

User Manual



© Copyright 2012 Rapid Insight® Inc. All Rights Reserved

TABLE OF CONTENTS

ABOUT ANALYTICS 5
This Document 5
Hardware Requirements 5
Additional Resources 5
PROGRAM ELEMENTS 6
Analytics Module 6
Scoring Module 6
VRE (Variable Reduction Engine) Module 6
ANALYTICS MODULE 7
Select and Load a Dataset 7
GENERAL DESCRIPTION 7
LOAD SAVED ANALYSIS AND DATASET 7
OPEN DATASET / SELECT DATASET TO OPEN 7
LOAD VARIABLES 9
LOAD SAVED ANALYSIS ON CURRENT DATASET 9
CREATE ANALYSIS 10
ANALYTICS MAIN MENU 10
GENERAL DESCRIPTION 10
FILE MENU 10
OPTIONS MENU 11
REPORTS MENU 11
Control Panel 12
GENERAL DESCRIPTION 12
VARIABLE CONTROL PANEL TABLE 12
Y VARIABLE 13
EXPORT TO EXCEL 13
Variable Statistics 14
GENERAL DESCRIPTION 14
VARIABLE STATISTICS GRID 14
VARIABLE STATISTICS LEFT SIDEBAR 15
Analyze Data 15
GENERAL DESCRIPTION 15
ANALYZE DATA RECORDS GRID 16
ANALYZE DATA LEFT SIDEBAR 16
Univariate Analysis 17

GENERAL DESCRIPTION 17
UNIVARIATE ANALYSIS VARIABLE LIST 17
UNIVARIATE ANALYSIS CHART AREA 17
Multivariate Analysis 18
GENERAL DESCRIPTION 18
MULTIVARIATE UPPER TOOLBAR CONTROLS 19
MULTIVARIATE LOWER TOOLBAR CONTROLS 20
Automated Mining 21
GENERAL DESCRIPTION 21
ANALYSIS VARIABLES 21
RELATED VARIABLES 22
NON-RELATED VARIABLES 22
CHANGE P-VALUE 22
AUTOMINE IT 22
VIEW RELATIONSHIPS 22
Variable Relationships 22
GENERAL DESCRIPTION 22
VARIABLE RELATIONSHIPS TOOLBAR CONTROLS 22
VARIABLE RELATIONSHIPS MAIN MENU 24
Create Variables 25
GENERAL DESCRIPTION 25
SINGLE VARIABLE OPERATIONS 25
CREATE A BINARY VARIABLE 25
BINNING 26
MULTIPLE VARIABLE OPERATIONS 26
USE MEMORIZED MODEL'S EQUATION TO CREATE A VARIABLE 26
CREATE A CONTINUOUS VARIABLE 27
Frequency Analysis 27
GENERAL DESCRIPTION 27
SINGLE VARIABLE CATEGORY FREQUENCIES 27
TWO WAY CROSS FREQUENCIES 27
Means Analysis 27
GENERAL DESCRIPTION 27
Correlation Analysis 28
GENERAL DESCRIPTION 28
Profiling Analysis 29
GENERAL DESCRIPTION 29
SELECT A BINARY VARIABLE TO PROFILE 29
CHANGE P-VALUE 29

AUTO-PROFILE 29
VIEW PROFILE GRAPHICALLY 29
PROFILE REPORT 29
EXPORT TO EXCEL 30
Model 30
GENERAL DESCRIPTION 30
REGRESSION TYPE 30
MISSING VALUES DEFAULT HANDLING 30
P-VALUES FOR STEPWISE 31
SAMPLING 31
MODELING OPTIONS 31
BEGIN REGRESSION ANALYSIS 33
Building a Model to Predict Y-Variable 33
GENERAL DESCRIPTION 33
REGRESSION OPTIONS 33
VARIABLE LIST 34
MEMORIZE MODELS 34
REGRESSION RESULTS 34
MODEL DECILE CHART 34
Compare Models 35
GENERAL DESCRIPTION 35
What If? 35
GENERAL DESCRIPTION 35
SELECT A MODEL / MODEL VARIABLES 35
Reports 36
GENERAL DESCRIPTION 36
DOCUMENT EDITING CONTROLS 36
Notes 37
GENERAL DESCRIPTION 37
SCORING MODULE 38
Select and Load a Dataset 38
GENERAL DESCRIPTION 38
SCORE TO 38
SELECT DATASET 38
SELECT SCORING MODEL 38
Contact Us 38

ABOUT ANALYTICS

Analytics is a powerful desktop application for analyzing analytic data sets and building predictive models. Predictive modeling is an analytic approach that identifies patterns and relationships between data variables. These relationships are statistically validated and then used to build analytic predictive models. Such models are used to evaluate new data and apply a probability scores to each record, predicting the likelihood of a specific event or condition to occur. This product can be used alone or in conjunction with Veera® (Rapid Insight's data processing and reporting tool).

This Document

This User Manual is intended to provide in-depth descriptions of all aspects of the program's function and features.

Hardware Requirements

Systems intending to host Analytics must meet the following hardware requirements:

Minimum System Requirements

- Windows XP SP3, or later
- .NET Framework 4.0 and its requirements
- 2 GHz 32-bit (x86) or 64-bit (x64) processor
- 1 GB of system memory
- 20 MB available hard drive space
- 32 MB of graphics memory

Recommended System Requirements

- Windows 7 x64
- .NET Framework 4.0 and its requirements
- 2.8 GHz 64-bit (x64) processor
- 4 GB of system memory
- 1 GB available hard drive space
- 512 MB of graphics memory

Additional Resources

For more specialized information, try one of these other Analytics documentation resources:

- Installation Guide; provides step-by-step instructions for installing the program
- Getting Started Guide; a quick reference and tutorial - includes several "How To…" examples
- In-program Help; accessed from the Analytics toolbar menu

PROGRAM ELEMENTS

Analytics is actually comprised of two standard and one optional modules that appear under the PROGRAMS > RAPID INSIGHT INC > ANALYTICS section of the Windows START menu. Though all of these are typically installed, licensing determines which are made available to a specific user.

Analytics Module

The module labeled ANALYTICS is the core of the program. This is where all aspects of data analysis, predictive model building and reporting are accomplished.

Scoring Module

Scoring is a module that is available to all standard, licensed users. It applies a predictive model developed with the Analytics program to new data. This module assigns a probability score to each record based upon its unique combination of values.

VRE (Variable Reduction Engine) Module

The Variable Reduction Engine (VRE) is an optional module whose purpose is to work-around limits imposed by ODBC and OLEDB drivers on some file-based datasets. VRE works in conjunction with the main Analytics program to permit the use of datasets with more than 255 variables.

ANALYTICS MODULE

Select and Load a Dataset

GENERAL DESCRIPTION

This is always the first window encountered when opening the Analytics module. This is where the dataset to be analyzed is selected. It is also where any stored analysis may be recalled and applied to the dataset.

DATA SELECTION CONSIDERATIONS

Analytics always performs its analysis and model-building upon a single dataset. This dataset may be in the form of a single structured data file (such as an Excel or CSV file) or a single table in a database (such as Oracle or MS SQL Server). All data to be used must be contained within this dataset – no imbedded or implied relational links will be recognized.

