inspection of all vehicles. The initial review of 5311 vehicles used in this research was compiled in late 2000 from the most recent set of tri-annual reviews. The database contained all 484 grant vehicles operating within the state at the time and included information such as vehicle year, mileage, passenger capacity, make, model, assigned condition ratings, and whether the vehicle was equipped with a wheelchair lift. The assigned vehicle condition rating was a composite rating based on on-site inspections of the points shown below.

- Engine Starting Trouble
- Engine Running Condition
- Interior Condition (upholstery damage, seats missing)
- A/C Condition
- Wheelchair Lift Operation
- Exterior Condition
- Mileage

From the inspection, a vehicle was assigned a rating number using the one to five values shown in Table 4-1.

Table 4-1. Condition ratings

Bad	=	1	Vehicle needs immediate replacement
Poor	=	2	Vehicle should be replaced
Fair	=	3	Vehicle is acceptable
Good	=	4	Vehicle has no outstanding problems
Excellent	=	5	Vehicle is in new condition

The inspection provided the performance measurement needed to develop goals and a prediction model for the asset management system. The remaining data collected were used as potential variables that could possibly help predict the condition of the vehicle.

Regression variable selection

To determine future vehicle replacement and accruement needs, the asset management system required a model to predict the conditions of vehicles in the future. Condition rating was chosen as the indicator of whether the vehicle was in need of replacement. Thirty-four data points were eliminated from the vehicle inventory due to their unusually low or high ratings that appeared to be the result of extraneous factors, such as relatively new vehicles that were given poor ratings because of engine or air conditioning failures. A listing of the outliers is included in Appendix A-2. The remaining 450 transit vehicle data entries were then analyzed using the Minitab Statistical Analysis Software, Release 14. Condition rating was used as the dependent variable and the sixteen independent variables selected for investigation during the model building phase of the study (Table 4-2).

Table 4-2. Complete set of independent variables for regression analysis

Age
Total Miles Traveled
Miles per year on paved roads
Miles per year on unpaved roads
Percent minority population in county of operation
Total population in county of operation
Percent single person households in county of operation
Wheelchair Accessibility
Percent income less than $15,000 in county of operation
Percent population greater than 65 in county of operation
Percent work in county
Percent that work out of county
Percent population less than 18 in county of operation
Percent of population that work in county of operation
Percent Commuters in county of operation
Percent person in poverty in county of operation

The vehicle model was not included as an independent variable due to the fact that most vehicles operating within the 5311 program were Ford chassis cutaway vans. Using Minitab, a regression analysis was performed on the data with an ANOVA to determine the significance of the predictors in the model. Predictors were eliminated by two criteria, the first being the significance of the predictor to the regression (p-value) and the second being the variance inflation factor (VIF) (Table 4-3).

Table 4-3. Results of variable selection regression analysis

Variables	T	P	VIF
Age	-8.16	0.000	7.5
Total Mileage	-2.28	0.012	14.2
mile/yr pav	0.2	0.542	5.2
mile/yr unp	-2.52	0.003	4.7
lift eq	-3.14	0.000	1.1
% Income<$15,000	0.24	0.968	29.8
Population	-2.78	0.012	11.7
% Population>65	1.19	0.083	2.0
% Population<18	-0.25	0.536	24.2
% 1 Person Households	2.68	0.034	15.7
% Minority	-1.95	0.125	34.9
% Work In County	-2.61	0.006	3.8
% Commuters On Pt	2.88	0.010	34.3
% Persons In Poverty	-0.7	0.820	81.5

The p-value for each variable was based on a t-test to determine if the variable coefficient was equal to zero, where higher p-values implied more likelihood that the coefficient of the variable was zero. This test was performed at the 95% significance level, which suggested that any variable with a p-value greater than 0.05 was considered insignificant to the regression. The VIF indicates how much a variable contributes to the overall variance of the equation. A VIF of five or above is considered to be high, suggesting that the variable introduces an unusually large amount of unexplained variance to the model. Using these two criteria together, variables that had a calculated VIF of seven or greater and a p value larger than 0.05 was removed from the regression. One variable, “work out of county,” was removed from the regression due to an interaction problem. Overall, these two criteria eliminated eight of the sixteen variables (Table 4-4).

Table 4-4. Variables remaining after initial regression analysis

Age
Total Miles Traveled
Miles per year on unpaved roads
Wheelchair Accessibility
Population
Percent population greater than 65 in county of operation
Percent of population that work in county of operation
Percent Commuters in county of operation

A best subsets analysis was performed on the remaining eight variables to determine the combination that provided the lowest error and variation, and the highest R^2(adj) value. The results are shown in Table 4-5.

Table 4-5. Best subset analysis results

$R^2_{(adj)}$	68	66	68	68	68	68	69	68	69	69	69
C_P	16	35	14	14	8.6	13	8.1	9.9	7.8	8.8	9
Variables											
Age	X	X	X	X	X	X	X	X	X	X	X
Total Mileage	X		X	X	X	X	X	X	X	X	X
Mile/Year Unpaved		X	X		X	X	X	X	X	X	X
Lift Equiped	X	X	X	X	X	X	X	X	X	X	X
Population										X	X
% Population Greater than 65				X	X		X	X	X	X	X
% Work in County								X	X		X
% Commuters on Pt						X	X		X	X	X

To simplify the table, the variables "lift equipped" and "age" were included in all the best subsets tests. The best subset analysis is interpreted by the $R^2_{(adj)}$ and C_P value. The $R^2_{(adj)}$ value quantifies how well the model represents the data; the higher the percentage, the greater the variability explained in the model (Montgomery et al., 2001). Similarly, lower C_P values imply less variance introduced into the model by the regressors (Montgomery et al., 2001). According to best subset analysis, the best model contained the seven independent variables shown in Table 4-6. These variables were again analyzed using multiple linear regression analysis, p-values and VIF values using the same criteria as earlier (Table 4-7).

Table 4-6. Remaining variables after subset analysis

Age
Total Miles Traveled
Miles per year on unpaved roads
Wheelchair Accessibility
Percent population greater than 65 in county of operation
Percent of population that work in county of operation
Percent Commuters in county of operation

Table 4-7 shows all variables to be significant except for "work in county." Another regression analysis was conducted without the regressor work in county. It indicated the variable "percent commuters" was not significant to the regression. This left "age," "total mileage," "mile/yr unpaved," "lift equipped" and "population greater than 65" as predictors to the model. A regression analysis was performed on the remaining variables, and showed that all regressors were statistically significant to the regression.

Table 4-7. Regression analysis on seven-variable model

Variables	T	P	VIF
Age	-14.4	<0.0001	3
Total Mileage	-3.79	<0.0001	4.5
Mile/yr unp	-3.32	0.001	2.9
lift eq	-3.72	<0.0010	1
% POPULATION>65	2.87	0.004	1.2
% WORK IN COUNTY	-1.53	0.126	1.4
% COMMUTERS ON PT	2.02	0.044	1.3

Variables were tested for multicollinearity by plotting each variable versus the other in Minitab (Montgomery et al., 2001). The total mileage and miles per year unpaved were of particular concern because they originated from the same data. "Miles per year unpaved" was determined by using the total mileage traveled and age. The graph showed no distinct linear relationship between the data, suggesting that multicollinearity was not an issue with the variables (Figure 4-1). This graph showed little correlation between the two mileage regressors.

