COMPUTATIONAL IMPLICATIONS FOR THE LEAST SQUARES' PARAMETERS OF THE SIMPLE LINEAR REGRESSION MODEL UNDER DATA TRANSFORMATION.

Igabari, J. N.
Department of Mathematics and Computer Science, Delta State University, Abraka
Email: jn_igabari@yahoo.com

ABSTRACT

The Least squares method of parameter estimation is considered important due to its relative simplicity, optimal properties and wide economic applications. This paper examines the computational implications for the estimates of regression parameters for a simple linear regression model when there are changes in units of measurement, leading to a new set of data which is a scaled form of the original data. Expressions for the Least Squares estimates of the regression parameters are derived as well as their precision for the new data in terms of the original data.

Keywords: Scaling Factor, Expected Values, Precision, Least Squares, Regression.

INTRODUCTION

Regression and correlation are two common concepts in the measurement of association between two factors (or variables). According to Greene (2005) regression studies the dependence of a factor Y (the dependent variable) on another factor X (the independent variable), while correlation measures the interdependence between X and Y. So while correlation may be regarded as a commutative concept in a light sense, that is not true for regression. The simple linear regression model is a relationship between two variables, where one is dependent and the other explanatory; and both are assumed related with a linear function of the form [Dougherty, 1992]:

$Y_i = a + bX_i$ (1)

where a and b are parameters of the function, and our aim is to obtain their numerical estimates, a and b, respectively. This form implies (or assumes) that there is a one-way causation between X and Y. Among the methods used to derive the estimates of the parameters of the relationship (1) from sample observations, the least squares method clearly stands out for the following reasons [Koutsoyannis, (2001)].

a) the optimal properties of the least squares
b) the fairly simple computational procedure compared to others econometric methods
c) its standing as a method that has been used in a wide range of economic relationships with fairly satisfactory results
d) the singular fact that majority of other techniques involve the application of the least squares method in direct or modified form.

In real life situation, it is a common occurrence to encounter changes in units of measurement for economic quantities. Examples abound in currency conversion and stock market quotations on the international scene, and changes in measuring units in the local scene or a general scaling of a variable by a factor for ease of computation. If estimates of the regression parameters were already made before any of the above mentioned interventions, how are the parameters affected between the original and scaled data. What is the effect of the scale factor on the expected value, variance and covariance of the regression parameters and by implication the regression line?

METHODOLOGY:

Consider the bi-variate data set (X_i , Y_i), i = 1(1) n, given in the form as

$$\left.\begin{matrix} x_1 & x_2 & x_3 & x_4 & . & . & . & x_n \\ y_1 & y_2 & y_3 & y_4 & . & . & . & y_4 \end{matrix}\right\} \quad .(2)$$

By method of least squares, the regression model of Y on X is

Y = a + bX + Î **(3)**

where Î is the random error term, and estimates of b and a are given respectively as

$$b = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 \;-\; \left(\sum x_i\right)^2} \qquad (4a)$$

$$a = \bar{y} - b\bar{x} \qquad (4b)$$

and s^2 = SÎ^2 $_{/ (n-2)}$ (5)

where Î = y_i –a - bx_i

Also from literature [Greene (2005), Koutsoyiannis (2001), Nduka (1999)], we know that

E (a) = a, E (b) = b (6)

$$\text{Var(a)} = \frac{\sigma^2 \sum x_i^{\,2}}{n\sum (x_i - \bar{x})^2} \qquad (7)$$

$$\text{Var(b)} = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \qquad (8)$$

$$\text{Cov(a, b)} = \frac{-\sigma^2 \,\bar{x}}{\sum (x_i - \bar{x})^2} \qquad (9)$$

Now suppose the observations in (2) are scaled by the respective factors p and q, (p ¹ 0, q ¹ 0), for X and Y, so that the tabular form becomes

$$\left.\begin{matrix} px_1, & px_2, & px_3, & px_4, & . \; . \; ., & px_n \\ qy_1, & qy_2, & qy_3 & qy_4, & . \; . \; ., & qy_n \end{matrix}\right\} \qquad (10)$$

[see (Igabari, 2006)].

Let Y = a* + b* X + Î (11)

be the regression model of Y on X resulting from (11), with a* and b* as least squares estimates for a* and b* respectively. We may now establish a relationship between the parameters of (3) and (11) as follows. Using the least squares method on (10) we have

$$b^* = \frac{n\sum pqx_i y_i - pq\sum x_i \sum y_i}{np^2\sum x_i^2 - p^2\left(\sum x_i\right)^2}$$

$$= \frac{q}{p}\left\{\frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}\right\}$$

$$\backslash\; b^* = \frac{qb}{p} \qquad (12)$$

Also a* = $\bar{y}$ - b* $\bar{x}$

$$= \frac{\sum qy_i}{n} - \frac{qb}{p} \cdot \frac{\sum px_i}{n} = q(\bar{y} - b\,\bar{x})$$

\ a* = qa (13)

