Một phương pháp chung để tính các kiểu hồi-quy thẳng giữa hai biến-số

Tô Đồng

Tóm lược

Khoảng mấy thập niên gần đây, việc so sánh kết qủa của một chất sinh- học do hai cách định-phân khác nhau hay do hai dạng của cùng một phương pháp đã trở thành vô cùng quan trọng và khó khăn. Sự kiện này cần thiết để lượng định và kiểm chứng giá trị của các kỹ thuật thử nghiệm hay trị liệu, nhất là khi các kỹ thuật ấy xuất hiện quá nhiều trên thị trường và cần duyệt xét lại. Người ta phải so sánh kết qủa y của một phương pháp đang tìm kiếm hay đang hoàn chỉnh với kết qủa x của một phương pháp hiện đang được dùng và đã được cơ quan y tế có thẩm quyền chấp thuận. Thông thường sự liên-hệ giữa y, biến-số phụ-thuộc, và x, biến-số độc-lập, là một đường thẳng. Theo cách cổ điển, đường chuẩn định hay hồi-quy y = ax + b này được tạo ra phải có khoảng cách sai biệt bình phương nhỏ nhất. Khoảng cách này là sai-số giữa y đo-được và y tính-được. Điều này chỉ có giá trị nếu trị số của x thật đúng. Trên thực tế, cả y và x đều có sai-số, nên cách tính cổ điển này đã được thay thế bởi các kiểu hồi-quy chính xác hơn giữa hai biến-số y và x, nay được coi như những biến-số phụ-thuộc. Tuy nhiên, cách tính mới phải nhờ vào sự đa-năng và kỳ-diệu của máy điện-toán. Bài này từ cách cổ điển căn bản trong thống kê, như về độ lệch tiêu chuẩn, khoảng tin cậy, tính ra những kiểu hồi-quy như hồi-quy Deming, hồi-quy thẳng góc, bằng hình-học giải-tích hay lượng-giác-học. Trên thực tế, các công thức Deming hay thẳng góc được áp dụng để phân-tích các dữ-kiện lâm sàng của hàng trăm hay hàng ngàn cặp y, x. Những trắc-nghiệm này thường được ấn định theo định-luật liên-hệ của ngành thống-kê học. Để dễ dàng theo dõi cách tính này, một thí-dụ số-học rất đơn giản với tám cặp y, x đã được trình bày trong phần phụ-lục, cùng với những thảo chương điện toán liên hệ.

Thuật Ngữ:

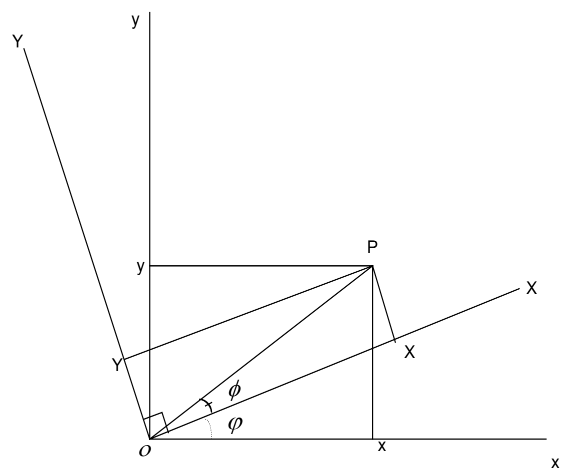
Regression	= Sự hồi-quy, sự chuẩn định
Regression linear	= Sự hồi-quy thẳng, sự chuẩn định thẳng
Orthogonal regression	= Sự hồi-quy thẳng góc
Deming regression	= Sự hồi-quy Deming
Error	= Sai số
Variable	= Biến số
Dependent variable	= Biến số phụ thuộc
Independent variable	= Biến số độc lập
Standard deviation	= Độ lệch tiêu chuẩn
Confidence interval	= Khoảng tin cậy
Variance	= Phương sai
Correlation coefficient	= Hệ số tương quan
Slope a	= Độ dốc a
y-intercept b	= Điểm cắt y tại b, tung độ cắt b

Unified Calculation Method for Linear Regressions

Dong To. D. Sc.

