

# User's guide to (S)MATR: Standardised Major Axis Tests & Routines

## Version 1.0, Copyright 2003

Programmed by: Daniel S Falster 2003, [dfalster@rna.bio.mq.edu.au](mailto:dfalster@rna.bio.mq.edu.au)

Other contributors: David I Warton, Ian J Wright

Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia.

<http://www.bio.mq.edu.au/ecology/SMATR/>

### Background

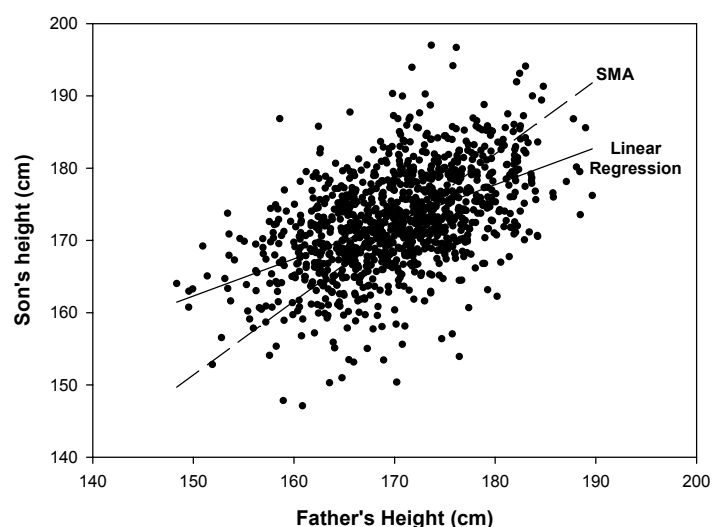
Bivariate relationships assuming a linear relationship between two variables are widely used in the biological sciences. Most frequently, linear regression (or ordinary least squares regression, model I regression) is invoked. Linear regression assumes:

1. The Y variable is normally distributed with constant variance
2. There exists a linear relationship between Y and X of the form:  $Y = \alpha + \beta x + \varepsilon$
3. That the purpose of the line is to predict Y from given values x.

A line is fitted to observed data that minimises the distance between the observed data points and the predicted line in the Y dimension only. Such lines are useful when one is interested in predicting one variable (y) from another (x).

Standardised major axis estimation (SMA) involves a similar procedure; however, residuals from the line are estimated in both the X and Y dimensions. SMA is appropriate when the purpose is to estimate the relationship between two variables, usually when the slope of the relationship is of primary interest. A linear regression is inappropriate for these situations because the slope of a linear regression is always flatter than expected, considering the theoretical relationship between two variables. This is known as “regression to the mean”, and is the property which gives regression its name (Galton 1886, Pearson and Lee 1903). Regression to the mean is entirely appropriate for prediction, but is misleading when the purpose is to estimate the line-of-best-fit relating two variables. For example, if two variables are scattered around the one-to-one line (i.e. the line-of-best-fit has a slope of 1), the linear regression slope will be flatter than 1 (and how much flatter it is depends how tightly the variables are correlated). If a large number of samples were taken, the regression slope would remain flatter than 1, and would not approach 1, despite the fact that the data are evenly scattered around a line of slope 1 (figure 1).

**Figure 1:** The term “regression” originates from early investigations into inheritance of body-size among humans (Galton 1886, Pearson and Lee 1903). The data are heights of 1078 father-son pairs, with a correlation of  $r = 0.50$ . A linear regression fitted to the data has a slope of 0.51 and can be used to predict the average height of a son from a father. Since the slope is  $< 1$ , tall fathers are predicted to have on average shorter sons, and short fathers taller sons. In contrast, the SMA through the data has a slope of 1.02, indicating that across the population tall fathers are associated with proportionately taller sons and vice versa.



The Standardised Major Axis (or SMA, also known as Reduced Major Axis) is a variation on the major axis (MA), which gets its name from being the major or principal axis through the ellipse fitted to bivariate normal data (McArdle 1988, Sokal and Rohlf 1995). The standardised major axis is the major axis fitted to standardised data, which is then rescaled to the original axes. A SMA slope is easily calculated as the ratio of the standard deviations of y and x, or equivalently, as the linear regression slope divided by r, the correlation coefficient. Consequently SMA slopes tend to be steeper than model I slopes.

Several situations may arise where the slope is of interest. Most common are biological scaling (allometric) relationships. Here we are interested in how one variable scales with another. Scaling relationships are typically of the form:  $Y = \alpha X^\beta$ , where  $\beta$  is the scaling exponent or slope of the relationship (Huxley 1932, McMahon 1973, Niklas 1994, West et al. 1997, Enquist 2002).

Conveniently, the relationship is linear after log transformation:

$$\log Y = \log \alpha + \beta \log X$$

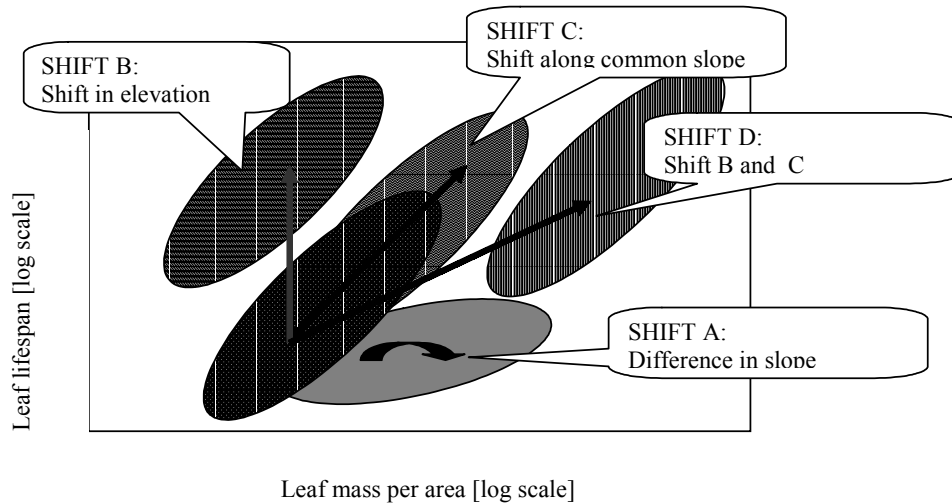
The scaling exponent  $\beta$  can be estimated as the SMA slope for  $\log y$  vs  $\log x$ . Further, if  $\beta$  is not significantly different to 1, then the relationship between the two variables is said to be isometric (i.e. a doubling of variable is associated with a doubling in the other). Importantly, if  $\beta \neq 1$ , then a graph between y and x without log transformations on both axes will be non-linear.

