Travel Demand Forecasting
Interdisciplinary Lively Application Project

Title: Travel Demand Forecasting and Analysis for South Texas

Authors: Lei Yu and Della D. Bell

Department of Transportation Studies and Department of Mathematics, Texas Southern University, 3100 Cleburne Avenue, Houston, Texas 77004

Editor: David C. Arney

Mathematics Classification: Algebra, Calculus, Statistics

Disciplinary Classifications: Transportation Planning

Prerequisite Skills:
1. Solving Linear Equations
2. Solving Quadratic Equations
3. Regression analysis concept
4. Elementary Optimization

Physical Concepts Examined:
1. Trip Attractions
2. Trip Distributions
3. Trip Modal Choice
4. Trip Route Choice

Materials:
1. Problem Statement; Student
2. Sample Solution; Instructor
3. Notes for the Instructor

Computing Requirements:
1. Microsoft Excel or Similar Tools;
2. Ability to Use Calculator

Contents
1. Problem Statement
2. Sample Solution
3. Instructor Notes

1. Problem Statement

Introduction

Travel demand forecasting is the most important phase in the urban transportation planning process. The purpose of travel demand forecasting is to predict the travel demands on the roads in order to estimate the likely transportation consequences of transportation alternatives (including a do-nothing alternative) that are being considered for implementation. Usually, travel demand forecasting is performed using a 4-step sequential model described as follows:

Step 1: Trip Generation: Should I make a trip?
Step 2: Trip Distribution: Where should I go?
Step 3: Modal Choice: What mode of transportation should I use?
Step 4: Trip Assignment: Which route in the network should I take?

Step 1 determines how many trips will be produced by each residential area and how many trips will be attracted by each commercial site. Step 2 determines where each trip generated in Step 1 will go. For example, for each residential area, there may exist multiple choices for the shopping purpose. The percentage of the total trips that will be attracted by each shopping site must be determined. Step 3 determines what mode of transportation each trip will use. The choices of transportation may include bus, light-rail (used only in the city), subway, private automobile, taxi, bicycle, walk and so on. Finally, Step 4 determines which route each trip will use.
By the end of the travel demand forecasting process, the traffic volumes (the number of vehicles per unit time) on the roads will be produced, which provide useful information about the congestion on the streets. Thereafter, transportation planners can select the best transportation projects by reviewing the resulting levels of congestion from a series of transportation alternatives. The 4-step travel demand forecasting process is usually carried out using various mathematical models.

PART 1: Step 1 - Trip Generation

General Information

The objective of a trip-generation model is to forecast the number of trips that will begin from or end in each travel zone within the region for a typical day of the target year. The most widely used trip generation models are regression models which are expressed by the following two linear equations:

\[ P_i = a_0 + a_1 X_1 + a_2 X_2 + ... + a_r X_r \]  
\[ A_i = a_0 + a_1 X_1 + a_2 X_2 + ... + a_r X_r \]

\( P_i \) represents the total number of trips produced by a residential zone \( i \) and \( A_i \) represents the total number of trips attracted by commercial activities in zone \( i \). \( X_1 \) to \( X_r \) are a series of independent variables, which are usually derived from the urban travel surveys. For example, \( X_1 \) may represent the total zonal population, \( X_2 \) may represent the average household income, and \( X_3 \) may represent the average auto ownership of each zone. \( a_0 \) to \( a_r \) are constant values/coefficients of the independent variables, which are usually derived through the regression analysis.

It should be noted that the above regression equations for trip productions and attractions can be non-linear in many cases. However, since the regression analysis for non-linear equations is often performed by converting the non-linear form to a linear form first, the understanding of solving the linear equations is essential.

Example: Trip Attractions in South Texas

A survey of the commercial activities was conducted for five zones in South Texas. Each zone represents a portion of the land that is used as the basic unit in an urban planning process. The data were collected based on three types of employment: manufacturing, retail and services, and others. The resulted zonal employment of three different commercial types and their respective trip attractions are listed in
the following table. The first column in Table 1 represents the Zone number. The second column is the total number of people employed in manufacturing related companies for each zone, while the third column is the total number of people employed in the retail and services sector. The fourth column is the total number of people employed in all companies other than manufacturing and retail and services, while the fifth column represents the total number of people employed for all companies. The sixth and also the last column lists the surveyed number of trips that are attracted by each zone. Variables $X_1$, $X_2$, $X_3$, and $X$ are used to express the employment types, while the variable $Y$ is used to express the trip attractions.

