Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 13, Number 11 (2017), pp. 7749–7758

http://www.ripublication.com/gjpam.htm

Estimation of the Parameters of a Linear regression System Using the Simple Averaging Method

Kikawa R. Cliff[1]

Independent Researcher,
Mathematical & Applied Statistics,
Buddo-Nakasozi, Wakiso District,
Uganda, East Africa.

Kalema M. Billy

Department of Informatics,
Tshwane University of Technology,
Republic of South Africa.

Abstract

In an effort to estimate the parameters of a linear regression model, the ordinary least-squares (OLS) technique is usually employed. In this article a new approach based on the reasoning of averaging individual model gradients is proposed. A real data set of coffee sales in relation to shelf space is used to examine the performance of simple averaging method (SAM) and OLS. Model performance of OLS and SAM is assessed using the root mean square error, mean absolute error and grapphical methods. It is shown that both methods yield comparable results based on a one dimensional regression problem. In this case, SAM is reommended since it has less strigent conditions for application as opposed to the OLS method.

AMS subject classification: 62-07, 62J05, 62H12.
Keywords: Linear Regression, Parameter estimation, Simple Averaging Method, Ordinary Least-Squares, One dimensional regression.

[1]Corresponding author.

1. Introduction

Economic, Social and Scientific variables are often times connected by linear associations with the assumption that the model is linear in parameters. In the field of econometrics and statistical modeling, regression analysis is a conceptual process for ascertaining the functional relationship that exists among variables ([1] & [2]).

2. Ordinary Least-Squares

The ordinary least-squares (OLS) method is a technique used to estimate parameters of a linear regression model by minimizing the squared residuals that occur between the measured values or observed data and the expected values ([3]).

The OLS method is usually studied and applied in the context of a linear regression problem, where a contrast in the regressand variable Y, can partially be accounted for by a contrast in the other variables, known as the regressors or covariables X.

2.1. Least-Squares Estimation

Consider a simple linear regression model,

$$Y = f(X) + e \tag{1}$$

The best prediction of Y in the context of mean squared error, when the value of X is given, would be the mean of $f(X)$ of the variable Y given the covariable X. In this case Y is a function of X plus the residual term e.

2.2. Parameter estimation in the OLS framework

The basic idea behind OLS estimation is, "It may seem unusual that when several people measure the same quantity, they usually do not obtain the same results. In fact, if the same person measures the same quantity several times, the results will vary. What then is the best estimate for the true measurement?"

"The method of least squares gives a way to find the best estimate, assuming that the errors (i.e. the differences from the true value) are random and unbiased" [5].

2.3. Algebraic formulation of OLS

Suppose there exist a variable y explained by several explanatory variables $x_1, x_2, \ldots, x_n$, and it is now of interest to find the behavior of y over time. A theory may suggest that the behavior of y can be well identified by a given function f of the variables $x_1, x_2, ..., x_k$. Then, $f(x_1, ..., x_k)$ may be observed as a "systematic" component of y if no other variables can further explain the behavior of the residual $y = f(x_1, ..., x_k)$ [7]. In the linear regression context, the function f is specified as a linear function and the unknown coefficients (parameters) can be identified in the OLS framework [6].

2.4. Simple Linear regression

The linear specification is

$$\beta_0 + \beta_1 x_1,$$

where β_0 and β_1 are unknown model coefficients. Therefore,

$$y = \beta_0 + \beta_1 x_1 \tag{2}$$

2.5. Least Squares Estimation

Assuming that a data set $\{(x_1, y_1), (x_2, y_2), \ldots, x_n, y_n)\}$ is available, and if a sample of n subjects is to be conisedered, observing values y of the response variable and x of the predictor variable. We would like to choose as estimates for β_0 and β_1, the values b_0 and b_1 that 'best fit' the sample data. Defining the fitted equation to be of the form;

$$\hat{y} = b_0 + b_1 x, \tag{3}$$

where the values of b_0 and b_1 are chosen such that they "best fit" the sample data. The estimates b_0 and b_1 are the values that minimize the distances of the data points to the fitted line, the line of "best fit". Now, for each observed response y_i, with a corresponding predictor variable x_i, we obtain a fitted value $\hat{y}_i = b_0 + b_1 x_i$. So, it is required to minimize the sum of the squared distances of each observed response to its fitted value. That is, we want to minimize the **error sum of squares, SSE**, where