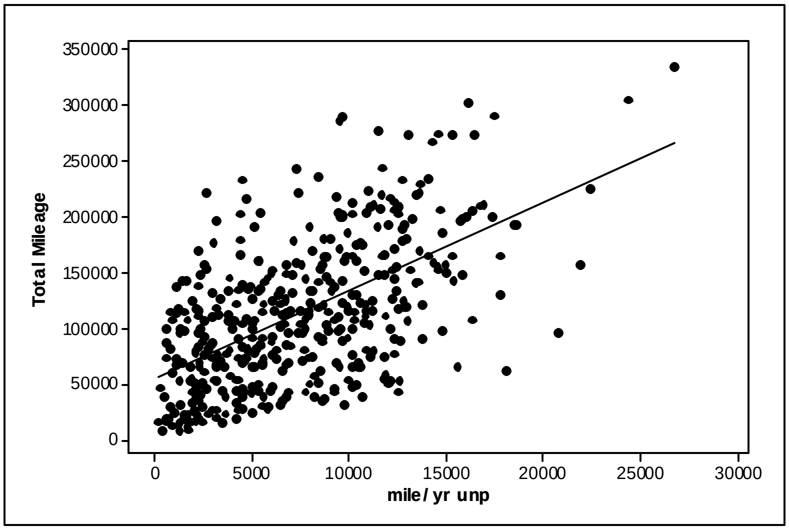


Figure 4-1. Scatterplot of total mileage versus mile/yr unpaved

Model creation

Upon completion of the regressor selection, regression analysis was run again to determine the slopes and intercept to use in the multiple linear regression prediction model. The equation obtained from the regression analysis is shown in equation 4-1.

$$\begin{aligned}\text{Condition Rating} = {} & 4.40 - 0.230\ \text{Age} - 0.000003\ \text{Total Mileage} \\ & - 0.000021\ \text{mile/yr unp} - 0.217\ \text{lift eq} \\ & + 3.73\ \%\ \text{Population}>65 \end{aligned} \qquad (4\text{-}1)$$

Results from the statistical analysis demonstrated that age was the strongest predictor of vehicle condition, followed total mileage and miles traveled per year on unpaved roads. The remaining variables contributed slightly to the overall regression. The five-variable model had an $R^2_{(adj)}$ of 68.4 percent, indicating that roughly 68 percent of the variability in the condition rating data was explained by this regression equation. The $R^2_{(pred)}$ describes the prediction capabilities of the model, estimating how well the model will predict future values. The regression analysis model showed an $R^2_{(pred)}$ of 67.85 percent, suggesting that the model will correctly predict 67 percent of future values.

Model validation

Validating and investigating the adequacy of the model is an important part of regression analysis. Applying statistical tests and investigating graphs of variables, residuals and fits provides a measure of the applicability of the model to real world circumstances and can identify the need for alternative methods of analysis to improve the system.

The first step in inspecting the model is the review of the three basic assumptions made in linear regression analysis. The first assumption, that the errors are normally distributed, can be inspected by taking a probability plot of the residuals (Figure 4-2).

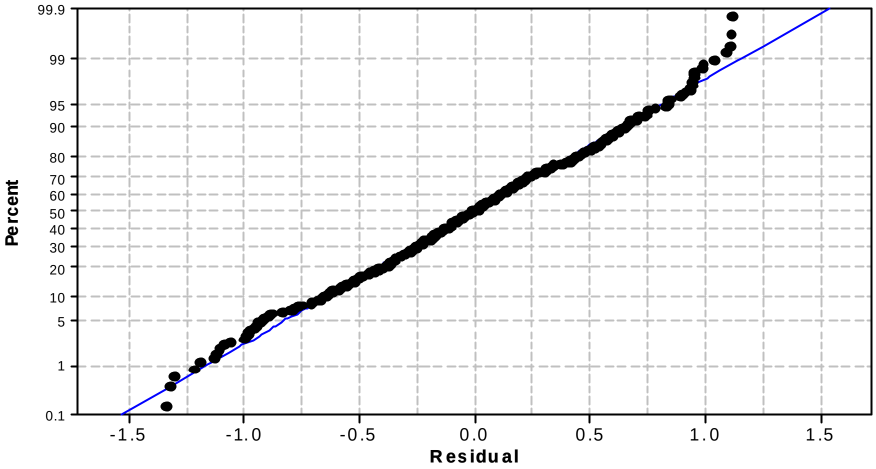


Figure 4-2. Normal plot of residuals

If the line represents a normal distribution, points should be dense toward the center of the line with the gradual decrease as the points move away. In the prediction model, the points show a deviation from the normal line at the ends, typically referred to as "tailing." To further investigate the normality of the vehicle data, an Anderson-Darling goodness of fit test was performed on the residuals. The Anderson-Darling test is used to determine the quality fit of data (Montgomery et al., 2001). The distribution is considered statistically normal when the p-value is greater than 0.05. The p-value of the Anderson-Darling test was 0.249 for the prediction model (Table 4-8).

Table 4-8. Anderson Darling test results

Mean	-4.05E-15
StDev	0.4961
N	450
Anderson Darling Test	0.468
P-Value	0.249

The test suggested that the distribution was normal, providing the statistical evidence needed to support the assumption of normality. The second assumption, the variability of the error is constant, was investigated by plotting the residuals versus the fitted values in an X-Y scatter plot (Figure 4-3).

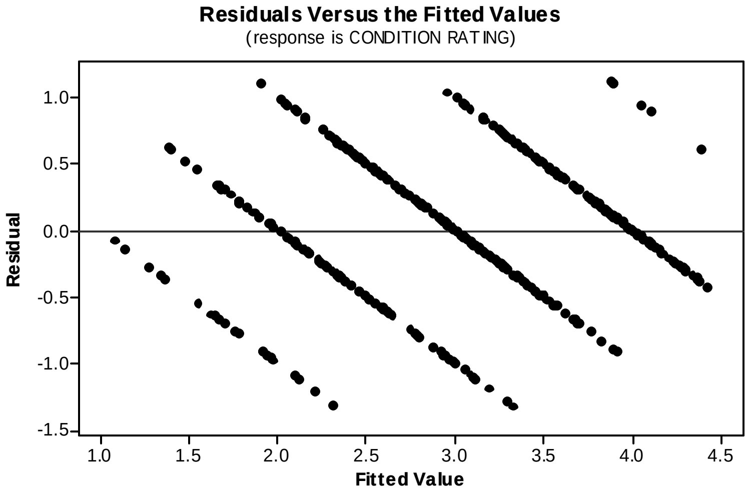


Figure 4-3. Residuals versus fitted value plot

Constant data will be distributed evenly across the plot with minimal clumping indicating that the variance is consistent across the fitted value range. The procurement model plot, Figure 4-3, shows diagonal bands across the centerline suggesting that there could be a variability issue with the data. Further research dismissed this concern by determining that the parallel bands were due to data composed only of integer values.

