MATEMATIKA, 2005, Jilid 21, Bil. 2, hlm. 101–112
©Jabatan Matematik, UTM.

Comparing Least-Squares and Goal Programming Estimates of Linear Regression Parameters

[1]**Maizah Hura Ahmad,** [1]**Robiah Adnan,** [1]**Lau Chik Kong &** [2]**Zalina Mohd Daud**
[1]Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia
4130 UTM, Skudai, Johor, Malaysia
[2]ATMA, Kuala Lumpur, Malaysia

Abstract A regression model is a mathematical equation that describes the relationship between two or more variables. In regression analysis, the basic idea is to use past data to fit a prediction equation that relates a dependent variable to independent variable(s). This prediction equation is then used to estimate future values of the dependent variable. The least-squares method is the most frequently used procedure for estimating the regression model parameters. However, the method of least-squares is biased when outliers exist. This paper proposes goal programming as a method to estimate regression model parameters when outliers must be included in the analysis.

Keywords Method of least squares, outliers, goal programming.

1 Introduction

Predicting future values of a variable is a crucial management activity. The statistical method most widely used in making predictions is regression analysis. In the regression approach past data on the relevant variables are used to develop and evaluate a prediction equation. For prediction to make much sense, there must be some connection between the variable one is predicting (the dependent variable) and the variables one is using to make the prediction (the independent variable). Prediction requires a unit of association. This means that there should be an entity that relates the two variables. With time-series data, the unit of association may simply be time. For cross-sectional data, an economic or physical entity should connect the variables.

A regression model is a mathematical equation that describes the relationship between two or more variables. In linear regression analysis, the basic idea is to use data to fit a prediction equation that relates a dependent variable y to independent variable(s) x, with an assumption that the relation is, in fact, linear. In simple linear regression analysis, the estimated model is $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$ ($\widehat{y}$ denotes the estimate value of y for any value of x). In the first-order multiple regression, the estimated model is $\widehat{y} = \widehat{\beta}_0 + \sum_{i=1}^{n} \widehat{\beta}_i x_i$.

The least-squares method is the most frequently used procedure for estimating the regression coefficients, β's. Although this method is one of the best known and probably most widely used, it is sensitive to outliers [1]. If an outlier must be kept in the data, a method other than least squares should be used [2].

The current study attempts to use linear goal programming to estimate linear regression coefficients when outlier(s) must be included in the estimation of the model parameters. In this paper, the methods (least-squares and goal programming) will be described first. These methods will be used to estimate regression model parameters. The resulting prediction equations will then be used to predict future values of y.

2 Least-Squares Method

The least squares method is a computational technique for determining the 'best' equation describing a set of points where best is defined geometrically [3]. In the study of the relationship between two variables, the polynomial, $p(x)$ that can describe n data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where x is the independent variable and y is the dependent can be written as

$$p(x) = \sum_{k=0}^{m} \beta_k x^k$$

where $\beta_0, \beta_1, \ldots, \beta_m$ are to be determined. The least squares method will choose as 'solutions' those β_k's that minimize the sum of squares of the vertical distances from the data points to the presumed polynomial. Let the residual from each data point be denoted by e, i.e., $e_i = y_i p(x_i)$. The 'best' polynomial $p(x)$ is the one whose coefficients minimize the function L, where

$$L = e_1^2 + e_2^2 + \cdots + e_n^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - p(x_i)]^2.$$

If $p(x)$ is a linear polynomial, the least-squares estimates of the regression coefficients are the values $\widehat{\beta}_0$ and $\widehat{\beta}_1$ obtained by minimizing

$$L = \sum_{i=1}^{n} [y_i - p(x_i)]^2 = \sum_{i=1}^{n} [y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i)]^2. \tag{1}$$

By differentiating (1) partially with respect to $\widehat{\beta}_0$ and $\widehat{\beta}_1$; equating the partial derivatives to zero; and then solving the system of equations by using determinants or the method of elimination [3], it can be shown that the estimated slope is

$$\widehat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{2}$$

and the estimated y-intercept is

$$\widehat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i - \widehat{\beta}_1 \sum_{i=1}^{n} x_i}{n} = \overline{y} - \widehat{\beta}_1 \bar{x}. \tag{3}$$