$$\mathbf{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_{i=1}))^2.$$

Using some little calculus, the estimates can be computed as

$$b_1 = \frac{\sum_i^n (x_{i=1} - \bar{x})(y_i - \bar{y})}{i = 1 \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{SS_{xy}}{SS_{xx}},$$

and

$$b_0 = \bar{y} - \hat{\beta}\bar{x} = \frac{\sum_{i=1}^n y_i}{n} - b_1 \frac{\sum_{i=1}^n x_i}{n}.$$

Definition 2.1. A slope of a line is an averge change on the vertical axis, $y - axis$ due to a change on the horizontal axis, $x - axis$.

This can algebraically be represented as

$$slope = \frac{y_n - y_{n-1}}{x_n - x_{n-1}} = \frac{\Delta y}{\Delta x} \tag{4}$$

Condition [CD-1] The change in magnitude of the intervention variable X must not be equal to zero. That is, $\Delta x \neq 0$ in Equation (4)

3. Simple Linear Regression

This is a regression method applied in parametric statistics to analyse the average outcome of a regressand Y, that varies in regard to the magnitude of a convariable X. In simple linear regression, one regressor, x is classified to explain the response of the regressand y. The identification is

$$y_j = \beta_{oj} + \beta_{1j}x_j + e_j, \tag{5}$$

where β_o and β_1 are the unknwon parameters, e_j are the random errors resulting from the identification, $j = 1, 2, \ldots, n$.

Lemma 3.1. Given the identification (5), and that **[CD-1]** holds. Then, the parameter β_1 can be estimated using the simple averaging method (SAM) as

$$\beta_1 = \frac{1}{n^*}\sum_{j=1}^{n}\beta_{1j}; \quad for \ \ n^* = n-1 \ \text{ and } \ j = 1, \ldots, n.$$

Proof. Assuming that a dataset $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ is available, from

Table 1: Tabulated data set

x_j	y_j	$\beta_{1,j}$
x_1	y_1	-
x_2	y_2	$\beta_{1,1}$
x_3	y_3	$\beta_{1,2}$
$\vdots$	$\vdots$	$\vdots$
x_n	y_n	$\beta_{1,n}$

From Table 1, it can be deduced from definition 2.1 that

$$\beta_{1,j} = \frac{y_n - y_{n-1}}{x_n - x_{n-1}},$$

for $j = 1, 2, \ldots, n$. The average of the respective slopes,

$$\bar{\beta}_{1,j} = \left(\frac{y_2 - y_1}{x_2 - x_1} + \frac{y_3 - y_2}{x_3 - x_2} + \cdots + \frac{y_n - y_{n-1}}{x_n - x_{n-1}}\right)/(n-1),$$

hence,

$$\hat{\beta}_1 = \frac{\beta_{1,1} + \beta_{1,2} + \cdots + \beta_{1,n}}{n-1},$$

which can be presented as

$$\hat{\beta}_1 = \frac{1}{n^*}\sum_{j=1}^{n}\beta_{1j}; for \ \ n^* = (n-1) \ \text{ and } \ j = 1, \ldots, n,$$

and this completes the proof. ■

Lemma 3.2. Given the identification (5), and that **[CD-1]** holds. Then, the constant β_o can be estimated using the simple averaging method as

$$\beta_o = \frac{1}{n^*}\sum_{j=1}^{n}\left(y_j - \hat{\beta}_1 x_j\right) \text{ for } n^* = n-1 \text{ and } j = 1,\dots,n.$$

Proof. Substituting the result of lemma (3.1) into identification (5) it can be ascertained that

$$\beta_{o,j} = y_j - \beta_1 x_1,$$

then

$$\beta_{o,1} = y_1 - \hat{\beta}_1 x_1$$
$$\beta_{o,2} = y_2 - \hat{\beta}_1 x_2$$
$$\vdots$$
$$\beta_{o,n} = y_n - \hat{\beta}_1 x_n.$$

