P/

UNIVERSITI TEKNOLOGI MARA

University Publication Centre (UPENA)

ESTEEM

Academic Journal UiTM Pulau Pinang

Volume 2 2005 ISSN 1675-7939

A Study on Compressive Strength of Concrete Cointaining MK7003 with Various Water Binder Ratio — Clotilda Petrus, Amer Yusuff, Ahmad Fairuz Othman, Ahmad Ikhwan Naim Mohd Zin

Designing Reinforced Earth Wall for Seismic Areas — Iqraz Nabi Khan, Swami Saran, Caroline Peter Diman

Study on Nylon Strip-soil Interaction — Iqraz Nabi Khan, Swami Saran

Sampling of Beach Sand in Kerteh, Terengganu — M.F. Ahmad, N.A. Zakaria, M.R. Selamat

A Study on the Relationship between Correlation Coefficient and Inverse Regression — Ng Set Foong, Teoh Sian Hoon

A New Finite Difference Scheme Based on Centroidal Mean Averaging for the Goursat Problem — Mohd Agos Salim Nasir, Ahmad Izani Md Ismail

A Comparative Study between Cu and Pb in the Speciation Study of Heavy Metals in a Stabilization Pond Treating Household Wastewater — Lim Boon Tik

Design of Road Humps in Residential Area — Muhammad Akram Adnan, Teoh Sian Hoon, Lim Boon Tik

A Study of Temperature Variation in Closed Cars — Nor Hanim Abdul Rahman, Mohd Rozaiman Aziz, Suzana Ab. Rahim

© ESTEEM Academic Journal is jointly published by the Universiti Teknologi MARA, Pulau Pinang and Pusat Penerbitan Universiti (UPENA), Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.

ESTEEM

Academic Journal UiTM Pulau Pinang

Volume 2 2005 ISSN 1675-7939

Editorial Team i
Foreword by The Chief Editor v

1. A Study on Compressive Strength of Concrete Cointaining MK7003 with Various Water Binder Ratio 1
Clotilda Petrus, Amer Yusuff, Ahmad Fairuz Othman & Ahmad Ikhwan Naim Mohd Zin

2. Designing Reinforced Earth Wall for Seismic Areas 9
Iqraz Nabi Khan, Swami Saran & Caroline Peter Diman

3. Study on Nylon Strip-soil Interaction 27
Iqraz Nabi Khan & Swami Saran

4. Sampling of Beach Sand in Kerteh, Terengganu 37
M.F. Ahmad, N.A. Zakaria & M.R. Selamat

5. A Study on the Relationship between Correlation Coefficient and Inverse Regression 45
Ng Set Foong & Teoh Sian Hoon

6. A New Finite Difference Scheme Based on Centroidal Mean Averaging for the Goursat Problem 55
Mohd Agos Salim Nasir & Ahmad Izani Md Ismail

7. A Comparative Study between Cu and Pb in the Speciation Study of Heavy Metals in a Stabilization Pond Treating Household Wastewater 63
Lim Boon Tik

8. Design of Road Humps in Residential Area 75
Muhammad Akram Adnan, Teoh Sian Hoon & Lim Boon Tik

9. A Study of Temperature Variation in Closed Cars 87
Nor Hanim Abdul Rahman, Mohd Rozaiman Aziz &
Suzana Ab. Rahim

10. Motivating ESL Students Using Vocabulary Enrichment Games 95
Emily Jothee Mathai & Suchithra Nair

11. Penggunaan Teknologi Maklumat untuk Membantu Pembelajaran: Di Manakah Tahap Kesediaan Pelajar? 113
Peridah Bahari & Salina Hamed

12. Masjid Batu dan Masjid Papan, Kampung Pertama: Penjelas Kemelut Masyarakat Setempat 123
Zulkifli Dahalan

13. The Language of Genders in the "Labour of Love" 135
Suzana Ab. Rahim

Foreword

Welcome to ESTEEM Volume 2. In this issue, we address a gamut of topics from the engineering disciplines to language education. We hope that ESTEEM, by publishing articles from a diverse range of disciplines, will encourage debate and exchange among researchers from assorted academic backgrounds.