The next step in the model checking process was the significance of regression test. This test examined the model adequacy by testing if the slopes of the regressors were equal to zero (Montgomery et al., 2001). A zero slope indicates that one or more of the regressors has no linear relationship with the response, and thus is not significant to the model and can be removed. The p-value must be less than 95 percent to fail to reject the hypothesis that the slopes are not equal to zero. The calculated p-value in the created procurement model was zero, suggesting that the regression in the model was significant.

The lack-of-fit statistical test was used to check the model's straight-line fit characteristics. The test assumed that all other assumptions of regression, normality and equal variance were met (Montgomery et al., 2001). To determine the straight-line fit, a pure error lack-of-fit test was conducted on the data using the Minitab software. Pure error lack-of-fit tests replicates data points (with the same values as the regressors) against each other to determine if the assumption of a straight-line fit is valid. The use of replicate values in the test provides a “model-independent estimate” of the variance (Montgomery et al., 2001). The lack of replicates in the mileage data required us to limit the lack-of-fit test to three regressors: “age,” “lift equipped,” and “percent population greater than 65.” The results of the test suggested that the data was not a straight-line fit, indicating a lack-of-fit problem. The results of this test are heavily dependent on the characteristics of the regressors. The absence of the Total Mileage regressor in the test for lack-of-fit was part of the problem because vehicles were not defined entirely by their age. Mileage varied with the age of the vehicle as seen in Table 4-9.

Table 4-9. Range of condition ratings compared to years

Condition Rating	Age	Total Mileage
1	9	219486
2	9	150764
3	9	99125

A second type of lack-of-fit test was applied in Minitab using data-subsetting. It is used to test the straight-line fit when there is a lack of replicates in the data (Montgomery et al., 2001). This test showed a possibility of lack-of-fit, but was within statistically acceptable limits for the model developed (see appendix).

The last test was concerned with the consistency between the model and the data. The data was split into two groups and regression models were created for each set of data (Montgomery et al., 2001). The corresponding regression coefficients were compared to determine if the model was consistent across the data (Table 4-10).

Table 4-10. Regressor coefficient comparison

Data	?0	ß1	?2	ß3	ß4	ß5
Full Model	4.4	-0.23	-0.000003	-0.000021	-0.217	3.73
Prediction	4.847	-0.225	-0.000004	-0.000010	-0.330	4.620
Validation	4.220	-0.226	-0.000003	-0.000030	-0.101	3.500

Regressor coefficient values that are similar show consistency for the overall model. The values in Table 4-10 showed similar values for the regression coefficients of the two “split” test models, suggesting that the overall model was adequate in representing the populations.

Analyzing the intercept (ß$_0$) provides insight into the model’s ability to predict data accurately, especially with deterioration models. In the performance rating system used for this asset management system, the vehicle condition ratings began at five and gradually deteriorated to zero. The model’s highest point, at the intercept point, should be close to five if the model represents the rating system accurately. The full model showed a value of 4.4 for the intercept. The most likely reason for the intercept being so low was the lack of data points in the new rating category (Figure 4-4).

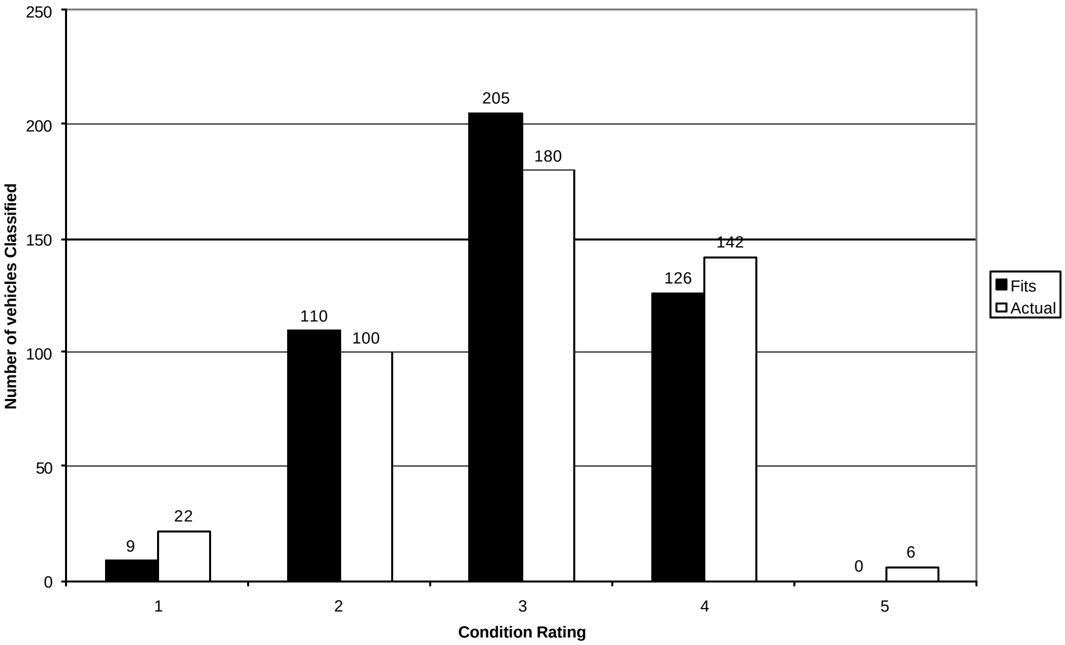


Figure 4-4. Fits and actual fleet categorized vehicles

This assumption was tested by creating "dummy entries" that represented new vehicles with no mileage, simulating vehicles just added to the fleet. Regression analysis was applied to the new data set and the intercept increased significantly, suggesting the lack of points in the condition = five rating group contributed to the lower intercept value.

Another method of validation used a separate set of data to determine how well the model predicted values outside the initial set. The data used to validate the regression model were obtained from the incomplete 2003 tri-annual review, which included 224 transit vehicles. The new data composed a validation data set, for which comparisons were made between actual ratings and predicted ratings calculated from the regression model. The errors were predicted for each point and the overall $R^2_{(pred)}$ was then calculated. The $R^2_{(pred)}$ value should be close to or greater than the $R^2_{(adj)}$ value from the original model. The $R^2_{(pred)}$ from the validation data was 67%, close to the 68.4% $R^2_{(adj)}$ of the original regression analysis. This analysis suggests that the model is valid for future data.

The last model characteristic investigated was the regressor coefficients. In the case of deterioration models, most if not all of the coefficients should reduce the intercept over time. As expected, age, total mileage and mileage unpaved all had negative values, indicating that these values lowered the condition rating as they grew larger. Adding a wheelchair lift to a vehicle requires a large cut in the side of the van, weakening the structure and adding more maintenance issues. Due to this vehicles equipped with wheelchair lifts are typically in worse condition than similar non-wheelchair equipped vans, thus the negative on this value was also expected. The positive regressor coefficient "percent population greater than 65," was the only variable that increased the condition rating. The reason for this positive value most likely comes from increased use of public transportation among the elderly.