Hence,

$$\hat{\beta}_o = \frac{\beta_{o,1} + \beta_{o,2} + \dots + \beta_{o,n}}{n-1}$$
$$= \left\{\left(y_1 - \hat{\beta}_1 x_1\right) + \left(y_2 - \hat{\beta}_1 x_2\right) + \cdots + \left(y_n - \hat{\beta}_1 x_n\right)\right\}/(n-1).$$

implying that,

$$\hat{\beta}_o = \frac{1}{n^*}\sum_{j=1}^{n}\left(y_j - \hat{\beta}_1 x_j\right)$$

and this concludes the proof. ■

Lemma 3.3. Given the identification (5), and that **[CD-1]** holds. Knowning that the error e_j is a function of β_o and β_1 defined as

$$e_j = \left(\beta_o, \beta_1\right) = y_j - \left(\beta_o + \beta_1 x_j\right),$$

the error term e_j can be estimated using the simple averaging method as

$$\hat{e_j} = \frac{1}{n^*}\sum_{j=1}^{n}\left\{y_j - \left(\hat{\beta}_o + \hat{\beta}_1 x_j\right)\right\} \text{ for } n^* = n-1 \text{ and } j = 1,\dots,n.$$

Proof. From the results of lemmas (3.1 and 3.2), and substituting in identification (5) it can be observed that

$$\hat{e_j} = y_j - \left(\hat{\beta}_o + \beta_1 x_j\right)$$

then

$$\hat{e_1} = y_1 - \left(\hat{\beta}_o + \beta_1 x_1\right)$$

$$\hat{e_2} = y_2 - \left(\hat{\beta}_o + \beta_1 x_2\right)$$

$$\vdots$$

$$\hat{e_n} = y_n - \left(\hat{\beta}_o + \beta_1 x_n\right).$$

Hence,

$$\hat{e_j} = \frac{\hat{e_1} + \hat{e_2} + \ldots + \hat{e_n}}{n-1}$$

$$= \left\{\left[y_1 - \left(\hat{\beta}_o + \hat{\beta}_1 x_1\right)\right] + \left[y_2 - \left(\hat{\beta}_o + \hat{\beta}_1 x_2\right)\right] + \cdots + \left[y_n - \left(\hat{\beta}_o + \hat{\beta}_1 x_n\right)\right]\right\} / (n-1)$$

.

$$\hat{e_j} = \frac{1}{n^*} \sum_{j=1}^{n} \left[y_j - \left(\hat{\beta}_o + \hat{\beta}_1 x_j\right)\right] \text{ for } n^* = n-1 \text{ and } j = 1, \ldots, n,$$

and this completes the proof. ■

4. Multiple Linear Regression

The multiple linear regression model [4] is an extension of a simple linear regression model, Section 3, to assimilate two or more regressor variables in a prediction equation for a response variable.

The response of y (the regrassand) can be classified when a linear function with k regressor variables is identified;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + e(\beta_1, \ldots, \beta_k), \tag{6}$$

where $e(\beta_1, \ldots, \beta_k)$ denotes the error of this identification.

Theorem 4.1. [multivariate model estimation] Given the identification (5), and that **[CD-1]** holds. The simple averaging method (SAM) estimates for the weights β_0, β_1, $\ldots$, β_k and the error $e(\beta_1, \ldots, \beta_k)$ are as follows,

1. $\hat{\beta}_1 = \bar{\beta}_j$ for $j = 1, \ldots, n$.

2. $\hat{\beta}_k = \bar{\beta}_j$ for $j = 1, \ldots, n$.

3. $\hat{\beta}_0 = \bar{\psi}_j$
   where, $\bar{\psi} = \dfrac{1}{(n-1)} \sum \left\{\left[y_j - \left(\bar{\beta}_1 x_1 + \cdots + \bar{\beta_{k,n}} x_{k,n}\right)\right]\right\}$

4. $\hat{e} = e(\beta_1, \ldots, \beta_k) = \bar{\gamma}$
   where, $\bar{\gamma} = \frac{1}{(n-1)} \sum \{[y_j - (\bar{\beta_0} + \bar{\beta}_{1,j} x_{1,j} + \cdots + \bar{\beta}_{k,n} x_{k,n})]\}$

Proof. Without loss of generality, the results can directly be verified from lemmas, 3.1, 3.2 and 3.3. ■

5. Empirical tests

Empirical tests of this line of reasoning (SAM) were performed and the results compared with those from OLS, in the simple linear regression framework. For simple linear regression, consider the following data set from coffee sales [8].