I would like to thank our advisor, Prof. Madya Mohd Zaki Abdullah for his distinctive imprint on this edition. His leadership of the journal in its 2nd year of growing impact and reputation has been outstanding. His vision, commitment to excellence, and attention to detail are widely recognized by the Penang academic community as determining factors in the journal's success so far. We will do our best to continue and expand on this tradition of excellence.

Since its launch in 2003, ESTEEM is indeed fortunate to have a dynamic Editorial Team. These people have provided the journal with an outstanding service of reviewing submissions for publications. The journal follows the established policy of a blind review process consisting of at least two peer reviewers per submission. We depend upon their knowledge and judgement in advancing the scope and utility of this journal. Without their support and enthusiasm none of this would have been possible. Also, my thanks to all the contributors, both the successful and not so successful.

Our vision of the ***ESTEEM*** journal is that it should be the journal that belongs to you, the academic and research community. This includes all engineers and academicians working to unravel the mysteries of research, teaching and learning, in all its facets. We wish the journal to be responsive to your needs and your interests. Please feel free to contact any of us in the editorial board to give us your ideas and suggestions for the development of the journal. We look forward to working with you all in expanding this emerging venue for communicating high quality research on the many aspects of academia.

Finally, I would like to take this opportunity to invite all authors and readers to contact me at **esteem@ppinang.uitm.edu.my** to share their comments and advice on how to further enhance the journal's value to the wider research community in knowledge and how to move ESTEEM to the next level of excellence.

The Chief Editor
May, 2005

A Study on the Relationship between Correlation Coefficient and Inverse Regression

Ng Set Foong
Teoh Sian Hoon

ABSTRACT

This paper focuses on the use of an inverse regression model. The discussion provides an idea on how correlation coefficient, r, in a regression analysis plays an important role in estimation of the value of independent variable X based on the known value of Y. The construction of an inverse regression mode is further illustrated. This issue is discussed in detail with analysis using SPSS software.

Introduction

In many scientific experiments or social researches, most of the researchers would like to predict the value of a dependent variable (*Y*) from knowledge of the values of one or more independent variables (*X*). For example, a chemical engineer might be interested in the relationship between the yield of a chemical process and the temperature as well as the reaction time, or a physician may want to know the relationship between a patient's blood pressure and the dosage of a new drug (William, 2003). In the case of one independent variable, the relationship between the variables *Y* and *X* can be represented as a simple linear regression model. The basic model is

$$Y = \beta_0 + \beta_1 x + \varepsilon \qquad (1)$$

The β_0 and β_1 are the parameters while e represents the error term. Once β_0 and β_1 are estimated, the prediction equation is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad (2)$$

The $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least squares estimators of the parameter $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively (Robert, 1992; Thomas, 1997).

In a simple linear regression model, the strength of the relationship between *X* and *Y* is measured by the (sample) correlation coefficient, r_{xy}. The possible values of r_{xy} range from -1 to $+1$. If r_{xy} has the value of $+1$, then the variables *X* and *Y* have a strong positive correlation. On the other hand, the value of -1 for represents a strong negative correlation between *X* and *Y*. A zero correlation between *X* and *Y* would signify that it is meaningless to construct the regression model using the variable *X* (John, 1998).

In regression analysis, the value of dependent variable *Y* can be predicted once the regression model is constructed. However, in some cases, the researcher would like to estimate the value of variable *X* based on the known value of *Y*. In this case, is the prediction equation (2) suitable? The issue is the correlation coefficient actually plays an important role and it is not commonly highlighted. Therefore, this issue would be discussed in detail in the following section with analysis using SPSS software.

The prediction equation (2) is the model when *Y* is regressed against *X*. In the case where the researcher is interested to predict *X* based on a given value of *Y*, then the prediction equation (2) can be modified to equation (3) provided that the correlation coefficient r_{xy} is close to 1 (Thomas, 1997). Thus, for a given value of *Y*, *X* is estimated as

$$\hat{X}_c = -\frac{\hat{\beta}_0}{\hat{\beta}_1} + \frac{1}{\hat{\beta}_1} Y \tag{3}$$

It is noticed that the coefficient of *Y* and the constant term is formed from the combination of $\hat{\beta}_0$ and $\hat{\beta}_1$. Therefore, if the prediction equation (2) is constructed and the value of $\hat{\beta}_0$ and $\hat{\beta}_1$ are available before hand, then the equation (3) can be easily formed. However, the estimation from equation (3) would be almost accurate only if the correlation coefficient $r_{xy} \approx 1$. For this condition, we would discuss it from theory and data analysis using two types of data.