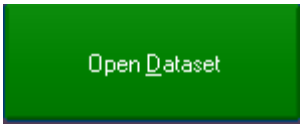
Which data types are available is often dependent upon whether the 32-bit drivers or providers for that type are installed on the computer. Drivers (OLE DB, ODBC, ADO.net or native drivers) are freely available for download from the corresponding manufacturer. Data types Analytics regularly supports include:

- Access
- DB2
- DSN (ODBC Connections)
- Excel
- FoxPro
- MySQL
- Oracle
- OtherDBMS
- OtherFile
- Paradox
- Rapid Insight Dataset
- SAS
- SPSS
- SQL Server
- Text/CSV
- UDL
- xBase
- XML



LOAD SAVED ANALYSIS AND DATASET

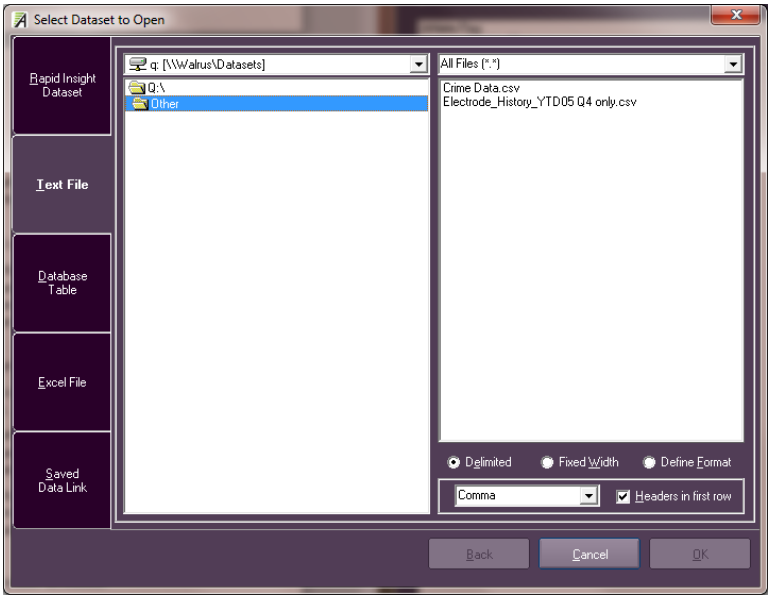
Users selecting this option will identify a saved Rapid Insight analysis file (*.ria) and open it along with the dataset used to originally develop it. The dataset file/table must be in its original location, must be named the same and must contain the same variables. The program will automatically open to the screen where the original analysis was saved.



OPEN DATASET / SELECT DATASET TO OPEN

Clicking this button opens a pop-up window “Select Dataset to Open” where the selection of a dataset to be analyzed is made. This pop-up window is typically divided into several sections.

Data Type Categories



The broad data categories Analytics accesses appear as buttons along the left-hand edge of the window. Clicking a button will allow the user to search for files or tables of that corresponding data type. These data categories are

- RAPID INSIGHT DATASET. This accesses a proprietary file type (.rids) provided by Rapid Insight.
- TEXT FILE. This type covers .asc, .csv, .tab and .txt data files. In addition to selecting the file, the user may specify the way in which the file is read by the program. Options include whether the fields are fixed width or delimited, the type of delimiter, if any, and whether the first row contains headers. The user may further specify the format of the file by selecting the "Define Format" radio button and clicking OK.
- DATABASE TABLE. This section is divided into 4 data type sub-categories.
    - FILE BASED. This allows access to file-based database types such as MS Access, FoxPro and Paradox, as well as formatted data types such as XML.
    - OLE DB. This section provides access to server-based datasets, such as MS SQL Server and Oracle, using an OLE DB data driver. The user must specify details of the connection such as the database name and user credentials.
    - ODBC. This section provides access to server-based datasets, such as MS SQL Server and Oracle, using an ODBC data driver. The user must specify details of the connection such as the database name and user credentials.
    - View/Query. This option is currently unavailable
- EXCEL FILE. Files using either the older .xls or newer .xlsx formats are both accessible.
- SAVED DATA LINK. This accesses data from files using the extension .udl

Data Location

This field allows the user to specify the location of the data source on the local machine or on a network drive.

Data Type Options

Data type options appear in the upper right of the window under a drop down control. It allows the user to define the specific file extension.

Matching Data Files / Tables

This field displays the files/tables that meet the data type and data type option in the selected location. The user may select the dataset to open from this list by clicking on it then clicking the OK button.

LOAD VARIABLES

Once a new dataset is selected using Open Dataset, the following fields and controls are enabled.

Sampling

This control is only active prior to clicking the Load Variables button. It allows the user the option of analyzing just a subset (1 to 100%) of the records originally found in the dataset. If the Random Order Rows box is checked, the records to analyze will be selected randomly, otherwise the will selected in the order encountered in the file/table.



Variables to Exclude from Loading

This field is only active prior to clicking the Load Variables button. Variables that the user has determined should not be part of the analysis are selected from the list on the left and "moved" to this field using the Exclude button. Variables may be returned to the list of loadable columns by selecting them and clicking the Include button.

Mark Variables as Identifiers

This field is only active prior to clicking the Load Variables button. Variables that serve as unique record identifiers in the dataset should not be part of the analysis. To keep them from loading, the user selects them from the list on the left and "moves" them to this field using the Identifier button. Variables may be returned to the list of loadable columns by selecting them and clicking the Other button.

Status

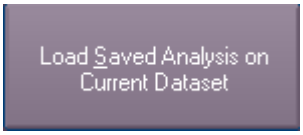
This field reports the record count and number of variables found in the original dataset.

Load Variables Button



This button temporarily replaces the Load Saved Analysis and Dataset option. Clicking the button loads the records into memory.

LOAD SAVED ANALYSIS ON CURRENT DATASET



Users selecting this option will identify a saved Rapid Insight analysis file (*.ria) and superimpose it on the current dataset. The dataset file/table should, at a minimum, contain the same variables used to originally create the analysis. The program will automatically open to the Control Panel tab.

CREATE ANALYSIS



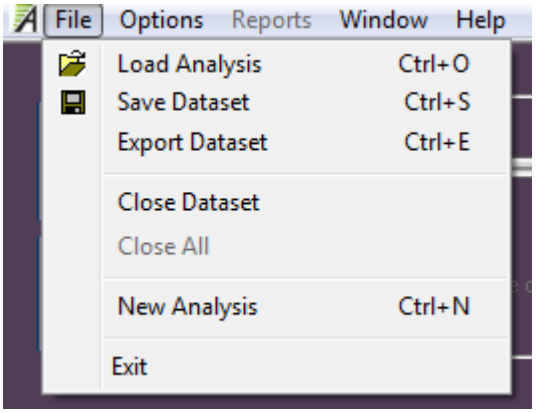
After clicking the "Load Variables" button, the "Create Analysis" button becomes available to begin a new analysis. Once clicked, the program will automatically open to the Control Panel tab.

ANALYTICS MAIN MENU

GENERAL DESCRIPTION

The main menu of Analytics allows us to: load, save, and export analyses, control the physical appearance of Analytics, change default data settings, launch the Report bar, toggle between various windows, and access the Help menu.

FILE MENU



Load Analysis

After loading the selected dataset, the "Load Analysis" button is available under the "File" menu. If you have previously conducted an Analytics analysis on the uploaded dataset, you can apply the analysis by selecting the appropriate .ria file here.

Save Analysis

To save all parts of the current analysis, click "Save Analysis" under the "File" menu at any time.

Export Analysis

Because we have the ability to set filters, cap outliers, and add calculated variables to a dataset in Analytics, sometimes it is necessary to export this data to a .csv file. To do this, click the "Export Analysis" button under the "File" menu. Here you will be prompted to select which variables you'd like to export and name the file you'd like to save the analysis to.

Close Analysis

To close a single analysis at any time, click the "Close Analysis" button from the "File" menu. You will be prompted to save any unsaved analyses at this time.

Close All

To close all running analyses at any time, click the "Close All" button from the "File" menu. You will be prompted to save any unsaved analyses at this time.