<table>
<thead>
<tr>
<th>Zone</th>
<th>Manuf. $X_1$</th>
<th>Ret&amp;Ser $X_2$</th>
<th>Others $X_3$</th>
<th>Total $X$</th>
<th>Attraction $Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6820</td>
<td>2547</td>
<td>115</td>
<td>9482</td>
<td>9428</td>
</tr>
<tr>
<td>2</td>
<td>111</td>
<td>1899</td>
<td>0</td>
<td>2010</td>
<td>2192</td>
</tr>
<tr>
<td>3</td>
<td>228</td>
<td>87</td>
<td>259</td>
<td>574</td>
<td>330</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>127</td>
<td>0</td>
<td>127</td>
<td>153</td>
</tr>
<tr>
<td>5</td>
<td>2729</td>
<td>813</td>
<td>294</td>
<td>3836</td>
<td>3948</td>
</tr>
</tbody>
</table>

Table 1: The number of people employed in each sector and the total trip attractions

This example intends to let students excise the concept of developing trip attraction equations through the regression analysis technique. In practice, the developed attraction equations are used to predict the actual trip attractions given the projected employment numbers for a future year. Trip productions should be predicted in the same way. To simplify the problem, this example is only designed to practice the regression analysis process, but not to predict the future trip attractions. In the Part 2: Step 2 of the next section, trip productions and attractions are assumed to have been predicted beforehand.

**Requirement 1**: Determine a single linear regression equation between dependent variable $Y$ and each of independent variables $X$, $X_1$, $X_2$, and $X_3$ individually.

**Requirement 2 (Optional)**: Determine a multiple linear regression equation between dependent variable $Y$ and independent variables $X_1$, $X_2$, and $X_3$. (Use of Microsoft Excel Data Analysis Tool is suggested.)
Requirement 3: Select the equations from Requirements 1 and 2 that are acceptable for use in trip generation analysis. (At a minimum, a minimum value of R-squared, such as 0.65, should be set as the criteria and compared with the value of R-squared for each equation. Optimally, an F-test and/or t-test should be conducted.)

PART 2: Step 2 - Trip Distribution

General Information
The objective of a trip distribution model is to determine the total number of trips between all pairs of zones $i$ and $j$, where $i$ is the trip-producing zone and $j$ is the trip-attracting zone of the pair. The rationale of trip distribution is as follows: all trip-attracting zones $j$ in the region are in competition with each other to attract trips produced by each zone $i$. Everything else being equal, more trips will be attracted by zones that have higher levels of “attractiveness.”

A widely used trip distribution model is the gravity model, which gets its name from the fact that it is conceptually based on Newton’s law of gravitation. Newton’s law of gravitation states that the force of attraction between two bodies is directly proportional to the product of the masses of the two bodies and inversely proportional to the square of the distance between them. It can be described by the following equation:

$$ F = k \frac{M_1 M_2}{r^2} \quad (3) $$

The variation of the Newton’s law of gravitation to the trip distribution takes the following form:

$$ T_{ij} = k \frac{P_i A_j}{W_{ij}} \quad (4) $$

$T_{ij}$ is the total number of trips between zones $i$ and $j$, $P_i$ and $A_j$ are the total trip productions for zone $i$ and attractions for zone $j$, which can be derived from the trip generation step. $A_j$ may also be the relative attractiveness of Zone $j$. For a given starting zone, $k$ is a constant value. $W_{ij}$ is the travel impedance, which can be defined as either the travel time or the travel distance. Rewriting Equation (4) in a form that is expressed by the trip productions and a probability factor results in the following Equation (5).
$T_{ij} = P_i \left( \frac{A_i F_{ij}}{\sum_x A_x F_{ix}} \right)$ where $F_{ij} = \frac{1}{W_{ij}^c}$ (5)

$F_{ij}$ is called the travel-time factor (or friction factor.) The term contained by the parentheses is a probability factor, which represents the proportion that trips produced by Zone $i$ are attracted by Zone $j$. This term is usually expressed by $p_{ij}$. Usually, the value of exponent $c$ in Equation (5) must be determined by calibration before the gravity model is applied in the trip distribution process.

**Example: Trip Distribution in South Texas**

In a small town in South Texas, the land uses are divided into four zones as shown in Figure 1. Zone 1 is a residential zone, which does not have any commercial sites. Zone 2 and Zone 4 are commercial zones, which do not include any residential land uses. Zone 3 is a combined residential and commercial zone, which has both residential and commercial land uses. Zones 2, 3 and 4 compete in attracting trips that are produced by the households in Zones 1 and 3.

Based on a trip-generation analysis, the total trip productions and the relative zonal attractiveness for the target-year 2005 are generated and illustrated by Table 2. The inter-zonal and intra-zonal travel times are found to be the Table 3. A calibration of the gravity model has also found that $c$ in the gravity Equation (5) equals 2.0.