Application of Linear Regression Based Procurement Model

The multiple linear regression model developed to predict the condition of vehicle assets was used in a data extrapolation exercise extending five years from the year the vehicles were originally reviewed. An aggregate of the fleet condition was calculated from data resulting from the regression model predictions. The extrapolations were applied to three funding scenarios to demonstrate an application of the procurement model. The funding scenarios were developed using an average vehicle price for the five year analysis, with fixed funding over the five year cycle (Table 4-11).

Table 4-11. Funding scenarios

Funding	Number of Vehicles Purchased Annually
3 million/year	75
2 million/year	50
1 million/year	25

Five assumptions were made in the scenarios:

- All funding levels did not include vehicle maintenance cost
- All vehicles were similar models, cutaway vans with an averaged five-year price of $40,000
- Amounts were in total dollars, absent the 80:20 match requirements
- All vehicles purchased were for replacement only
- All vehicles were replaced by identical vehicles (lift equipped, etc)

Each year, the vehicles with the lowest condition ratings were replaced in each scenario, with the assumption that no additional vehicles were added to the fleet. The results of the funding scenarios are shown in Table 4-12 and Figure 4-5.

Table 4-12. Results of funding scenarios (regression)

Funding	Condition Rating Fleet Average for Year				
	2001	2002	2003	2004	2005
$3 million	3.577778	3.681154	3.789746	3.882572	3.95904
$2 million	3.411111	3.418468	3.404558	3.385535	3.435525
$1 million	3.244444	3.127746	3.009722	2.894155	2.821809

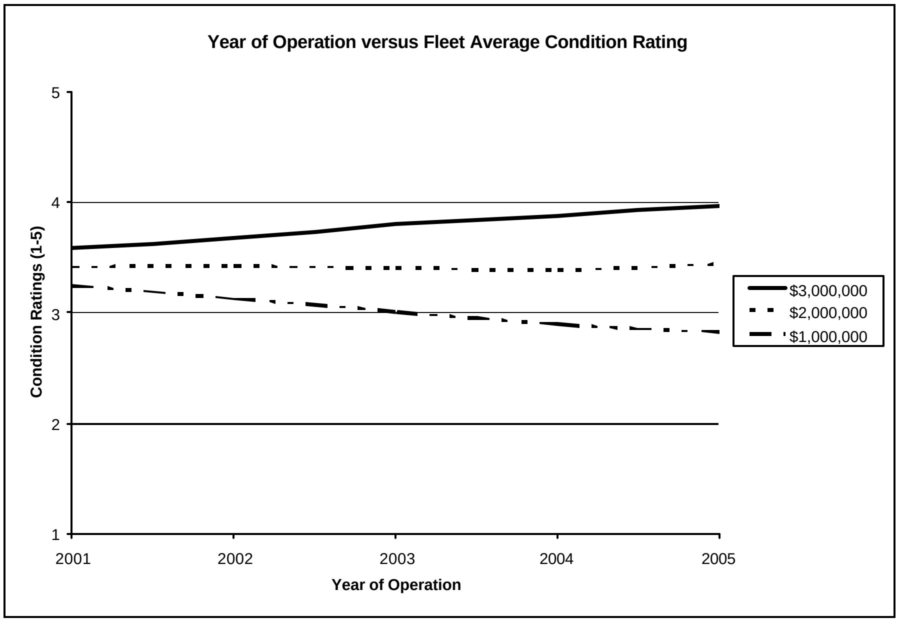


Figure 4-5. Years of operation versus fleet average condition rating

The funding scenarios exhibited the anticipated linear pattern, due to the linear regression used to develop the model and the averaging of the condition ratings. For the $3 million budget, the increase in the average condition rating each year was a result of adding replacement of vehicles with higher condition ratings than the vehicle that were removed form the fleet. In the funded scenario, at no point was a vehicle replaced in fair (condition rating three) or better condition.

Summary

The regression analysis provided a statistically valid model to predict future conditions of individual vehicles and the overall fleet quality. The quality was defined by an integer-only variable, the condition rating. This variable was a result of a series of reviews of agencies, where each vehicle in the 5311 fleet was rated. Linear regression analysis was then used to develop a model to predict the condition rating based on vehicle and socioeconomic characteristics. The developed prediction equation was used to predict future values, which were extrapolated annually for period of five years. Funding scenarios where developed and new vehicles were introduced to the fleet to replace the vehicles with the lowest condition ratings. To determine the effect that specific budgets would have on the overall fleet quality, the average fleet condition rating was calculated for each year in the funding scenarios. These could then be compared using graphs and tables to determine the scenario best suited to ALDOT’s needs and resources.

Chapter 5
Discriminant Analysis

As an alternative to regression analysis, discriminant analysis is used to classify data points into specific groups (Johnson and Wichern, 2002). The condition rating values can be treated as categorical data and assigned by using multiple probability equations to statistically determine each point's category. The same regressors from the multiple regression analysis were used to determine the probability equations for the discriminant analysis. The assumptions found in the multiple regression analysis applied to discriminant analysis, as well as the testing of the individual variables.

Probability equations

Minitab statistical analysis software was used to conduct the discriminant analysis and categorize the data points. Full Minitab output for the discriminant analysis conducted for this model is included in the appendix. The same data that was used in the regression analysis was analyzed with discriminant analysis. The linear discriminant function (ldf) developed in the analysis had an APER of 0.433 (Figure 5-1).

```
Summary of classification
                            True Group
Put into Group      1      2      3      4      5
1                  17     24      3      0      0
2                   5     54     41      0      0
3                   0     21    102     23      0
4                   0      0     29     78      2
5                   0      1      5     41      4
Total N            22    100    180    142      6
N correct          17     54    102     78      4
Proportion      0.773  0.540  0.567  0.549  0.667

N = 450          N Correct = 255           Proportion Correct =
0.567
```

Figure 5-1. Minitab results for discriminant analysis

The APER was determined by subtracting the proportion correct from one (Johnson and Wichern, 2002). Categories one and five had the most correctly classified points, and the highest proportion of correctly categorized data was in category one. The values were also cross-validated to provide a conservative view of the predictive abilities of the current discriminant function (Figure 5-2).

