NATIONAL BIOLOGICAL SERVICE
U.S. DEPARTMENT OF THE INTERIOR

NBS INFORMATION BULLETIN

No. 45 1995

Simple QBASIC Program Used to Compare Biological Regressions

Increased emphasis on the status and trends of ecological systems by the National Biological Service will inevitably increase the use of regression analysis of biological data. Simple linear regression modeling can be used to study trends at the molecular, cellular, organismal, or ecosystem levels. Regression analysis can also be used to compare more than one trend or relation by comparing the slopes and elevations of several lines simultaneously.

Regression analysis of fisheries data is most often facilitated by computers. Software capable of statistical analysis, such as the latest versions of SAS, SigmaStat, SPlus, Statgraphics-Plus, and SYSTAT, are powerful and efficient at calculating and analyzing regressions. These software packages, however, can be expensive and may not be readily available to all computer users.

The Microsoft disk operating system (MS-DOS) versions 5.0 or greater provide the user with the capability of running and writing programs in QBASIC. The QBASIC language is similar to older versions of BASIC and many good programming instruction manuals are available. QBASIC was selected because of its availability and ease in use. A 133-line program was written in MS-DOS QBASIC (available upon request) to compare slopes and elevations of two or more regression lines. Prerequisites for running an analysis include tabulation of each (x,y) pair in each data set, determining the number of regression lines (k) to be analyzed, determining the number of (x,y) pairs in each regression set, and selecting an alpha level (commonly used values include 0.05, 0.01, and 0.001) for the analysis.

Testing for Differences in Slopes

For sample sets consisting of paired (x,y) data, simple linear regression equations can be expressed as $y = ax + b$, where a is the slope and b is the y-intercept. When the slopes of two or more regression lines are being compared, the null hypothesis (H_o) becomes H_o: $\beta_1 = \beta_2 = \ldots \beta_k$, where β is the population slope and k is the number of regression lines. The alternative hypothesis (H_a) can be stated as at least one of the k regression lines not having the same slope.

Calculations from each of the k sets of paired (x,y) data consists of Σx^2, Σxy, Σy^2 (used to calculate the total sum of squares [SS], and the residual SS), and degrees of freedom (DF). Other calculated values required include the pooled residual SS (SS_p), pooled DF (DF_p), common residual SS (SS_c), and common DF (DF_c). The test statistic, F, is calculated as:

$$F = \frac{(SS_c - SS_p) / (k - 1)}{(SS_p / DF_p)}$$

The hypothesis is then tested with the desired alpha (commonly 0.05), with a numerator DF equal to (k - 1) and a denominator DF equal to DF_p.

Testing for Differences in Elevations

If the null hypothesis, H_o: $\beta_1 = \beta_2 = \ldots \beta_k$, is not rejected, then it may be desirable to determine if all k regression lines are indeed from the same population by testing for differences among the k elevations. However, if the null hypothesis is rejected, then it is meaningless to test for differences in elevations, and this test should not be done. To test for differences in elevations, the null and alternative hypotheses can be stated as all k regression lines have the same elevation, and at least one of the k regression lines do not have the same elevation, respectively.

For this test it is necessary to calculate a total residual SS (SS_t) by computing Σx^2, Σxy, and Σy^2 using all the X and Y values from all k regression lines. The test static, F, is calculated according to Zar (1974) as:

$$F = \frac{(SS_t - SS_c) / (k - 1)}{(SS_c / DF_c)}$$

The hypothesis is tested with the desired alpha (commonly 0.05), with a numerator DF equal to (k - 1) and a denominator DF equal to DF_c.

If H_o is rejected when testing slopes or when testing elevations, a multiple range test, such as the Newman-Keuls test, can be used to determine which of the k regression lines are different. Computer software is available for multiple range testing and is therefore not included in this QBASIC program.

Example Application of the QBASIC Program

The fish pathogenic bacteria *Cytophaga columnaris* produces a cartilage-degrading enzyme called chondroitin AC lyase. The kinetics of this enzyme can be studied by determining chondroitin sulfate (the enzyme substrate) concentration over time using a microtiter assay. To determine the effect of pH on the assay system, three tests were run at three different pH levels (pH 5, pH 8, pH 11). In this example, x was time in hours and y was chondroitin sulfate concentration in µg/mL. The following paired (x,y) values are representative of the type of data obtained; For pH 5: (0.5, 122), (1.0, 106), (1.5, 56), (2.0, 26); For pH 8: (0.5, 113), (1.0, 61), (1.5, 31), (2.0, 25); For pH 11: (0.5, 129), (1.0, 112), (1.5, 109), (2.0, 100).

Required information to run the program includes the selected alpha value (enter 0.05), How many regression lines will be analyzed (enter 3), How many paired (x,y) values in regression 1 (enter 4), How many paired (x,y) values in regression 2 (enter 4), How many paired (x,y) values in regression 3 (enter 4). Output from a successfully run program includes regression equations for each data set (pH 5: y = -67.6x + 162; pH 8: y = -58.8x + 131; pH 11: y = -18x + 135), F-statistic for slope comparison (6.911579), and F-statistic for elevation comparison (9.88). The table values for slope comparison ($F_{0.05(1),2,6}$) and elevation comparison ($F_{0.05(1),2,8}$) are 5.14 and 4.46, respectively. The null hypothesis of equal slopes is rejected, indicating that pH has an effect on the kinetics of chondroitin AC lyase.

For further information contact

Jeffrey D. Teska
Leetown Science Center
National Fish Health Research Laboratory
1700 Leetown Road
Kearneysville, West Virginia 25430
(304)725-8461