![Figure 1: A 4-zone network](image_url)
Table 2: Trip productions $P_i$ and attractiveness $A_j$ for the 4-zone network

<table>
<thead>
<tr>
<th>Zone</th>
<th>Productions</th>
<th>Attractiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20000</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>15000</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 3: Travel time $W_{ij}$ between Zone $i$ and Zone $j$

<table>
<thead>
<tr>
<th>$i \setminus j$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

Requirement 4: Use the gravity model Equation (5) to estimate the interchange trips between zones.

Requirement 5: Estimate the total trip attractions for Zones 2, 3 and 4.

PART 3: Step 3 - Modal Choice

General Information

In a typical travel situation, trip-makers can select between several travel modes. These include driving, riding with someone else, taking the bus, walking, riding a motorcycle and so forth. A modal choice (sometimes called modal split) is concerned with the trip-maker’s behavior regarding the selection of travel mode.

The most widely used modal choice model is the multinomial Logit Model. The multinomial logit model calculates the proportion of travelers that will select a specific mode $K$ according to the following relationship:

$$p(K) = \frac{e^{U_k}}{\sum_x e^{U_x}}$$

$U_k$ is a utility function, which measures the degree of satisfaction that people derive from their choices. The magnitude of $U_k$ depends on the characteristics of each choice, which are called attributes, and on the characteristics of the individual making that choice, which are called socioeconomic status. The utility function is typically expressed as the linear weighted sum of the independent variables or their transformation as the follow:
\[ U_k = a_k + a_1 X_1 + a_2 X_2 + \ldots + a_r X_r \]  \hspace{1cm} (7)

\( a_k \) is the calibrated mode-specific constant for the mode \( K \) and \( X_i \) is the \( i \)th attribute weighted by the model parameter \( a_i \). The actual attributes may include such variables as the access and egress time, waiting time, line-haul time (actual time spent in the vehicle), out-of-pocket cost, level of service and convenience.

**Example: Modal Choice in South Texas**

In the same town as used in PART 2, people who go from Zone 1 to Zone 4 have two choices of transportation: private automobile and local bus, which is operated by a local public transportation authority. A calibration process has resulted in the following utility function for the auto and bus:

\[ U_k = a_k - 0.05X_1 - 0.02X_2 - 0.025X_3 - 0.001X_4 \]

where

- \( X_1 \) = waiting time, in minutes
- \( X_2 \) = line-haul time, in minutes
- \( X_3 \) = access time, in minutes
- \( X_4 \) = out-of-pocket cost, in cents

The interchange trips between Zone 1 and Zone 4 for the target-year 2005 has been forecasted in the PART 2. The target-year service attributes of the two competing modes have been estimated as Table 4:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automobile</td>
<td>0</td>
<td>15</td>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td>Local Bus</td>
<td>15</td>
<td>35</td>
<td>10</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 4: Target-year service attributes for automobile and local bus

**Requirement 6:** Assuming that the calibrated mode-specific constants \( a_k \) are –0.14 for the automobile and –0.60 for the bus, use the logit model Equation (6) to estimate the target-year market shares of the automobile and bus and the resulting fare revenue of the bus system.

**Requirement 7:** A proposal is submitted to build a subway transit system (ST) between Zone 1 and Zone 4 for this town. A study has projected that the service attributes of the proposed subway system will be:

\[ X_1 \text{ (ST)} = 5 \quad X_2 \text{ (ST)} = 30 \quad X_3 \text{ (ST)} = 10 \quad X_4 \text{ (ST)} = 100 \]
Based on professional experience, the mode-specific constant for the subway system is -0.45. Estimate the new market shares of three modes that will result from implementing the subway system proposal and the effect on the revenues of the public transportation authority, which operates the bus and subway systems.

Requirement 8: What other factors need to be considered to effectively decide if a subway should be built? Explain.

PART 4: Step 4 - Trip Assignment

General Information
The last step in the sequential travel demand forecasting process is concerned with the trip-makers’ choice of path between pairs of zones by travel mode and with the resulting vehicular flows on the multimodal transportation network. The trip assignment problem can be described as follows: given traffic demand departure rates for each Origin Destination (OD) pair in the network, which are usually derived from the Steps 1-3 of the travel demand forecasting process, determine the likely path choices and path use probabilities for each OD pair in the network, and predict the resulting traffic flows and link travel times on each of the individual links that make up the network.

Most of the methods for solving the trip assignment problem are based on two trip assignment principles first presented by Wardrop. These principles are:

Wardrop’s 1st principle: each driver may strive to choose the path that will minimize his own travel time through the network.

Wardrop’s 2nd principle: each driver may choose the path that will minimize the total network travel time of all drivers.

When the network trips are such assigned that no driver can reduce his own travel time by switching from his current route to an alternative, the traffic flows are said to be in a user-equilibrium state. This state coincides with Wardrop’s 1st principle. Similarly, when the network trips are such assigned that no driver can reduce the total network travel time by switching from his current route to an alternative, the traffic flows are said to be in a system equilibrium state. This state coincides with Wardrop’s 2nd principle.