In some situations, researchers may be interested in comparing bivariate functional relationships among two or more ‘groups’ of data. For example, Wright et al. (2002) compared the relationship between leaf mass per area and leaf lifespan among groups of coexisting species from different sites contrasted on rainfall and nutrients (Figure 2). The first question asked in the comparison of estimated lines is whether a common slope is observed among the groups. Until recently statistical algorithms were not available for comparing SMA lines among groups of data. The purpose of this program is to provide a toolkit for researchers to compare SMA lines among groups, using algorithms recently derived by Warton and Weber (2002). Three relevant tests are possible (illustrated in figure 2):

1. Test whether groups differ in the slope of a relationship between two variables (Fig 2 - shift A)
2. Given a common slope, test whether groups are vertically separated, i.e. test the fitted lines differ in elevation (Fig 2 - shift B).
3. Given a common slope and elevation, test whether groups are separated along the SMA axis (Fig 2- shift C).

Figure 3 outlines the series of questions that can be asked to arrive at the various outcomes. As an add-on, the program also provides a test for comparing slopes fitted to individual groups to a nominated value.

The program uses the likelihood ratio developed by Warton and Weber (2002) to test for a common slope among any number of groups and datapoints. If no significant difference in slope can be detected tests for shifts in elevation and along the common slope are accomplished using standard ANOVA routines (Sokal and Rohlf 1995) after appropriate transformations. The equivalent test using linear regression is known as ANCOVA.



**Figure 2.** Schematic illustrating possible differences between groups in leaf lifespan – leaf mass per area relationships. Each oval represents the scatter observed within a given group.

### Assumptions

Several assumptions are required for the calculation of P-values and confidence intervals of the statistics calculated. Different tests involve slightly different assumptions, and some refer specifically to “fitted values” (calculated as  $y + bx$ ) vs “residuals” ( $y - bx$ ), where  $b$  is the estimated slope assuming the null hypothesis (the common slope estimate in multiple sample tests, or the hypothesised slope in one-sample tests). The fitted values measure how far along the SMA axis a point is, and the residuals measure how far away from the SMA axis the point is. The five types of assumptions are:

1. *Independence of observations*: All datapoints should come from independent observations, which can be guaranteed through random sampling
2. *Linearity*: A linear relationship is assumed to exist between X and Y variables.
3. *Equal variance*: In all tests, either the residuals or the fitted values along the SMA axis are assumed to have the same variance for all observations in a fitted line (although in the case of common slope testing, the variance can be different for the different lines).
4. *Normality*: Either residuals or fitted values are assumed to be normally distributed.
5. *Equal variance across groups*: In some tests the variance of residuals/fitted values must be the same across all groups.

The last of these assumptions only applies to some hypothesis tests of SMA axes, and the last three only apply to specific variables. The following table summarises which assumptions are required of which variables for which tests.

test	independence	linearity	equal variance along line	normality	equal variance across groups
one-sample test	yes	yes	residuals	residuals	no
equal slopes	yes	yes	residuals	residuals	no
equal elevations	yes	yes	residuals	residuals	residuals
shift along axis	yes	yes	fitted values	fitted values	fitted values

### Importance of assumptions

These assumptions are similar to those made in linear regression, and the robustness of SMA statistics to failure of these assumptions is similar to the robustness of linear regression (see (Miller 1986) for more information).

Independence of observations is an important assumption, although this assumption is easily satisfied by randomly selecting observations from a larger population. If the dataset involves exhaustively sampling an area, then this assumption is not necessarily satisfied and it would have been better to use a larger study area from which observations could be randomly sampled and independence could reasonably be assumed.

Normality is of little concern except in small sample sizes ( $<20$  total). If  $n > 20$ , the only practical concern regarding normality is that the variable of interest not be long-tailed i.e. not have outliers, and these can be checked for on residual plots.

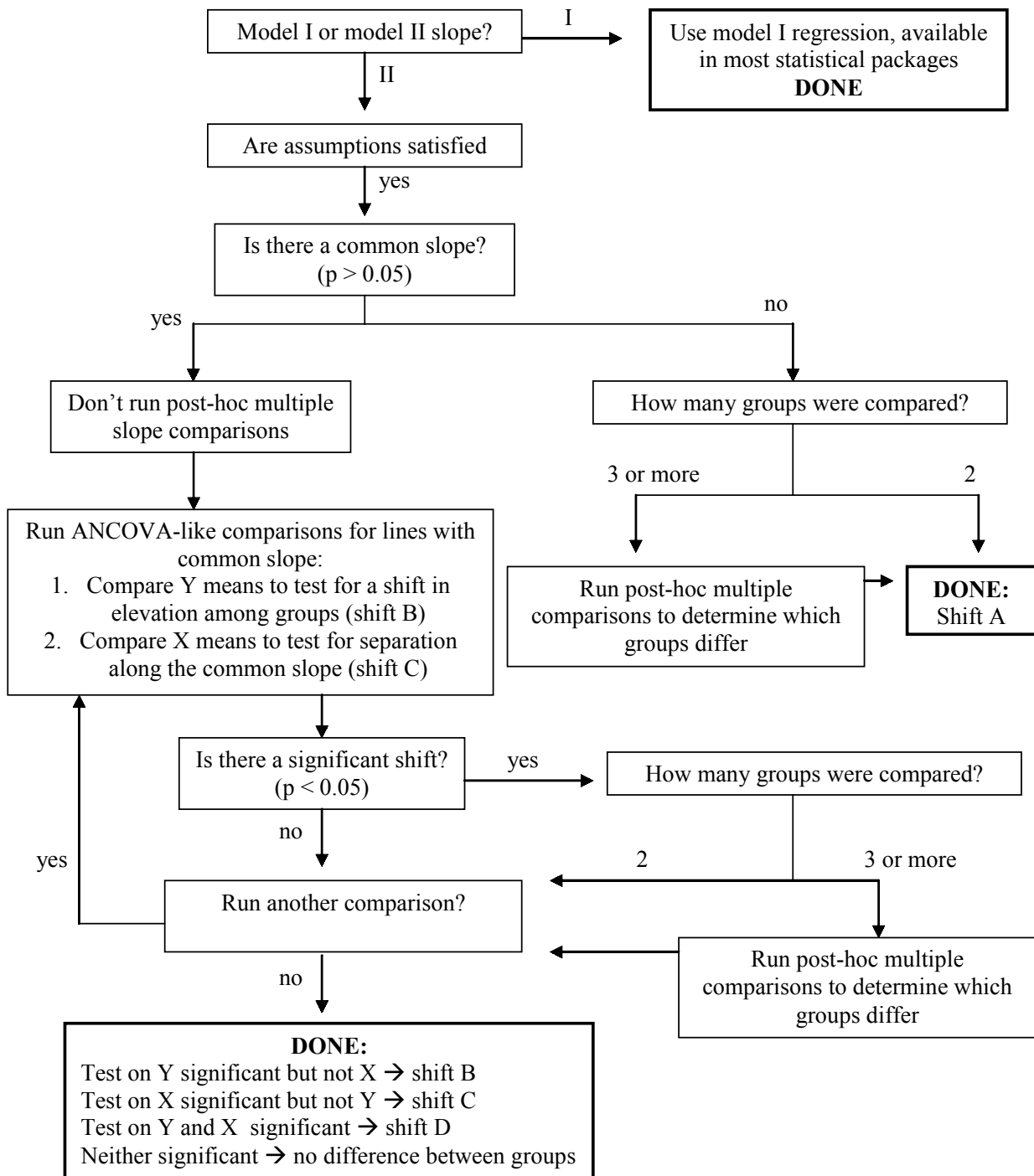
The remaining assumptions – linearity and equal variance – are important and failure of these can invalidate inference using P-values and confidence intervals. The robustness of tests for common slope (both for ordinary linear regression and SMA techniques) are currently under investigation, following our findings that for some allometric datasets, tests for common slope do not appear to have the desired levels of robustness.