In the study of the relationship between one dependent variable y and two independent variables x_1 and x_2, the model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ for a given data set

$$(y_1, x_{11}, x_{21}), (y_2, x_{12}, x_{22}), \ldots, (y_n, x_{1n}, x_{2n}), \text{ where } n \geq 3.$$

The best fitting curve $p(x)$ has the least squares error

$$L = \sum_{i=1}^{n} [y_i - p(x_{1i}, x_{2i})]^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})]^2.$$

Solving the following,

$$\begin{aligned}
\frac{\partial L}{\partial \widehat{\beta}_0} &= (-2) \textstyle\sum_{i=1}^{n} [y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_{1i} + \widehat{\beta}_2 x_{2i})] = 0, \\
\frac{\partial L}{\partial \widehat{\beta}_1} &= (-2) \textstyle\sum_{i=1}^{n} x_{1i} [y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_{1i} + \widehat{\beta}_2 x_{2i})] = 0, \\
\frac{\partial L}{\partial \widehat{\beta}_2} &= (-2) \textstyle\sum_{i=1}^{n} x_{1i} [y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_{1i} + \widehat{\beta}_2 x_{2i})] = 0,
\end{aligned}$$

and expanding the above equations, we have

$$\sum_{i=1}^{n} y_i = \widehat{\beta}_0 \sum_{i=1}^{n} 1 + \widehat{\beta}_1 \sum_{i=1}^{n} x_{1i} + \widehat{\beta}_2 \sum_{i=1}^{n} x_{2i} \tag{4}$$

$$\sum_{i=1}^{n} x_{1i} y_i = \widehat{\beta}_0 \sum_{i=1}^{n} x_{1i} + \widehat{\beta}_1 \sum_{i=1}^{n} x_{1i}^2 + \widehat{\beta}_2 \sum_{i=1}^{n} x_{1i} x_{2i} \tag{5}$$

$$\sum_{i=1}^{n} y_i x_{2i} = \widehat{\beta}_0 \sum_{i=1}^{n} x_{2i} + \widehat{\beta}_1 \sum_{i=1}^{n} x_{1i} x_{2i} + \widehat{\beta}_2 \sum_{i=1}^{n} x_{2i}^2 \tag{6}$$

These least squares estimates of the coefficients $\widehat{\beta}_0, \widehat{\beta}_1$ and $\widehat{\beta}_2$ can be obtained by solving (4), (5) and (6) simultaneously.

3 Linear Goal Programming (GP)

Linear GP problems are GP problems where each objective function is linear. GP was first developed and introduced by A. Charnes and W.W. Cooper in 1961 and further refined by Y. Ijiri in 1965. According to Charnes and Cooper [4], GP extends the linear programming formulation to accommodate mathematical programming with multiple objectives.

GP's objective function is always minimized and must be composed of deviational variables only. It minimizes the deviations of the compromise solution from target goals, weighted and prioritized.

In the formulation, two types of variables are used: decision variables and deviational variables. There are two categories of constraints: structural/system constraints and goal constraints, which are expressions of the original functions with target goals set a priori and positive and negative deviational variables.

The general GP model can be expressed as follows:

$$\left.\begin{array}{l}\text{Minimize } Z=\displaystyle\sum_{i=1}^{m}(d_i^- + d_i^+)\\ \text{Subject to the linear constraints:}\\ \text{Goal constraints: } \left(\displaystyle\sum_{i=1}^{n} a_{ij}x_j\right)+d_i^- - d_i^+ = b_i, i=1,2,\dots,m\\ \text{System constraints:} \displaystyle\sum_{i=1}^{n} a_{ij}x_j \begin{bmatrix}\le\\=\\\ge\end{bmatrix} b_i, i=m+1,\dots,m+p\\ \text{with } x_j, d_i^-, d_i^+ \ge 0,\ \text{for } i=1,2,\dots,m \text{ and } j=1,2,\dots,n.\end{array}\right\} \tag{7}$$