The trip assignment problems can be formulated as non-linear mathematical programs based on the above two Wardrop’s assignment principles. The
following will use a simple two-node and two-link network to describe the trip assignment models based on the Wardrop’s principles.

It is assumed that there are only two nodes and two links in the network as shown below. Node 1 is an origin where the trips generate and Node 2 is a destination where the trips sink. There are two routes available for driving from Node 1 to Node 2 through either Link 1 or Link 2. The link travel time on either Link 1 or Link 2 is a function of the traffic flow on the link. In other words, link travel time will vary with the change of the number of vehicles on the link. If it is assumed that \( r_{12} \) vehicles per hour go from Node 1 to Node 2, the objective of the trip assignment process is to determine how much traffic of the \( r_{12} \) will use Link 1 and how much will use Link 2.

The link travel time functions are expressed as follows:

\[
\begin{align*}
tt_1 &= f(X_1) \\
tt_2 &= f(X_2)
\end{align*}
\]

where \( X_1 \) and \( X_2 \) are the total number of vehicles per hour on Link 1 and Link 2, and \( tt_1 \) and \( tt_2 \) are the travel times on Link 1 and Link 2. To perform the trip assignment for this simple network, two formulations can be made based on the Wardrop’s two principles.

**Based on Wardrop’s 1st Assignment Principle** - Based on the Wardrop’s 1st assignment principle, the travel times on Link 1 and Link 2 must be equal after all trips are assigned. Otherwise, some vehicles on the higher travel time link will automatically switch to the lower travel time link, as each individual driver seeks
to minimize his/her own travel times. Based on this logic and the assumption that there are a total of \( r_{12} \) vehicles per hour going from Node 1 to Node 2, the following equations hold:

\[
f(X_1) = f(X_2) \quad (10)
\]

\[
X_1 + X_2 = r_{12} \quad (11)
\]

Equations (10) and (11) contain only two variables. Therefore, \( X_1 \) and \( X_2 \) can often be solved for using these two simultaneous equations if \( f(X_1) \) and \( f(X_2) \) are linear equations. Accordingly, the link travel times can also be calculated using Equation (8) or (9).

**Based on Wardrop’s 2nd Assignment Principle** - Based on the Wardrop’s 2nd assignment principle, the total travel times of Link 1 and Link 2 should be minimum. Otherwise, some vehicles will switch the link in order to reduce the total travel times. The total travel times for Link 1 and Link 2 are expressed as follows:

\[
TT = X_1, f(X_1) + X_2, f(X_2) \quad (12)
\]

Therefore, the following optimization formulation can be derived:

\[
\text{min } X_1, f(X_1) + X_2, f(X_2) \quad \text{subject to } X_1 + X_2 = r_{12} \quad (13)
\]

The solution of Equation (13) will result in the values of \( X_1 \) and \( X_2 \). Thereafter, the individual Link travel times as well as the total link travel times can be calculated based on Equations (8), (9) and (12).

**Example: Route Choice in South Texas**

In the same example network as the assignment in PART 2, it is assumed that the automobile users driving from Zone 1 to Zone 4 have two available routes to use, similar to what has been illustrated in Figure 2. During the morning peak hour, the travel times using the two routes are functions of the vehicles using each route, as indicated by Equations (8) and (9). It is assumed that travel time functions for Route 1 and Route 2 are expressed by the following equations:

\[
\begin{align*}
\text{tt}_1 &= 10 + 0.02X_1 \text{ minutes} \\
\text{tt}_2 &= 15 + 0.01X_2 \text{ minutes}
\end{align*}
\]
where $tt_1$ and $tt_2$ are the actual travel times on Route 1 and Route 2 and $X_1$ and $X_2$ are the number of vehicles using Route 1 and Route 2 per hour. It is estimated from the trip distribution and modal split process that during the morning peak hour, about 2000 vehicles per hour will leave Zone 1 heading for Zone 4.

**Requirement 9**: Based on the Wardrop’s first user equilibrium principle and Equations (10) and (11), formulate the trip assignment problem. Solve the formulation and estimate the number of vehicles that will use Route 1 and the number of vehicles that will use Route 2. Calculate the resulting travel times on each route and the total vehicle travel times between Zone 1 and Zone 4.

**Requirement 10**: Based on the Wardrop’s second system optimal trip assignment principle and Equation (13), formulate the trip assignment problem. Solve the formulation and estimate the number of vehicles that will use Route 1 and the number of vehicles that will use Route 2. Calculate the resulting travel times on each route and the total vehicle travel times between Zone 1 and Zone 4.