Discussion

For the prediction equation $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$, the least squares estimators of β_0 and β_1 are shown as equation (4) and (5) respectively, where $\overline{Y}$ and $\overline{X}$ are average value of *Y* and *X*.

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-x)^2} \tag{4}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{5}$$

As discussed previously, the strength of relationship between X and Y is measured by the (sample) correlation coefficient, r_{xy}. The formula of r_{xy} is shown in equation (6) below.

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} \tag{6}$$

In order to compare the accuracy of the estimation by equation (3), we would construct an inverse regression model that is regressing X on Y. Regression X on Y produces the inverse regression estimator

$$\hat{X}^+ = \hat{C}_0 + \hat{C}_1 Y \tag{7}$$

The $\hat{C}_0$ and $\hat{C}_1$ are least squares estimator for parameters and the formula of $\hat{C}_0$ and $\hat{C}_1$ are shown as equation (8) and (9) respectively.

$$\hat{C}_1 = \frac{s_{xy}}{s_{yy}} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (y-\bar{y})^2} \tag{8}$$

$$\hat{C}_0 = \bar{X} - \hat{C}_1 \bar{Y} \tag{9}$$

In the case where the condition of $r_{xy} \approx 1$ is fulfilled, $\sum (Y-\bar{Y})^2$ can be expressed as the function shown in the equation (10).

If $r_{xy} \approx 1$,

$$r_{xy} \approx 1$$

$$\frac{\sum (\boldsymbol{x}-\bar{\boldsymbol{x}})(\boldsymbol{y}-\bar{\boldsymbol{y}})}{\sqrt{\sum (\boldsymbol{x}-\bar{\boldsymbol{x}})^2 \sum (\boldsymbol{y}-\bar{\boldsymbol{y}})^2}} \approx 1$$

then
$$\sum (y-\bar{y})^2 \approx \frac{\left[\sum (x-\bar{x})(y-\bar{y})\right]^2}{\sum (x-x)^2} \tag{10}$$

Substituting the expression of $\sum(Y-\bar{Y})^2$ in equation (10) into (8) would get the approximated value of $\hat{C}_1$ as shown in the following equation (11).

$$\begin{aligned}
\hat{C}_1 &= \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2} \\
&\approx \sum(x-\bar{x})(y-\bar{y})\frac{\sum(x-\bar{x})^2}{\left[\sum(x-\bar{x})(y-\bar{y})\right]^2} \\
&= \frac{\sum(x-\bar{x})^2}{\sum(x-\bar{x})(y-\bar{y})} \\
&= \frac{1}{\hat{\beta}_1} \qquad (11)
\end{aligned}$$

With the approximated value of $\hat{C}_1$ shown in equation (11), we could get an approximated value of $\hat{C}_0$.

$$\begin{aligned}
\hat{C}_0 &= \bar{X} - \hat{C}_1 Y \\
&\approx \bar{X} - \frac{1}{\hat{\beta}_1}\bar{Y} \\
&= -\frac{1}{\hat{\beta}_1}\left(\bar{Y} - \hat{\beta}_1\bar{X}\right) \\
&= -\frac{1}{\hat{\beta}_1}\left(\hat{\beta}_0\right) \\
&= -\frac{\hat{\beta}_0}{\hat{\beta}_1}
\end{aligned}$$

Thus, the results from the workings above are summarized below.

If $r_{xy} \approx 1$, then $\hat{C}_0 \approx -\frac{\hat{\beta}_0}{\hat{\beta}_1}$, $\hat{C}_1 \approx -\frac{1}{\hat{\beta}_1}$.

It is noticed that these approximated values of the estimators are exactly the coefficient of Y and the constant term in the equation shown in the equation (3).

Therefore, from the theoretical point of view, the estimation of $\hat{X}_c$ will almost equal to the prediction from the inverse regression estimator $\hat{X}^+$ if $r_{xy} \approx 1$.