I. Introduction:

The classical ordinary least squares (OLS) method, for regression of dependent variable y vs. independent variable x, can be applied for all linear regression models. If both y and x are having constant gaussian error, the Deming regression of type dependent/dependent variables should be considered. In that case, the rotation of axes of an angle j determined by the ratio of standard deviations of x and y, tanj = sx /sy, restores Deming regression to an OLS model. With Orthogonal regression minimizing the sum of squares of the perpendicular distance from data point P(yi, xi) to the regression line, tanj equals to the slope of the OLS model. The rotation angle j described in the following figure, is used in the calculation of all error-in-variables regression models by back conversion of established formulae of the OLS method.



II. Conversion between systems:

The equation of all linear regressions of y vs. x or y = a + bx, will become after the rotation of axes as stated in the preceding paragraph, Y = A + BX. In that new coordinates system of notation, uppercase alphabets are used to distinguish with the corresponding lowercase alphabets of the original coordinates system. For simplicity in the formula description, the subscripts i=1 to n determinations of all summation operators are omitted.

The Deming method minimizes the sum of squared distances from data points to the regression line, in the direction defined by the angle j. Therefore, in the new system, with the regression line Y = A + BX, the sum of the squares d2 of the deviations parallel to Y-axis equals to:

$$\Sigma\, d^2 \quad = \Sigma\, (A+BX-Y)^2$$

This sum is a function F of A and B, or $F(A, B) = \Sigma\,(A+BX-Y)^2$. The conditions for F being a minimum (or maximum) are $\partial F/\partial A = 0$ and $\partial F/\partial B = 0$.

Since $\quad \partial F/\partial A = \Sigma\, \partial F/\partial A(A+BX-Y)^2 = \Sigma\, 2(A+BX-Y)$

and $\quad \partial F/\partial B = \Sigma\, \partial F/\partial B(A+BX-Y)^2 = \Sigma\, 2X(A+BX-Y)$

Then $\quad \Sigma\,(A+BX-Y) \quad = 0$

and $\quad \Sigma\, X(A+BX-Y) = 0$

i.e. $\quad \Sigma\, Y \quad = nA+B\Sigma X \qquad (1)$

and $\quad \Sigma\, XY = A\Sigma X + B\Sigma X^2 \qquad (2)$

By induction, the classical normal equations (1) and (2), and any other derived computational formulae of the OLS can be applied to the new system. For example, the standard errors of the estimate $S_{Y.X}$, of the Y-intercept S_A and of the slope S_B are calculated by the usual formulae:

$$S_{Y.X} = \sqrt{\frac{\Sigma Y^2 - A\Sigma Y - B\Sigma XY}{n-2}} \qquad (3)$$

$$S_A = S_{Y.X}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{(n-1)S_X^2}} \qquad (4)$$

$$S_B = \frac{1}{\sqrt{n-1}}\frac{S_{Y.X}}{S_X} \qquad (5)$$

All preceding equations could be solved by converting the new coordinates X and Y to the original coordinates x and y, using:

$$X = p\cos\theta$$
$$Y = p\sin\theta$$
$$x = p\cos(\theta + \varphi)$$
$$y = p\sin(\theta + \varphi)$$

or

$$x = p\cos\theta\cos\varphi - p\sin\theta\sin\varphi = X\cos\varphi - Y\sin\varphi$$
$$y = p\sin\theta\cos\varphi + p\cos\theta\sin\varphi = X\sin\varphi + Y\cos\varphi$$

Therefore, it is sufficient to apply all OLS formulae for the new system of coordinates by replacing X and Y with the following identities:

$$X = x\cos\varphi + y\sin\varphi$$
$$Y = -x\sin\varphi + y\cos\varphi$$

<u>1. Slope and y-intercept:</u>

The first two equations (1) and (2) providing the y-intercept A and the slope B of the desired regression line could be converted to:

$$\Sigma(-x\sin\varphi + y\cos\varphi) = nA + B\Sigma(x\cos\varphi + y\sin\varphi)$$

or
$$nA + B(\cos\varphi\Sigma x + \sin\varphi\Sigma y) + \sin\varphi\Sigma x - \cos\varphi\Sigma y = 0 \qquad (6)$$

and
$$\Sigma(x\cos\varphi + y\sin\varphi)(-x\sin\varphi + y\cos\varphi)$$
$$= A\Sigma(x\cos\varphi + y\sin\varphi) + B\Sigma(x\cos\varphi + y\sin\varphi)^2$$

or
$$A(\Sigma x\cos\varphi + \Sigma y\sin\varphi) +$$
$$B(\cos^2\varphi\,\Sigma x^2 + 2\sin\varphi\cos\varphi\,\Sigma xy + \sin^2\varphi\,\Sigma y^2)$$
$$+(\sin^2\varphi - \cos^2\varphi)\Sigma xy + \sin\varphi\cos\varphi\,(\Sigma x^2 - \Sigma y^2) = 0 \qquad (7)$$