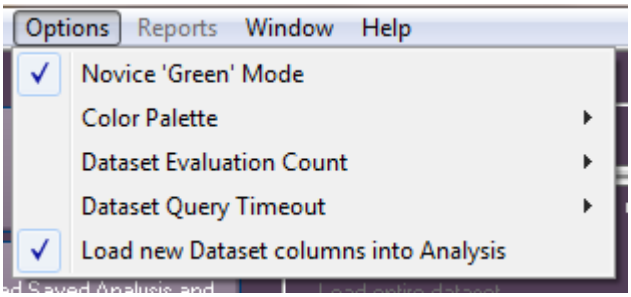
New Analysis

To perform a new analysis on the current dataset, click the "New Analysis" button from the "File" menu. To toggle between analyses, we select the one we wish to view from the "Win

Exit

To exit the Analytics program at any time, click the "Exit" button from the "File" menu or the . You will be prompted to save any unsaved analyses at this time.

OPTIONS MENU



Novice Green Mode

The Novice Green Mode applies to color green to the button that is the next logical step for the user, and is a good way to be guided through the process of analysis.

Color Palette

If you wish to change the color palette of Analytics, you may do so by selecting a new palette from the "Color Palette" drop-down menu. Note that selecting "Windows Desktop" directs the palette to match whatever the Windows Desktop's colors are set to.

Dataset Evaluation Count

The Dataset Evaluation Count drop-down menu allows us to change the number of records taken into account to determine the datatype of each column variable. 5,000 is the default number of records used to determine datatype; note that the larger this setting is, the more precise the datatype evaluation will be, though increasing this setting may take more time to load. Manually changing a column's datatype can be done by using the "Define Format" option when selecting the dataset or selecting a datatype from the "Type" drop-down menu on the Variable Control Panel.

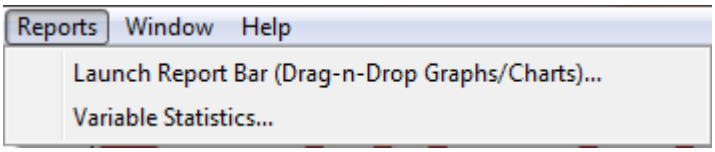
Dataset Query Timeout

The amount of time selected in the "Dataset Query Timeout" drop-down menu corresponds to the amount of time Analytics will allow a server to take before responding. If the server does not respond within the specified amount of time, Analytics will perform a timeout.

Load New Dataset Columns Into Analysis

The "Load New Dataset Columns Into Analysis" option grants Analytics the ability to load in any columns that have been added to the original data source post-analysis into the current analysis.

REPORTS MENU



Launch Report Bar

The report bar gives us a handy place to drag and drop any charts or tables we wish to capture anytime during our analysis. When we have secured all of the graphs and charts we want, we can export them all to a report or a PowerPoint presentation by selecting either toggle button on the bottom of the report bar and clicking "Export". Clicking "Clear All" will erase all collected diagrams from the report bar.

Variable Statistics

Clicking the "Variable Statistics" auto-generates a report that includes the number of dataset and sample records, dataset variables, analysis variables, and the statistics that go along with each analysis variable, including: mean, minimum and maximum values, variable type, number of missing values,

standard deviation, number of distinct values, and any filters or caps you may have applied during analysis.

Control Panel

GENERAL DESCRIPTION

The control panel tab provides us with the tools to make decisions about our data based on our own criteria. For example, if we want to examine a specific period of time, we need to filter out records that do not fall within the given time span. We also have the ability to change variable types and model availability here. Importantly, the Control Panel is the tab that is used to designate and edit your Y-variable.

Variable Control Panel

Variable	Type	Outlier Cap		Missing Handling	Model Availability	Filtering	
Athlete Flag	binary	n/a		no missings	If Related to Y		
Campus Housing Flag	binary	n/a		no missings	If Related to Y		
Campus Visit Flag	binary	n/a		no missings	If Related to Y		
Days between app date and term start	continuous			no missings	If Related to Y		
Gender	categorical	n/a		no missings	If Related to Y		
HS_GPA	continuous			no missings	If Related to Y		
Legacy Indicator	categorical	n/a		no missings	If Related to Y		
SAT - Math	continuous			no missings	If Related to Y		
Scholarship Flag	binary	n/a		no missings	If Related to Y		
State	categorical	n/a		default	If Related to Y		
Student ID	continuous			no missings	If Related to Y		

VARIABLE CONTROL PANEL TABLE

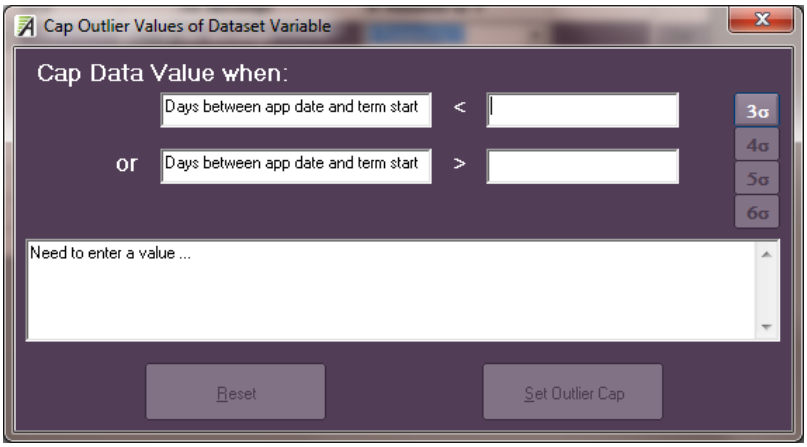
The Variable Control Panel table contains information about each of the variables used in analysis.

Variable Column

The variable column names each variable to be used in analysis.

Type Column

The type column describes each variable's type or distribution. If Analytics incorrectly determines the variable type, the user can change the type by choosing the best option from the drop-down menu.



Outlier Cap Column

Often in a large dataset there are variables that contain one or more atypical values, or infrequent observations. To ensure the quality of data used in analysis, users may wish to set a cap which each value of a given

variable cannot exceed or dip below. To do this, simply click the "Set" button in the next column and enter the range of values that is appropriate for each variable. When satisfied with the settings, click "Set Outlier Cap" to save and exit back to the Control Panel.

Missing Handling Column

Missing Handling identifies missing values within each variable. If missing values are found, several options are available from the drop-down menu:

- default (which can be defined on the "Model" tab),
- Fill with Zero,
- Fill with Variable Mean, or
- Exclude Entire Record

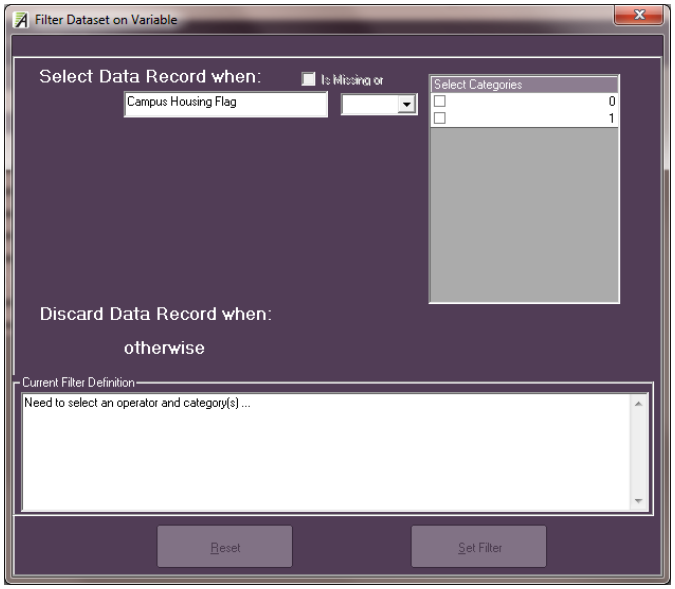
It is important to keep in mind that excluding multiple records can greatly reduce your sample size for analysis; do so with caution.