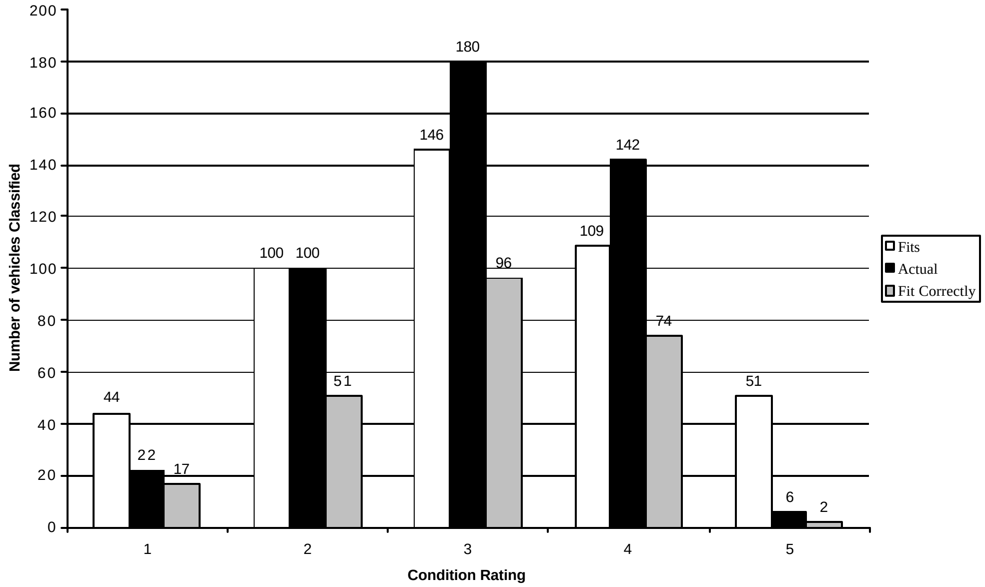


Figure 5-2. Fits and actual fleet categorized vehicles

Cross-validation splits the data and uses one set for comparison against the other set, enabling a realistic estimate of the discriminative properties of the model (Johnson and Wichern, 2002). The APER with cross validation suggested an error rate of 0.473. The lowest proportion correct occurred in group five, predicting 0.333 percent of the categories. The lack of well-defined points in category five (new vehicles) caused less-than-optimum classification in the higher ranges. The linear discriminant function from the analysis, Table 5-1, showed the constants and weights used to determine the probabilities used in category assignments.

Table 5-1. Linear discriminant function for groups

Variable	Condition Ratings				
	1	2	3	4	5
Constant	-52.87	-43.91	-38.15	-31.75	-33.42
Age	4.01	2.82	1.99	0.96	0.69
Total Mileage	0.0000308	0.0000324	0.0000199	0.0000001	-0.0000121
Mile/yr unp	0.0002136	0.0001351	0.0000119	-0.0000403	-0.0000357
Lift eq	4.82	4.62	3.26	2.44	3.47
% POP>65	400.67	416.74	439.44	444.38	463.24

Application of a discriminant analysis based procurement model

The next step in the use of discriminant analysis was the application of the model to a procurement schedule. The same assumptions were made with the procurement model as were made with the previous regression model. The three funding scenarios were the same as in the regression model application. The only change to the procurement model was an extra step that analyzed agencies to determine the greatest need for vehicle replacement based on agency condition rating averages and size. The results of the application of the discriminant analysis model are shown in Table 5-2 and Figure 5-3.

Table 5-2. Results of funding scenarios (discriminant analysis)

Funding	Condition Rating Fleet Average for Year				
	2001	2002	2003	2004	2005
$3 million	3.33	3.60	3.77	4.02	4.24
$2 million	3.11	3.18	3.20	3.38	3.37
$1 million	2.88	2.73	2.60	2.48	2.37

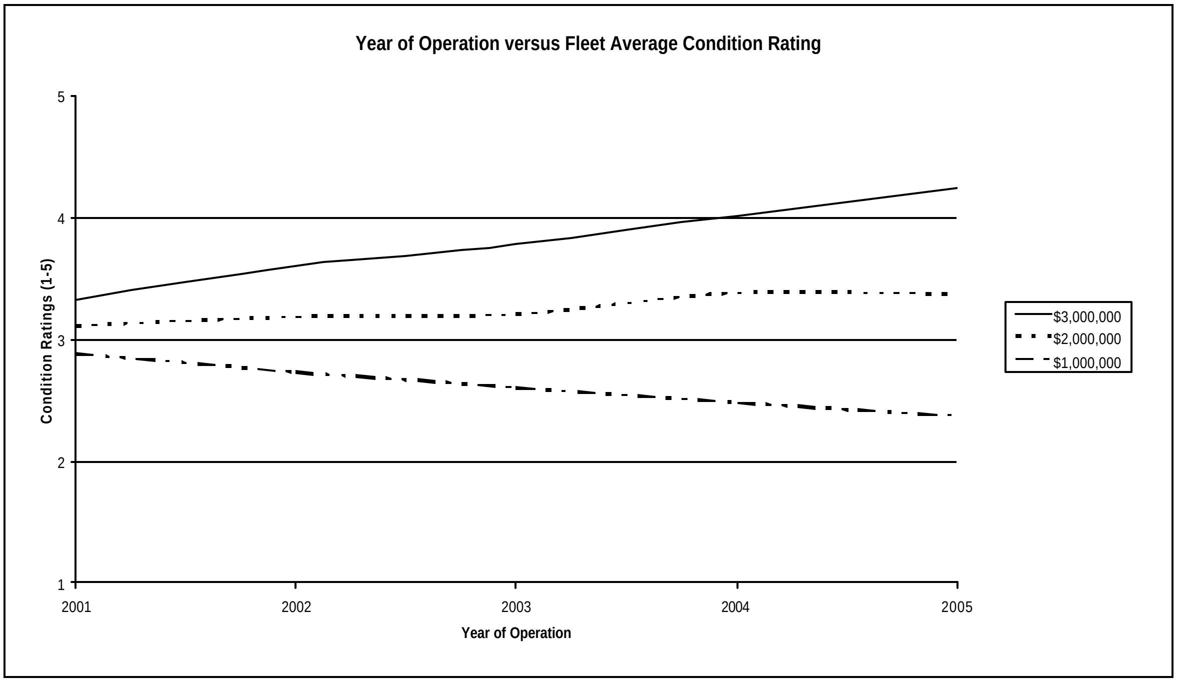


Figure 5-3. Year of operation versus fleet average condition rating

The procurement model based on discriminant analysis demonstrated linear orientation with slight curvature due to the categorical nature of the analysis. There was greater variation between the funding scenarios for this model than the regression model. As with the regression analysis, it took approximately two million dollars per year to maintain the average fleet rating at the current condition level.

Summary

As a valid alternative to the use of regression analysis for categorical data, discriminant analysis categorized data into groups based on the location of each point compared to the centroid of each category. The advantage of using this form of analysis in the fleet prediction stems from the assignment of points to specific categories, reducing the need for interpolation of points that fall between categories. Discriminant analysis was applied to the 5311 vehicle data and used to develop a linear discriminant function to categorize the data points into the condition ratings. The categorization function was then used in a procurement model to predict the effect of funding scenarios on the overall condition of the fleet. The scenarios predicted that overall fleet conditions would remain constant at a funding level of two million dollars per year.

Chapter 6
Conclusions

The successful development of the asset management system in this study enabled ALDOT to effectively store, maintain and analyze Sections 5310/5311 fleet and agency data. The GIS interface allowed users to access the system through multiple query methods including a visual interface. Custom macros and SQL database interlinks enabled ALDOT personnel to summarize, view and print data in various forms from a simplified interface.

A prediction model was developed to estimate future procurement needs and overall fleet quality. The model was based on existing ALDOT data and census information to reduce the need for outside data survey and for additional time and expense associated with more intricate models.

Database

The underlying database was the key to the quality of the analysis and the asset management system. The database was updated to ensure data integrity for ALDOT, with review of over 1,900 vehicle entries and data correction if needed. Input and output routines were simplified to reduce errors and to make the asset management system more user friendly and efficient for ALDOT personnel. GIS enabled advanced queries to the database and allowed advanced analysis using spatial characteristics. The use of advanced queries along with accurate vehicle, agency and review summaries in the database can potentially lead to enhanced management strategies.