Since the solution of equations (6) and (7) gives A, B, the values of a and b will be derived by back conversion as follows, by matching the equation Y = A + BX:

$$-x\sin\varphi + y\cos\varphi = A + B(x\cos\varphi + y\sin\varphi)$$

or

$$y = \frac{A}{\cos\varphi - B\sin\varphi} + \frac{B\cos\varphi + \sin\varphi}{\cos\varphi - B\sin\varphi}\,x$$

with its equivalent y = a + bx,

$$a = \frac{A}{\cos\varphi \text{ - } B\sin\varphi} \tag{8}$$

$$b = \frac{B\cos\varphi + \sin\varphi}{\cos\varphi \text{ - } B\sin\varphi} \tag{9}$$

2. Standard errors:

An element of (3) to calculate the standard error of the estimate $S_{Y.X}$ is:

$$G = \Sigma\, Y^2 - A\Sigma Y - B\Sigma XY$$

or

$$G = \Sigma(-x\sin\varphi + y\cos\varphi)^2 - A\Sigma(-x\sin\varphi + y\cos\varphi)$$
$$- B\Sigma(x\cos\varphi + y\sin\varphi)\,(-x\sin\varphi + y\cos\varphi)$$
$$G = \sin^2\varphi(\Sigma x^2 + B\Sigma xy) + \cos^2\varphi(\Sigma y^2 - B\Sigma xy) +$$
$$\sin\varphi\cos\varphi(B\Sigma x^2 - B\Sigma y^2 - 2\Sigma xy) + \sin\varphi\, A\Sigma x - \cos\varphi\, A\Sigma y$$

Then

$$S_{Y.X} = \sqrt{\frac{G}{n - 2}} \tag{10}$$

To calculate the standard errors for A and B, the standard deviation of X or S_X must be calculated first, using the classical definition:

$$S_X = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n - 1}} \tag{11}$$

If $\quad H = \Sigma\,(X - \overline{X})^2,$

then $\quad H = \Sigma\,(X^2 - 2X\overline{X} + \overline{X}^2)$

$$H = \Sigma\, X^2 - 2n\overline{X}^2 + n\overline{X}^2$$

$$H = \Sigma X^2 - n\overline{X}^2$$
$$H = \Sigma X^2 - (\Sigma X)^2/n$$
$$H = \Sigma (x\cos\varphi + y\sin\varphi)^2 - (\Sigma(x\cos\varphi + y\sin\varphi))^2/n$$
$$H = \Sigma(x^2\cos^2\varphi+y^2\sin^2\varphi+2xy\sin\varphi\cos\varphi) - (\Sigma x\cos\varphi+\Sigma y\sin\varphi)^2/n$$

or
$$H = \Sigma (x^2\cos^2\varphi + y^2\sin^2\varphi+ 2\, xy\sin\varphi\cos\varphi) - [(\Sigma x)^2\cos^2\varphi + (\Sigma y)^2\sin^2\varphi + 2\Sigma x\, \Sigma y\, \sin\varphi\, \cos\varphi]/n$$
$$H = \sin^2\varphi(\Sigma y^2 - (\Sigma y)^2/n) + \cos^2\varphi(\Sigma x^2 - (\Sigma x)^2/n) + 2\sin\varphi\cos\varphi\, (\Sigma xy - \Sigma x\, \Sigma y /n)$$

