In order to make decisions and provide guidance to clients, engineers often analyze large databases of measurements to uncover trends and behaviors. These datasets are of a wide range of formats and organizations, and a large part of the analysis can involve moving the data into an appropriate format for analysis. One rich dataset developed by researchers at Texas A&M University is for toll road price and usage in Seoul, South Korea. The purpose of this assignment is to analyze a set of time series data for the toll usage and to compare the toll data with other economic factors, including the price of gas, the average household income, and the cost of public transportation (all data is provided in the datasets). Section 1 provides more background on the Seoul dataset and on economic models for traffic data. The detailed analyses required by this assignment are specified in Section 2. Section 3 details the contents of the memorandum.

1 Toll Roads in Seoul, South Korea

One important reason for imposing tolls on roadways is to help reduce traffic congestion by giving an incentive to drivers to carpool, drive at non-toll times, or to seek alternate routes. While there are correlations between traffic volumes and toll rates, many other factors also influence driver decisions, including gas prices, availability of alternate routes, road construction, family income, and many others. Hence, management and analysis of a toll structure for a roadway can require a complex analysis.

In Seoul, South Korea, tolls were imposed on two roadways between the Han River and the Central Business District (CBD) beginning in November, 1996. These were at the Namsan 1st and Namsan 3rd tunnels. Both roadways are supplied by similar suburb regions and are the main gateways to the CBD. Figure 1 shows a map of these tunnels with respect to the Han River and the CBD.

The toll structure provided several different incentives to drivers. Vehicles with more than two passengers were admitted toll-free. In addition, vehicles participating in a “do not drive one day
in seven” program received coupons for some tolls. Other vehicles paid a nearly constant toll rate of 2000 Korean Won (approximately US $2.50).

1.1 Provided Datasets

Three datasets are provided describing the toll usage in Seoul, South Korea from January, 1997, through December, 2007. The first two datasets are named Tunnel_1.txt and Tunnel_3.txt, and present data for the Namsan 1st and Namsan 3rd tunnels, respectively. The columns in these two data files present data for the year, the month (1 through 12), number of vehicles paying cash, number of vehicles paying with credit card, number of vehicles using coupons, number of vehicles admitted toll-free, total number of vehicles, and the toll price in 2007 market value.

A third dataset, Other_data.txt, presents several other economic measures for the same time period. The columns in this dataset are the year, month (1 through 12), number of work days per month (accounting for holidays, etc.), average income per household in 2007 market value, the price of gas in 2007 market value, the price of public transportation in 2007 market value, and the number of registered vehicles in the Seoul metropolitan area.
1.2 Price Elasticity of Demand

An important measure indicating to what degree the toll price affects driver decision is the elasticity of demand, defined as the percentage change in demand resulting from a percentage change in price. Although the toll rate has remained constant since 1996, the value of the toll to users has changed due to inflation. In order to take this effect into account, the data presented here have been converted to 2007 prices using the consumer price index (CPI) for Seoul. Thus, all prices in the database are Korean Won in 2007 equivalent market values, and the elasticity of demand can be estimated based on these prices.

If the elasticity is assumed constant, then the demand function can be specified as

\[ V = \alpha t^\beta \]  

(1)

where \( V \) is the volume of traffic in vehicles per workday, \( t \) is the toll price in 2007 market values, and \( \alpha \) and \( \beta \) are curve-fitting parameters. The elasticity is then evaluated as \( \beta \).

2 Assignment

Use the provided data to complete several different analyses of traffic usage of these tollways in Seoul, South Korea, as described in the following sections.

2.1 Monthly Statistics for Each Tunnel

Tollway and traffic volume are often correlated with seasonal variations, especially as is related to holidays. In order to visualize the seasonal variation of the toll volume at each tunnel, plot the mean, maximum, minimum, and standard deviation of total number of vehicles for each month in the dataset separately for each tunnel. Figure 2 presents the desired result of this step of the analysis.