Model Availability Column

A variable's Model Availability column determines whether a column will or will not appear in any predictive model formula. Users may select an option from the drop-down list:

- "Exclude" a variable, even if it is related to the Y-variable,
- Make the variable "Available" even if it is not related to the Y-variable, and
- Allow the software to include the variable only "If Related to Y"

If the variable is data types text, date, or constant, the column will be marked as "n/a" and it won't be available for use in a model.



Filtering Column

The filtering column selects or discards records for analysis based on user-entered criteria. To filter, select a variable and click the "Set" button at the end of the row. Use drop down boxes in the Filter Dataset on Variable window to define the filtering criteria. Click "Set Filter" to save the filter and exit back to Control Panel.

Y VARIABLE

Select one variable to be the variable to be modeled (the "Y Variable") by either highlighting it then clicking the green downward arrow or by double-clicking the variable. Only binary and continuous variables may be selected.

EXPORT TO EXCEL

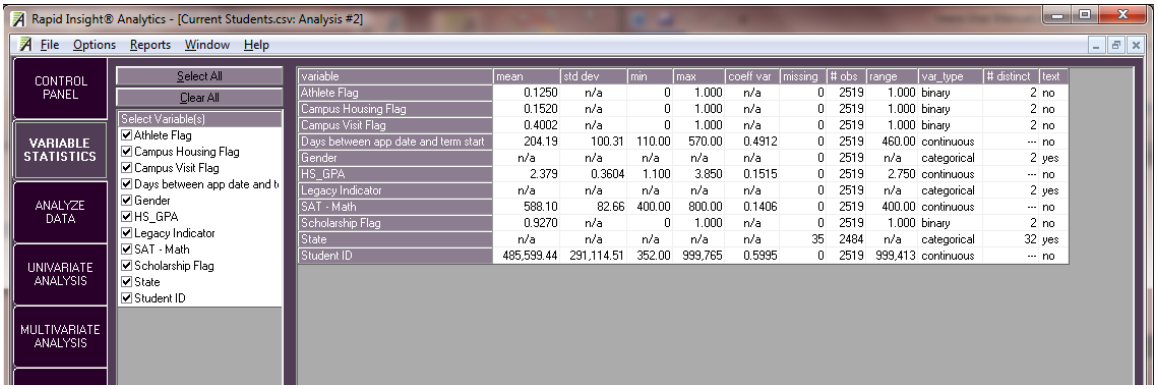


Save a copy of the Control Panel to Excel by clicking the "Export to Excel" button.

Variable Statistics

GENERAL DESCRIPTION

The Variable Statistics tab allows users to view simple statistics for each of the variables, giving a clear and concise way of gaining insight into the dataset. It is a good idea to check these statistics because problems such as data omissions and outlying points are often detected here.



variable	mean	std dev	min	max	coeff var	missing	# obs	range	var_type	# distinct	text
Athlete Flag	0.1250	n/a	0	1.000	n/a	0	2519	1.000	binary	2	no
Campus Housing Flag	0.1520	n/a	0	1.000	n/a	0	2519	1.000	binary	2	no
Campus Visit Flag	0.4002	n/a	0	1.000	n/a	0	2519	1.000	binary	2	no
Days between app date and term start	204.19	100.31	110.00	570.00	0.4912	0	2519	460.00	continuous	···	no
Gender	n/a	n/a	n/a	n/a	n/a	0	2519	n/a	categorical	2	yes
HS_GPA	2.379	0.3604	1.100	3.850	0.1515	0	2519	2.750	continuous	···	no
Legacy Indicator	n/a	n/a	n/a	n/a	n/a	0	2519	n/a	categorical	2	yes
SAT - Math	588.10	82.66	400.00	800.00	0.1406	0	2519	400.00	continuous	···	no
Scholarship Flag	0.9270	n/a	0	1.000	n/a	0	2519	1.000	binary	2	no
State	n/a	n/a	n/a	n/a	n/a	35	2484	n/a	categorical	32	yes
Student ID	485,599.44	291,114.51	352.00	999,765	0.5995	0	2519	999,413	continuous	···	no

VARIABLE STATISTICS GRID

This grid contains information about each variable's mean, standard deviation, minimum value, maximum value, data type, coefficient of variation, number of missing values, number of observations, number of distinct values, range, and whether a variable is text or not.

Mean

The mean is the average value of all of the observations of one variable.

Std Dev (standard deviation)

The standard deviation represents the "spread" of data around its mean. A higher value indicates a more spread out data set than a lower value.

Min (minimum)

The minimum value represents the smallest observation found.

Max (maximum)

The maximum value represents the largest observation found.

Coeff Var (coefficient variation)

This represents the ratio of the standard deviation to the mean. The larger this number is, the larger the standard deviation is relative to the mean.

Missing

This number denotes the number of missing (null) observations for each variable.

 # Obs (total # of observations)

This is the total number of non-null observations for each variable.

Range

The range represents the difference between the highest and lowest observation for each variable.

Var_type (variable type)

This represents the type of variable (continuous, binary, text, etc.).

 # Distinct (categories)

This represents the number of unique categories within each variable.

Text

This is a yes/no indicator for the presence of a text variable.

VARIABLE STATISTICS LEFT SIDEBAR

The left sidebar allows you to see summary statistics only for the selected variables. The statistics table adds or removes variables from view, making it easy to place variables next to each other to see how they compare.

Select All / Clear All / Select Variables

Select All displays all of the variables and their statistics at once; Clear All clears all variables from the table, while Select Variables individually displays variables by clicking the checkbox next to each one.



Variables Statistics Report

A report of the variable statistics selected can be generated and saved/exported by clicking on the “Variable Statistics Report” button below the list of variables.



Export to Excel

The statistics can be exported as a spreadsheet to Excel by clicking the “Export to Excel” button below the list of variables.



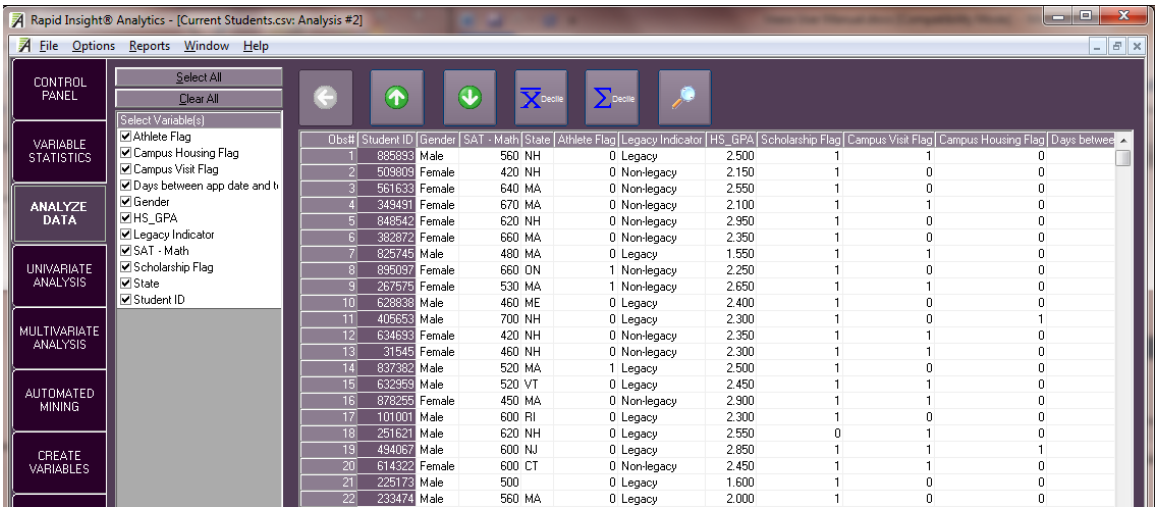
Analyze Data

GENERAL DESCRIPTION

Use the “Analyze Data” tab to sort, search, find sums and means (by both decile and category), as well as view of the entire dataset.