### Checking assumptions

Assumptions of linearity, normality and equal variance can be checked informally from a residual plot (a plot of residuals vs fitted values):

1. *Linearity assumption*: check that there is no non-linear trend on the residual plot.
2. *Normality*: check that there are no outlier residuals/fitted values on residual plot.
3. *Equal variance along axis*: there should be similar amounts of scatter in residuals at different fitted values. (e.g. A cone shape in the residual plot suggests that this assumption is not satisfied. Most texts have some schematics showing what it looks like when this assumption is/isn't satisfied.)
4. *Equal variance across different groups*: you could construct box plots of residuals/fitted values (whichever for whichever test) to check that the spread of values is similar for different groups.

**Figure 3:** Schematic diagram illustrating the flow of logic in an investigation comparing model II bivariate trait relationships among two or more groups of data.



## Instructions

The program is a dos-based Windows console program designed to run in a PC Windows environment (compiled on Windows 98). It can be run by double clicking from windows explorer, or using the run command under the start menu. The program may be compiled to run in other operating environments (eg. UNIX, Mac) by downloading the source code from:

<http://www.bio.mq.edu.au/ecology/>

During operation, a series of questions are asked of the user requiring either yes-no answers or numerical input.

**Input** – data can be imported from a tab delimited text file only. This file must be located in the same directory as the program itself. The first line must include variable names. Many variables can be present, including various grouping or filter variables. The file can include rows with missing data, but only where the missing data cell is blank, or has a single space character (standard Excel output). The file can have any name, but the default is for “input.txt”. There are no limits on variable name length, although short are more user-friendly.

**Output** – results are output to screen and file (default name “output.txt”). Only abbreviated results are printed to screen. If multiple procedures are run in a single session, the output of each procedure is printed successively into the one output file if no new name is specified. Alternatively, a new file can be created for each procedure. If the program is restarted, old output files with the same name will be replaced. Output files can be opened in standard text editors (NOTEPAD, WordPad) or in a spreadsheet such as MS Excel.

**Program Flow** – an opening menu prompts for information concerning input/output files, the number of iterations to be run, required confidence intervals, and other details. When complete press L to load data. This opens a second menu, which prompts for specific detail regarding the test to be run:

- **Variables** : enter variable number not name
- **Filter**: several variables can be included as filters; these can be either categorical or scalar. Using filters wisely allows multiple tests to be run without the need to create a new dataset for each test
- **Group**: at least one grouping variable must be included. Several variables may be selected in anyone test. Note: all subgroups will be compared directly; nested grouping structures are not possible.
- **Transformation**: if your data requires log transformation this can be achieved by choosing log transformation for either the X, Y or both axes. If your data is already log-transformed you should choose the default “linear” option.
- **H0 slope**: a slope value against which the fitted slope for each group is compared (default =1). This is useful if you wish to compare slopes against a specified *a-priori* value; however, this option does not affect other comparisons between groups.

To save time in this menu, you can enter successive commands without pressing return, e.g.

gc – clears grouping variables

x1 – selects variable #1 as the x variable

When ready, press L to load the data and run the common slope tests. Following this, you will be prompted as to whether you wish to:

1. Run post-hoc multiple slope comparisons among groups
2. Run “ANCOVA” comparisons among groups with a common slope.

You should use the schematic in Figure 3 to determine whether these tests are necessary.

Eventually you will return to the initial menu screen. You can reload the same data and run an additional analysis by pressing L. Previously selected variables, groups and filter settings are stored between successive analyses to facilitate multiple analyses with similar data. However, if you change the filter variable in a subsequent analysis (eg add an extra filter), the filter options will need to be re-entered (you are prompted for this). In addition, if this filter was previously scalar, the transformed categorical values will be shown instead. Remember also, consecutive analyses will be all output to the same output file, unless a new name is specified.

**File Output** – full output details are only available in the output file. Successive analyses in a given file are delimited by a line break. For each analysis the following is indicated:

- the x and y variables selected, including any transformations used
- the critical p-value, confidence intervals and number of iterations
- any filters variables used, including the relevant inclusion/ exclusion criteria (indicated in brackets: 1 = include, 0 = remove).
- the total number of data points included and removed (due to missing data or filters), and the name of groups with less than the minimum group size

### SMA RESULTS

The next output summarises the SMA regression lines fitted to individual groups, including:

1. Group name, number datapoints,  $R^2$ , and a p-value. The  $R^2$  and p-value are identical to those reported when testing for a significant correlation between two variables
2. The SMA Slope followed by the lower and upper confidence intervals on this value, and the estimated intercept for the regression line. Also indicated is the model I slope for comparison.
3. Ymean, Xmean – the average of the y and x values for each group
4.  $H0\_b$ , F, p: the output of a test comparing the SMA slope of each group to the specified value ( $H0\_b$ ), where F is the F-statistic for the test (with  $df=1, n-2$ ), and p is p-value.  $P < 0.05$  indicates a slope significantly different from  $H0\_b$ .

