Timber Theft Program Users Guide

Version 2016_0314

Washington Office
Forest Management Service Center
Fort Collins, Colorado

Table of Contents

INTRODUCTION

The Timber Theft Program is designed to help the user predict standing tree volumes from stumps using regression analysis in a timber theft case. The Timber Theft program is designed to work with the Cruise Processing Program but is a self-contained stand-alone program.

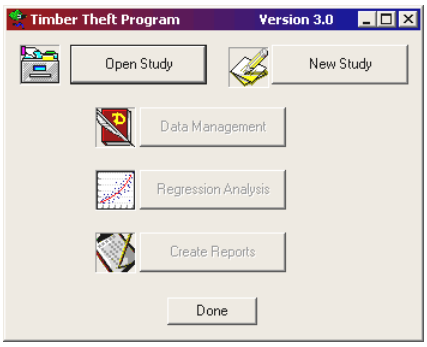
The Timber Theft program is distributed free of charge and can be downloaded from the Forest Management Service Center web site (http://www.fs.fed.us/fmsc/measure/cruising/index.shtml). The program is distributed as a zip file. Once the files have been extracted, the program is ready to run. The Timber Theft program is designed to run under the Win95, Win97/98, and Win2000 operating systems.

GETTING STARTED

The Timber Theft program is distributed as a zip file with the file name TimberTheft.zip. Save this file to your hard drive and extract the files. You should have the following files:

TimberTheft.exe	Application executable
TimberTheft.doc	Documentation file
sample.theft	Test data set

To run the program, execute the file TimberTheft.exe.



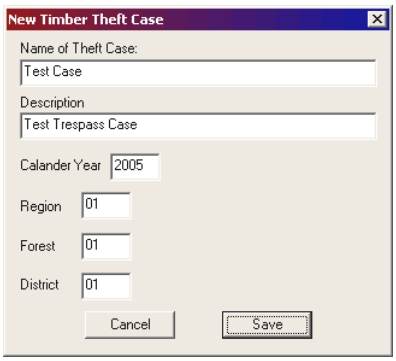
The main window of the Timber Theft program consists of six buttons, three of them are disabled initially. The New Study button will allow you to create a new study, the Open Study button will prompt you top open an existing study, and the Done button will exit you from the program.

OPEN STUDY

Selecting this option brings up the standard Windows open file dialog box displaying files with a ".Theft" extension. Select the study you wish to open by clicking on it and then pressing "open" in the lower right corner. You may be required to locate the study if it is not in the current folder. Once an existing study has been opened, a pop-up box will display the Study Name. The main window enables the Data Management, Regression Analysis, and Create Reports buttons.

DEFINE A NEW STUDY

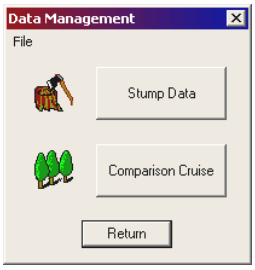
Selecting this option brings up the New Timber Theft Case window. The fields in this window describe and identify the new study.



Once you have filled in the fields, select “done” and you will be prompted for a location to save the new study. A window informing you that the new study has been saved then appears. Once a study has been saved, the main window enables the Data Management, Regression Analysis, and Create Reports buttons.

DATA MANAGEMENT

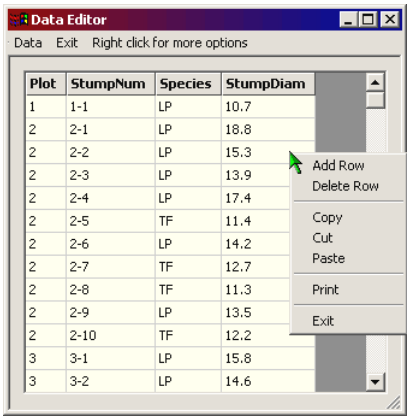
The data management window allows the user to import, enter, and edit stumps or comparison cruise data. Click the Stump Data button to import, add, or edit stump data. Click the Comparison cruise button to import, add, or edit comparison cruise data.



To import data from the previous versions of the Timber Theft/Local Volume Table Program (RegressVol), select File – Import VST file from the menu at the top of the dialog box. A standard windows file dialog box will appear. Select the VST file to import and click Open to import the data.