**Requirement 11**: Compare the results from Question 1 and Question 2. Check if these results are consistent with the Wardrop’s trip assignment principles. Comment on your findings.
2. Sample Solution

PART 1: Step 1 – Trip Generation

Trip Attractions in South Texas

Requirement 1: Using the Excel Tool – Data Analysis – Regression, the linear regressions for Y-X, Y-X₁, Y-X₂, and Y-X₃ result in the following output Table 5 to Table 8.

Requirement 2: Using the Excel Tool – Data Analysis – Regression, the linear regression for Y-X₁, X₂ and X₃ results in the following output Table 9 and Table 10.

Requirement 3: Assuming that the value of R-square higher than 0.65 is acceptable for the analysis, the following equations are selected for the trip attraction analysis:

\[
Y = 0.163903 + 1.001321X \quad R^2 = 0.998124
\]
\[
Y = 705.9511 + 1.266307X_1 \quad R^2 = 0.954894
\]
\[
Y = 40.22623 + 2.896011X_2 \quad R^2 = 0.69364
\]
\[
Y = -321931 + 0.969715X_1 + 1.105403X_2 + 0.858613X_3 \quad R^2 = 0.998761
\]

If F-test and t-test are conducted to further examine the validity of each equation, the following equations are found acceptable statistically:

\[
Y = 0.163903 + 1.001321X \quad R^2 = 0.998124
\]
\[
Y = 705.9511 + 1.266307X_1 \quad R^2 = 0.954894
\]
\[
Y = 119.7625 + 1.001821X_1 + 1.013371X_2 \quad R^2 = 0.99817
\]
SUMMARY OUTPUT

\textit{Regression Statistics}

\begin{itemize}
  \item Multiple R 0.999061
  \item R Square 0.998124
  \item Adjusted R Square 0.997498
  \item Standard Error 190.2617
  \item Observations 5
\end{itemize}

\textit{ANOVA}

\begin{tabular}{lrrrr}
\textbf{df} & \textbf{SS} & \textbf{MS} & \textbf{F} & \textbf{Significance F} \\
\hline
Regression & 1 & 57775542 & 57775542 & 1596.03 & 3.45E-05 \\
Residual & 3 & 108598.6 & 36199.53 & & \\
Total & 4 & 57884141 & & & \\
\end{tabular}

\begin{tabular}{lrrrrrr}
\textbf{Coefficients} & \textbf{Standard Error} & \textbf{t Stat} & \textbf{P-value} & \textbf{Lower 95\%} & \textbf{Upper 95\%} \\
Intercept & 0.163903 & 0.001401 & 0.99897 & -372.28 & 372.6074 \\
Total & 1.001321 & 39.95035 & 3.45E-05 & 0.921556 & 1.081087 \\
\end{tabular}

\textbf{Table 5:} Simple linear regression result for trip attraction and total employment

SUMMARY OUTPUT

\textit{Regression Statistics}

\begin{itemize}
  \item Multiple R 0.977187
  \item R Square 0.954894
  \item Adjusted R Square 0.939859
  \item Standard Error 932.9017
  \item Observations 5
\end{itemize}

\textit{ANOVA}

\begin{tabular}{lrrrr}
\textbf{df} & \textbf{SS} & \textbf{MS} & \textbf{F} & \textbf{Significance F} \\
\hline
Regression & 1 & 55273224 & 55273224 & 63.51013 & 0.004122 \\
Residual & 3 & 2610917 & 870305.6 & & \\
Total & 4 & 57884141 & & & \\
\end{tabular}

\begin{tabular}{lrrrrrr}
\textbf{Coefficients} & \textbf{Standard Error} & \textbf{t Stat} & \textbf{P-value} & \textbf{Lower 95\%} & \textbf{Upper 95\%} \\
Intercept & 705.9511 & 117.0305 & 0.001401 & -956.267 & 2368.169 \\
Manuf. & 1.266307 & 0.025064 & 39.95035 & 0.921556 & 1.081087 \\
\end{tabular}

\textbf{Table 6:} Simple linear regression result for trip attraction and manufacturing employment
SUMMARY OUTPUT

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R Square</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

ANOVA

<table>
<thead>
<tr>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>40150774</td>
<td>40150774</td>
<td>6.792412</td>
</tr>
<tr>
<td>Residual</td>
<td>3</td>
<td>17733367</td>
<td>5911122</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>57884141</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Lower 95.0%</th>
<th>Upper 95.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>40.22623</td>
<td>1631.45</td>
<td>0.024657</td>
<td>-5151.78</td>
<td>5232.234</td>
<td>-5151.78</td>
<td>5232.234</td>
</tr>
<tr>
<td>Ret&amp;Ser</td>
<td>2.896011</td>
<td>1.11119</td>
<td>2.606226</td>
<td>-0.64029</td>
<td>6.432316</td>
<td>-0.64029</td>
<td>6.432316</td>
</tr>
</tbody>
</table>

Table 7: Simple linear regression result for trip attraction and retail and service employment