### TEST FOR COMMON SLOPE ACROSS GROUPS

Next comes the estimated common slope and a p-value for a test of heterogeneity in slopes among groups, as derived by Warton and Weber (2002). If  $p > 0.05$ , there is no evidence that the group slopes are heterogenous and the common slope estimate can be used.

### POST-HOC MULTIPLE COMPARISON OF SLOPES AMONG GROUPS

Post-hoc multiple slope comparisons are only necessary if the test for a common slope indicated significant heterogeneity among groups (i.e.  $p < 0.05$ ). The purpose of the post-hoc tests is to identify which groups differ. This is achieved using the same procedures of estimating a common slope and testing for heterogeneity of slopes, as above, but including only two groups at a time. The output therefore consists of pairwise comparisons between groups, including:

1. A matrix of individual pairwise comparisons. Data in brackets are the common slope estimate for the two groups, the test statistic and p value ( $df$  for all tests = 1).
2. A second matrix listing p-values only
3. A list of the groups ordered by slope, with p-values substituted for an asterix if  $p > \text{critical } p$  (default:  $p_{\text{crit}}=0.05$ ).

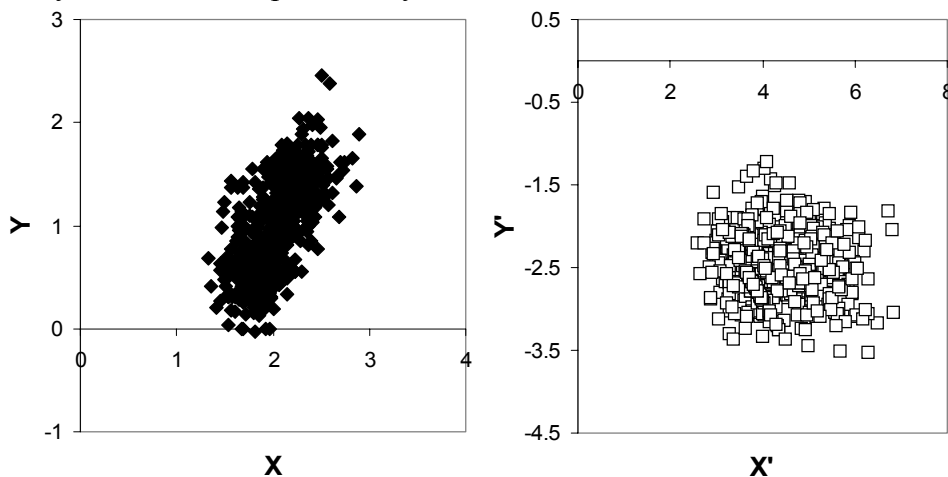
In interpreting the multiple comparisons of SMAs, a correction on the p-values may be required. This is because for any given test, if there are no differences between groups, there is still some chance or “Type I error” (5% for  $p_{\text{crit}}=0.05$ ) of statistical significance. When conducting many tests, the Type I error increases roughly proportionally to the number of tests. So, for example, when comparing 5 groups, there are 10 multiple comparisons, and at  $p_{\text{crit}}=0.05$  there is about a 50%

chance of incorrectly saying that two groups differ. Applying a correction to p-values is a way of keeping the overall (family-wise) Type I error low, by interpreting individual p-values more conservatively (reducing the per-comparison Type I error). Unfortunately, there is no single-best method for overcoming this problem. A particularly conservative approach may be to apply a Bonferroni correction, although other more flexible techniques exist (Day and Quinn 1989). We have not provided any correction here, leaving this decision to the discretion of the user.

### ANCOVA COMPARISON OF GROUP MEANS

ANCOVA-like comparisons, testing for shifts between groups in elevation or along a common slope, are computed using standard ANOVA tables after the data has been transformed to remove the effect of correlation between x and y. The transformed variables (called  $x'$ ,  $y'$ ; Figure 4) are uncorrelated, and reflect the spread of data along ( $x'$ ) and away from ( $y'$ ) the fitted line (see appendix 2 for calculations).

**Figure 4:** Scatter plots indicating the effect of transforming a dataset with a positive correlation ( $r^2=0.43$ ) and slope of 1.71 (left) before testing for shifts in elevation or along the common slope among groups. The transformed variables ( $y'$ ,  $x'$ ) data are uncorrelated (right ;  $r^2=0.00$ ). Variation in  $x'$  corresponds to the spread along the common slope in the initial dataset, while variation in  $y'$  indicates the spread away from the fitted line.



The initial output in this section indicates the common slope together with the grand mean (mean across all data points in all groups) of x and y variables prior to vector rotation. Following this a table is provided summarising the properties of individual groups:

- the intercept of the group: the intercept of the fitted line for each group, assuming the common slope. The equation for each line is then  $Y = \text{intercept} + \text{common\_slope} * X$
- YgrandX: Provides the y value expected for each group at the grand mean of x. This is useful for making comparisons between groups at a common point
- XgrandY: as above, but for x
- Ymean, Xmean: the mean value of y and x in the untransformed data
- DAS: distance along slope – the mean value of  $X'$  for each group, the transformed x data

If testing for shifts in either  $y'$  or  $x'$ , standard ANOVA output is provided. An alternate option allows you to save the rotated data to a file for use in alternate tests (this file includes the initial



data plus the rotated data. The number of rows may differ to your original dataset if missing cells or filtered data were removed.

#### Running post-hoc multiple comparisons

Following a significant ANOVA result (indicating a significant shift in elevation or along the slope) you can run post-hoc multiple comparisons to identify which groups differ. But before this is done, Levene's test for equality of variances (Levene 1960) is given. Post-hoc output using either the Tukey-Kramer method (equal variance) or the Games-Howell method (unequal variance) is then provided in the same format to the post-hoc slope comparisons (Day and Quinn 1989).