Stump Data

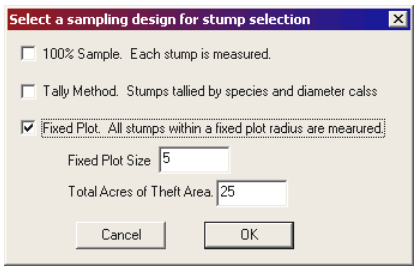
To add and/or edit stump data, select the Stump Data button. The stump data will be displayed in a grid-based data editor. You may add and/or edit any of the data shown in this form. A right mouse click will display a context menu containing several options: Add Row, Delete Row, Copy, Cut, Paste, Print, and Exit. Add row will add a row of data to the end of the data set. Delete row will delete the row you right clicked on. Data will be saved on Exit.



To import stump data, select Data – Import - Cruise Object from the menu at the top of the form. If you collected the stump data using the FScruiser data entry program, you may import the stump data into the Timber Theft Program provided you entered the stump diameters in the Diameter at Root Collar Outside Bark (DRCOB) field. Find the appropriate cruise object and click the Open button to import the stump diameters. If stump data already exists in the table, you will be asked to either append or replace the existing data.

<u>Stump Cruise Type</u>

The first time you click on the Stump Data button, you will need to define the method used to determine the measured stumps. There are three sampling methods to choose from: 100%, Tally Method, and Fixed Plot Sample.



For a 100% sample every stump in the timber theft area is measured. Stump count is assumed to be '1' in this method.

For the Tally Method, stump diameters are tallied by species and diameter class. The Stump Count for each measured stump is number tallied. This number needs to be recorded for each measured stump diameter and species class.

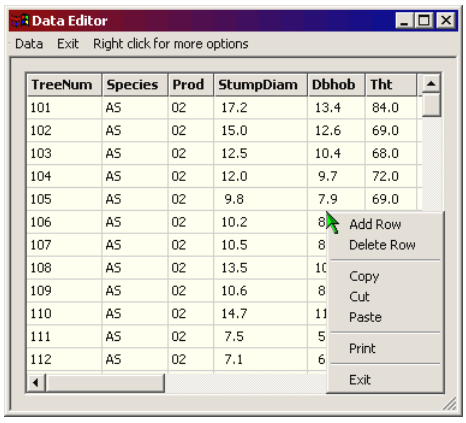
A Fixed Plot Sample uses fixed plots of a predetermined size to determine the measured stump diameters. You will be asked to provide the following information for this design:

Fixed Plot Size,
Total Acres Involved.

The Expansion Factor will be calculated from this information and will be displayed in the last field. This number will need to be added to the stump count for all of the measured stumps entered. All null plots must be recorded and added to the data grid with just plot number given a value (all other fields remain blank).

Comparison Cruise Data

To add and/or edit comparison cruise data, click the Comparison Cruise button. The comparison cruise data will be displayed in a grid-based data editor. You may add and/or edit any of the data shown in this form. A right mouse click will display a context menu containing several options: Add Row, Delete Row, Copy, Cut, Paste, Print, and Exit. Add row will add a row of data to the end of the data set. Delete row will delete the row you right clicked on. Data will be saved on Exit.



To import comparison cruise data, select Data – Import from the menu at the top of the form. There are currently two options for importing comparison cruise data:

Cruise Object: If you collected the comparison cruise data using the FScruiser data entry program, you may import the stump data into the Timber Theft Program provided you entered the stump diameters in the Diameter at Root Collar Outside Bark (DRCOB) field and process the file through Cruise Processing first. Find the appropriate cruise

object and click the Open button to import the comparison cruise data. If comparison cruise data already exists in the table, you will be asked to either append or replace the existing data.

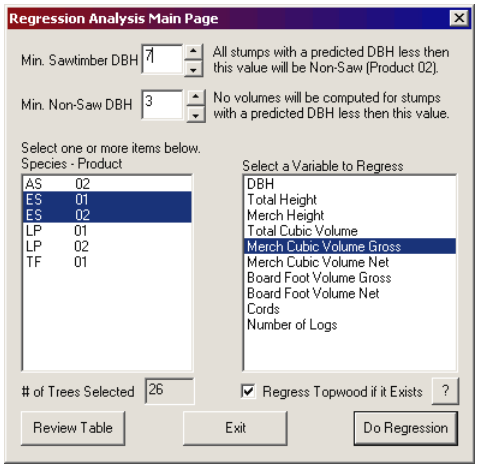
Text Files: The Text Files option will allow you to import a .CSV file created when the CSV10 Report is requested from the Cruise Processing (Version 2) program.