In (3.1) there are m goals, p system constraints and n decision variables, where

Z : objective function
a_{ij} : the coefficient associated with variable j in the ith goal
x_j : the jth decision variable
b_i : the associated right hand side value
d_i^- : negative deviational variable from the ith goal (underachievement)
d_i^+ : positive deviational variable from the ith goal (overachievement)
$d_i^+ \times d_i^- = 0,\ d_i^+, d_i^- \ge 0$

Before solving a GP problem, the goals need to be ranked. In priority GP, the objectives can be divided into different priority classes. Here it is assumed that no two goals have equal priority. The goals are given ordinal rankings and are called *preemptive priority factors.* These *preemptive priority factors* have the relationship

$$P_1 >>> P_2 >>> \cdots >>> P_k >>> P_{k+1}$$

where >>> means “very much greater than”. This priority ranking is absolute. Therefore, the P_1 goal is so much more important than the P_2 goal and P_2 goal will never be attempted until the P_1 goal is achieved to the greatest extent possible.

The priority relationship implies that multiplication by n, however large it may be, cannot make the lower-level goal as important as the higher goal (that is,$P_j > nP_{j+1}$). In formulating a GP model having prioritized goals, those preemptive priority factors are incorporated into the objective function as weights for the deviational variables.

Using equation (7), the preemptive GP model can be presented as:

$$\left.\begin{array}{l}\text{Minimize } Z=\displaystyle\sum_{i=1}^{m}P_k(d_i^- + d_i^+)\\ \text{Subject to the linear constraints:}\\ \text{Goal constraints: } \displaystyle\sum_{i=1}^{n} a_{ij}x_j+d_i^- - d_i^+ = b_i, i=1,2,\dots,m\\ \text{System constraints:} \displaystyle\sum_{i=1}^{n} a_{ij}x_j \begin{bmatrix}\le\\=\\\ge\end{bmatrix} b_i, i=m+1,\dots,m+p\\ \text{with } x_j, d_i^-, d_i^+ \ge 0,\ \text{for } i=1,2,\dots,m \text{ and } j=1,2,\dots,n.\end{array}\right\} \tag{8}$$

where there are m goals, p system constraints, k priority levels and n decision variables with P_k = the priority factor of the kth goal

GP problems can be solved using the graphical (involving only two decision variables) and the modified simplex method.

4 Data Sets

The data used in this study were taken from [3] and [5] (see Appendix I). [3] and [5] used the data to determine least-squares prediction equations and to predict values of y with given values of x's. There were outlier(s) in the data and they were retained in the analysis. Outliers are data points that lie apart from the rest of the other points. They are unusually small or large data values, presumed to come from a different distribution for the majority of the data set, and can have profound influence on statistical analysis that finally leads to erroneous conclusions [6].

In the current study, outliers were detected using the box plot technique (see Appendix II). The box of the plot was determined by locating the median, the lower (Q_1) and upper quartiles (Q_3). A box was drawn around the median with the lower and upper quartiles as the box endpoints. The interquartile range (iqr), given by

$$iqr = \text{upper quartile} - \text{lower quartile}$$

was then calculated. An observation was a mild outlier if it is more than $1.5iqr$ away from the closest end of the box (the closest quartile). An outlier is extreme if it is more than $3iqr$ from the closest end of the box.

5 Estimating Model Parameters Using the Least-Squares Method

Only the first 20 observations were used to determine the least-squares prediction equations.

Data Set 1

$$\sum_{i=1}^{20} x_i = 2731.8, \quad \sum_{i=1}^{20} y_i = 2654, \quad \sum_{i=1}^{20} x_i^2 = 484531.16, \quad \sum_{i=1}^{20} x_i y_i = 444011.2.$$

Using equations (2) and (3), we have

$$\widehat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{20(444011.2) - 2731.8(2654)}{20(484531.16) - (2731.8)^2} = 0.7316.$$

and

$$\widehat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n} = \frac{2654 - 0.7316(2731.8)}{20} = 32.7708.$$