ANALYZE DATA RECORDS GRID

The data records grid contains all of our observations for each variable, ready to be sorted.



Sort Ascending / Sort Descending / Reset Sort Order



Sort any column by highlighting it and clicking on the green "up" arrow above the grid to sort ascending, or the green "down" arrow above to sort descending. Reset the sort order by clicking the green left-pointing arrow. To sort by multiple columns, drag the columns so that they are next to each other in the order you wish to sort. Click the first column, hold the shift key down and click the second column, then click either the ascending or descending sort button.

Decile and Take Means By Selected Column



To find deciles and take means of each decile of any select column, highlight the column and click the "Decile Mean" button above the grid. To return to the main grid, press the green "reset sort order" arrow.

Decile and Sum By Selected Column



To find deciles and take the sum of each decile of any select column, highlight the column and click the "Decile Sum" button above the grid. To return to the main grid, press the green "reset sort order" arrow.

Find



Search for any data value by clicking the "Find" button or using CTRL+F. The find window gives the option of searching by just the column and row of a highlighted cell by clicking the "Column/Row Only" button, or searching the whole grid by selecting the "Everywhere" button. Searches are done by row or by column, depending on which option is selected.

ANALYZE DATA LEFT SIDEBAR

The left sidebar displays only the selected variables. The table adds or removes variables from view, making it easy to place variables next to each other to see how they compare.

Select All / Clear All / Select Variables



Use Select All to see all of the variables at once; Clear All to clear all variables from the table, or Select Variables individually by clicking the checkbox next to each one.

Export to Excel

The results of a decile aggregation of the records can be exported as a spreadsheet to Excel by clicking the "Export to Excel" button below the list of variables.



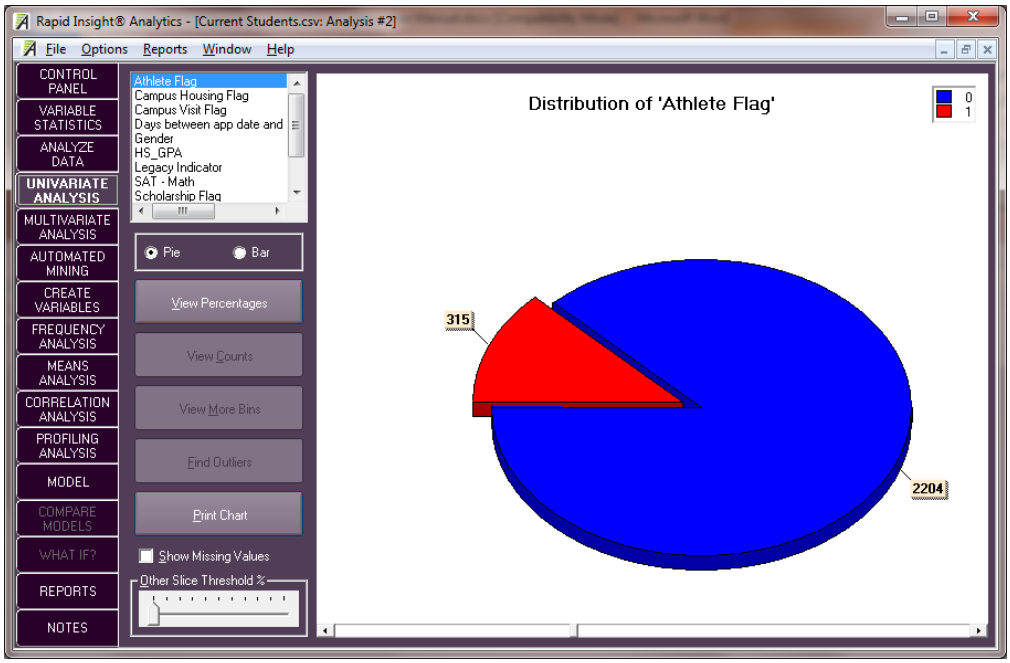
Univariate Analysis

GENERAL DESCRIPTION

The Univariate Analysis tab graphically displays only one variable at a time. View the distribution of any variable in the dataset by simply selecting the variable from the list.

UNIVARIATE ANALYSIS VARIABLE LIST

To perform a univariate analyses for a specific variable, select it from the variable list.



UNIVARIATE ANALYSIS CHART AREA

The chart area displays one chart at a time for each selected variable and chart type.

Pie / Bar (chart types)

If a variable is a categorical or binary variable, its distribution will automatically be shown as a pie chart. If the variable is continuous, its distribution will be shown as a bar chart. Users can choose to select the preferred chart type.



View Percentages / View Counts

Add the counts or percentages for each distribution by clicking the "View Counts" or "View Percentages" buttons.



View More Bins



Choose "View More Bins" to spread data out into more categories to see the effect that a different categorization setup has on each variable.

Find Outliers



If a continuous variable has values that are atypical (outliers), click on the "Find Outliers" button to view them. To view the effect that outliers have on each chart, choose "View Effect on Chart" to view a new chart which does not include the outlier values. Note: to actually exclude outliers from the analysis, set a column filter on the Control Panel tab.

Print Chart



Print out individual charts by clicking "Print Chart".

Show Missing Values



Check "Show Missing Values" to see a pie slice or bar for the missing values.

OTHER Slice Threshold % Slider

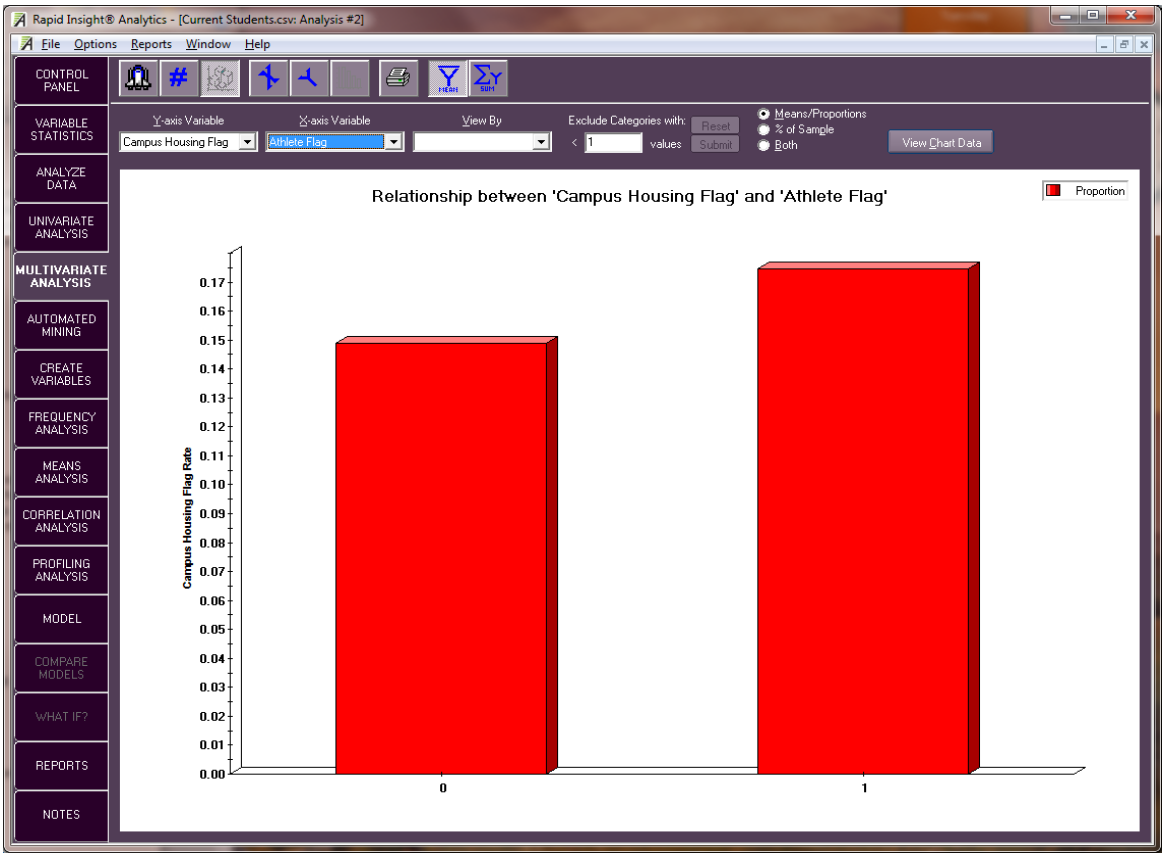
Categorical charts with a number of very small slices may be modified to combine the small slices into one large slice by sliding the "Other Slice Threshold %" slider bar. The slice that contains the combined smaller slices is always yellow and does not show a count or percentage. To see a pie slice or bar for the missing values, check the "Show Missing Values" box.