SUMMARY OUTPUT

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R Square</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

ANOVA

<table>
<thead>
<tr>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>446360.4</td>
<td>446360.4</td>
<td>0.023314</td>
</tr>
<tr>
<td>Residual</td>
<td>3</td>
<td>57437780</td>
<td>19145927</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>57884141</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Lower 95.0%</th>
<th>Upper 95.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2889.568</td>
<td>2870.336</td>
<td>1.006701</td>
<td>-6245.13</td>
<td>12024.27</td>
<td>-6245.13</td>
<td>12024.27</td>
</tr>
<tr>
<td>Others</td>
<td>2.399937</td>
<td>15.71792</td>
<td>0.152688</td>
<td>-47.6216</td>
<td>52.42143</td>
<td>-47.6216</td>
<td>52.42143</td>
</tr>
</tbody>
</table>

Table 8: Simple linear regression result for trip attraction and others employment
**Table 9:** Multiple linear regression result for trip attraction and manufacturing, retail and service, and others employment

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R Square</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Lower 95.0%</th>
<th>Upper 95.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-32.1931</td>
<td>-0.1133</td>
<td>0.928177</td>
<td>-3642.53</td>
<td>3578.144</td>
<td>-3642.53</td>
<td>3578.144</td>
</tr>
<tr>
<td>Manuf.</td>
<td>0.969715</td>
<td>0.079029</td>
<td>12.27035</td>
<td>0.051768</td>
<td>1.97387</td>
<td>-0.03444</td>
<td>1.97387</td>
</tr>
<tr>
<td>Ret&amp;Ser</td>
<td>1.105403</td>
<td>0.217178</td>
<td>5.089861</td>
<td>0.123503</td>
<td>3.864894</td>
<td>-1.65409</td>
<td>3.864894</td>
</tr>
<tr>
<td>Others</td>
<td>0.858613</td>
<td>1.243212</td>
<td>0.690641</td>
<td>0.615216</td>
<td>16.65505</td>
<td>-14.9378</td>
<td>16.65505</td>
</tr>
</tbody>
</table>

**Table 10:** Multiple linear regression result for trip attraction and manufacturing and retail and service

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R Square</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Lower 95.0%</th>
<th>Upper 95.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>119.7625</td>
<td>0.775115</td>
<td>0.519369</td>
<td>-545.0381</td>
<td>784.563</td>
<td>-545.0381</td>
<td>784.563</td>
</tr>
<tr>
<td>Manuf.</td>
<td>1.001821</td>
<td>0.054921</td>
<td>18.24112</td>
<td>0.002992</td>
<td>1.238128</td>
<td>0.765515</td>
<td>1.238128</td>
</tr>
<tr>
<td>Ret&amp;Ser</td>
<td>1.013371</td>
<td>0.147178</td>
<td>6.876347</td>
<td>0.020501</td>
<td>1.647455</td>
<td>0.379286</td>
<td>1.647455</td>
</tr>
</tbody>
</table>
PART 2: Step 2 – Trip Distribution
Trip Distributions in South Texas

Requirement 4:
For the origin \( i = 1 \), the trip production \( P_i = 20000 \). Based on the Equation (5) and Table 2, the friction factor \( F_{ij} \) can be calculated as follows:

\[
\begin{align*}
F_{11} &= \frac{1}{W_{11}^2} = \frac{1}{5^2} = 0.04 \\
F_{12} &= \frac{1}{W_{12}^2} = \frac{1}{10^2} = 0.01 \\
F_{13} &= \frac{1}{W_{13}^2} = \frac{1}{10^2} = 0.01 \\
F_{14} &= \frac{1}{W_{14}^2} = \frac{1}{20^2} = 0.0025
\end{align*}
\]

The interchange trips between Zone 1 and other zones can then be calculated based on the Equation (5). To simplify the calculation process, the following table is used to perform the calculation.

<table>
<thead>
<tr>
<th>( j )</th>
<th>( A_j )</th>
<th>( F_{ij} )</th>
<th>( A_j F_{ij} )</th>
<th>( P_{ij} )</th>
<th>( T_{ij} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.01</td>
<td>0.03</td>
<td>0.462</td>
<td>9240</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.01</td>
<td>0.02</td>
<td>0.308</td>
<td>6160</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0.0025</td>
<td>0.015</td>
<td>0.230</td>
<td>4600</td>
</tr>
</tbody>
</table>

\[ \Sigma = 0.065 \quad 1.000 \]

For the origin \( i = 3 \), the trip production \( P_i = 15000 \). A similar calculation table to the above can be used to calculate the interchange trips between Zone 3 and other zones.