**CAUTION:** The ANCOVA routines may be subject to inflated Type I errors under situations where groups are separated along the common slope by a large amount (i.e. they differ in Xmean). Since our estimate of the common slope inevitably differs from the true common slope (this is the case for any population estimate), the transformation prior to ANCOVA testing (depicted in figure 4) will be imperfect. The effect arises because groups of data located further along the common slope (higher Xmean) will have a different mean value for y' to groups further back along the common slope. Consequently, a significant shift in elevation may be reported even if one does not exist. This problem is not unique to the methods presented here and exists for standard ANCOVA testing in linear regression. The severity of the error will depend on the accuracy of the common slope estimate and the magnitude of differences in group means for x. The common slope estimate can be improved with increasing sampling. Otherwise p-values close to 0.05 should be interpreted with caution in instances where (graphically) there is a large shift along the x-axis but apparently no shift in elevation. We are hoping to develop a solution to this problem in the future.

#### Statistical options -

- **Statistic:** Currently only routines for SMA tests are offered. We intend to include MA tests shortly.
- **Minimum group size:** this sets the limit on the number of datapoints below which entire groups are excluded from the analysis. Three is the absolute minimum value recommended. Usually 8 or more points are required for a reasonable estimate of slope, although this depends on the  $r^2$  of the relationship.
- **Iterations:** The common slope statistic used is that recommended by Warton & Weber (2002). However, significance is determined here via permutation testing (details below), not via comparison to the chi-squared distribution as suggested. Subsequent testing by us indicated that chi-squared test was susceptible to unacceptable levels of type II error. This has been overcome by the use of permutation testing. The recommended number of iterations is least 1000 if your critical level is  $p = 0.05$ , or 5000 if your critical level is  $p = 0.01$  (Manly 1997). You can always do more iterations, but the program run time will be greater. Beyond 10000 iterations measurement error will likely cancel out any increased resolution in significance. A good way to get a feel for how much error there is in estimating P-values by permutation is to run the test several times and see how much the P-values change, and whether that affects interpretation (even with 5000 permutations, p moves around a bit).
- **Critical p-value:** this determines the significance level below which a null result is rejected. This influences the construction of tables in post-hoc comparisons.
- **Confidence intervals:** 95% confidence intervals are given on slope estimates for individual groups, and for the common slope. This value can be varied if desired.

## Reporting

Below are some suggested phrases for reporting statistics calculated by the program. These suggestions draw on some published examples from Wright *et al.* (2001) and Wright *et al.* (2002).

### Methods:

- Slopes are referred to as Standardised Major Axis (SMA) slopes, because they equate to a line passing through the major axis of roughly elliptic cloud of standardised data. Other terms that may be used are Reduced Major Axis (RMA) slope or Geometric Mean slope.
- SMA routines and tests were calculated using the program (S)MATR (See below for reference).
- Bivariate trait relationships were analysed by fitting Standardised Major Axis (SMA) lines to log scaled variables. SMA techniques provide a superior estimate of the line summarising the relationship between two variables (i.e. the main axis along which two variables are correlated) to that of ordinary linear regression, because the residual variance is minimised in both x and y dimensions, rather than the y dimension only (McArdle 1988, Sokal and Rohlf 1995).
- The SMA slope, calculated on log-transformed variables, give the proportional relationship between variables. Slopes were first fitted across the species within each group, with confidence intervals (95%) calculated following (Pitman 1939).
- To compare observed relationships among groups, we tested for statistical differences in the slope and intercept of group SMA relationships. An SMA slope common to both groups was estimated following (Warton and Weber 2002), using a likelihood ratio method. The significance of this estimate was determined by testing for significant heterogeneity among group slope estimates by permutation (Manly 1997). After fixing the position of individual points along the estimated common slope, residuals were permuted among groups 1000 times, with the common slope and test statistic recalculated after each iteration. This method is analogous to that proposed by (Freedman and Lane 1983) for linear regression, which has been shown to maintain close to exact significance levels in small samples for linear models (Anderson and Robinson 2001).
- The ability to calculate common slopes allows one to test for elevation (i.e. intercept) differences between individual slopes, as in standard analyses of covariance (ANCOVA). Where significant heterogeneity in group slopes could not be detected, we tested for shifts in elevation and shifts along the common SMA (Wright et al. 2001) by transforming the data such that the common slope was 0, and testing for differences in group means of  $y'$  and  $x'$  using one-sample ANOVA, where  $y'$  and  $x'$  are  $y$  and  $x$  after data is transformed by an amount determined by the common slope  $\beta$  ( $y' = y - \beta x$  &  $x' = y + \beta x$ ).
- Where significant shifts in elevation or along the common slope were detected, post-hoc multiple comparisons were carried out using the Tukey-Kramer method (equal variance among groups) or the Games-Howell method (for unequal variance), as recommended by (Day and Quinn 1989). Both tests maintain an experimentwise significance level of 0.05. (i.e. no further correction is required to allow for the fact that multiple comparisons have been made)

### Results:

- For each group, report the estimated SMA slope, the confidence interval of this slope, the intercept and the  $r^2$  of the relationship
- Where  $p > 0.05$ , group slopes are said to be not significantly heterogeneous and a common slope can be estimated. Report the common slope, confidence intervals and estimated  $r^2$ . Where  $p < 0.05$ , slopes are significantly heterogeneous, and the common slope estimate cannot be used.
- Where a significant shift in elevation or along the common slope is detected, standard ANOVA statistics can be reported ( $F=3.24$ ,  $df = 12$ ,  $p = 0.07$ ).

## Citation

If you appreciate the effort that has gone into providing this software, please credit the application and its authors in any published work that makes use of results from (S)MATR. (S)MATR is an electronic publication found at the following web address:

<http://www.bio.mq.edu.au/ecology/SMATR>

Possible formats for citation, in order of preference are:

1. Falster, DS, Warton, DI, Wright IJ (2003) (S)MATR: Standardised major axis tests and routines. Version 1.0. <http://www.bio.mq.edu.au/ecology/SMATR>

2. (in text citation): "...calculated using (S)MATR (Version 1, Falster DS, Warton DI & Wright IJ <http://www.bio.mq.edu.au/ecology/SMATR>)...."

3. (in text citation): "...calculated using (S)MATR (Version 1, Falster DS, Warton DI & Wright IJ unpublished)...."