Multivariate Analysis

GENERAL DESCRIPTION

The Multivariate Analysis tab allows us to analyze two or more variables simultaneously, illustrating how pairs of variables may be correlated.



MULTIVARIATE UPPER TOOLBAR CONTROLS

This toolbar modifies the charts that the Multivariate Analysis tab creates. Control whether charts are created from the sums or means of variables. It add labels to charts to make them easier to read, changes their formatting to make them more aesthetically pleasing, and print charts.

Use Lift Index on Y-Axis

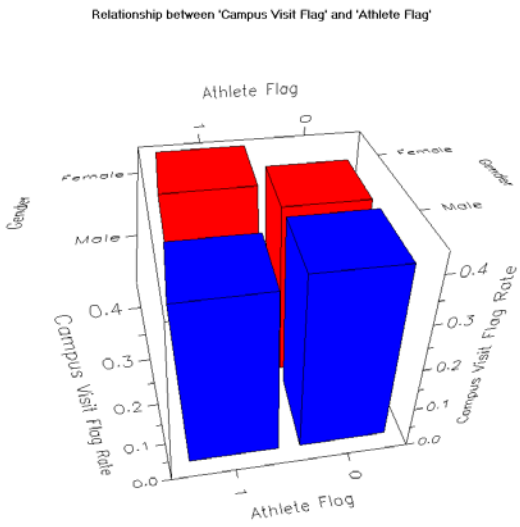


To view the lift index for each entry on the Y-axis, rather than the Y-value, click the "Use Lift Index" button. The lift index shows the Y-axis indexed to one, and for any value that falls above or below 1.0, we know that the variable mean for that combination of data is above or below the population mean.

Show Counts



By clicking the "Show Counts" button, the user can view counts as a label for each bar or point on our chart.



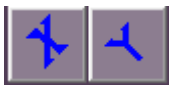
3-D Toggle Button



Clicking the 3-D Toggle Button allows users to switch between 2-D and a rotating, 3-D view of a chart. This button is available only when the X-axis and View By variables are both either binary or categorical. To

stop the auto-rotation, click anywhere on the chart and the chart will freeze. To see other views, drag the scroll bars on below and to the right of the chart.

Angle X-Axis / Y-Axis Annotations



These options angle the X-axis and Y-axis labels to improve legibility.

Sort Bars by Height



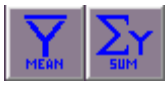
The "Sort Bars by Height" button is available if the X-variable is binary or categorical. This button orders the bars from largest to smallest, left to right.

Print Current Chart



Print the chart in view by clicking the "Print Current Chart" button.

Chart Y-Axis with Means / Sums

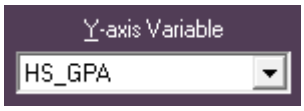


Choosing either the "Chart Y-Axis with Sums" or "Chart Y-axis with Means" button determines how the Y-axis will be distributed.

MULTIVARIATE LOWER TOOLBAR CONTROLS

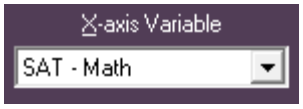
The lower controls on the Multivariate Analysis tab select and sort which variables to compare. Users can also exclude categories, and view these charts by their means or the percentage of the sample that they represent. Data records represented in each chart may also be viewed.

Y-axis Variable



To look for relationships between variables manually, first select a Y-axis variable to compare with multiple X-axis variables. This way, users can visually compare all variables against each other. The Y-axis variable can only be continuous or binary. If it is binary, the Y-value is expressed as a rate.

X-axis Variable



Choose each X-axis variable to compare to the Y-axis variable via the drop-down menu. If the X-axis variable is continuous, the output will be in the form of a point graph. If it is categorical or binary, the output will be in the form of a bar graph.

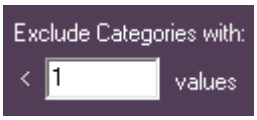
View By



Select a variable to "View By" to view it along with an X-axis and Y-axis variable all on the same graph. If the X-axis variable and View By variable are both either binary or categorical, the resulting chart will be 3-D. To switch to a 2-D view, use the 3-D toggle button on the upper toolbar.

Exclude Categories With:

Exclude those categories that don't have enough values for analysis by filling in the "Exclude Categories With:" field. Simply enter the number of records a category must represent to be included in each graph and click Submit. This option is available if the X-axis variable is categorical or binary.



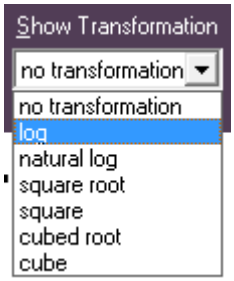


Granularity Slider

Use the Granularity Slider to increase or decrease the granularity of the graph.

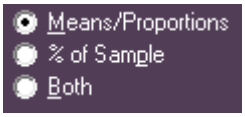
Show Transformation

Select a transformation under the "Show Transformation" drop-down menu to see if data is improved by applying a transformation to it.



Means / Proportions / % of Sample / Both

The "Means/Proportions" button shows the means or proportions of each X-axis variable as Y-axis. The "% of Sample" button shows the percentage of the sample each category represents as Y-axis. The "Both" button shows the means or proportions as shown with the "Means/Proportions" button as well as a blue line representing the percent of sample that fell into each category. These options are available for a categorical or binary variable X-variable.

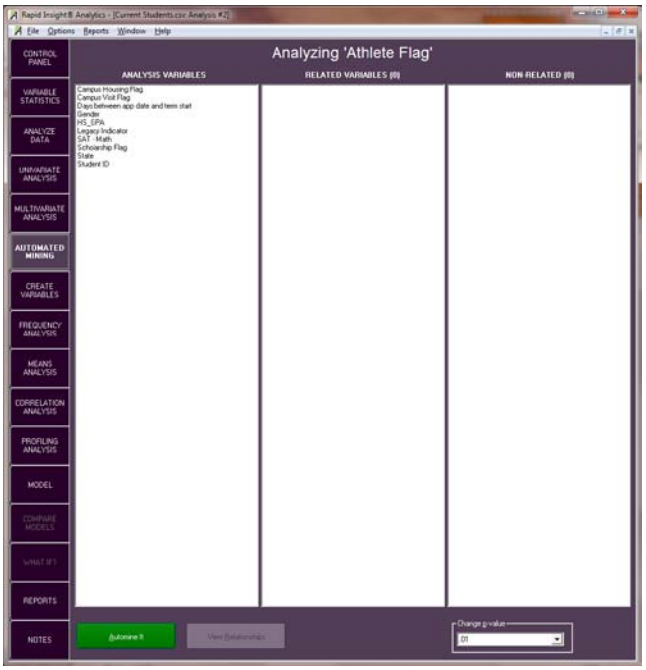


View Chart Data



Click the "View Chart Data" button to see the sum, mean, min, max, and count for each point or bar used in the chart. We are able to save this as an Excel file by clicking "Save".