REGRESSION ANALYSIS

The regression analysis portion of the Timber Theft/Local Volume Table Program uses an interactive approach to develop appropriate regression models for all variables of interest. The objective of the regression routine is to give the user the freedom to explore several regression models and select the model that best describes the variable of interest. The best-fit determination will be based on some basic statistical and visual interpretations of the data (see appendix A).

Regression Analysis Main Page

The Regression Analysis Main Page is displayed when the Regression Analysis button is clicked. This form prompts the user to select a species and dependent variable for the regression analysis.



Minimum Diameter Limits:

These fields are used to assign some minimum merchantability standards to the predicted Dbh of each stump in a Timber Theft Study. These fields are saved when the Regression Results are saved, so it is possible to set different diameter limits for each species or species grouping. To change the diameter limits for a species with saved regression results, you will need to re-run those regressions.

Minimum Sawtimber DBH: The Minimum Sawtimber DBH sets the break point between sawtimber and non-sawtimber trees. When volumes are predicted for the measured stumps, the program will keep a separate tally for sawtimber-sized stumps and non-sawtimber-sized stumps based on the predicted DBH of each stump. If all stumps for a species are to be considered non-sawtimber, set the value to 99 to encompass all stumps.

Minimum Non-Saw DBH: The Minimum Non-Saw Dbh sets the minimum merchantability limit for non-sawtimber sized trees. All stumps with a predicted Dbh of less then this value will be assigned zero volumes. If this value is set to zero, all trees will be considered to have some product of value.

Select a Species - Product

A list box is provided that contains each species and product code combination from the comparison cruise data. Use the mouse to select one or more species – product code combinations by left-clicking the species code. All highlighted species codes will be used in the regression analysis. To unselect a species code, use the left mouse button to unselect a highlighted species code. The number of comparison cruise trees in the selected species-product combination will be displayed at the bottom of the list box in the field next to the '# of Trees Selected' label.

Select a Variable to Regress

Before you can run a regression, you need to select an independent variable. There are currently 10 dependent variables available.
DBH – Diameter at Breast Height. This variable is used to assign stump volume by product or by diameter class. If multiple products exist for a single species, a single regression equation for the species as a whole is recommended (select all products for a single species before regressing on DBH).

Total Height – Height of the tree from ground to tip. If multiple products exist for a single species, a single regression equation for the species as a whole is recommended (select all products for a single species before regressing on Total Height).

Merch Height – Height to a specific top diameter (in feet or number of logs) is sometime used instead of total height. If multiple products exist for a single species and the height is determined to different top diameters for each product, regressing on each product separately is recommended if a sufficient sample size is available for each product.

Total Cubic Volume – The mainstem cubic volume from ground to tip, excluding branches. If multiple products exist for a single species, a single regression equation for the species as a whole is recommended (select all products for a single species before regressing on Total Cubic).

Merch Cubic Volume Gross – The cubic foot volume from the stump to some predetermined minimum top diameter. No deductions for defect are included. If multiple products exist for a single species, regressing on each product separately is recommended if a sufficient sample size is available for each product.

Merch Cubic Volume Net – The cubic foot volume from the stump to some predetermined minimum top diameter with deductions for any seen defect, hidden defect, and/or expected cull and breakage. If multiple products exist for a single species, regressing on each product separately is recommended if a sufficient sample size is available for each product.

Board Foot Volume Gross - The board foot volume from the stump to some predetermined minimum top diameter. No deductions for defect are included. If

multiple products exist for a single species, regressing on each product separately is recommended if a sufficient sample size is available for each product.

Board Foot Volume Net - The board foot volume from the stump to some predetermined minimum top diameter with deductions for any seen defect, hidden defect, and/or expected cull and breakage. If multiple products exist for a single species, regressing on each product separately is recommended if a sufficient sample size is available for each product.

Cords – The volume from the stump the some predetermined minimum top diameter expressed in cords. No deductions for defect are included. If multiple products exist for a single species, regressing on each product separately is recommended if a sufficient sample size is available for each product.

Number of Logs – The number of logs calculated between the stump and a predetermined minimum top diameter. If multiple products exist for a single species, regressing on each product separately is recommended if a sufficient sample size is available for each product.