We would be most grateful if you would kindly send a reprint of any paper based on your use of the program to [dfalster@rma.bio.mq.edu.au](mailto:dfalster@rma.bio.mq.edu.au) or:

Daniel Falster,  
Ecology Lab, Biological Sciences  
Macquarie University, NSW, 2109 Australia

## Distribution

This program is freeware, provided free of charge. You are welcome to use (S)MATR in any way for your own research, as long as such use is acknowledged as outlined above. By downloading and using (S)MATR, you are agreeing not to distribute (S)MATR in any commercial form.

You are welcome to distribute copies of (S)MATR to other users, but for the benefit of the users we would request that you inform us (D Falster) if you are currently using a copy of the program. This way we can keep users informed of new versions or bugs.

## Support

Please refer to the manual or references first for further statistical support. If you do not find the answer here, please check the website or contact us via email at: [dfalster@rma.bio.mq.edu.au](mailto:dfalster@rma.bio.mq.edu.au)

Bugs can be reported to the same address.

## Programming

(S)MATR is written in C++ and was compiled using the Dev C++ v4.0 compiler for windows available at <http://www.bloodshed.net/>.

## APPENDIX 1: Frequently Asked Questions

### 1. When will routines be available for Major Axis tests?

We hope to include major axis routines shortly. Practically, it is appropriate to use SMA routines for most biological datasets

### 2. Why don't the confidence intervals overlap zero?

An SMA slope is calculated as the ratio of the standard deviation in Y to standard deviation in X. Logically this can never be zero and the default slope for random data (no relationship between X & Y, similar variance in X and Y) is 1. The slope differs from 1 when the data is stretched more along one axis than the other.

### 3. There's only a weak relationship in my data, why are the confidence intervals about the predicted slope so tight?

An SMA reflects the nature of a relationship between two variables, while an  $r^2$  reflects the strength of the relationship. These are two independent properties characterising the relationship between two variables. Narrow confidence intervals are possible, even when  $r^2$  is low, with sufficient sampling effort (i.e. large n). With a large sample size you have increased certainty about both the strength of the relationship ( $r^2$  - even its low you are sure about that) and the nature of the relationship (i.e. slope).

The confidence intervals do not affect any of the statistics testing for a common slope, or the ANCOVA comparisons among groups. They are provided for further interpretation only.

Confidence intervals in (S)MATR are calculated according to (Pitman 1939). Full details are given in appendix 2. This method provides an exact calculation of the required intervals. An alternate method commonly used involves approximating the confidence interval about the estimated slope by applying the same envelope as is calculated in ordinary linear regression (Niklas 1994, Sokal and Rohlf 1995). In practice we have found the latter method gives a significantly wider confidence envelope, especially for weak relationships.

### 4. Can (S)MATR be run on different operating platforms?

Presumably yes, but we haven't tried. (S)MATR is programmed in C++, and given the right compiler it could be compiled under any number of operating environments. Contact D Falster for more information.

## APPENDIX 2: Statistics – details of calculation

The SMA routines presented below are those outlined by Warton & Weber (2002). P-values for the F-distribution were obtained using the stand-alone mathematical library available with the R program (<http://www.r-project.org/>) under the GNU General Public Licence (June 1991; <http://www.gnu.org/copyleft/gpl.html>).

If there are  $g$  groups, and the  $i$ th group contains  $n_i$  observations ( $x_{ij}, y_{ij}$ ) denotes the  $j^{\text{th}}$  observation in group  $i$ . To find the  $i$ th SMA slope we require:

$$Sxx_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = \sum_{j=1}^{n_i} x_{ij}^2 - \frac{1}{n_i} \left( \sum_{j=1}^{n_i} x_{ij} \right)^2 \quad \text{where } \bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad \rightarrow \text{Sum squares in } x$$

$$Syy_i = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \sum_{j=1}^{n_i} y_{ij}^2 - \frac{1}{n_i} \left( \sum_{j=1}^{n_i} y_{ij} \right)^2 \quad \text{where } \bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad \rightarrow \text{Sum squares in } y$$

$$Sxy_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i) = \sum_{j=1}^{n_i} x_{ij}y_{ij} - \frac{1}{n_i} \left( \sum_{j=1}^{n_i} x_{ij} \right) \left( \sum_{j=1}^{n_i} y_{ij} \right) \quad \rightarrow \text{Sum of cross-products}$$

then the SMA slope estimate for group  $i$   $\hat{\beta}_i = \text{sign}(S_{xyi}) \sqrt{\frac{S_{yyi}}{S_{xxi}}}$ , and  $R^2 = \frac{S_{xyi}^2}{S_{xxi} \cdot S_{yyi}}$ ,  $df = n_i - 2$

Upper and lower confidence intervals about the slope are calculated according to (Pitman 1939):

$$\text{Lower } \hat{\beta}_i = \min(\hat{\beta}_i \cdot (\sqrt{B+1} - \sqrt{B}), \hat{\beta}_i \cdot (\sqrt{B+1} + \sqrt{B}))$$

$$\text{Upper } \hat{\beta}_i = \max(\hat{\beta}_i \cdot (\sqrt{B+1} - \sqrt{B}), \hat{\beta}_i \cdot (\sqrt{B+1} + \sqrt{B}))$$

Where  $B = \text{finv}(1 - \text{CI}/100, 1, df) * (1 - R^2)/df$ ;  $df = n - 2$ ; CI is the confidence level (eg. 95); and  $\text{Finv}$  is the inverse of the cumulative probability distribution for the F statistic.

The intercept for each line is calculated as:  $Y_{\text{mean}} - \beta_i * X_{\text{mean}}$

### Regression ANOVA table:

The significance of the correlation coefficient is calculated as per standard linear regression and correlation. (note: the correlation is identical for linear regression and SMA fitting). An ANOVA on the predicted Y values (the “treatment” group) is calculated to determine whether there is a significant treatment effect (i.e. whether the predicted values explain a significant proportion of the variance). The p value is then calculated from the F-statistic, a ratio of the variance explained by the regression to variance unexplained (Sokal and Rohlf 1995). The df for this are 1 and  $n - 2$ . The p-value is obtained by taking the inverse of the cumulative probability distribution of the F-statistic with 1,  $n - 2$  df. Only the p-value is provided (as for ordinary correlation statistics).