<table>
<thead>
<tr>
<th>( j )</th>
<th>( A_j )</th>
<th>( F_{3j} )</th>
<th>( A_j F_{3j} )</th>
<th>( P_{3j} )</th>
<th>( T_{3j} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.01</td>
<td>0.03</td>
<td>0.176</td>
<td>2640</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.04</td>
<td>0.08</td>
<td>0.471</td>
<td>7065</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0.01</td>
<td>0.06</td>
<td>0.353</td>
<td>5295</td>
</tr>
</tbody>
</table>

\[ \Sigma = 0.17 \quad 1.000 \]

For Origins 2 and 4, since there are no trip productions, the interchange trips between these two zones and other zones become zero. Therefore, the final interchange trips can be summarized into the following table:

<table>
<thead>
<tr>
<th>( i ) ( \setminus ) ( j )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>9240</td>
<td>6160</td>
<td>4600</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2640</td>
<td>7065</td>
<td>5295</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Requirement 5: If a row in the above interchange trip table is summed, it should be equal to the total trip productions for that zone. On the other hand, if a column in the table is summed, the total trip attractions for that can be derived. Therefore, the total trip attractions for Zones 2, 3 and 4 are calculated as follows:

\[ A_2^* = 9240 + 2640 = 11880 \]
\[ A_3^* = 6160 + 7065 = 13225 \]
\[ A_4^* = 4600 + 5295 = 9895 \]

PART 3: Step 3 – Modal Choice

Modal Choice in South Texas

Requirement 6: Based on the trip distribution analysis in PART 2, the interchange trips between Zone 1 and Zone 4 are 4600. Given the utility function and the attribute values by Table 3, the utilities for automobile and bus can be calculated as

\[ U(Auto) = -0.14 - 0.05(0) - 0.02(15) - 0.025(5) - 0.001(80) = -1.095 \]
\[ U(Bus) = -0.60 - 0.05(15) - 0.02(35) - 0.025(10) - 0.001(50) = -2.35 \]

Therefore, the proportions of trips that will select the automobile and bus can be estimated as follows using the Equation (6):

\[ p(Auto) = \frac{e^{-1.095}}{e^{-1.095} + e^{-2.35}} = 78\% \]
\[ p(Bus) = \frac{e^{-2.35}}{e^{-1.095} + e^{-2.35}} = 22\% \]

The market shares of the automobile and bus are calculated as

\[ T_{14}(Auto) = (0.78)(4600) = 3588 \text{ trips/day} \]
\[ T_{14}(Bus) = (0.22)(4600) = 1012 \text{ trips/day} \]

The fare revenue of the bus system = (1012 trips/day)($0.50/trip) = $506/day

Requirement 7: By introducing the subway system, the utilities for automobile and bus will keep the same as those calculated in the above Question 1, which are \( U(Auto) = - \)
1.095 and $U(Bus) = -2.35$. The utility of the proposed subway system is calculated as follows:

$$U(\text{Subway}) = -0.45 - 0.05(5) - 0.02(30) - 0.025(10) - 0.001(100) = -1.65$$

Therefore, the proportions of trips that will use each mode are estimated as:

$$p(\text{Auto}) = \frac{e^{-1.095}}{e^{-1.095} + e^{-2.35} + e^{-1.65}} = 54\%$$

$$p(\text{Bus}) = \frac{e^{-2.35}}{e^{-1.095} + e^{-2.35} + e^{-1.65}} = 15\%$$

$$p(\text{Subway}) = \frac{e^{-1.65}}{e^{-1.095} + e^{-2.35} + e^{-1.65}} = 31\%$$

The new market shares after the introduction of the subway system are calculated as follows:

$$T_{14}(\text{Auto}) = (0.54)(4600) = 2484 \text{ trips/day}$$

$$T_{14}(\text{Bus}) = (0.15)(4600) = 690 \text{ trips/day}$$

$$T_{14}(\text{Subway}) = (0.31)(4600) = 1426 \text{ trips/day}$$

The new fare revenue of both bus and subway systems = (690 bus trips/day)($0.50/bus trip) + (1426 subway trips/day)($1.00/subway trip) = $1,771/day. The revenue increase of the public transportation authority, which operates both the local bus and subway systems, due to the introduction of the subway system is $1,265 ($1,771 - $506). It should be noted that the increase of the fare revenue only does not justify the construction of the subway. In reality, a comprehensive cost-benefit analysis should be conducted before the decision is made.