Regress Topwood if it Exists

There is also an option to create equations for determining volume and number of logs for any existing topwood. Topwood is not regressed directly because there is little correlation between DBH and topwood. Instead, topwood is added to the mainstem volume and the regression is made on the total volume or total number of logs. To determine the amount of topwood, the mainstem regression result (predicted volume or number of logs) is subtracted from the topwood regression result (predicted total volume or total number of logs) to provide the topwood volume or number of logs in the topwood section of the tree.

If Regress Topwood if it Exists is checked, the program will first provide the regression analysis for the mainstem volume or number of logs. After the appropriate mainstem model is selected, a second regression analysis is provided for the total mainstem plus topwood volume or number of logs.

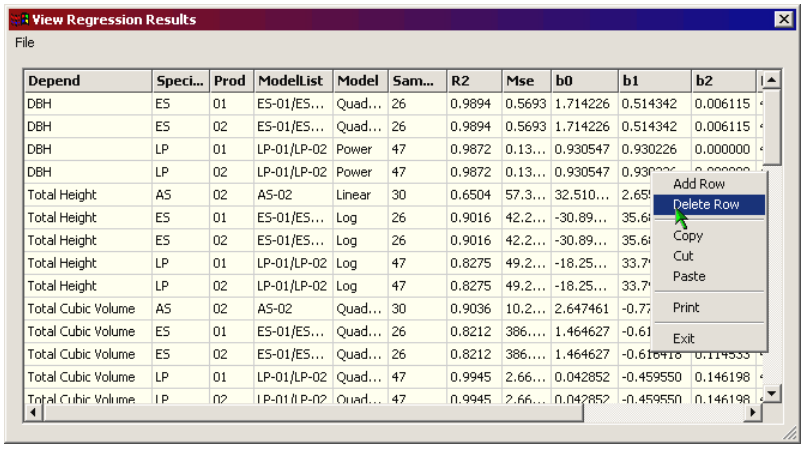
If topwood equations are to be computed and multiple products exist for a single species, regressing on each product separately is recommended if a sufficient sample size is available for each product. Non-sawtimber products will generally not have any topwood volume, but if lumped in with sawtimber products containing topwood volume, some topwood volume might be computed for the non-sawtimber products. The TimberTheft program does have internal checks that will set any topwood volume or number of logs to zero for all nonsawtimber products.

Review Table

The Review Table button will display a grid containing all of the regression analysis that you have saved to this point. Although the table is displayed in a Read-Only format, which is not editable, you may still delete rows by using the pop-up context menu. Simply right click on the row you wish to delete and select Delete Row from the menu. You may save this table in a CSV format (which can be imported into Excel) by selecting File – Export to CSV from the menu at the top. This provides the user with an easy way to import the regression coefficients into an Excel spreadsheet for further analysis.

<u>Do Regression</u>

Once your minimum diameter ranges have been set, your Species – Product combinations have been selected, and your regression variable has been selected, click on the Do Regression button to begin the regression analysis.



You may exit from the Regression Analysis section at any time by selecting either the Exit button at the bottom of the page.

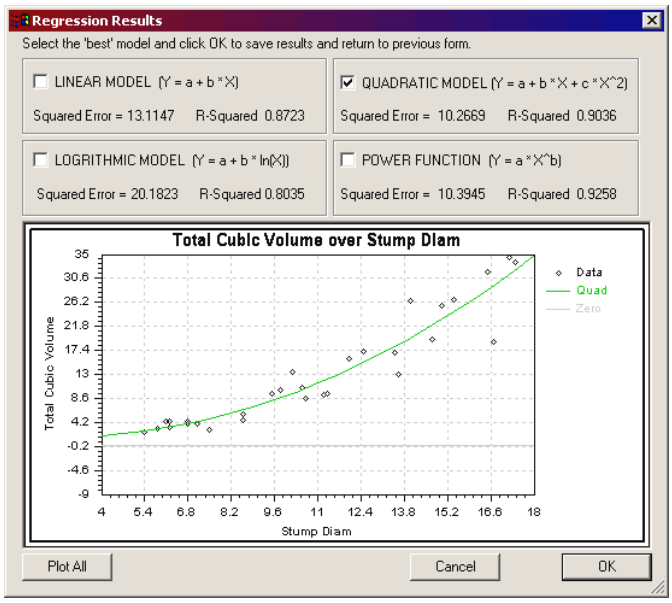
Regression Result Page

The Timber Theft/Local Volume Table Program produces the results of four regression models defined as follows:

Linear Model	Simple linear regression model of the form Y = a + bX.
Quadratic Model	Multivariate regression model of the form Y = a + bX + c(X2).
Natural Log Model	Log linear regression model of the form Y = a + b(log(X))
Power Function	Exponential regression model of the form $Y = aX^b$.