Data Analysis

In data analysis, we would take two datasets with different values of correlation coefficient r_{xy}. The r_{xy} for the first data set (Example 1) is very close to 1 (that is 0.997) while the r_{xy} for the second data set (Example 2) is very far from 1 (that is 0.244). We would then compare the accuracy of $\hat{X}_c$ from equation (3) with the inverse regression estimator $\hat{X}^+$ from equation (7).

Example 1

Regression methods were used to analyze the data from a study of the relationship between the level of prior knowledge (X) and the gain score in mathematics (Y).

Table 1: Data Set for Example 1

X	6	7	8	9	10	11	12	13	14	15
Y	13.1	15.0	17.2	19.3	21.0	23.0	24.2	27.0	27.9	31.3

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

where

$$\hat{\beta}_1 = 1.938$$
$$\hat{\beta}_0 = 1.549$$
$$r_{xy} = 0.997$$

Due to certain reason, the value of X is to be estimated for a given value of Y. Therefore, it is suggested to apply the model from equation (3) where the coefficient and the constant term can be calculated from the known $\hat{\beta}_0$ and $\hat{\beta}_1$.

Thus, for a given value of *Y*, *X* is estimated as

$$\hat{X}_c = -\frac{\hat{\beta}_0}{\hat{\beta}_1} + \frac{1}{\hat{\beta}_1}Y$$

$$\hat{X}_c = -\frac{1.549}{1.938} + \frac{1}{1.938}Y$$

$$\hat{X}_c = -0.799 + 0.516Y \tag{12}$$

For the purpose of comparison, we would use the values of *Y* given in Table (1) to estimate value of *X* using model in equation (12).

Table 2: Estimated Value of $\hat{X}_c$ for Example 1

Y	13.1	15.0	17.2	19.3	21.0	23.0	24.2	27.0	27.9	31.3
$\hat{X}_c$	5.9606	6.941	8.0762	9.1598	10.037	11.069	11.6882	13.133	13.5974	15.3518

In fact, in order to regress *X* on *Y*, an inverse regression model can be constructed that produces the inverse regression estimator

$$\hat{X}^+ = \hat{C}_0 + \hat{C}_1 Y$$

Using the data set in Table (1), regression analysis takes the variable *X* as a dependent variable while taking variable *Y* as an independent variable is done by using SPSS. The values of the estimators $\hat{C}_0$ and $\hat{C}_1$ are recorded from SPSS output.

$$\hat{C}_1 = 0.513$$
$$\hat{C}_0 = -0.739$$

Thus, the inverse regression model is

$$\hat{X}^+ = -0.739 + 0.513Y \tag{13}$$

The value of *X* would again be estimated by using the values of *Y* from the table (1) by using the model in equation (13).

Table 3: Estimated Value of $\hat{X}^+$ for Example 1

Y	13.1	15.0	17.2	19.3	21.0	23.0	24.2	27.0	27.9	31.3
$\hat{X}^+$	5.9813	6.956	8.0846	9.1619	10.034	11.06	11.6756	13.112	13.5737	15.3179

The comparison between the values of $\hat{X}_c$ and $\hat{X}^+$ for the corresponding value of Y is shown in the Table (4) below.

Table 4: Comparison between $\hat{X}_c$ and $\hat{X}^+$ for Example 1

Y	$\hat{X}_c$	$\hat{X}^+$	Relative error = $\left\lvert \frac{\hat{x}_c - \hat{x}^+}{\hat{x}^+} \right\rvert \times 100\%$
13.1	5.9606	5.9813	0.35%
15.0	6.9410	6.9560	0.22%
17.2	8.0762	8.0846	0.10%
19.3	9.1598	9.1619	0.02%
21.0	10.0370	10.0340	0.03%
23.0	11.0690	11.0600	0.08%
24.2	11.6882	11.6756	0.11%
27.0	13.1330	13.1120	0.16%
27.9	13.5974	13.5737	0.17%
31.3	15.3518	15.3179	0.22%

In Example 1, the correlation coefficient $r_{xy} = 0.997$ which is very close to 1. Table (4) shows that $\hat{X}_c$ is almost close to $\hat{X}^+$. The relative errors are all less than 0.5%. Therefore, it is shown that the model in equation (3) could be used to estimate X effectively if $r_{xy} \approx 1$.

Example 2

Table (5) shows the data set in Example 2. In this case, the correlation coefficient r_{xy} is far from 1, that is 0.244. The least squares estimators of the parameters are recorded from SPSS output.