Finally:

$$S_X = \sqrt{\frac{H}{n-1}} \tag{12}$$

As with S_x , the standard deviation of Y or S_Y can be calculated by the classical definition:

$$S_Y = \sqrt{\frac{\Sigma(Y-\overline{Y})^2}{n-1}} \tag{13}$$

If
$$K = \Sigma (Y - \overline{Y})^2,$$

then
$$K = \Sigma Y^2 - (\Sigma Y)^2/n$$
$$K = \Sigma (-x\sin\varphi + y\cos\varphi)^2 - (\Sigma(-x\sin\varphi + y\cos\varphi))^2/n$$
$$K = \Sigma(x^2\sin^2\varphi+y^2\cos^2\varphi-2xy\sin\varphi\cos\varphi) - (-\Sigma x\sin\varphi+\Sigma y\cos\varphi)^2/n$$

or
$$K = \Sigma (x^2\sin^2\varphi + y^2\cos^2\varphi - 2\, xy\sin\varphi\cos\varphi) - [(\Sigma x)^2\sin^2\varphi + (\Sigma y)^2\cos^2\varphi - 2\Sigma x\, \Sigma y\, \sin\varphi\cos\varphi]/n$$
$$K = \sin^2\varphi(\Sigma x^2 - (\Sigma x)^2/n) + \cos^2\varphi(\Sigma y^2 - (\Sigma y)^2/n) + 2\sin\varphi\cos\varphi(- \Sigma xy + \Sigma x\, \Sigma y /n)$$

Note that the formula giving S_y equals to that of S_x when replacing $x = -x$, $y = y$, $\cos\varphi = \sin\varphi$ and $\sin\varphi = \cos\varphi$.

Finally:

$$S_Y = \sqrt{\frac{K}{n-1}} \tag{14}$$

The covariance of X and Y , S_{XY} can be calculated as:

$$S_{XY} = \frac{\Sigma XY - (\frac{\Sigma X \Sigma Y}{n})}{n-1} \tag{15}$$

The main components in the formula (15) are:

$$\begin{aligned}
\Sigma XY &= \Sigma(x\cos\varphi + y\sin\varphi) \times (-x\sin\varphi + y\cos\varphi) \\
&= \Sigma(-x^2\cos\varphi\sin\varphi + xy\cos\varphi^2 - xy\sin\varphi^2 + y^2\sin\varphi\cos\varphi) \\
&= \cos\varphi\sin\varphi\,(-\Sigma x^2 + \Sigma y^2) + (\cos\varphi^2 - \sin\varphi^2)\Sigma xy \\
\Sigma X\Sigma Y &= \Sigma(x\cos\varphi + y\sin\varphi) \times \Sigma(-x\sin\varphi + y\cos\varphi) \\
&= (\Sigma x\cos\varphi + \Sigma y\sin\varphi) \times (-\Sigma x\sin\varphi + \Sigma y\cos\varphi) \\
&= \cos\varphi\sin\varphi\,(-(\Sigma x)^2 + (\Sigma y)^2) + (\cos\varphi^2 - \sin\varphi^2)\Sigma x\Sigma y
\end{aligned}$$

Knowing S_{YX} and S_X , the values of S_A and S_B can be calculated from equations (4) and (5).

3. Confidence Intervals:

The confidence intervals around the y-intercept and the slope can be constructed as A $\pm$ t(1-α/2, n-2) $\times$ S_A and B $\pm$ t(1-α/2, n-2) $\times$ S_B. All corresponding values in the old system can be calculated respectively by analytical geometry or by trigonometric conversion. For example, y-intercept a and slope b calculated from A and B by (8) and (9) are identical to:

$$a = A\sin\varphi \times \tan(\varphi + \operatorname{atan}B) + A\cos\varphi$$

$$b = (B + \tan\varphi) / (1 - B \times \tan\varphi).$$

The same method is used for calculation of S_a , S_b, as well as for related intervals.

4. Correlation coefficients:

The correlation coefficients R in the new system can be calculated as:

$$R = \frac{S_{XY}}{S_X \quad S_Y} \tag{16}$$

They are dependent on new coordinates system or rotation angle φ and therefore becoming variable with regression models. On the contrary, the correlation coefficients r in the original coordinates system is invariable to any model. An equivalent formula can be used:

$$r = \frac{n\Sigma xy - \Sigma x\Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2]\,[n\Sigma y^2 - (\Sigma y)^2]}} \tag{17}$$

III. Discussion:

This simple rotation approach provides an unified method to calculating linear regressions using all established formulae of ordinary least squares method. In that OLS model, a constant gaussian error of the dependent y variable and an error-free of the independent x variable are assumed. By analogy or by induction, all derived equations can be applied for other types of regressions: the Deming model in which measurement errors are observed for both y and x, and the Orthogonal model where both y and x are speculated as having the same error. In another word, only the direction to calculate least squares must be defined to work with the error-in-variables regression models.