These four regression models are the most commonly used regression models and describe a wide variety of shapes and curves. For each model, the program displays the Mean Squared Error (MSE), the R-squared variable, and the regression equation. Each regression line can be plotted against the actual fitting data by selecting the model (check the box next to the model name). To plot all of the regression lines at the same time, click on the Plot All button at the bottom. The title of the graph will display the regression variable you have selected.

The regression analyses for this program are done utilizing a matrix class library called MatClass developed by C.R. Birchenhall, Dept of Econometrics and Social Statistics, University of Manchester, UK. The results of several MatClass regression analyses were tested against the results of SAS regression analyses and showed no noticeable differences in the output statistics.



Selecting the Best Model

Now that we have completed a regression analysis on four different models, how do we determine which model has the Best-Fit? The Timber Theft Program lets you evaluate the models using both statistical and visual criteria. The Best-Fit model will be the model that performs the best both statistically and visually. First, let us define what we mean by Statistical and Visual Evaluation.

Statistical Evaluation compares the R-Squared and Mean Squared Errors between the given models to find the model that gives the best mathematical fit. This can be done from the information provided at the top of this page.

Visual Evaluation examines the plot of the regression line(s) through the data to determine whether the model displays any bias, inconsistencies, or illogical behavior such as predicting negative values. This can be done by examining the plots at the bottom of the page.

So what constitutes a Best-Fit model? Ideally, it is the model: with the highest R-Squared value, with the smallest MSE, that shows no bias or pattern in the residual plot, and has a regression line that runs cleanly through the middle of the data points. If there are two models that both look good, have high R-Squares, have a small MSE, and have good looking plots, then make the decision based on the smallest MSE value. If the model with the highest R-Square and smallest MSE shows a bias in the plot, then choose the next highest R-Square and next smallest MSE combination, provided the plot looks appropriate. Remember, you are looking for the “best-fit” model, not necessarily the one with the highest R-Squared value.

Also, be aware that the R-Squared for the “best-fit” model might be less then 0.50. This may be especially true with predicting total heights or net volumes. This is acceptable. There is no “magic” R-Squared value to shoot for. Collecting data on more trees will not necessarily increase the R-Squared value. If there is a lot of variability in the data, as is

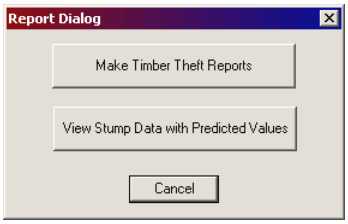
typical with net volumes or total tree height, you will never get a great R-Square value. For more information, see Appendix A.

Note: The model with the smallest MS Error is checked by default. This does not mean it is the best model, but rather the model to explore first.

Save Equation

Once the Best-Fit equation has been identified, check the box to the left of the equation name. Click on the OK button to save the generated equation for later reference or use. Clicking the OK button will return you to the Regression Analysis Main Page once the data has been saved, so you can select a new species and dependent variable combination. The Cancel button will return you to the Regression Analysis Main Page without saving any regression results.

REPORTS



The Timber Theft program has several reports that are used to display and summarize your regression results. After all of the regression analysis has been completed, click on the Reports button on the main window to generate the available reports. There are currently two options available:

Make Timber Theft Reports – Will produce the following reports:
Summary Report – Descriptive Attributes
Fixed Plot Error Calculations
Summary Report – Volume Attributes
Appraisal Report – Volume Attributes by Diameter Class
Regression Summary Report for Stump Diameters
Warning Report – Negative Predicted Volumes

View Stump Data With Predicted Values – Produces a table with all the predicted values for each stump in the Stump Data table.

Make Timber Theft Reports

The Timber Theft Reports button will produce several pages of reports and will display them in a report viewer. The reports can be saved or printed from within this report viewer.

Summary Report – Descriptive Attributes

This report will apply the regression equations to the stump diameter data and will display for each species and product the minimum predicted DBH, minimum stump

diameter for the data range, maximum stump diameter for the data range, estimated number of stumps, average DBH, Average Total Height, and Average Height PP The minimum predicted DBH is used to break out the stumps by product.

If a regression equation was not saved for a particular species and product, it will be displayed as zero. If any value was predicted to be less then zero, a warning message is displayed at the bottom of this report. This usually means your cruise data failed to enclose the range of your diameter data.