Developed Model

The two forms of analysis used to predict future condition ratings were both considered valid for overall fleet prediction. For individual vehicle condition ratings, the regression model provided more-detailed information on vehicle condition and was the more-accurate statistical predictor of the two approaches. Discriminant analysis was useful in understanding the overall condition of the fleet as well as the categorizing the data in specific groups without the added step of rounding the fits and grouping of the data after analysis. Discriminant analysis also provided a clearer understanding of the error involved in the initial categorization process by showing the groups and their respective misclassifications using confusion matrices.

The model used in the asset management system was based on the linear regression analysis. The regression model predicted more conservative condition ratings compared to the discriminant analysis and was considered the best solution for future budget predictions.

Introduction of other variables, such as maintenance, could improve the model characteristics. A maintenance management system is being created for ALDOT and could provide the needed data for a logical maintenance predictor in future research. Both models suffered from a lack of data in the new vehicle region (condition rating five). Vehicles were added during the process of the tri-annual review but were not included in the overall database and therefore are not shown in the final vehicle analysis.

Procurement model

Prediction models add an important facet to an asset management system. ALDOT can use them to estimate future vehicle procurement needs, future budget allocations, vehicle lifespan and fleet quality. Budgetary allocation analysis was integrated into the system to determine the effects of state assistance and the federal grants process.

Closure

The asset management system produced in this project provided ALDOT with a system to enhance management of its transit assets. By combining a traditional database with GIS, mathematical and statistical analysis and SQL programming, an efficient tool was developed to aid in decision making. The new system provides summaries, future predictions, vehicle quality predictions, advanced reporting and agency data reports in digital or paper format, simplifying the process of maintaining, upgrading, analyzing, and accessing asset data.

Chapter 7
References

Anderson, M.D. and A.B. Sandlin. "A Rural Transit Vehicle Management System and Condition Predictor Model," *Journal of Public Transportation*. Vol. 4, No. 1, Pp. 59-72. November 2001.

CAMBRIDGE. Cambridge Systematics, *Pontis Bridge Maintenance System. http://www.camsys.com/ponti03.htm* 2003.

FHWA. US Department of Transportation. *Asset Management Primer*. December 1999.

GASB34. Governmental Accounting Standard Board Statement 34, Basic Financial Statements and Management's Discussion and Analysis for State and Local Governments. Washington, DC, 1999.

ISTEA. *Intermodal Surface Transportation Efficiency Act*. United States Department of Transportation: Federal Highway Administration. 1991.

Johnson, R.A. and D.W. Wichern. *Applied Multivariate Statistical Analysis* 5th ed. Upper Saddle River, New Jersey: Prentice Hall 2002.

Karlaftis, M.G. and K.C. Sinha. "Modeling Approach for Transit Rolling-Stock Deterioration Prediction." *Journal of Transportation Engineering*, May/June 1997, 223-228.

Kurt, C.E., P. Weaver and D.A. Kroeger. "GIS-based Integrated Rural and Small Urban Asset Management System." Midwest Transportation Consortium published under. Department of Transportation, Washington, DC, 2003.

.

Ludwig, Ann. "Systems Planning for Capital Asset Management: A Case Study of the New Jersey Transportation Facilities and Equipment Management System (PTMS)." *Transportation Research Record* 1604, 1997, 109-117.

Montgomery, D. C., G.G. Vining and E.A. Peck. *Introduction to linear analysis* 3rd ed. New York: Wiley Publishing 2001.

Rice, J.A. *Mathematics Statistics and Data Analysis* 2nd ed. Belmont, CA: Duxbury Press 1995.

STATSOFT. Statsoft Inc. *Discriminant Function Analysis*. http://www.statsoft.com/textbook/stathome.html. 2003.

TEA21. *Transportation Efficiency Act of the 21st Century*. United States Department of Transportation: Federal Highway Administration. 2001.

Appendix
Regression Analysis

```
Regression Analysis: CONDITION RATING versus age, Total Mileage, ...

* % WORK OUT COUNTY is highly correlated with other X variables
* % WORK OUT COUNTY has been removed from the equation.

The regression equation is
CONDITION RATING = 3.97 - 0.225 age - 0.000003 Total Mileage
                   + 0.000006 mile/yr pav - 0.000033 mile/yr unp
                   - 0.207 lift eq - 0.06 % INCOME<$15,000
                   - 0.000001 POPULATION + 3.26 % POPULATION>65
                   - 2.65 % POPULATION<18
                   + 9.84 % 1 PERSON HOUSEHOLDS
                   - 1.00 % MINORITY - 1.22 % WORK IN COUNTY
                   + 66.8 % COMMUTERS ON PT
                   - 0.56 % PERSONS IN POVERTY


Predictor               Coef     SE Coef      T      P   VIF
Constant               3.966       1.602   2.47  0.014
age                 -0.22527     0.02592  -8.69  0.000   7.5
Total Mileage    -0.00000345  0.00000137  -2.52  0.012  14.2
mile/yr pav       0.00000594  0.00000973   0.61  0.542   5.2
mile/yr unp      -0.00003265  0.00001079  -3.03  0.003   4.7
lift eq             -0.20651     0.05605  -3.68  0.000   1.1
% INCOME<$15,000      -0.064       1.585  -0.04  0.968  29.8
POPULATION       -0.00000131  0.00000052  -2.52  0.012  11.7
% POPULATION>65        3.265       1.876   1.74  0.083   2.0
% POPULATION<18       -2.652       4.277  -0.62  0.536  24.2
% 1 PERSON HOUSEHOLDS   9.844       4.633   2.12  0.034  15.7
% MINORITY           -1.0038      0.6528  -1.54  0.125  34.9
% WORK IN COUNTY     -1.2233      0.4386  -2.79  0.006   3.8
% COMMUTERS ON PT      66.76       25.73   2.59  0.010  34.3
% PERSONS IN POVERTY  -0.559       2.460  -0.23  0.820  81.5

S = 0.491748   R-Sq = 70.3%   R-Sq(adj) = 69.3%

PRESS = 112.825   R-Sq(pred) = 68.11%
```