Thus, the least-squares prediction equation is

$$\widehat{y}_i = 32.765 + 0.7316x_i. \tag{9}$$

Data Set 2

$$\sum_{i=1}^{20} x_{1i} = 193.7, \quad \sum_{i=1}^{20} x_{1i}^2 = 2481.57, \quad \sum_{i=1}^{20} x_{2i} = 194, \quad \sum_{i=1}^{20} x_{2i}^2 = 2206,$$
$$\sum_{i=1}^{20} y_i = 665, \quad \sum_{i=1}^{20} x_{1i}y_i = 7127.2, \quad \sum_{i=1}^{20} x_{1i}x_{2i} = 2111.5, \quad \sum_{i=1}^{20} x_{2i}y_i = 6959.$$

Using equations (4), (5) and (6) we have

$$\widehat{\beta}_0 n + \widehat{\beta}_1 \sum_{i=1}^{n} x_{1i} + \widehat{\beta}_2 \sum_{i=1}^{n} x_{2i} = \sum_{i=1}^{n} y_i, \text{ implying}$$
$$20\widehat{\beta}_0 + 193.7\widehat{\beta}_1 + 194\widehat{\beta}_2 = 665; \tag{10}$$
$$\widehat{\beta}_0 \sum_{i=1}^{n} x_{1i} + \widehat{\beta}_1 \sum_{i=1}^{n} x_{1i}^2 + \widehat{\beta}_2 \sum_{i=1}^{n} x_{1i}x_{2i} = \sum_{i=1}^{n} x_{1i}y_i, \text{ implying}$$
$$93.7\widehat{\beta}_0 + 2481.57\widehat{\beta}_1 + 2111.5\widehat{\beta}_2 = 7127.2; \tag{11}$$
$$\widehat{\beta}_0 \sum_{i=1}^{n} x_{2i} + \widehat{\beta}_1 \sum_{i=1}^{n} x_{1i}x_{2i} + \widehat{\beta}_2 \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} x_{2i}y_i, \text{ implying}$$
$$194\widehat{\beta}_0 + 2111.5\widehat{\beta}_1 + 2206\widehat{\beta}_2 = 6959. \tag{12}$$

Solving equations (10), (11) and (12) simultaneously, we obtain $\widehat{\beta}_0 = 22.2$, $\widehat{\beta}_1 = 1.1$ and $\widehat{\beta}_2 = 0.0167$. Thus, the least-squares prediction equation is

$$\widehat{y}_i = 22.2 + 1.1x_{1i} + 0.0167x_{2i}. \tag{13}$$

6 Converting a Least-Squares Problem Into a GP Problem

From the least-squares method,

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i \tag{14}$$

and the residual is

$$y_i - \widehat{y}_i = e_i \quad \text{or} \quad \widehat{y}_i = y_i - e_i \tag{15}$$

Therefore equation (14) equals to equation (15). Thus,

$$y_i - e_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i \text{ implying } y_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i + e_i. \tag{16}$$

The general GP model is as follows:

$$\begin{aligned}
&\text{Minimize } Z = \sum_{i=1}^{m}(d_i^+ + d_i^-)\\
&\text{Subject to the linear constraints:}\\
&\text{Goal constraints: } \Big(\sum_{i=1}^{n} a_{ij}x_j\Big) + d_i^- - d_i^+ = b_i, i = 1, 2, \ldots, m \qquad (17)\\
&\text{System constraints:} \sum_{i=1}^{n} a_{ij}x_j \begin{bmatrix} \le \\ = \\ \ge \end{bmatrix} b_i, i = m+1, \ldots, m+p\\
&\text{with } x_j, d_i^-, d_i^+ \ge 0, \text{ for } i = 1, 2, \ldots, m \text{ and } j = 1, 2, \ldots, n.
\end{aligned}$$

From (16) and (17), let $e_i = d_i^- - d_i^+, y_i = b$. In GP, $\widehat{\beta}_{i-1}$ is treated as the decision variable. Thus, the value of x_i in GP equals the value of $\widehat{\beta}_{i-1}$.

7 Estimating Model Parameters Using GP

Formulation of GP models were based on the first 20 observations in the data set. In this study, the GP solutions were obtained by using Windows-based software package known as QM (Quantitative Methods) for Windows.