Table 5: Data Set for Example 2

X	4	4	5	5	6	6	7	7	8	8
Y	5.2	5.6	8.1	7.9	3.2	3.0	1.0	1.2	12.0	12.4

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

where

$$\hat{\beta}_1 = 0.67$$
$$\hat{\beta}_0 = 1.94$$
$$r_{xy} = 0.244$$

Similar to the analysis done in Example 1, *X* is estimated by the model presented in equation (3). Here, for a given value of *Y, X* is estimated as

$$\hat{X}_c = -\frac{\hat{\beta}_0}{\hat{\beta}_1} + \frac{1}{\hat{\beta}_1} Y$$

$$\hat{X}_c = -\frac{1.94}{0.67} + \frac{1}{0.67} Y$$

$$\hat{X}_c = -2.90 + 1.49Y \tag{14}$$

The Table (6) shows the value of *X* using model in equation (14). The values of *Y* is taken from Table (5).

Table 6: Estimated Value of $\hat{X}_c$ for Example 2

Y	5.2	5.6	8.1	7.9	3.2	3.0	1.0	1.2	12.0	12.4
$\hat{X}_c$	4.848	5.444	9.169	8.871	1.868	1.570.	-1.41	-1.112	14.98	15.576

As discussed previously, an inverse regression model can be constructed as

$$\hat{X}^+ = \hat{C}_0 + \hat{C}_1 Y$$

From SPSS output, the values of $\hat{C}_0$ and $\hat{C}_1$ are presented as following

$$\hat{C}_1 = 0.089 \text{ and } \hat{C}_0 = 5.47$$

For the purpose of comparison the value of *X* is estimated by the inverse regression model

$$\hat{X}^+ = 5.47 + 0.089Y \tag{15}$$

Table 7: Estimated Value of $\hat{X}_c$ for Example 2

Y	5.2	5.6	8.1	7.9	3.2	3.0	1.0	1.2	12.0	12.4
$\hat{X}_c$	5.933	5.968	6.191	6.173	5.755	5.737	5.559	5.577	6.538	6.574

Table (8) below shows the comparison between the values of and for the corresponding value of *Y*.

Table 4: Comparison between $\hat{X}_c$ and $\hat{X}^+$ for Example 2

Y	$\hat{X}_c$	$\hat{X}^+$	Relative error = $\left\lvert \frac{\hat{x}_c - \hat{x}^+}{\hat{x}^+} \right\rvert \times 100\%$
5.2	4.848	5.933	18.285%
5.6	5.444	5.968	8.786%
8.1	9.169	6.191	48.104%
7.9	8.871	6.173	43.704%
3.2	1.868	5.755	67.540%
3.0	1.570	5.737	72.634%
1.0	-1.410	5.559	125.364%
1.2	-1.112	5.577	119.940%
12.0	14.980	6.538	129.122%
12.4	15.576	6.574	136.948%

The relative errors presented in the table above are very high. It shows that the values of $\hat{X}_c$ are very different from $\hat{X}^+$. In this example, the r_{xy} is 0.244, that is very far from 1. Therefore, it is shown that the model in equation (3) is not suitable when the value of r_{xy} is not very close to 1.

Conclusion

The prediction model $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ is constructed when the dependent variable *Y* is regressed against independent variable *X*. However, inverse regression model is required when we want to predict the value of *X* based on the known value of *Y*. Through the discussion in this paper, the inverse regression model can be formed from the direct modification of the prediction model only if the correlation coefficient r_{xy} is very close to 1. Nevertheless, further discussion on the accuracy of the inverse regression model is needed.

References

John O. R., Sastry G.P. & David A. D. (1998). *Applied Regression Analysis- A Research Tool,* New Jersey: Spring-Verlag.

Robert V. H. & Johannes L. (1992). *Applied Statistics For Engineers and Physical Scientists,* Singapore: Macmillan Publishing Company.

Thomas P. R. (1997). *Modern Regression Methods,* New York: John Wiley & Sons, Inc.

William M. & Terry S. (2003). *Regression Analysis: A Second Course in Statistics*, Prentice Hall.

NG SET FOONG & TEOH SIAN HOON, Faculty of Science and Mathematics, UiTM Penang