The values calculated by this rotation method are identical to that obtained by Deming with the ratio of variances λ as weight factor. Although there are many definition/calculation methods related to Orthogonal regressions, the results are almost the same if the ratio σ_x / σ_y of standard deviations of x and y equals 1, the weight factor λ=1 or the rotation angle $\varphi = 45^0$.

Useful information could be drawn from the new coordinates system generating new values for standard errors and correlation coefficients. For example, the correlation coefficients R in the temporary new system are becoming alterable with types of regression, while the coefficients r in the old system using same coordinates are invariable to OLS, Deming and Orthogonal models. It is then possible to look for the highest correlation by changing the rotation angle φ, i.e. the direction of least squares, without any assumption of error inherent to x and y. By induction, this angle giving the highest correlation reflects the relative precision of the two supposedly equivalent assay methods giving x and y results.

The ranges constructed around slope and y-intercept with this method, assuming homoscedasticity, are more conservative than estimates using computer-intensive methods such as jackknife or maximum likelihood solutions, but deviations are probably acceptable. However, this estimation method provides consistency of scientific reasoning about regressions.

All short coding statistical methods for regression are applicable in this approach, and various programs can be applied. For simplicity, an example of a data set defined by Σx, Σx^2, Σy, Σy^2, Σxy, σ_x , σ_y and having 8 paired data points is given. Appendix 1 provides results of all regression models with detail. In addition, the OLS regression line of x on y as the reverse of y vs. x is equally presented, noting that a, b will be -A, -B by transformation. Appendix 2 represents a SAS program applicable to all data sets. It is obvious that Deming is the best amongst known linear regression methods applied to this example. The highest correlation, however, is found at a rotation angle of 63.43 ^{o}, corresponding to a new correlation coefficient R of -0.9643. Appendix 3 presents Excel graphs of all three regressions.

IV. Conclusion:

A unified method to calculate all error-in variables regression models is proposed. It can make use of all formulae of the classical ordinary least squares method, for linear regression of dependent variable y vs. independent variable x. The conversion between systems is based on analytical geometry and/or trigonometry and provides consistency of scientific reasoning about regressions.

V. References:

1. Incorrect least-squares regression coefficients in method-comparison analysis.
   P. Joanne Cornbleet and Nathan Gochman.
   Clin. Chem. : 25/3, 432-438 (1979).

2. A new biometrical procedure for testing the equality of measurements from two different analytical methods.
   H. Passing.
   J. Clin. Chem. Clin. Biochem: 21, 709-720 (1983).

3. Fitting straight lines when both variables are subject to error.
   John Mandel.
   Journal of Quality Technology: 16/1, 1-14 (1984).

4. Estimation of the linear relationship between the measurements of Two Methods with Proportional Errors.
   Kristian Linnet.
   Statistics in Medicine: 9, 1463-1473 (1990).

5. The performance of Deming regression analysis in case of misspecified analytical error in method comparison studies.
   Kristian Linnet.
   Clin. Chem. : 44/5, 1024-1031 (1998).

6. Beyond Anova. Basics of Applied Statistics.
   Rupert G. Miller, Jr.
   First Edition, Chapman & Hall (1997).

Appendix 1

Numerical example

Data set ABC

x:	y:
100	35
45	29
44	22
35	18
30	14
25	12
20	9
15	8

Data characteristics

$\Sigma x=$	314
$\Sigma y=$	147
$\Sigma x^2=$	17336
$\Sigma y^2=$	3359
$\Sigma xy=$	7423
$S_X=$	26.7568
$S_y=$	9.6944
n=	8

OLS

Tan (rotation angle) = 0.0000

Rotation angle (radian) = 0.0000

Normal equations:

$$A + 39.2500\,B - 18.3750 = 0$$

$$A + 55.2102\,B - 23.6401 = 0$$

Results:

n =	8		
r =	0.9105		
Slope b =	0.3299		
Y-int a =	5.4268		
b range =	0.1802	to	0.4795
a range =	–1.540	to	12.394

Deming

Tan (rotation angle) = 2.7600

Rotation angle (radian) = 1.2232

Normal equations:

$$A + 30.6464\,B + 30.6431 = 0$$

$$A + 39.7097\,B + 41.5086 = 0$$

Standard errors:

$S_{Y.X}$ = 6.5024

S_B =	0.1379		
S_A =	4.8121		

Confidence ranges:

t =	2.4469		
B range =	–1.536	to	–0.861
A range =	–5.677	to	17.872

Results:

n =	8		
r =	0.9105		
R =	–0.9625		
Slope b =	0.3623		
Y-int a =	4.1541		
b range =	0.2335	to	0.5622
a range =	–3.868	to	12.176

Orthogonal

Tan (rotation angle) = 0.3299

Rotation angle (radian) = 0.3186

Normal equations:

$$A + 43.0307\,B - 5.1536 = 0$$
$$A + 59.2050\,B - 5.2508 = 0$$

Standard errors:

$S_{Y.X}$ =	4.1077		
S_B =	0.0550		
S_A =	2.7786		

Confidence ranges:

t =	2.4469		
B range =	–0.129	to	0.1407
A range =	–1.904	to	11.694

Results:

n =	8		
r =	0.9105		
R =	0.0445		
Slope b =	0.3366		
Y-int a =	5.1647		
b range =	0.1930	to	0.4935
a range =	–2.009 to	12.338	

OLS (x vs. y)

Tan (rotation angle) = 1.0e+39

Rotation angle (radian) = 1.5708

Normal equations:

$$A + 18.3750\,B + 39.2500 = 0$$
$$A + 22.8503\,B + 50.4966 = 0$$

Results:

n =	8

r =	0.9105		
Slope b =	2.5130		
Y-int a =	–6.9267		
b range =	1.3730	to	3.6531
a range =	–30.287 to	16.434	