Data Set 1

The GP model for data set 1 is as follows:

$$\begin{aligned}
\text{Minimize}\quad & Z = P_1\sum_{i=1}^{20}(d_i^+ + d_i^-)\\
\text{Subject to}\quad & 164.2x_1 + x_2 + d_1^- - d_1^+ = 181\\
& 156.9x_1 + x_2 + d_2^- - d_2^+ = 156\\
& 109.8x_1 + x_2 + d_3^- - d_3^+ = 115\\
& 111.4x_1 + x_2 + d_4^- - d_4^+ = 132\\
& 87x_1 + x_2 + d_5^- - d_5^+ = 96\\
& 161.8x_1 + x_2 + d_6^- - d_6^+ = 170\\
& 230.9x_1 + x_2 + d_7^- - d_7^+ = 193\\
& 106.5x_1 + x_2 + d_8^- - d_8^+ = 110\\
& 97.6x_1 + x_2 + d_9^- - d_9^+ = 94\\
& 79.7x_1 + x_2 + d_{10}^- - d_{10}^+ = 77\\
& 118.7x_1 + x_2 + d_{11}^- - d_{11}^+ = 106\\
& 248.8x_1 + x_2 + d_{12}^- - d_{12}^+ = 204\\
& 102.4x_1 + x_2 + d_{13}^- - d_{13}^+ = 98\\
& 64.2x_1 + x_2 + d_{14}^- - d_{14}^+ = 76\\
& 89.4x_1 + x_2 + d_{15}^- - d_{15}^+ = 89
\end{aligned}$$

$$\begin{aligned}
& 78.9x_1 + x_2 + d_{16}^- - d_{16}^+ = 86 \\
& 387.8x_1 + x_2 + d_{17}^- - d_{17}^+ = 310 \\
& 135x_1 + x_2 + d_{18}^- - d_{18}^+ = 141 \\
& 82.9x_1 + x_2 + d_{19}^- - d_{19}^+ = 90 \\
& 117.9x_1 + x_2 + d_{20}^- - d_{20}^+ = 130 \\
\text{with} \quad & x_i, d_i^-, d_i^+ \geq 0, \ i = 1, 2, \ldots, 20.
\end{aligned}$$

There is only one goal, P_1 for data set 1. The goal is to predict elemental carbon based on carbon aerosols. The GP results are $x_1 = 0.7215$ and $x_2 = 30.1839$. Thus, the prediction equation is

$$\widehat{y}_i = 30.1839 + 0.7215x_i. \tag{18}$$

Data Set 2

The GP model for data set 2 is as follows:

$$\begin{aligned}
\text{Minimize} \quad & Z = P_1 \sum_{i=1}^{20} (d_i^+ + d_i^-) \\
\text{Subject to} \quad & 6.5x_1 + 3x_2 + x_3 + d_1^- - d_1^+ = 27 \\
& 6.5x_1 + 2x_2 + x_3 + d_2^- - d_2^+ = 29 \\
& 13.0x_1 + 15x_2 + x_3 + d_3^- - d_3^+ = 41 \\
& 8.1x_1 + 13x_2 + x_3 + d_4^- - d_4^+ = 36 \\
& 4.0x_1 + 6x_2 + d_5^- - d_5^+ = 22 \\
& 11.5x_1 + 13x_2 + x_3 + d_6^- - d_6^+ = 40 \\
& 18.0x_1 + 17x_2 + x_3 + d_7^- - d_7^+ = 52 \\
& 10.0x_1 + 12x_2 + x_3 + d_8^- - d_8^+ = 39 \\
& 7.1x_1 + 4x_2 + x_3 + d_9^- - d_9^+ = 27 \\
& 6.5x_1 + 10x_2 + x_3 + d_{10}^- - d_{10}^+ = 28 \\
& 7.0x_1 + 5x_2 + x_3 + d_{11}^- - d_{11}^+ = 24 \\
& 7.3x_1 + 11x_2 + x_3 + d_{12}^- - d_{12}^+ = 29 \\
& 7.5x_1 + 12x_2 + x_3 + d_{13}^- - d_{13}^+ = 33 \\
& 7.5x_1 + 12x_2 + x_3 + d_{14}^- - d_{14}^+ = 35 \\
& 4.9x_1 + 9x_2 + x_3 + d_{15}^- - d_{15}^+ = 27 \\
& 3.7x_1 + 6x_2 + x_3 + d_{16}^- - d_{16}^+ = 19 \\
& 9.1x_1 + 12x_2 + x_3 + d_{17}^- - d_{17}^+ = 36 \\
& 23.0x_1 + 13x_2 + x_3 + d_{18}^- - d_{18}^+ = 43 \\
& 23.5x_1 + 10x_2 + x_3 + d_{19}^- - d_{19}^+ = 40 \\
& 9.0x_1 + 9x_2 + x_3 + d_{20}^- - d_{20}^+ = 38 \\
\text{with} \quad & x_i, d_i^-, d_i^+ \geq 0, \ i = 1, 2, \ldots, 20.
\end{aligned}$$

