

MINITAB Release 14

Statistical software for Windows

■ John A. Wass, Ph.D.

MINITAB is a statistical software package in wide use by academe and industry. It includes modules for general univariate statistics, multivariate statistics, experimental design (DOE), reliability and survival analysis, quality assurance, time series analysis, nonparametric statistics,

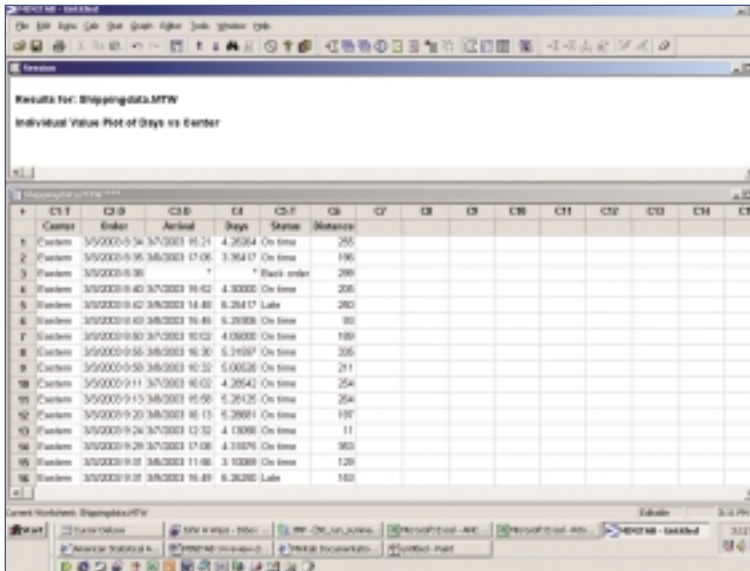


Figure 1: MINITAB Workspace

power/sample size calculations, and graphics. The package was specifically developed to relieve the need for manual drudgery from lengthy hand calculation by students taking introductory statistics classes and quickly found use in a variety of disciplines.

The new version has many enhanced features in the following areas: ease of use, data and file management, graphics and tables. New items also can be found in many of its routine and advanced statistical tests, greatly extending the programs' already considerable capabilities.

The analytic level makes this suitable for the needs of a wide range of users from

undergraduates to bench scientists, social scientists and engineers.

System requirements

The software is made for Windows and will perform well on the following versions: 98, Me, NT4, 2000 and XP. A computer with at least a 300 MHz clock and 64 MB of RAM is recommended but, for large calculations and intricate graphics, more is better. The software can be downloaded from the Internet or from a CD-ROM and takes about 85 MB of hard disk space.

Documentation

The only major disappointment, was that the two-volume *User's Guide* supplied with version 13 is no longer available. A handbook may be purchased from the vendor, but the breadth and depth of the manuals were always appreciated and, to the fans of these books, were much easier to navigate than on-line help areas or help buttons displaying short bursts of information.

The good news is that the short *Meet MINITAB* manual still is provided and has been rewritten to tell an interesting story. By introducing data from a fictional business, the manual walks the user through sections of the program while demonstrating a typical analysis. By investigating various aspects of this business from a supplied datasheet, the user learns about calling and saving files, graphing and analyzing data, and creating a report. The next section of this review will make use of this interesting format to acquaint the reader with some of these features.

Work with MINITAB: A sample problem

Using simple menu-driven commands familiar to all Windows users, the file is brought into the spreadsheet in the main work area. The main window includes a session window (where the commands and results are visual-

