DEPARTMENT OF FORESTRY · CANADA

Timber Talks

Prepared by V. H. Phelps, Forest Research Laboratory, 506 W. Burnside Road, Victoria, B.C.

VALIDITY OF PREDICTED VALUES

No. 38

Biological problems are complex and their investigation is usually dependent on the measurement of several dissimilar variables. A valid evaluation of such data and an interpretation that is meaningful for fulfillment of the objectives of the investigation, requires that variations and correlations amongst and between the measurable parameters be accurately determined. Regressions derived from such data are utilized to obtain predicted values. The predictions are expressed as mathematical equations and their accuracy determined statistically. Two statistical procedures commonly used are to determine the standard error of estimate (SE_E) and the coefficient of determination (R^2). The former expresses the variation amongst estimated values and the latter represents the proportion of the variation in the dependent variable that is associated with the equation. Degree of precision of regression equations as expressed in this manner is not always readily understood by all persons but might be more easily understood if described pictorially.

A technique, useful to indicate the degree of confidence that can be placed on predicted values, is the examination of scatter-patterns made by plotting the difference between measured and predicted values against the corresponding predicted value. Four basic scatter-patterns may be obtained.

Fig. 1 indicates that the regression equation is satisfactory and that the predicted values may be considered valid. The validity of the values increase as the band becomes narrower. Less confidence can be placed if the pattern is scattered otherwise. The wedge band (Fig. 2) indicates that the variance between observed and predicted values is not constant and that before determining the equation, weighted least squares or transformation on the observed values are required in the regression analyses. The ascending band (Fig. 3) shows that an error in analyses and departure from the fitted regression is such that negative residuals correspond to low predicted values and positive residuals correspond to high predicted values. The scattered pattern can be caused by omitting a constant term in the regression analyses. The arch band (Fig. 4) shows the equation model to be inadequate and requires additional terms and/or transformation on the observed values.

REPORT: Confidence on Predicted Values in Regression Analyses in Forestry Research. Yam (Jim) Lee, Forest Research Laboratory, Victoria, B. C.