Simple linear regression

A marketer is interested in the relation between the width of the shelf space for her brand of coffee (x) and weekly sales (y) of the product in a suburban supermarket (assume the height is always at eye level). Marketers are well aware of the concept of 'compulsive purchases', and know that the more shelf space their product takes up, the higher the frequency of such purchases. She believes that in the range of 3 to 9 feet, the mean weekly sales will be linearly related to the width of the shelf space. Further, among weeks with the same shelf space, she believes that sales will be normally distributed with unknown standard deviation σ (that is, σ measures how variable weekly sales are at a given amount of shelf space). Thus, she would like to fit a model relating weekly sales y to the amount of shelf space x her product receives that week. That is, she is fitting the model: $y = \beta_0 + \beta_1 x_1$

Table 2: Coffee sales for n=12

Shelf Space (x)	Weekly Sales (y)	Shelf Space (x)	Weekly Sales (y)
6	526	6	434
3	421	3	443
6	581	9	590
9	630	6	570
3	412	3	346
9	560	9	672

Fitting the model $y = \beta_0 + \beta_1 x_1$ to the data in Table 2 we obtain the following results:

Fitted models

$$y_{OLS} = 307.9167 + 34.5833x \tag{7}$$

$$y_{SAM} = 315.4545 + 33.1667x \quad (8)$$

Models (7) and (8) where fitted using OLS method using STATA statistical software and SAM respectively.

Predicting ability

To make informed conclusions about the two methods of estimation, i.e. OLS and SAM, we investigated how well each model 7 and 8 was able to predict the observed values, 3

Table 3: Predicated Coffee sales using the OLS (Pred_OLS) and SAM (Pred_SAM)

Space_x	Observed_y	Pred_OLS	Pred_SAM
6	526	515.4165	514.4547
3	421	411.6666	414.9546
6	581	515.4165	514.4547
9	630	619.1664	613.9548
3	412	411.6666	414.9546
9	560	619.1664	613.9548
6	434	515.4165	514.4547
3	443	411.6666	414.9548
9	590	619.1664	613.9548
6	570	515.4166	514.4547
3	346	411.6666	414.9546
9	672	619.1664	613.9548

Table 3 shows the predicted values of the coffee sales using the OLS and the SAM, at this point it is difficult to tell which of the two approaches has a better predictive power [11] or which is a better model. More statistical tests are performed to evaluate the models.

Model evaluation

The root mean square error (RMSE) is widely employed as a standard statistical metric to assess model performance [9] and [10] in air quality, economics and climatic studies. The mean absolute error (MAE) is another important metric that is also widely used in model evaluations. It is stated that, while they have both been extensively employed to examine model performance for several years, there is no consensus on the most appropriate metric for model errors [9]. In this work the models 7 and 8 are evaluated using both the RMSE and MAE, Table 4

From Table 4 it is observed that both models have comparable performance since the error values from RMSE and MAE are approximately the same to with in the nearest whole number.

We further explore the visual methods of ascertaing model performance using a scatter plot and fitting the two lines, [1]. In this way it may be easier to ascertain which line fits the data set better.