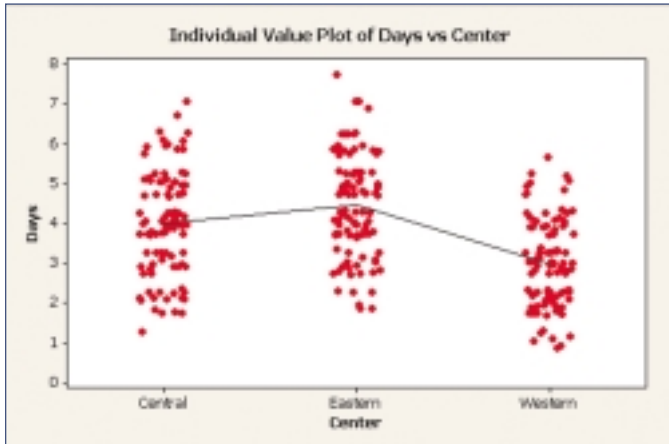


Figure 2: Means of Groups Plot

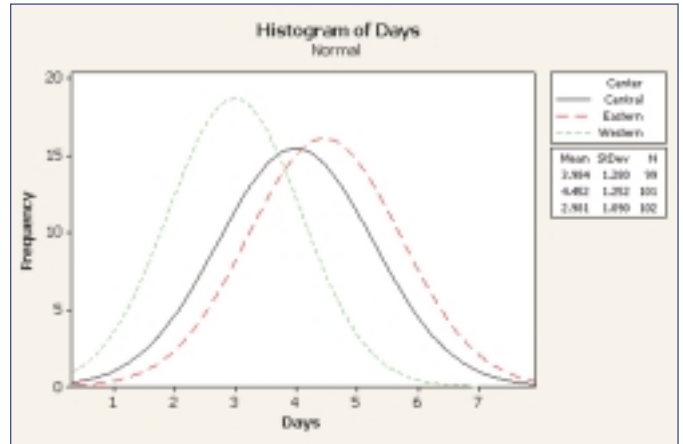


Figure 3: Distribution Plots

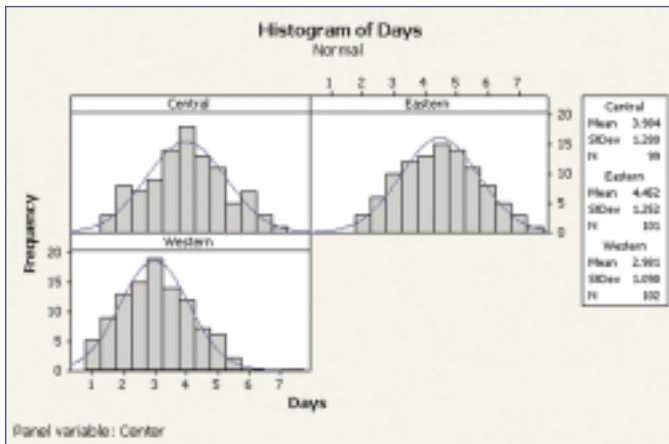


Figure 4: Histograms

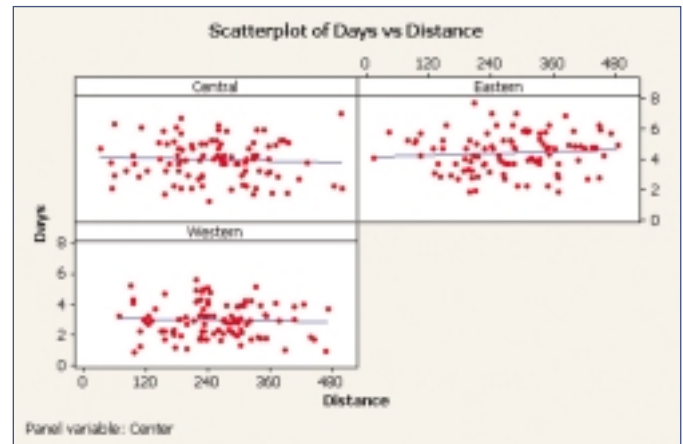


Figure 5: Scatterplots

ized, a data window (spreadsheet), and a menu/toolbar (Figure 1). In addition, special windows come up with the various graphics as they are generated. These areas are easily resized and edited to user preference. Now, to actually look at the data...

There are six columns containing data on the shipping center, order date, delivery date, delivery days, on-time status and miles from shipper to destination. This illustrates the three data input types that MINITAB accepts, i.e., numeric, text and date/time. Before doing any statistical analysis, it is usually wise to generate a few simple graphics to assess trends, aberrancies and relationships

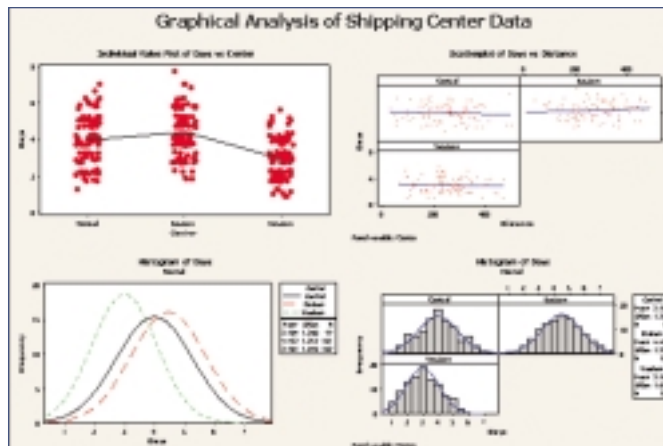


Figure 6: Consolidated Plot

between variables. Some exclude this step, and that is an omission that may return to haunt the analyst. At this point, if nothing else, the graphs may serve as an excellent reality check to ensure data adequacy.