Fixed Plot Error Calculations

If fixed plots were used to sample for stumps, this report will be included. The program will use the variability in plot-to-plot basal areas to determine the error for the cruise. With a normal cruise, the variability in plot-to-plot volume is used, but as there is no volume information available, basal area was considered the best variable of interest to use. The calculations use the same formulas detailed in the Timber Cruising Handbook (FSH 2409.12 Chap. 34) for fixed plot, substituting Basal area (0.005454 * Diameter Squared) for tree volume.

Summary Report – Volume Attributes

This report is broken up into two sections; a Primary Product Report and a Secondary Product Report. The Secondary product report will only be displayed if a regression equation of topwood exists. This report displays by stump species and estimated stump product the sum of total cubic volume, gross cubic volume, net cubic volume , gross board foot volume, net board foot volume, and estimated total number of logs. Each category will be displayed only if a regression equation exists.

Appraisal Report – Volume Attributes by Diameter Class

This report creates a table for each species that lists the number of stumps and the sum of the volumes by predicted DBH class. The report also displays the minimum and maximum stump diameters for each predicted DBH class. This report can only be created if a regression equation was created with DBH as the dependent variable.

This report is broken up into two sections; a Primary Product Report and a Secondary Product Report. The Secondary product report will only be displayed if a regression equation of topwood exists.

If a regression equation was not saved for a particular volume, it will be displayed as zero.

Regression Summary Report for Stump Diameters

The regression report displays the results of your regression analysis. For each dependent variable, the report displays the list of species-product combinations that

went into the regression analysis, the sample size, the minimum and maximum stump diameters in the comparison cruise data, the R-squared value, The Mean Square Error, and the actual equation with coefficients.

Warning Report – Negative Predicted Volumes

The last page(s) of the report contain a list of volume warnings. If any of the predicted volumes were computed to be less then zero, the stump ID along with the predicted volume is listed in this report. The volume will be set to zero before it is summed in the reports. A lot of stumps with negative volumes is typically an indicator of one of two things:

An inappropriate regression model, in which case you should reexamine your choice of regression model (redo the regression analysis for this species).

The comparison cruise data set does not contain the full range of stump diameters found in the trespass area. Additional comparison cruise trees should be measured to insure the range of stump diameters in the trespass area is covered by the range of stump diameters in the comparison cruise.

View Stump Data With Predicted Values

This option will display a table with all the predicted values for each stump in the Stump Data table. Negative predicted volumes will be displayed as negative volumes in this table. All volumes are expanded the number of stumps. If a fixed plot sampling method was used, the program will expand the stumps by the computed expansion factor as described in the Timber Cruising Handbook (FSH 2409.12 Chap. 34) and list the expansion factor in the Count column.

The table can be exported as a CSV file by selecting File – Export to CSV from the menu at the top. The CSV file can be opened in Excel for additional analysis.

APPENDIX A: REGRESSION ANALYSIS

Description: The statistical term regression, as used in this document, is comparable to the word prediction. Regression analysis can be described as a statistical tool that utilizes the relation between two or more variables so that one variable can be predicted from the other, or others (Neter and Wassermann, 1974). For example, if we know the relationship, by means of regression analysis, between stump diameter and gross cubic foot volume, we can predict the gross cubic foot volume of a tree once the stump diameter of that tree is known.

Before we can start talking about the mechanics of regression, we need to define some terms. The predictor variable is called the independent variable, and is usually denoted by the letter X. The target variable, or the variable we wish to predict, is called the dependent variable and is usually denoted by the letter Y. In the previous example, stump diameter would be the independent variable while gross cubic foot volume would be the dependent variable. We will define more terms through out the discussion.

How regression works

Often, the first step is to plot the dependent variable over the independent variable. This is done to provide some visual evidence of whether the two variables are related. If there is a relationship between the two variables, the plotted points will show a pattern. If the relationship is very strong, the pattern will be very distinct. If the relationship is weak, the plotted points will be more spread out and the pattern will be less definite. In a nutshell, regression analysis plots a single line that best describes that pattern. The line could be straight or curved, but it is always defined by a single equation. A line (assuming a straight line for now) can be describe by the equation:

$Y = a + bX$

where: a = the y-intercept or where the line crosses the y-axis.
b = slope of the line or how much Y changes with each change in X.