Factor	df	SS	MS	F
Variance explained by regression	1	$\sum (\hat{y} - \bar{y})^2$	SS/df	$\frac{MS_{\text{regr}}}{MS_{\text{error}}}$
Error	$n - 2$	$SS_{\text{tot}} - SS_{\text{reg}}$	SS/df	
Total	$n - 1$	$\sum (y - \bar{y})^2$		

### Comparison to hypothesised slope

Also given is an F-statistic and p-value for a test comparing the slope for each group to a specified value,  $b_0$ . The following calculations are used:

$$Var(y + bx) = \frac{S_{YY} + 2.S_{YX}.b_0 + S_{XX}.b_0^2}{2.abs(b_0)} \rightarrow \text{variance explained by line of slope } b_0$$

$$Var(y - bx) = \frac{S_{YY} - 2.S_{YX}.b_0 + S_{XX}.b_0^2}{2.abs(b_0)} \rightarrow \text{variance unexplained by line of slope } b_0$$

$$Cov(y + bx, y - bx) = \frac{S_{YY} - S_{XX}.b_0^2}{2.abs(b_0)} \rightarrow \text{measure of trend in a residual plot}$$

Then  $r^2 = Cov^2/(Var(y+bx)Var(y-bx))$   $\rightarrow$  correlation in residual plot

and  $F = r^2 * (n-2) / (1-r^2)$

The p-value is obtained by taking the inverse of the cumulative probability distribution of the F-statistic with 1,n-2 df.

### **Common Slope Calculation and test**

The common slope  $\hat{\beta}$  satisfies:  $0 = \sum_{i=1}^g n_i \left( \frac{1}{\lambda_{1i}} + \frac{1}{\lambda_{2i}} \right) (S_{YYi} - \hat{\beta}^2 S_{XXi})$  (Warton and Weber 2002)

Where  $\lambda_{1i} = (\hat{\beta}^2 S_{XXi} + 2\hat{\beta} S_{XYi} + S_{YYi}) / 2\hat{\beta}$  and  $\lambda_{2i} = (\hat{\beta}^2 S_{XXi} - 2\hat{\beta} S_{XYi} + S_{YYi}) / 2\hat{\beta}$

$\hat{\beta}$  is then calculated using an iteratively reweighted least squares algorithm:

Given present slope estimate  $\hat{\beta}^k$ , find weights  $w_i^k = n_i \left( \frac{1}{\lambda_{1i}^k} + \frac{1}{\lambda_{2i}^k} \right)$ . Then the next slope

$$\text{estimate } \hat{\beta}^{k+1} = \sqrt{\frac{\sum w_i^k S_{YYi}}{\sum w_i^k S_{XXi}}} \cdot \text{sign}(\sum w_i^k S_{XYi})$$

Initial estimates of  $w_i^0$  are found using  $\hat{\beta}_0 = \sqrt{\frac{\sum S_{YYi}}{\sum S_{XXi}}}$  to find  $\lambda_{1i}^0$  and  $\lambda_{2i}^0$ . After 100 iterations,

if the slope does not converge on a stable estimate ( $\hat{\beta}^{k+1} - \hat{\beta}^k < 0.0001$ ), then new initial starting points are tried, using the slope estimates of individual groups as starting points for  $\hat{\beta}_0$ .

95% confidence intervals for the common slope are calculated as follows:

- A slope and  $R^2$  is fitted to all data assuming no grouping structure. To do this variance and covariance is summed across groups 1...g and inserted into formulas above.
- CIs for this slope are estimated as above, but with  $df = n-g-1$  where n is the total number of data points and g is the number of groups.
- The interval calculated for this slope is then applied to the common slope.

### Test Statistic

The test statistic is:  $-2 \log \Lambda = \sum_{i=1}^g (n_i - 2.5) \log \frac{\lambda_{1i} \lambda_{2i}}{|S_i|}$

where  $\lambda_{1i}$  and  $\lambda_{2i}$  are the final estimates defined above, and  $|S_i| = S_{XXi} S_{YYi} - S_{XYi}^2$ .

Under  $H_0$  (no heterogeneity in slopes amongst groups):  $-2 \log \Lambda \approx \chi_{g-1}^2$

Then if  $-2 \log \Lambda > \chi_{g-1}^2$  there is significant heterogeneity of slopes between groups (Warton and Weber 2002). However, the chi-squared distribution has subsequently been thought to overestimate type 1 errors, so the test statistic is recalculated using permutation. For  $n-1$  permutations,

$$p = \frac{\sum_{i=1}^{n-1} (\text{recalculated} > \text{observed}) + 1}{n} \quad (\text{Manly 1997})$$

The minimum possible p-value is then  $1/n$ . Re-sampling is conducted using the random\_shuffle algorithm in the C++ <algorithm> library. In each permutation the position of each data point along the common slope is retained, while the residuals (distance between the actual point and its position along the line) are permuted among groups.

The intercept for group  $i$ , given common slope, is:  $\alpha_i = \bar{y}_i - \hat{\beta} \bar{x}_i$

Then, the position along the line of the point  $(x_{ij}, y_{ij})$  is given by:

$$\hat{y}_{ij} = 0.5(y_{ij} + \hat{\beta}x_{ij} + \alpha) \quad \text{and} \quad \hat{x}_{ij} = \frac{y_{ij} + \hat{\beta}x_{ij} - \alpha_i}{2\hat{\beta}x_{ij}}$$

The residuals are then  $(x_{ij} - \hat{x}_{ij}, y_{ij} - \hat{y}_{ij})$ . These are permuted amongst groups, and the new data points are calculated as:

$$x_{ij}^p = \hat{x}_{ij} + (x_{ij} - \hat{x}_{ij})^*, \quad y_{ij}^p = \hat{y}_{ij} + (y_{ij} - \hat{y}_{ij})^* \quad \text{where } * \text{ refers to a permuted residual.}$$

A common slope is then fitted to the new data and the test statistic calculated as before. This method is analogous to that proposed by Freedman and Lane (1983) for linear regression, which has been shown to maintain close to exact significance levels in small samples for linear models (Anderson and Robinson 2001).

### Post-hoc multiple comparisons

Use same method as above, but fitting a common slope to pairs of groups. Significance is determined by permutation as above.