Table 4: Results of OLS and SAM model evaluations using the root mean square error (RMSE) and Mean absolute error (MAE)

Method	RMSE	MAE
OLS	47.1367	39.2362
SAM	47.2740	39.3409

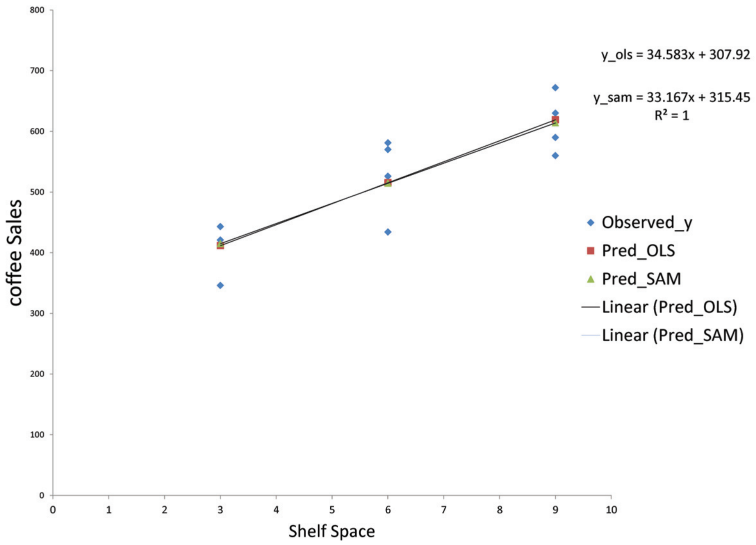

Figure 1: Coffee Sales against shelf space, for n=12

6. Conclusion

In this paper, a new method for estimating parameter coefficients of linear regression models is proposed (Simple Averaging Method, SAM). Theorems are derived and a condition stated for the application of SAM. A real data set is used to test the capability of SAM in predicting values of coffee sales relative to the shelf space as discussed in the example. The efficacy of SAM is compared with the ordinary least squares (OLS) approach by predicting the values of coffee sales, for a one dimensional regression problem, empirical example in 5. Model evaluation is conducted using the root mean square error (RMSE), mean absolute error (MAE) and graphical methods (scatter plot), Table 4 and Figure 1. SAM produces comparable results to OLS and it can be a better subsitute to OLS since it has less stringent assumptions in the univariate case so far studied. This work has put emphasis on a simple regression problem however the reasoning in the SAM can easily be extended to a multivariate estimation problem as Theorem 4.1 suggests.

7. Future work

1. To study SAM on the multivariate estimation problems.
2. To investigate the performace of SAM on ill-conditioned problems without using the usual approach of pre-conditioning.

References

[1] Al-Nasser, D.A. and Radaideh, A.C. (2008). Estimation of Simple Linear Regression Model Uisng L Ranked Set Sampling, *International Journal Of Open Problems in Computer Science and Mathematics*, **1**, 1, 19–38.

[2] Quandt, R.E. (1958). The Estimation of the Parameters of a Linear regression. System obeying Two Separate Regimes, *Journal of American Statistical Association*, **53**, 284, 873–880.

[3] Sara, A. (2005). *Encyclopedia of Statistics in Behavioral Science.*, John Wiley & Sons, Ltd, Chichester.

[4] Kikawa, C.R. and Kloppers, P.H. (2016). Multiple Linear Regression with Constrained Coefficients: Application of the Lagrange Multiplier, *South African Statistical Journal*, **50**, 303–312.

[5] Reid, F. (2000). The Mathematician on the Banknote: Carl Friedrich Gauss, *Parabola*, **36**, 2, 2–9.

[6] Kloppers, P.H., Kikawa, C.R., and Shatalov, M.Y., (2012). A new method for least squares identification of parameters of the transcendental equations, *International Journal of the Physical Sciences*, **7**, 5218–5223.

[7] Manski, C.F. (2000). Regression, *Journal of Economic Literature*, **29**, 34–50.

[8] http://www.stat.ufl.edu/ winner/qmb3250/notespart2.pdf, [Accessed on: 04/10/ 2016]

[9] Chai, T. and Draxler, R.R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature, *Geoscientific Model Development*, **7**, 1247–1250.

[10] Kikawa, C.R., Kalema, B.M., Kloppers, P.H., and Ssematimba, A. (2016). A Semi-Parametric Method for Generating Time Series Data: A Novel Approach for Bootstrapping the Residuals, *Transylvanian Review*, **24**, 9, 1325–1332.

[11] Zheng, B. and Agresti, A. (2000). Summarizing the predictive power of a generalized linear model, *Statist. Med.*, **19**, 1771–1781.