**Requirement 8:** Free thoughts.

**PART 4: Step 4 – Trip Assignment**

**Route Choice in South Texas**

**Question 1:** Based on the Wardrop’s first assignment principle, Equations (10) and (11), and the travel time functions for Route 1 and Route 2, the following linear simultaneous equations can be established:
\[
10 + 0.02X_1 = 15 + 0.01X_2 \\
X_1 + X_2 = 2000
\]

The above equations can be solved as \(X_1 = 833.33\) and \(X_2 = 1166.67\), which means that 833.33 vehicles will actually use Route 1 and 1166.67 vehicles will use Route 2. The travel times are

- Travel time on Route 1 = \(10 + 0.02(833.33) = 26.67\) minutes
- Travel time on Route 2 = \(15 + 0.01(1166.67) = 26.67\) minutes
- Total travel times = \((833.33)(26.67) + (1166.67)(26.67) = 53,340\) vehicle - minutes

Question 2: Based on the Wardrop’s second trip assignment principle, Equation (13), and the travel time functions, the following optimization formulation can be established:

\[
\begin{align*}
\text{min} & \quad TT = X_1(10 + 0.02X_1) + X_2(15 + 0.01X_2) \\
\text{subject to} & \quad X_1 + X_2 = 2000
\end{align*}
\]

where \(TT\) represents the total system travel times for all vehicles going from Zone 1 to Zone 4. The constraint can be re-written as \(X_2 = 2000 - X_1\). Substituting this into the objective function results in

\[
TT = X_1(10 + 0.02X_1) + (2000 - X_1)(15 + 0.01(2000 - X_1)) \\
= 0.03X_1^2 - 55X_1 + 70000
\]

By setting the derivative of the above equation to zero, the optimal value for \(X_1\) can be solved.

\[
\begin{align*}
\frac{dT}{dX_1} = (0.03)(2)X_1 - 45 &= 0 \\
X_1 &= 750 \\
\text{Therefore} \quad X_2 &= 2000 - 750 = 1250
\end{align*}
\]

Therefore, based on the Wardrop’s second assignment principle, 750 vehicles will use Route 1 and 1250 vehicles will use Route 2. The travel times are calculated as follows:
Travel time on Route 1 = 10 + 0.02(750) = 25 minutes
Travel time on Route 2 = 15 + 0.01(1250) = 27.5 minutes
Total travel times = (750)(25) + (1250)(27.5) = 53,125 vehicle - minutes

Question 3: Based on Wardrop’s first assignment principle, all the used routes should have equal travel times. In Question 1, both Routes 1 and 2 have travel time of 26.67 minutes, which indicates that this result conforms to the Wardrop’s first assignment principle. In Question 2, two routes have different travel times, which means that the result does not conform to the Wardrop’s first assignment principle.

Based on the Wardrop’s second assignment principle, the total travel times should be minimum. Question 2 resulted in the total travel times of 53,125 vehicle-minutes, which are less than the total travel times of 53,340 in Question 1. This means that although Wardrop’s second assignment principle does not result in equal travel times for all routes, it does result in minimum total travel times.
3. Notes for the Instructor

This project is designed to exercise some basic concepts in regression analysis, algebra, calculus, and optimization theory using a realistic example of transportation planning process. Understanding the travel demand forecasting process is critical in this project, while the mathematical calculation part for each assignment is not very difficult. The project is suitable for either individual or group work. With the understanding of the travel demand forecasting process concept by the students, the instructor can expand the assignment in each part to let students practice. The assignment in each part can also be used independently for exercise of mathematical topics. The requirements for PART 1 through PART 4 can also be used for a complete project for students. Some additional suggestions are as follows:

1. In PART 1, similar problem can be also designed for trip productions in which the zonal population, household income and the household automobile ownership become the independent variables while the trip production becomes the dependent variable. If a single independent variable is used, students can be asked to do the simple linear regression analysis manually. Non-linear regression equations can be considered to let students practice how to transform a non-linear equation to a linear equation.

2. In PART 2, variations of the requirement can be designed. More zones can be added to the analysis areas. Residential zones and commercial zones can be redesigned. The attractiveness can be assigned different values to test how the changes will affect the trip distribution results. The travel time table can use different values for each cell. The exponential power $c$ in the gravity equation can be assigned a different value other than 2 to examine how different values of $c$ will change people’s choices of commercial zones.

3. In PART 3, students can be asked to change the attribute values for automobile and local bus, for example the out-of-pocket cost or waiting time, to examine how the changes will affect the choice of mode. The coefficients of attributes in the utility function can also be redesigned so that a separate problem can be created. In addition to the introduction of the subway transit system, more modes of travel can be considered, such as motorcycle, walking, ride-sharing and so on. If more modes are added to the problem, values of various attributes for each added mode should be assumed.

4. In PART 4, three-route and four-route problems can be designed, which will make the mathematical equations more complicated. In designing additional route(s) to
the network, new travel time function(s) should be assumed. Essentially, different forms of the travel time functions reflect different conditions of the roads. For example, some roads have one lane and some roads have two lanes in each direction. Some roads may have a lot of parked vehicles on the roadside, which significantly impede the moving of the vehicles, while some roads may have uphill and downhill slopes or short radius on curves. The travel time functions can also be made non-linear, which will make the optimization more complicated.