There is also only one goal, P_1 for data set 2. The goal is to predict the seminar enrollment. The GP results are $x_1 = 0.5055, x_2 = 1.2615$ and $x_3 = 15.5055$.

From these results, we can write the predicted equation as

$$\widehat{y}_i = 15.5055 + 0.5055x_{1i} + 1.2615x_{2i}. \tag{19}$$

8 Comparison Between the Least-Squares Prediction Equation and Goal Programming Prediction Equation

Equations (9), (13), (18) and (19) were used to predict the last five observations in each data set. To compare the prediction values obtained using least-squares and goal programming, mean absolute percentage errors (MAPE) were calculated.

When choosing between competing models or when evaluating an existing model, measures that summarize the overall accuracy provided by the model(s) should be used [7]. Generally, the closer the predicted $\widehat{y}_i$ are to the actual y_i of the series, the more accurate the model is. Thus, the quality of a model can be evaluated by examining the series of prediction errors $(y_i - \widehat{y}_i)$. Since MAPE is measured as a percentage, it is particularly useful for comparing the performance of a model in different units.

The formula for MAPE is as follows:

$$\mathrm{MAPE} = \frac{\sum_{i=1}^{n}\left[\frac{|e_i|}{y_i}(100)\right]}{n} \tag{20}$$

where n is number of predictions. A large value of MAPE means that the value of the error is large. Tables 1 and 2 show the calculated MAPEs for the data sets.

9 Concluding Remarks

In both data sets, MAPE values using GP prediction equations were lower than those obtained using the least-squares prediction equations. Based on these results, it can be concluded that in the presence of outliers, the prediction equations obtained using the GP approach were more accurate than those obtained using the method of least-squares. This is because, using the GP approach, the problem can be restated as to minimize the sum of absolute residuals $|e_i|$ rather than the sum of the squares of the residuals e_i^2 as in the case of the least squares technique.

References

[1] C. G. Hugh and J.P. Ignizio, *Using Linear Programming for Predicting Student Performance,* Journal of Educational and Psychological Measurement, 32(2)(1972), 397-401.

[2] L.R. Ott and M. Longnecker, *An Introduction to Statistical Methods and Data Analysis, 5th. Edition,* Thomson Learning, Duxbury, USA, 2001.

[3] R.J. Larsen and M.L. Marx, *An Introduction to Mathematical Statistics and its Applications, 3rd. Edition,* Prentice Hall, Inc., New Jersey, 2001.

[4] A. Charnes and W.W. Cooper, *Management Models and Industrial Applications of Linear Programming, Vols. I and II,* John Wiley & Sons, New York, 1961.

[5] Ken. Black, *Business Statistics: Contemporary Decision Making, 3rd. Ed.,* South-Western College Publishing, 2001.

[6] N.C. Schwertman, M.A. Owens, and R. Adnan, *A Simple More General Boxplot Method for Identifying Outliers,* Computational Statistics & Data Analysis, 47(2004), 165-174.

[7] S. Makridakis, *Accuracy Measures: Theoretical and Practical Concerns,* International Journal of Forecasting, 9(1993), 527-529.

[8] T.C. Kalu, *An Algorithm for Systems Welfare Interactive Goal Programming Modelling,* European Journal of Operational Research, 116(1999), 508-529.