Automated Mining



GENERAL DESCRIPTION

The Automated Mining tab lets us take advantage of Rapid Insight Analytics' ability to quickly evaluate all variables against the Y-variable to find any statistical relationships that exist. In order to accomplish this, a Y-variable must first be chosen in the Control Panel tab.

ANALYSIS VARIABLES

The Analysis Variables column lists all of the variables from the Control Panel tab. The Y-variable is at the top of the Automated Mining tab and will not be included in this list.

RELATED VARIABLES

The Automine process automatically separates the variables according to whether they are statistically related to the selected Y-variable. Related variables appear in this column. Note that Automine runs independently of the Control Panel. For example, variables designated as "Exclude" in the Control Panel may still appear in the "Related Variables" column.

NON-RELATED VARIABLES

The Automine process automatically separates the analysis variables according to whether they are statistically related to the Y-variable or not. Non-related variables appear in this column.

CHANGE P-VALUE



The "Change P-Value" drop-down menu gives the ability to change our p-value according to how statistically significant we require the relationship between the X-variables and the Y-variable to be. The lower the p-value, the more significant the relation is.

AUTOMINE IT



To begin the Automine process, simply press the "Automine It" button. Each analysis variable is tested against the Y-variable to determine if a statistically significant relationship exists between the two variables.

VIEW RELATIONSHIPS



The "View Relationships" button becomes available after the data is automined. It allows users to visually compare each related variable to the Y-variable.

Variable Relationships

GENERAL DESCRIPTION

The "Variable Relationships" window can be reached by clicking the "View Relationships" button on the Automated Mining tab. This window allows us to view the relationships between our Y-variable and each statistically related X-variable as a graph.

VARIABLE RELATIONSHIPS TOOLBAR CONTROLS

This toolbar modifies the charts that the Variable Relationships window creates. Select and sort which variables to compare, and then add labels to charts to make them easier to read, changes their formatting to make them more aesthetically pleasing, and print them. Users can also exclude categories, and view these charts by their means or the percentage of the sample that they represent. Data records represented in each chart may also be viewed.

Use Lift Index on Y-Axis



To view the lift index for each entry on the Y-axis, rather than the Y-value, click the "Use Lift Index" button. The lift index shows the Y-axis indexed to one, and for any value that falls above or below 1.0, we know that the variable mean for that combination of data is above or below the population mean.

Show Values

Selecting the "Show Values" button labels each bar or point with its Y-value.



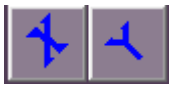
3-D Chart View



Clicking the 3-D Toggle Button allows users to switch between 2-D and a rotating, 3-D view of a chart. This button is available only when the X-axis and View By variables are both either binary or categorical. To stop the auto-rotation, click anywhere on the chart and the chart will freeze. To see other views, drag the scroll bars on below and to the right of the chart.

Angle X-Axis / Y-Axis Annotations

These options angle the X-axis and Y-axis labels to improve legibility.



Sort Bars by Height



The "Sort Bars by Height" button is available if the X-variable is binary or categorical. This button orders the bars from largest to smallest, left to right.

Show Regression



Clicking the "Show Regression" button superimposes a regression line fitted to the variable chosen. A regression line is only available when our X-variable is continuous.

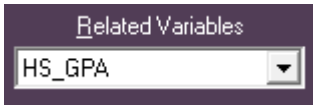
Print Current Chart



Print the chart in view by clicking the "Print Current Chart" button.

Related Variables

The "Related Variables" drop-down menu contains only the variables that Rapid Insight Analytics' chose as having a statistically significant relationship with the Y-variable during the Automine process.



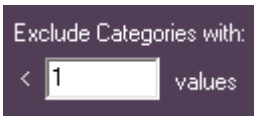
View By



Select a variable to "View By" to view it along with an X-axis and Y-axis variable all on the same graph. If the X-axis variable and View By variable are both either binary or categorical, the resulting chart will be 3-D. To switch to a 2-D view, use the 3-D toggle button on the upper toolbar.

Exclude Categories With:

Exclude those categories that don't have enough values for analysis by filling in the "Exclude Categories With:" field. Simply enter the number of records a category must represent to be included in each graph and click Submit. This option is available if the X-axis variable is categorical or binary.

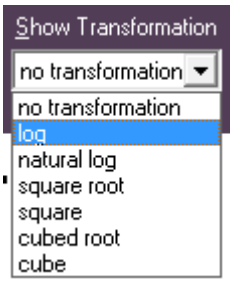


Granularity Slider



Use the Granularity Slider to increase or decrease the granularity of the graph.

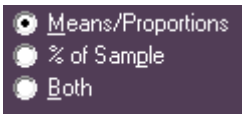
Show Transformation



Select a transformation under the "Show Transformation" drop-down menu to see if data is improved by applying a transformation to it.

Means / Proportions / % of Sample / Both

The "Means/Proportions" button shows the means or proportions of each X-axis variable as Y-axis. The "% of Sample" button shows the percentage of the sample each category represents as Y-axis. The "Both" button shows the means or proportions as shown with the "Means/Proportions" button as well as a blue line representing the percent of sample that fell into each category. These options are available for a categorical or binary variable X-variable.



View Chart Data

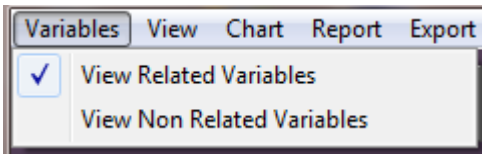


Click the "View Chart Data" button to see the sum, mean, min, max, and count for each point or bar used in the chart. We are able to save this as an Excel file by clicking "Save".

VARIABLE RELATIONSHIPS MAIN MENU

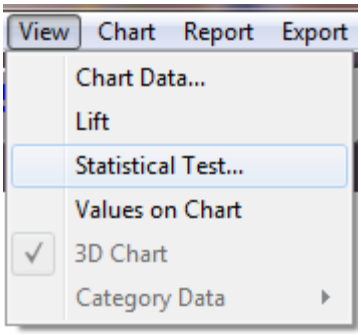
The Variable Relationships Menu allows us to view charts of each variable involved in Analysis. We can toggle between those variables deemed related or unrelated by the Automated Mining tab to see each variable and the relationship it has with our designated Y-variable.

Variables Menu - View Related/Unrelated Variables



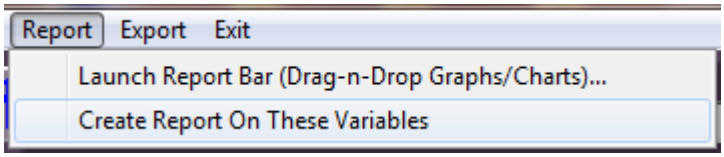
The "View Related/Unrelated Variables" lines allow us to toggle between graphs of variables that the Automated Mining tab has decided are unrelated to our selected Y-variable and those that are related to our Y-variable. We can gain more insight into these relationships by also selecting a "View By" variable from the drop-down menu.

