ORIGIN ≡ 0

RMA Regression

Prepared by:
Wm Stein

Linear Regression in the "usual" or "Type I" mode, as described by Sokal & Rohlf (*Biometry* 3rd Edition 1995) involves specifying in advance which of two variables is to be considered fixed and indepenent (X) versus dependent with natural variation (Y). This prior assignment is useful for prediction of Y from X, and for calculating confidence intervals, since all variation from a trend line is interpreted to occur in Y alone with a single Normally distributed error term ε. For futher details, I highly recommend reading Sokal & Rohlf's discussion of this in their chapters on Regression and Correlation. Type II regressions, by constrast to Type I, are useful primarily for finding "best fit" trend lines between variables each of which are considered to have natural variation. Rather than prediction of Y from X, the main object is to describe a relationship between variables in a bivariate sense.

Reduced Major Axis (RMA) Regression (also called Geometric Mean (GM) Regression, Standard Major Axis Regression, or Relation d'allometre - see SR p. 544) is ultimately derived from principal axes (as in PCA) of standardized variables. Variables are "standardized" so that each has a mean of zero and stadard deviation of one. Standardization allows use of the the simple methods of calculation for RMA Regression below without going the typical eigenvector/eigenvalue route.

Assumptions:

- **Type II regressions are symmetrical with regard to natural variation.**
  Each variable X & Y is considered to be derived from a population of possible values:
- **$Y_1, Y_2, Y_3, \ldots, Y_n$ is a random sample ~ $N(\mu_Y, \sigma_Y^2)$.**
- **$X_1, X_2, X_3, \ldots, X_n$ is a random sample ~ $N(\mu_X, \sigma_X^2)$**
- **Values of X_i matched to Y_i**

Model:

$Y = \alpha + \beta X$

here: **α is the y intercept of the regression line (translation)**
β is the slope of the regression line (scaling coefficient)
where: each variable has its own error term. This will not be analyzed explicitly since no predictions are formed.

Example: Sokal & Rohlf p. 546 Box 14.12

K := READPRN("c:/RData/SR14.12.txt")

$X := K^{\langle 0 \rangle}$ $\quad Y := K^{\langle 1 \rangle}$ **< assigning variables X & Y from matrix K**

$X_{bar} := \text{mean}(X)$ $\quad X_{bar} = 30.36364$ **< mean for X = first column in K**

$Y_{bar} := \text{mean}(Y)$ $\quad Y_{bar} = 76.54545$ **< mean for Y = second column in K**

$n := \text{length}\left(K^{\langle 0 \rangle}\right)$ $\quad n = 11$ **< number of paired observations**

K =

	0	1
0	14	61
1	17	37
2	24	65
3	25	69
4	27	54
5	33	93
6	34	87
7	37	89
8	40	100
9	41	90
10	42	97

Least Squares Estimation of the Regression Line:

Sums of Squares and Cross Products corrected for mean location:

$i := 0 .. n - 1$

$L_{xx} := \sum_i \left(X_i - X_{bar}\right)^2$ $\quad L_{xx} = 932.5455$ **< corrected Sum of squares of X**

$L_{yy} := \sum_i \left(Y_i - Y_{bar}\right)^2$ $\quad L_{yy} = 4188.7273$ **< corrected Sum of squares of Y**

$L_{xy} := \sum_i \left(X_i - X_{bar}\right) \cdot \left(Y_i - Y_{bar}\right)$ $\quad L_{xy} = 1743.8182$ **< corrected Sum of cross products**

Type I "Simple" Regression of Y on X:

Estimated Regression Coefficients for Y = $\alpha + \beta X$:

$b := \frac{L_{xy}}{L_{xx}}$ $b = 1.86996$ **< sample estimate of β**

$a := Y_{bar} - b \cdot X_{bar}$ $a = 19.76682$ **< sample estimate of α**

Estimated Values of Y:

$Y_{hat_i} := a + b \cdot X_i$

Residuals:

$\varepsilon_i := Y_{hat_i} - Y_i$

Mean Square for Error:

$$MSE := \frac{\sum_i (\varepsilon_i)^2}{n - 2}$$

< from ANOVA for Regression Standard Table

Standard Error of slope:

$s_b := \sqrt{\frac{MSE}{L_{xx}}}$ $s_b = 0.3325$ **< See SR p. 467 & 471 for calculation of s_b**

$Y_{hat} =$

	0
0	45.9462
1	51.5561
2	64.6457
3	66.5157
4	70.2556
5	81.4753
6	83.3453
7	88.9552
8	94.565
9	96.435
10	98.3049

$\varepsilon =$

	0
0	-15.0538
1	14.5561
2	-0.3543
3	-2.4843
4	16.2556
5	-11.5247
6	-3.6547
7	-0.0448
8	-5.435
9	6.435
10	1.3049

Type II RMA - Reduced Major Axis Regression:

Estimated Regression Coefficients:

$b_\nu := \sqrt{\frac{L_{yy}}{L_{xx}}}$ $b_\nu = 2.11937$ **< RMA regression estimate of slope**