By choosing the 'graph/individual value' plot on the menu bar, a dialog box appears with small graphics representing the choices available. We select the 'One Y/With Groups' option, place the proper variables in the resultant boxes, and customize by checking a 'Mean connect line' choice. Upon clicking, the resultant graph appears within seconds (Figure 2). Noticing that one shipping center has a lower mean delivery time than the other two and that within-group variability seems about the same, we can quickly determine data overlap by choosing 'Graph/Histogram/With Fit and Groups' from the main toolbar (Figure

3). Note that each graphic is automatically labeled from the existing column headers and that each graph may be modified by simply right clicking on that graph and selecting from a variety of features. Realizing that many of our most commonly used statistical tests assume a normal or near normal distribution of errors we can now use the histogram feature with a 'Multiple Graphs' option to check the data distribution (the usual surrogate for error distribution). Figure 4 displays the multiple distribution graphs. Note that these graphs also contain the mean, standard deviation, and N for each group, a very useful summary. Noting the distributions and lack of aberrancies, we can proceed to more illuminating graphics.

The aforementioned visualizations are basically background analyses, quick and dirty determinations of data distributions. Examining associations is many times very instructive and can serve as a guide for future analyses. Here we utilize the 'Graphs/Scatterplot' and 'Multiple graphs' options to quickly generate a plot of these associations (Figure 5). Note that there seems to be no significant relationship between the two variables for any of the three centers. We can place a note to that effect at the bottom of the graph. Now comes the really useful part.

We often times wish to condense information on a single page and this is possible with graphics in the MINITAB Layout Tool. By a few (very few) mouse clicks and drags, a composite graph (Figure 6) is created showing all of the individual graphics and numerics on a single page. Unfortunately, this composite is not automatically updated as the data changes, as are the individual graphs. These, however, are but a small sampling of the graphics available, and the interested reader can go to the Web site for a more exhaustive listing. Now that the preliminaries are complete, the actual statistical capabilities can be utilized.

Basic descriptive statistics for selected indicators are generated even more quickly than the graphics. The dialog box for this still does not possess a 'select all' button for the purists, but it is possible to store the complete selection as a preference so that all boxes are automatically checked. Once the results are perused, group comparisons may be made just as quickly by using the ANOVA or non-parametric Kruskal-Wallis test available on the 'Stat' menu. The ANOVA selection presents the analyst with not only the usual sum of squares table and 95 percent confidence intervals for the means, but a series of graphics (again, selected by the user) that visually compare groups and distributions. One choice, the 4-in-1, automatically produces the important normal probability plot, histogram and order plot for the residuals, as well as the residuals versus fitted values as a one page layout. This is another very quick way of assessing normality and data adequacy. From here, the more sophisticated quality control tests are run and further analysis done. Depending upon need, these can include: specialized regression (including partial least squares), reliability and survival analysis, multivariate analysis, time series analysis, cross-tabulation and chi-squared tests, and the very important power and sample size calculations. Again, for an exhaustive list of tests, consult the Web site.

As it is taking an increasingly important role in product development, the experimental design section (DOE) will be heavily utilized by many industrial engineers and scientists. A variety of designs are offered and, of most importance to the novice, the module will not only design an experiment, but will analyze it as well. This has been a glaring omission in many statistical packages as analysis of these designs is sometimes challenging even to the experienced practitioner. It easily can become a nightmare to the beginner. In addition, this version of MINITAB

includes density, distribution, and random number generators, which are heavily utilized by the more advanced users. These types also will find MINITAB's General Linear Models module to be sophisticated in several areas. It was quite pleasing to see the program produce error messages for empty cells, unbalanced nesting, colinearity and undeclared covariates. This module will be fully tested by many statisticians to assess the adequacy of the algorithms.

Summary

MINITAB has always had much to recommend it to the novice, but with increasing power and sophistication, it has rapidly evolved into a very useable analytic program. There are many forms of help and informational updates available to the newcomer on-line and the level of technical assistance is still quite good. For intense handholding, the Web site does list a telephone number and explanations for the statistical tests are included under the Stat Guide. A demo version is available on the Web for further information. At this point my only large addition to the wish list would be for a neural networks module as these are increasingly used for classification and prediction problems and specialized software for this is many times beyond the financial resources of small firms.

Availability

- Single copy \$1,195
- Upgrade \$395

Minitab Inc.

Quality Plaza, 1829 Pine Hall Road
State College, PA 16801-3008
800-448-3555 (U.S. and Canada)
814-238-3280; Fax: 814-238-1702
sales@minitab.com
www.minitab.com

John Wass is a statistician with GPRD Pharmacogenetics, Abbott Laboratories. He may be contacted at sceditor@scimag.com.