If we substitute our dependent and independent variables for the Y and X variables in the above equation, we will now have an equation for a regression line. To determine the regression line, we must fit or estimate the values for the y-intercept and the slope of the line. These estimated values are known as regression coefficients.

How do we estimate the y-intercept and the slope of the line? One way would be to simply draw a straight line through the center of the data points and then measure its intercept and slope. But there are two problems with this approach. First, different people would draw slightly different lines. Second, there is no guarantee that our line is the “best possible” line. To alleviate both these problems, we employ an objective approach to finding the best possible line.

Since we will be using our line for making predictions, we would like the predicted value of Y (or dependent variable) to be as close as possible to the actual, or observed, value

of Y for each observation. Equivalently, we would like the difference between the predicted value of Y and the actual value of Y to as small as possible for each observation. This difference between the predicted and the actual value of Y is called the residual and can be thought of as the vertical distance between the plotted data point and the regression line. Because we would like the residual to be small for each observation, we could try summing them to find the line that minimizes the sum of the residuals. Unfortunately, this is not an adequate criterion for choosing a "best fitting" line. Any line that passes through the exact center of the data (or passes through the point described by the mean of X and the mean of Y) has a sum of residuals equal to zero. For such a line, half the values are too small and half are too large and the resulting sum of the residuals equal zero.

To avoid the cancellation of positive and negative values, we square each residual before summing. Now, when we find a line that minimizes the sum of these squared residuals, we have found our "best fitting" line. This procedure is known as the method of least squares or least squares regression.

A regression line can be thought of as a moving average. It gives an average value of Y associated with a particular value of X. Some of the actual values of Y will be above the regression line, or moving average, and some will be below. Using the example defined earlier with gross cubic volume to stump diameter, our regression line, or moving average, would predict the average gross cubic foot volume for each given stump diameter.

The next question we will need to ask is how well did our regression line fit our data? Now that we have found what we have defined as our "best fitting" line using least squares regression, we want to know accurately does that line describe the relationship, or pattern, between our dependent and independent variables. One method is to examine the R-Squared or Coefficient of Variation. The R-Squared value is a measure, ranging from 0 to 1, of the percent of variability in the dependent variable that is explained by independent variable. Each dependent variable has some natural variability associated with it. A regression line is trying to account for as much of that natural variability as possible. The R-Squared value is the percent of the variability associated with the dependent variable that the independent variable can account for.

A high R-Squared value means the actual data points fall close to the regression line. If all the data points fall on the regression line, the R-Squared value would be 1.0. If there were no pattern in the data whatsoever, the R-Squared value would be 0. Do not confuse the R-Squared value with a level of confidence. There is no upper limit to shoot for when looking at an R-Squared value. Either a relationship between the dependent and independent variables exists or it does not. No amount of data will give you an R-Squared value of 0.95 if no relationship exists.

For example, let's take gross cubic volume as our dependent variable. There is some natural variation in gross cubic volume from tree to tree; small trees tend to have small gross cubic foot volumes, while big trees tend to have large gross cubic foot volumes. If

we select stump diameter as our independent variable and run a regression on gross cubic volume, we will probably end up with an R-Squared value close to 1 because stump diameter is a good indicator of tree size and tree size is a good indicator of gross cubic volume. Let's say we end up with an R-Squared of 0.9552. We can say that 95.52% of the variability in gross cubic foot volume was associated with stump diameter.

The Mean Square Error (MSE) is another variable we can use to evaluate how well our model fit the data. The MSE is the residual sum of squares divided by the degrees of freedom. The degree of freedom is defined as the total number of samples minus the number of variables in the regression equation minus one. With our linear equation given above, our degrees of freedom would be the number of samples minus one for the independent variable, minus one, or the number of samples minus two. The MSE is the variance of the regression model and can be used the same way a sampling variance can be used, including building confidence intervals around your regression line.

More often then not, the pattern displayed by plotting the dependent over the independent variable will show some sign of curvature. In this case, we would want to use an equation that can accommodate this curvature. The most common equations used to account for curvature in regression are the quadratic, natural logarithmic, and the exponential or power function. Choosing an appropriate regression equation model is just as important as choosing the right predictor variable. If the pattern displays an exponential trend, a linear equation will not fit the data as well as an exponential equation will. When comparing regression equations for the best model, compare the R-Squared value, the MSE, and plots of the residual values. The best model should have the highest R-Squared value, the lowest MSE, and display no discernible pattern in the residual plot.