Figure A-1. Variable selection regression analysis output

Table A-1. Data points removed because of inconsistencies

Residuals Removed From Analysis

CR	age	Total Mileage	mile/yr pav	mile/yr unp	lift eq	% POP>65
3	14	87288	3803	2432	0	0.14945
1	5	98370	16329	3345	0	0.11693
4	8	152828	11653	7450	0	0.14945
2	6	347382	30106	27791	0	0.14514
5	14	138362	3854	6029	0	0.16397
3	10	105142	8727	1787	0	0.11693
4	6	138880	17591	5555	0	0.14648
4	6	145434	10665	13574	0	0.12040
4	9	55101	3184	2939	0	0.14514
1	5	98370	13182	6492	1	0.13382
3	2	25784	5028	7864	0	0.16397
3	11	170296	11766	3716	0	0.14648
3	10	143181	6300	8018	1	0.12040
4	8	57081	3710	3425	0	0.14514
2	2	30776	7078	8310	1	0.09853
1	5	94856	9865	9106	0	0.14514
1	6	137277	15329	7550	1	0.13382
1	17	72013	1694	2542	0	0.14030
4	7	60331	7154	1465	1	0.11693
4	6	193133	24142	8047	1	0.12453
1	7	117866	11281	5557	1	0.13382
4	6	186606	23326	7775	1	0.12453
4	5	146811	17324	12039	1	0.09994
3	9	144745	12223	3860	0	0.14648
3	10	180273	10997	7031	0	0.14945
3	9	233344	13482	12445	0	0.14514
4	6	132515	13472	8613	0	0.14945
4	7	49505	4314	2758	1	0.14945
4	6	117394	14674	4891	1	0.12453
2	5	18233	1896	1750	0	0.14514
3	10	164205	10017	6404	0	0.14945
4	6	139030	14135	9037	0	0.14945
3	9	84434	8912	469	1	0.08953
3	2	51989	18196	7798	0	0.13035

```
Best Subsets Regression: CONDITION RA versus mile/yr unp, POPULATION, ...

Response is CONDITION RATING
The following variables are included in all models: age lift eq

                                                 %
                                               %
                                             %   C
                                               W O
                                             P O M T
                                             O R M o
                                         m   P K U t
                                         i P U   T a
                                         l O L I E l
                                         e P A N R
                                         / U T   S M
                                         y L I C   i
                                         r A O O O l
                                           T N U N e
                                         u I > N   a
                       Mallows           n O 6 T P g
Vars  R-Sq  R-Sq(adj)      C-p        S  p N 5 Y T e
   1  68.0        67.8      16.0  0.50405            X
   1  66.7        66.4      34.6  0.51418  X
   2  68.3        68.0      13.5  0.50212  X         X
   2  68.2        67.9      14.2  0.50253      X     X
   3  68.8        68.4       8.6  0.49889  X   X     X
   3  68.5        68.1      12.9  0.50129  X       X X
   4  68.9        68.5       8.1  0.49804  X   X   X X
   4  68.8        68.4       9.9  0.49902  X   X X   X
   5  69.1        68.6       7.8  0.49729  X   X X X X
   5  69.0        68.5       8.8  0.49785  X X X   X X
   6  69.2        68.6       9.0  0.49741  X X X X X X
```