### ANCOVA

For SMA tests the following transformation is used:

$$y' = y - bx \quad \& \quad x' = y + bx, \quad \text{where } b \text{ is the common slope.}$$

For MA tests, data are rotated in 2-dimensional space about the origin, by an amount corresponding to the size of the common slope:

$$y' = x \sin \theta + y \cos \theta \quad \& \quad x' = x \cos \theta - y \sin \theta \quad \text{where } \theta = -\arctan b; b = \text{common slope}$$

Then test for single group effect in one-way ANOVA. A significant shift along a common slope can be tested for by testing for a significant group effect in the rotated x data (x'). A significant elevation shift can be tested for by testing for a significant group effect in the rotated y data (y').

Standard ANOVA routines are used (Devore 1995), with significance determined using the inverse of the F-distribution (function available in the R statistical library).

Post-hoc multiple comparisons can be used to determine which groups differ from each other (as above for slope comparisons). Prior to this, Levene's test for equality of variance (Levene 1960) is used to determine whether variances are equal across groups:

Given a variable Y with sample of size N divided into k subgroups, where  $N_i$  is the sample size of the  $i$ th subgroup, the Levene test statistic is defined as:

$$W = \frac{(N - k) \sum_{i=1}^k N_i (\bar{Z}_{i.} - \bar{Z}_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_{i.})^2}$$

where  $Z_{ij} = |Y_{ij} - \bar{Y}_{i.}|$ , and  $\bar{Z}_{i.}$ ,  $\bar{Z}_{..}$  are the group means and grand mean of the  $Z_{ij}$ . Then  $p =$

$\text{Finv}(\alpha, k-1, N-k)$ .

Levene's is less powerful than some other tests, but less sensitive to departures from normality.

Depending on the output from Levene's test, the following methods for post-hoc comparisons (Day and Quinn 1989) are available:

- Tukey-Kramer method, for equal variance
- Games-Howell method for unequal variance

In each, all pairwise differences between group means are considered. The test statistic Q is assumed to follow the Studentized Range Distribution (Day and Quinn 1989). In the **Tukey-Kramer method** (Day and Quinn 1989):

$$Q = \frac{\bar{Y}_{i.} - \bar{Y}_{j.}}{SE_c / \sqrt{2}}, \text{ where } SE_c \text{ is the standard error: } SE_c = \sqrt{MS_d \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \text{ and } MS_d \text{ is the}$$

denominator of the ANOVA F statistic (i.e. Mean Squares within groups).  $df =$  degrees of freedom for  $MS_d = \sum n_i - k$ .  $p$  is then the inverse of the Studentized Range Distribution =  $\text{ptukey}(Q, 1, k, df, 0, 0)$ .

In the **Games-Howell method** (Day and Quinn 1989):

$$Q = \frac{\bar{Y}_{i.} - \bar{Y}_{j.}}{SE_c^* / \sqrt{2}}, \text{ where } SE_c^* \text{ is the standard error: } SE_c^* = \sqrt{(s_i^2 / n_i) + (s_j^2 / n_j)}$$

$$\text{and } s_i^2 = \frac{Sxx_i}{n_i - 1} = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \text{ i.e. variance of group } i.$$

$$df^* = \frac{(s_i^2 / n_i + s_j^2 / n_j)^2}{s_i^4 / (n_i^2 (n_i - 1)) + s_j^4 / (n_j^2 (n_j - 1))} \text{ (Satterwaite's adjusted degrees of freedom). } p \text{ is then}$$

the inverse of the Studentized Range Distribution =  $\text{ptukey}(Q, 1, k, df^*, 0, 0)$ , obtained by taking the inverse of the cumulative probability distribution of Tukeys' studentised range statistic, Q, with  $df = 1, k, df^*$



## References

- Anderson, M. J., and J. Robinson. 2001. Permutation tests for linear models. *Australian and New Zealand Journal of Statistics* **43**:75-88.
- Day, R. W., and G. P. Quinn. 1989. Comparisons of treatments after an analysis of variance in ecology. *Ecological Monographs* **59**:433-463.
- Devore, J. L. 1995. *Probability and statistics for engineering and the sciences*, 4th edition. Duxbury Press, Belmont.
- Enquist, B. J. 2002. Universal scaling in tree and vascular plant allometry: toward a general quantitative theory linking plant form and function from cells to ecosystems. *Tree Physiology* **22**:1045-1064.
- Freedman, D., and D. Lane. 1983. A nonstochastic interpretation of reported significance levels. *Journal of Business & Economic Statistics* **1**:292-298.
- Galton, F. 1886. Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland* **15**:246-263.
- Huxley, J. S. 1932. *Problems of Relative Growth*. Methuen, Methuen.
- Levene, H. 1960. Robust tests for equality of variances. Pages 278-292 *in* I. Olkin and P. Alto, editors. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Stanford University Press.
- Manly, B. F. J. 1997. *Randomization, Bootstrap and Monte Carlo methods in Biology*, 2nd edition. Chapman & Hall, London.
- McArdle, B. H. 1988. The structural relationship - regression in biology. *Canadian Journal of Zoology* **66**:2329-2339.
- McMahon, T. A. 1973. Size and shape in biology. *Science* **179**:1201-1204.
- Miller, R. G. J. 1986. *Beyond ANOVA, Basics of Applied Statistics*. John Wiley & Sons, New York.
- Niklas, K. 1994. *Plant Allometry: the scaling of form and process*. University of Chicago Press, Chicago.
- Pearson, K., and A. Lee. 1903. On the laws of inheritance in man: I. Inheritance of physical characters. *Biometrika* **2**:357-462.
- Pitman, E. T. G. 1939. A note on normal correlation. *Biometrika* **31**:9-12.
- Sokal, R. R., and F. J. Rohlf. 1995. *Biometry : the principles and practice of statistics in biological research*, 3rd edition. W. H. Freeman, San Francisco.
- Warton, D. I., and N. C. Weber. 2002. Common slope tests for bivariate structural relationships. *Biometrical Journal* **44**:161-174.
- West, G. B., J. H. Brown, and B. J. Enquist. 1997. A General Model For the Origin of Allometric Scaling Laws in Biology. *Science* **276**:122-126.
- Wright, I. J., P. B. Reich, and M. Westoby. 2001. Strategy shifts in leaf physiology, structure and nutrient content between species of high- and low-rainfall and high- and low-nutrient habitats. *Functional Ecology* **15**:423-434.
- Wright, I. J., M. Westoby, and P. B. Reich. 2002. Convergence towards higher leaf mass per area in dry and nutrient-poor habitats has different consequences for leaf life span. *Journal of Ecology* **90**:534-543.