[9] Sang M. Lee and M.J. Schniederjans, *A Multicriterial Assignment Problem: A Goal Programming Approach,* Interfaces, 13(4)(1983), 75-79.

Appendix I

Data Set 1

Carbon aerosols have been identified as a contributing factor in a number of air quality problems. In this set, mass (x) is an independent variable while elemental carbon (y) is a dependent variable.

Data Set 2

This set examines the relationship between seminar enrollments (y), the number of mailings (x_1)[$\times$1 000], and the lead time of mailings (x_2)[weeks] of seminar announcements.

Appendix II

Detecting Outliers

Data Set 1

The values (independent variable) were arranged from the smallest value to the largest value or vice versa as follows:

x_i : 64.2, 76.4, 78.9, 79.7, 82.9, 87.0, 89.4, 89.4, 97.6, 100.8, 102.4, 106.5, 108.1, 109.8, 111.4, 117.9, 118.7, 131.7, 135.0, 156.9, 161.8, 164.2, 230.9, 248.8, 387.8

The quantities needed for constructing the modified box plot are as follows:
Median = 108.1, Lower quartile = 88.2, Upper quartile = 145.95

$$\begin{aligned} iqr &= \text{upper quartilelower quartile} \\ &= 145.9588.2 = 57.75 \\ 1.5 \cdot iqr &= 1.5 \cdot 57.75 = 86.625 \\ 3 \cdot iqr &= 3 \cdot 57.75 = 173.25 \end{aligned}$$

Thus,

Upper edge of box (upper quartile) $+ 1.5 \cdot iqr = 145.95 + 86.625 = 232.575$
Lower edge of box (lower quartile) $- 1.5 \cdot iqr = 88.286.625 = 1.575$

So 248.8 and 387.8 are both outliers at the upper end, and there are no outliers at the lower end.

Since, Upper edge of box $+3 \cdot iqr = 145.95 + 173.25 = 319.2$, 387.8 is an extreme outlier and 248.8 is only a mild outlier.

The MINITAB box plot is as follows:

Data Set 2

For the number of mailings (x_1)

x_{1i} : 3.7, 4.0, 4.9, 5.0, 6.5, 6.5, 6.5, 6.8, 7.0, 7.0, 7.1, 7.2, 7.3, 7.5, 7.5, 8.1, 9.0, 9.1, 10.0, 11.5, 12.5, 13.0, 18.0, 23.0, 23.5

The quantities needed for constructing the modified box plot are as follows:
Median = 7.3, Lower quartile = 6.5, Upper quartile = 10.75

$$iqr = 10.756.5 = 4.25$$
$$1.5 \cdot iqr = 1.5 \cdot 4.25 = 6.375$$
$$3 \cdot iqr = 3 \cdot 4.25 = 12.75$$

Thus,

$$\text{Upper edge of box} + 1.5 \cdot iqr = 10.75 + 6.375 = 17.125$$
$$\text{Lower edge of box} - 1.5 \cdot iqr = 6.56.375 = 0.125$$

So, $18.0, 23.0$ and 23.5 are both outliers at the upper end.

Since, Upper edge of box $+3 \cdot iqr = 10.75 + 12.75 = 23.5$ there are no an extreme outlier for this data set.

The MINITAB box plot is as follows:

For independent variable lead time of mailings (x_2),

x_{2i} : 2, 3, 4, 5, 6, 6, 6, 6, 9, 9, 10, 10, 11, 12, 12, 12, 12, 12, 12, 13, 13, 13, 14, 15, 16, 17

The quantities needed for constructing the modified box plot are as follows:
Median = 12, Lower quartile = 6, Upper quartile = 13

$$iqr = 136 = 7$$
$$1.5 \cdot iqr = 1.5 \cdot 7 = 10.5$$
$$3 \cdot iqr = 3 \cdot 7 = 21$$

Thus,

$$\text{Upper edge of box} + 1.5 \cdot iqr = 13 + 10.5 = 23.5$$
$$\text{Lower edge of box} - 1.5 \cdot iqr = 610.5 = -4.5$$

So, there are not outliers for lead time of mailings, x_2.