Figure A-2. Best subset Minitab output

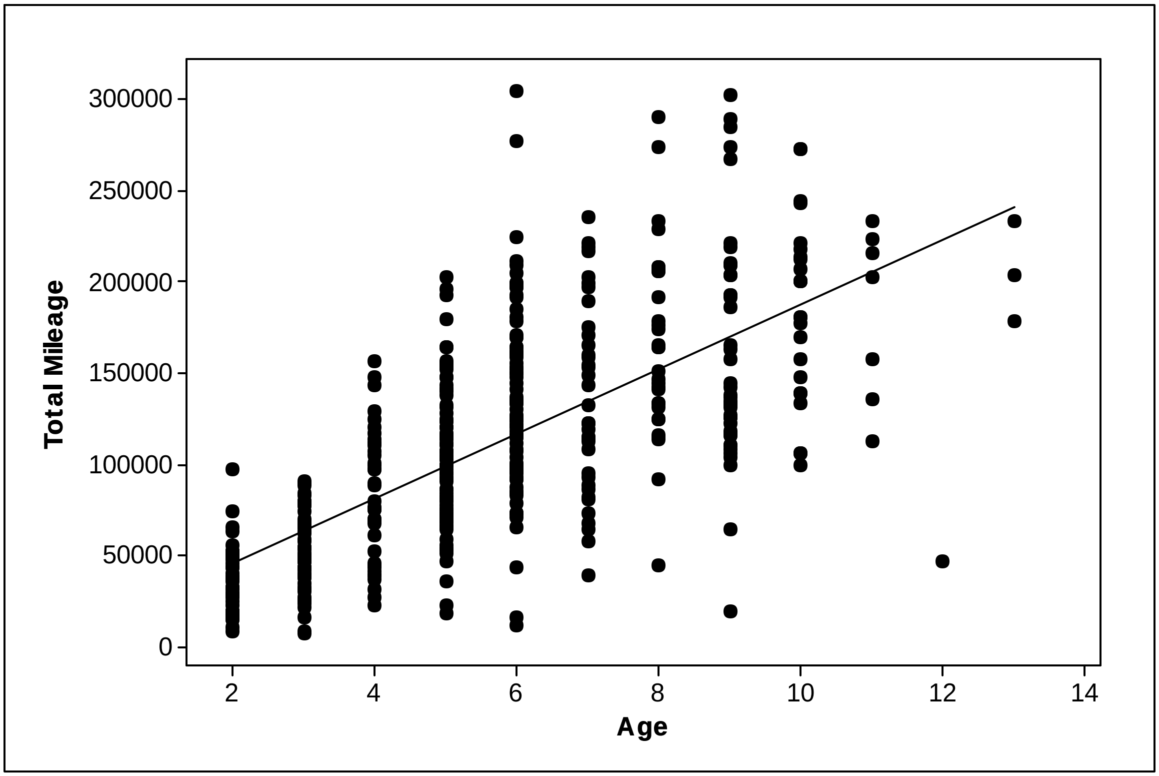


Figure A-3. Age versus total mileage scatterplot

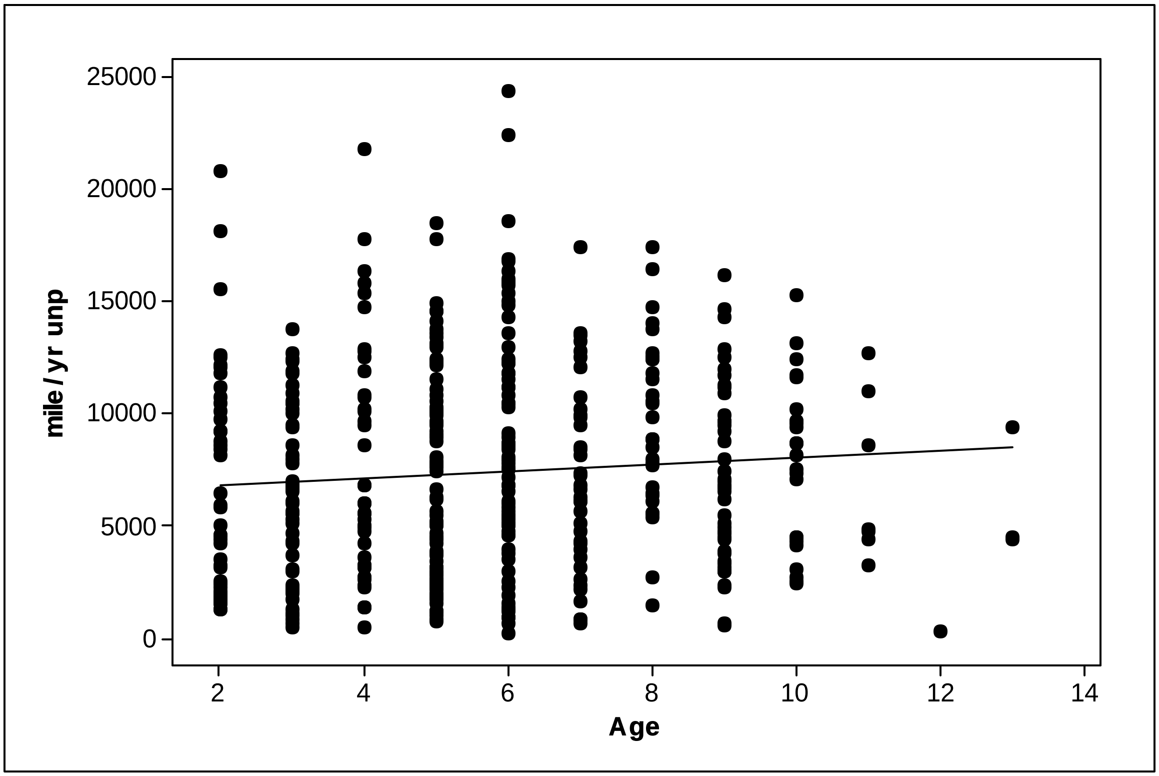


Figure A-4. Mile/yr unpaved versus scatterplot

```
Regression Analysis: CONDITION RATING versus age, Total Mileage, ...
```

```
The regression equation is
CONDITION RATING = 4.40 - 0.230 age - 0.000003 Total Mileage
                   - 0.000021 mile/yr unp - 0.217 lift eq
                   + 3.73% POPULATION>65

Predictor              Coef     SE Coef       T      P  VIF
Constant             4.4022      0.1921   22.92  0.000
age                -0.22972     0.01593  -14.42  0.000  2.8
Total Mileage   -0.00000328  0.00000074   -4.41  0.000  4.1
mile/yr unp     -0.00002089  0.00000762   -2.74  0.006  2.3
lift eq            -0.21713     0.05581   -3.89  0.000  1.0
% POPULATION>65       3.729       1.432    2.60  0.010  1.1

S = 0.498892   R-Sq = 68.8%   R-Sq(adj) = 68.4%

PRESS = 113.727   R-Sq(pred) = 67.85%

Analysis of Variance

Source           DF       SS      MS       F      P
Regression        5  243.269  48.654  195.48  0.000
Residual Error  444  110.509   0.249
Total           449  353.778

No replicates.
Cannot do pure error test.

Source           DF   Seq SS
age               1  216.840
Total Mileage     1   20.046
mile/yr unp       1    0.956
lift eq           1    3.741
% POPULATION>65   1    1.687

Lack of fit test
Overall lack of fit test is significant at P = 0.063
```

Figure A-5. Model creation regression analysis output

```
Regression Analysis: CONDITION RATING versus age, Total Mileage, ...

The regression equation is
CONDITION RATING = 4.85 - 0.225 age - 0.000004 Total Mileage
                        - 0.000015 mile/yr unp - 0.330 lift eq
                        + 4.62 % POPULATION>65

Predictor               Coef     SE Coef       T      P  VIF
Constant              4.8474      0.2682   18.07  0.000
age                 -0.22535     0.02140  -10.53  0.000  3.0
Total Mileage    -0.00000410  0.00000101   -4.07  0.000  4.3
mile/yr unp      -0.00001474  0.00000998   -1.48  0.141  2.2
lift eq             -0.32984     0.07352   -4.49  0.000  1.0
% POPULATION>65        0.669       1.997    0.33  0.738  1.1

S = 0.469133   R-Sq = 74.6%   R-Sq(adj) = 74.1%

PRESS = 50.8152   R-Sq(pred) = 73.25%

Analysis of Variance

Source           DF       SS      MS       F      P
Regression        5  141.784  28.357  128.84  0.000
Residual Error  219   48.199   0.220
Total           224  189.982

No replicates.
Cannot do pure error test.

Source           DF   Seq SS
age               1  125.626
Total Mileage     1   11.295
mile/yr unp       1    0.403
lift eq           1    4.436
% POPULATION>65   1    0.025

No evidence of lack of fit (P >= 0.1).
```

Figure A-6. Prediction model regression analysis output

```
Regression Analysis: CONDITION RATING versus age, Total Mileage, ...

The regression equation is
CONDITION RATING = 4.22 - 0.226 age - 0.000003 Total Mileage
                        - 0.000027 mile/yr unp - 0.101 lift eq
                        + 3.05% POPULATION>65

Predictor                Coef     SE Coef      T      P  VIF
Constant               4.0170      0.2737  14.68  0.000
age                  -0.22579     0.02351  -9.60  0.000  2.6
Total Mileage     -0.00000270  0.00000109  -2.49  0.014  3.9
mile/yr unp       -0.00002712  0.00001149  -2.36  0.019  2.4
lift eq              -0.10127     0.08347  -1.21  0.226  1.0
% POPULATION>65         6.200       2.048   3.03  0.003  1.2

S = 0.521610   R-Sq = 63.5%   R-Sq(adj) = 62.7%

PRESS = 63.3663   R-Sq(pred) = 61.21%

Analysis of Variance

Source           DF       SS      MS      F      P
Regression        5  103.775  20.755  76.28  0.000
Residual Error  219   59.585   0.272
Total           224  163.360

No replicates.
Cannot do pure error test.

Source           DF  Seq SS
age               1  91.335
Total Mileage     1   9.036
mile/yr unp       1   0.544
lift eq           1   0.366
% POPULATION>65   1   2.495

No evidence of lack of fit (P >= 0.1).
```

Figure A-7. Validation model regression analysis output

```
Discriminant Analysis: CONDITION RATING versus age, Total Mileage, ...
Linear Method for Response: CONDITION RATING
Predictors: age, Total Mileage, mile/yr unp, lift eq, % POPULATION>65

Group         1         2         3         4         5
Count        22       100       180       142         6

Summary of classification

                            True Group
Put into Group      1      2      3      4      5
1                  17     24      3      0      0
2                   5     54     41      0      0
3                   0     21    102     23      0
4                   0      0     29     78      2
5                   0      1      5     41      4
Total N            22    100    180    142      6
N correct          17     54    102     78      4
Proportion      0.773  0.540  0.567  0.549  0.667

N = 450           N Correct = 255    Proportion Correct = 0.567

Summary of Classification with Cross-validation
                            True Group
Put into Group      1      2      3      4      5
1                  17     27      3      0      0
2                   5     51     43      0      0
3                   0     21     96     23      0
4                   0      0     32     74      4
5                   0      1      6     45      2
Total N            22    100    180    142      6
N correct          17     51     96     74      2
Proportion      0.773  0.510  0.533  0.521  0.333
N = 450           N Correct = 240           Proportion Correct = 0.533

Linear Discriminant Function for Groups
                      1       2       3       4       5
Constant         -52.87  -43.91  -38.15  -31.75  -33.42
age                4.01    2.82    1.99    0.96    0.69
Total Mileage      0.00    0.00    0.00    0.00   -0.00
mile/yr unp        0.00    0.00    0.00   -0.00   -0.00
lift eq            4.82    4.62    3.26    2.44    3.47
% POPULATION>65  400.67  416.74  439.44  444.38  463.24
```

Figure A-8. Discriminant analysis Minitab output