In order to accomplish this task, the data need to first be organized by month (group all January data together, etc.). Once the data are grouped by month, compute the statistics using the built-in Matlab functions, \texttt{max()}, \texttt{min()}, \texttt{mean()}, and \texttt{std()}. Use the \texttt{help} command to learn more about these functions and how to use them.

For the figure, the mean values should be plotted versus the month number and connected by a solid line with circles at each data point. The ± one standard deviation should be plotted using error bars. You may use the built-in Matlab function \texttt{errorbar()} to perform this task. Finally, the maximum and minimum value for each month should be plotted as starred data points. Refer to the figure for added detail.

2.2 Time Series of Tunnel Usage

While the monthly statistics help identify seasonal variations, they do not uncover long-term trends. In order to understand year-to-year variations, the actual data time series must be analyzed. A second plot should summarize tollway usage for each tunnel by plotting the time series of total number of vehicles, numbers of vehicles paying cash, vehicles using coupons, and vehicles admitted toll-free. These four datasets should be plotted versus months since 1997. One subplot for each tunnel should be created, providing two rows and one column of plots in a single figure. For each
Figure 2: Monthly variation in the total number of vehicles for the Namsan 1st and Namsan 3rd toll tunnels. The solid line presents the mean value with error bars indicating ± one standard deviation. The data range is indicated by the stars; the figures summarize the data from 1997 to 2007.

2.3 Factors Influencing Toll Usage

Several factors influence a driver’s decision to pay a toll. For the third figure, plot the time series of gas price, average income per household, and cost of public transportation versus months since 1997 in a single figure. Because the household income is much greater than the other values, the datasets must be normalized in order to compare them in a single figure. To do this, divide each time series by the value on January 1, 1997. Hence, each line in the figure should originate from 1.0 on the vertical axis.

2.4 Elasticity of Demand

The elasticity of demand is calculated using Equation 1. Because the elasticity is not defined when the tollway first opens, it takes a few years before the elasticity is well-behaved. For this reason, only consider the data from January 1, 1998 through December 31, 2007.
First, plot the total number of vehicles per workday paying tolls (cash or credit, but not coupons or free) versus 2007 market price. Plot data for both tunnels in a single figure. Use different data point types to distinguish the two datasets, and do not connect any of the points with a line.

Second, fit Equation 1 to each dataset separately and plot the fitted line in the figure. Use matrix least-squares linear regression to compute the fitted line. Record the exponent $\beta$: this value is the elasticity.

Third, calculate the $r^2$ value (the coefficient of determination) for each fit.

2.5 Correlation Coefficient

In order to evaluate which other factors may contribute the most to drivers’ decisions to pay tolls, a cross-correlation analysis should be performed between the toll usage and the other data. The correlation coefficient is defined in Chapra (2008) in Equation (13.21). Compute the correlation coefficient between the following sets of data for each of the tunnels separately:

- The number of vehicles admitted toll-free (having more than two passengers) and the price of gas.
- The total number of vehicles paying tolls (cash and credit) versus the price of the toll

3 Memorandum

The results of your analyses will be reported in the form of a technical memorandum with supporting appendices where the details of the analysis are presented.

For the main body of the memorandum, present and discuss each of the four figures listed above. Use good technical English and present a critical analysis of the results. Your text should both state facts about the figure (e.g., “...in 2004 the price of gas increased showing a corresponding decrease in toll volume...”) and interpret the figure to draw conclusions (e.g., “...thus, gas price is an important predictor of traffic volume...”).

The memorandum should also present the numerical values of the elasticity of demand calculated for each tunnel with a corresponding, brief interpretation of their meaning.

Finally, after considering all of your analysis, make a recommendation to the toll road authority for what price the toll should be set at to maximize revenue in 2010-15, the years following the data set used for this analysis. Back up your recommendation with quantitative analysis.

The appendices should present all Matlab codes created to perform the analysis, relevant output from the Matlab Command Window, and any other analyses performed that support the main body of the memorandum.

4 Acknowledgements

Data provided by Professor Mark Burris.