$a_\nu := Y_{bar} - b_\nu \cdot X_{bar}$ $a_\nu = 12.19378$ **< RMA regression intercept. Note that it differs slightly from that given in SR.**

Estimated Values of Y:

$Y\nu_{hat_i} := a_\nu + b_\nu \cdot X_i$

Residuals:

$\varepsilon_{\nu_i} := Y\nu_{hat_i} - Y_i$

Mean Square for Error:

$$MSE_\nu := \frac{\sum_i (\varepsilon_{\nu_i})^2}{n - 2}$$

Standard Error of slope:

$Y\nu_{hat} =$

	0
0	41.8649
1	48.223
2	63.0586
3	65.1779
4	69.4167
5	82.1329
6	84.2522
7	90.6103
8	96.9684
9	99.0878
10	101.2072

$\varepsilon_\nu =$

	0
0	-19.1351
1	11.223
2	-1.9414
3	-3.8221
4	15.4167
5	-10.8671
6	-2.7478
7	1.6103
8	-3.0316
9	9.0878
10	4.2072

$s_b := \sqrt{\frac{MSE_\nu}{L_{xx}}}$ $s_b = 0.34273$ **< Note that my calculations use residuals in Y calculated from the RMA predicted values, rather than the Type I predicted values. This approach differs from that of SR where values are simply taken from Type I regression above. I'm not sure the value of this.**

Prototype in R:

General commands for Type I Regression:

COMMANDS:

```
> K=read.table("c:/2008Morphometrics/SR14.12.txt")
> K
```

```
  V1  V2
1  14  61
2  17  37
3  24  65
4  25  69
5  27  54
6  33  93
7  34  87
8  37  89
9  40 100
10 41  90
11 42  97
```

< Values of K returned

```
> X=K$V1
> Y=K$V2
> X
[1] 14 17 24 25 27 33 34 37 40 41 42
> Y
[1]  61  37  65  69  54  93  87  89 100  90  97
```

^ values of X & Y returned

```
> lm(Y~X)
Call:
lm(formula = Y ~ X)

Coefficients:
(Intercept)           X
     19.77         1.87
```

< Values returned from lm() call

```
> lsfit(X,Y)

$coefficients
Intercept         X
19.766816  1.869955

$residuals
 [1]  15.05381166 -14.55605381   0.35426009   2.48430493 -16.25560538
 [6]  11.52466368   3.65470852   0.04484305   5.43497758  -6.43497758
[11]  -1.30493274

$intercept
[1] TRUE

$qr
$qt
 [1] -253.8725521   57.1039561   -1.9757844   -0.2955843  -19.9351843
 [6]    5.1460158   -3.1737842   -8.1331841   -4.0925840  -16.4123840
[11]  -11.7321840

$qr
      Intercept            X
 [1,] -3.3166248 -100.70478909
 [2,]  0.3015113   30.53760722
 [3,]  0.3015113    0.08425008
 [4,]  0.3015113    0.05150357
 [5,]  0.3015113   -0.01398945
 [6,]  0.3015113   -0.21046849
 [7,]  0.3015113   -0.24321500
 [8,]  0.3015113   -0.34145453
 [9,]  0.3015113   -0.43969405
[10,]  0.3015113   -0.47244056
[11,]  0.3015113   -0.50518707

$qraux
[1] 1.301511 1.313476

$rank
[1] 2

$pivot
[1] 1 2

$tol
[1] 1e-07

attr(,"class")
[1] "qr"
```

Prototype in R: **Using R to calculate RMA Regression:**

COMMANDS:

```
> K=read.table("c:/2008Morphometrics/SR14.12.txt")
> X=K$V1
> Y=K$V2

> Xc=X-mean(X)
> Xc
[1] -16.363636 -13.363636  -6.363636  -5.363636
[5]  -3.363636   2.636364   3.636364   6.636364
[9]   9.636364  10.636364  11.636364
```

^ Values of X centered on the mean of X

```
> Lxx=sum(Xc*Xc)
> Lxx
[1] 932.5455
```

< Lxx value returned

```
> Yc=Y-mean(Y)
> Lyy=sum(Yc*Yc)
> Lyy
[1] 4188.727
```

< Lyy value returned

< Compare these values with our calculations above!

```
> bv=sqrt(Lyy/Lxx)
> bv
[1] 2.119366
```

< RMA Regression Slope returned

```
> av=mean(Y)-bv*mean(X)
> av
[1] 12.19378
```

< RMA Regression Intercept returned

```
> Yvhat=av+bv*X
> Yvhat
[1]  41.86491  48.22301  63.05858  65.17794
[5]  69.41668  82.13287  84.25224  90.61034
[9]  96.96844  99.08781 101.20717
```

< Estimated values of Y

```
> ev=Yvhat-Y
> ev
[1] -19.135086  11.223013  -1.941422  -3.822056
[5]  15.416677 -10.867125  -2.747759   1.610340
[9]  -3.031560   9.087806   4.207172
```

< Residuals