To find expressions for E[a*] and E[b*],

E[a*] = E[q a] = qE(a) from (13)

Hence E[a*] = q a (14)

Since E[a] = a from (6)

Similarly using (12) and (6) we get

$$E[b^*] = \frac{q}{b}\beta \qquad (15)$$

On the precision of the regression coefficients, we note from (7), (8) and (9) that Var(a), Var (b) and Cov(a,b) are directly proportional to the variance, s^2, of the error term Î$_i$ [Dougherty, 1992]. Similarly, Var(a*), Var(b*) and Cov(a*, b*) are proportional to s^2 and thus, by simplification we can easily see that

Var(a*) = Var(q a) = q^2 Var(a)

$$= (q\sigma)^2 \frac{\sum x_i^2}{n\sum (x_i - \bar{x})^2} \qquad (16)$$

Also Var(b*) = Var $\left(\frac{q}{p}b\right)$ = $\left(\frac{q}{p}\right)^2 Var(b)$

$$= \frac{(q\sigma)^2}{p^2\sum (x_i - \bar{x})^2} \qquad (17)$$

And lastly, Cov(a* , b*) = Cov(q a , $\frac{q}{p}$ b)

$$= \frac{q^2}{p}\ \text{Cov(a, b)}$$

$$= -\frac{q^2\,\sigma^2\,\bar{x}}{p\sum (x_i - \bar{x})^2} \qquad (18)$$

Illustration

We hereby illustrate the foregoing with the hypothetical data below, of two quantities X (independent) and Y (dependent).

Table 1: Hypothetical data of X and Y.

X	2	3	1	5	9
Y	4	7	3	9	17

Now SX = 70, SY = 40, SX^2 = 120, SXY = 230, S(X$_i$ – Y$_i$)^2 = 40

Given the model $Y_{i-} = b_0 + b_i X_i + \hat{I}_i$ and the estimate $\hat{y} = a + bX$ (least squares) then applying (4) and (5) on Table 1, we get a = 1, b = 1.75, and $s^2 = 0.5$ and by (7), (8) and (9) we also have that Var(a) = 0.3, Var(b) = 0.0125, and Cov(a, b) = -0.05 Now suppose that by scaling, data becomes as in table 2

Table 2: The scaled data of table 1

X	7	10.5	3.5	17.5	31.5
Y	0.4	0.7	0.3	0.9	1.7

i.e. X has a scaling factor of p = 3.5, while Y has a scaling factor of q = 0.1 .
Then, SX = 70, SY = 4 $SX^2 = 1470$, SXY = 80.5.

The new parameters of regression for Table 2 could be obtained either by using (4) and (5) on Table 2 as new data, or using (12) and (13) as scaled form of Table 1
Using (12) and (13) gives

$$a^* = q\,a = 0.1 \text{ x } 1 = 0.1$$

$$b^* = \frac{q}{p}\,b = \frac{0.1}{3.5} \text{ x } 1.75 = 0.05$$

which gives same result as using (4) and (5) on Table 2.
Also, using (16), (17) and (18) on results of Table 1, we have

$$\text{Var}(a^*) = q^2\,\text{Var}(a) = (0.1)^2 \text{ x } 0.3 = 0.003$$

$$\text{Var}(b^*) = \left(\frac{q}{p}\right)^2 \text{Var}(b) = \left(\frac{0.1}{3.5}\right)^2 \text{ x } 0.0125 = 0.0000102$$

Finally Cov(a^*, b^*) = 0.000143 using (18); which is same result as using (7), (8) and (9) directly on Table 2.

CONCLUSION

There is drastic reduction in computational efforts as long as we can establish scaling factors (p, q) between respective variables (X,Y) of the two sets of data, given that the regression parameters of one is known already. In summary, the Y-intercept, a, is affected by the factor q, the regression coefficient, b, by the factor q / p; the precision of a by q^2, the precision of b by $(q / p)^2$; and the Covariance by q^2 / p. Hence the parameters of the Simple Linear Regression Model are sensitive to scale, because changes in units of measurement will affect the parameter estimates. Here we have provided simple formula that can be used to obtain the corresponding estimates for the regression parameters when the unit of measurement is altered.

REFERENCES

Dougherty, C. (1992). *Introduction to Econometrics.* Oxford University Press, New York.

Greene, W. H. (2005). *Econometric Analysis.* 5th Edition. Pearson Ed. Inc. Singapore.

Igabari, J. N. (2006). *A Note on the Parameters of the Least Squares Simple Regression Model in the Presence of Scaling.* Proceedings of the Annual National Conference of Mathematical Association of Nigeria.

Igabari, J. N. and Emenonye, C.E. (2000). *An Outline on Statistical Inference.* Krisbec Publishers, Nigeria.

Koutsoyiannis, A (2001). *Theory of Econometrics.* 2nd Ed., Palgrave. New York.

Nduka, E. C. (1999). *Principles of Applied Statistics I: Regression and Correlation Analyses.* Crystal Publishers. Owerri.