Appendix 2
Statistical Analysis System

1. SAS program

```
/* DATA INPUT                                                                   */
DATA REG;
INPUT X Y;
DATALINES;
100 35
45 29
44 22
35 18
30 14
25 12
20 9
15 8
;
RUN;
PROC PRINT;
TITLE 'DATA SET = ABC';
RUN;


/* DATA SET CHARACTERISTICS                                                     */
DATA REG2;
SET REG;
XX =    X*X;
YY =    Y*Y;
XY =    X*Y;
RUN;
PROC MEANS NOPRINT;
VAR X Y XX YY XY;
OUTPUT OUT=REG3
N = N SUM=SUMX SUMY SUMXX SUMYY SUMXY STD = XSD YSD XXSD YYSD XYSD;
RUN;


/* CALCULATION OF SLOPE B AND Y-INTERCEPT A IN THE NEW COORDINATES              */
DATA REG4;
SET REG3;
PHI=ATAN (XSD/YSD);
L1 =    N;
M1 =    COS(PHI)*SUMX + SIN(PHI)*SUMY;
N1 =    SIN(PHI)*SUMX - COS(PHI)*SUMY;
L2 =    COS(PHI)*SUMX + SIN(PHI)*SUMY;
M2 =    COS(PHI)**2*SUMXX + 2*SIN(PHI)*COS(PHI)*SUMXY + SIN(PHI)**2*SUMYY;
N2 =    (SIN(PHI)**2 - COS(PHI)**2)*SUMXY + SIN(PHI)*COS(PHI)*(SUMXX-SUMYY);
A =     (M1*N2 - M2*N1)/(L1*M2 - L2*M1);
B =     (L1*N2 - L2*N1)/(L2*M1 - L1*M2);

/* CALCULATION OF STANDARD ERRORS IN THE NEW COORDINATES                        */
G =     SIN(PHI)**2*(SUMXX + B*SUMXY) +
        COS(PHI)**2*(SUMYY - B*SUMXY) +
        SIN(PHI)*COS(PHI)*(B*SUMXX - B*SUMYY - 2*SUMXY) +
        SIN(PHI)*A*SUMX - COS(PHI)*A*SUMY;
SEYX =  SQRT(G/(N-2));
H =     SIN(PHI)**2*(SUMYY - (SUMY)**2/N) +
        COS(PHI)**2*(SUMXX - (SUMX)**2/N) +
        2*SIN(PHI)*COS(PHI)*(SUMXY - SUMX*SUMY/N);
SEX =   SQRT(H/(N-1));
K =     SIN(PHI)**2*(SUMXX - (SUMX)**2/N) +
        COS(PHI)**2*(SUMYY - (SUMY)**2/N) +
```

```
        2*SIN(PHI)*COS(PHI)*(-SUMXY + SUMX*SUMY/N);
SEY =   SQRT(K/(N-1));
XBAR =  (SUMX*COS(PHI)+SUMY*SIN(PHI))/N;
SEA =   SEYX*SQRT((1/N)+XBAR**2/((N-1)*SEX**2));
SEB =   (SEYX/SEX)*(1/SQRT(N-1));


/* CALCULATION OF CONFIDENCE INTERVALS IN THE NEW COORDINATES             */
T = TINV(0.975, N-2);
HIA =   A + TINV(0.975, N-2)*SEA;
LOA =   A - TINV(0.975, N-2)*SEA;
HIB =   B + TINV(0.975, N-2)*SEB;
LOB =   B - TINV(0.975, N-2)*SEB;


/* BACK CONVERSION TO THE ORIGINAL COORDINATES                            */
R =     (N*SUMXY-SUMX*SUMY)/SQRT((N*SUMXX-SUMX**2)*(N*SUMYY-SUMY**2));
YINT =  A/(COS(PHI)-B*SIN(PHI));
SLOPE=  (B*COS(PHI)+SIN(PHI))/(COS(PHI) - B*SIN(PHI));

HA   =  HIA/(COS(PHI)-B*SIN(PHI));
LA   =  LOA/(COS(PHI)-B*SIN(PHI));

HB   =  (HIB*COS(PHI)+SIN(PHI))/(COS(PHI) - HIB*SIN(PHI));
LB   =  (LOB*COS(PHI)+SIN(PHI))/(COS(PHI) - LOB*SIN(PHI));
SA   = SEA/(COS(PHI)-B*SIN(PHI));
SB   = ((B+SEB)*COS(PHI)+SIN(PHI))/(COS(PHI) - (B+SEB)*SIN(PHI)) - SLOPE;
NSUMXY = COS(PHI)*SIN(PHI)*(-SUMXX+SUMYY) +
        (COS(PHI)**2-SIN(PHI)**2) *SUMXY;
        SUMXSUMY = COS(PHI)*SIN(PHI)*(-SUMX**2+SUMY**2) +
        (COS(PHI)**2-SIN(PHI)**2)*SUMX*SUMY;
NSXY =  (NSUMXY - (SUMXSUMY/N))/(N-1);
RNEW =  NSXY/(SEX*SEY);
RUN;
PROC PRINT;
RUN;


/* PRESENTATION OF RESULTS                                                */
DATA REG5;
SET REG4 (KEEP = N R RNEW SLOPE LB HB YINT LA HA);
RUN;

PROC TRANSPOSE OUT = REG6;
RUN;


PROC FORMAT;
VALUE $NNAME
'N      '='NUMBER OF PAIRED X, Y DETERMINATIONS'
'R      '='CORRELATION COEFFICIENT'
'RNEW   '='NEW CORRELATION COEFFICIENT'
'SLOPE  '='SLOPE OF THE REGRESSION LINE'
'HB     '='UPPER BOUND OF THE SLOPE'
'LB     '='LOWER BOUND OF THE SLOPE'
'YINT   '='Y-INTERCEPT'
'HA     '='UPPER BOUND OF THE Y-INTERCEPT'
'LA     '='LOWER BOUND OF THE Y-INTERCEPT';
PICTURE UNIFMT 1      = ' 0.0000'
14-HIGH = '0000';
OPTIONS NODATE NONUMBER;
PROC PRINT NOOBS LABEL DOUBLE;
FORMAT COL1 UNIFMT.;
FORMAT _NAME_ $NNAME.;
LABEL _NAME_  ='PARAMETER'
COL1    = 'VALUE';
TITLE1 'DEMING REGRESSION';
TITLE2 'DATA SET = DATA ABC';
RUN;
```

2. SAS output

```
                    DEMING REGRESSION
                   DATA SET = DATA ABC

PARAMETER                                    VALUE


NUMBER OF PAIRED X, Y DETERMINATIONS             8

CORRELATION COEFFICIENT                    0.91051

Y-INTERCEPT                                4.15409

SLOPE OF THE REGRESSION LINE               0.36232

UPPER BOUND OF THE Y-INTERCEPT             12.1762

LOWER BOUND OF THE Y-INTERCEPT              -3.868

UPPER BOUND OF THE SLOPE                    0.5622

LOWER BOUND OF THE SLOPE                    0.2335

NEW CORRELATION COEFFICIENT                -0.9625
```

Appendix 3
Excel Graphs