View Menu - Statistical Test



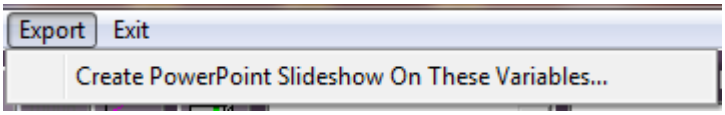
The "Statistical Test" line allows you to see the statistical test values for the variable shown. Depending on the type of variable selected, the output of this feature does change. Generally, a Z- or F-value is shown along with any pertinent information.

Report Menu - Create Report on these Variables



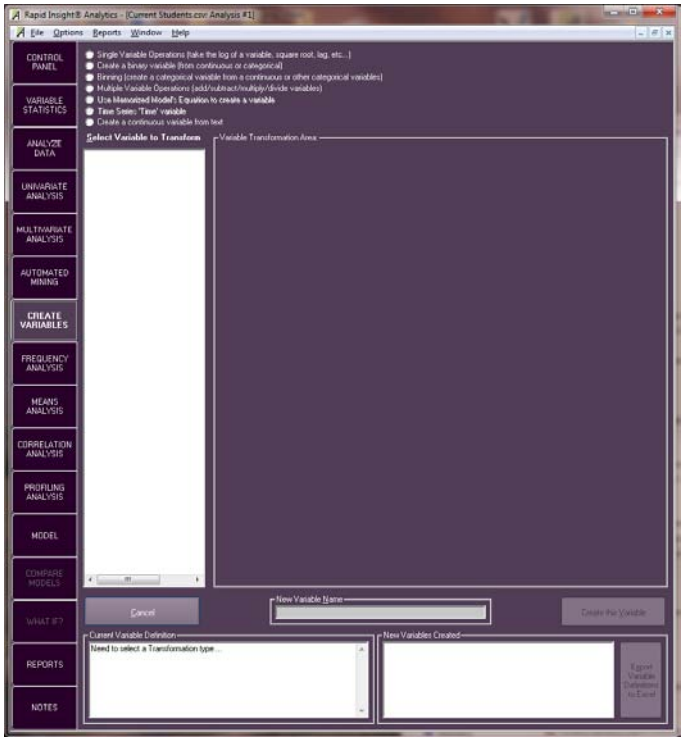
The "Create Report On These Variables" button will create a series of charts for the relationship between each predictor variable and the selected Y-variable that can be printed or saved as needed.

Export Menu - Create PowerPoint Slide Show on these Variables



This button creates a PowerPoint slideshow containing one chart per slide comparing each predictor variable to the selected Y-variable.

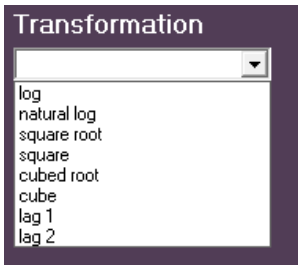
Create Variables



GENERAL DESCRIPTION

The Create Variables tab allows users to create new variables from existing variables. Choose what type of variable to create and rename variables in this tab. Once a transformation type is selected, only the variables that can be transformed in this way will be shown in the "Select Variable to Transform" window. Each created variable will be available in all other areas of the application, such as the Control Panel and Univariate Analysis tabs. We can view any created variable's definition in the "Current Variable Definition" box by highlighting that variable.

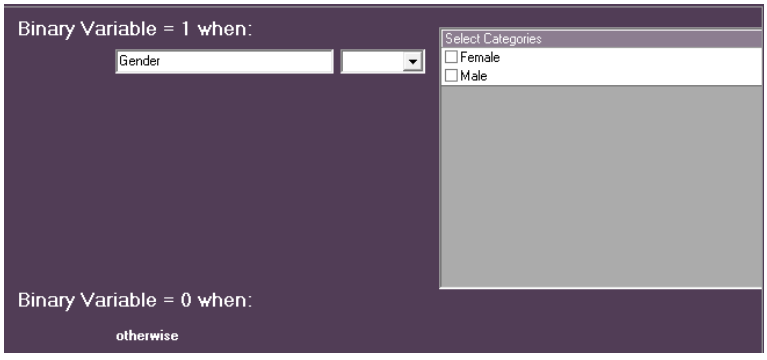
SINGLE VARIABLE OPERATIONS



After clicking the "Single Variable Operations" button and choosing which variable to transform, options for transformation from the "Transformation" drop-down menu appear – such as taking the log or square root of a continuous variable, sub string for text variables, and "days from" for date variables. After defining a new variable, name it in the "New Variable Name" field, and click "Create this Variable". It will then show up in the "New Variables Created" window.

CREATE A BINARY VARIABLE

Select "Create a binary variable" and choose a variable. The options will be different depending on the data type of the variable selected. Generally, users will be prompted to choose a category or an operation (<, = to,



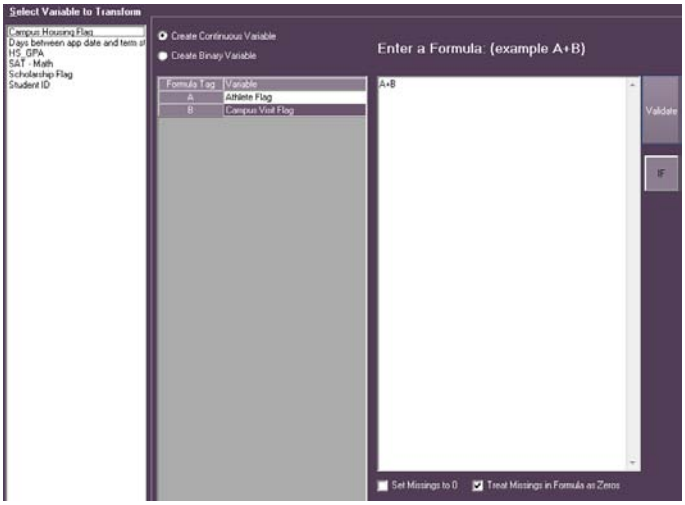
etc.) to sort values between the “0” and “1” categories.

BINNING

To create a categorical variable from a continuous or another categorical variable, select “Binning” and a variable. Depending on the data type of the variable selected, the next steps differ:

- Binning binary variables doesn’t usually make sense, but one helpful feature of binning is the “Is Missing” checkbox, which allows assigning a binary value to any missing data cells.
- For continuous variables, users are prompted to name our first category and define which values fall into this category. To move to the next category, click “Validate” and repeat each step. Be sure to name the variable in the “New Variable Name” field. When finished creating categories, click “Create this Variable” and the variable will save. Note that if there are any values which fall outside of the categories created, they will automatically be binned together into the category “other”.
- For categorical variables, name the first category and define which categories fall into the new category selecting “is” or “is not” from the drop-down menu, and then the categories to bin according to this guideline. Click “validate” to create a new category and repeat as necessary. When finished creating new categories, name the variable in the “New Variable Name” field, and save by clicking “Create this Variable”. Note that if there are any categories not included in the new categories, the data contained in them will be binned together into the category “other”.

MULTIPLE VARIABLE OPERATIONS



The multiple variable operations button allows users to create new variables by performing arithmetic operations and/or using logic statements to combine existing variables. Select all of the variables to work with from the “Select Variable to Transform” window. After selecting variables, choose to create either a continuous or binary variable. The variables selected are listed in the Variable column. The Formula Tag column assigns each variable a letter to be used in the “Enter a Formula” window. Enter any operations or logic statements desired to create a new variable. When finished, click “Validate”, fill in the “New Variable Name” field, and click the “Create this Variable” button. You can any of the new variables created by selecting a name from the “New Variables Created” window. The formula for this variable will then appear in the “Current Variable Definition” window.

USE MEMORIZED MODEL’S EQUATION TO CREATE A VARIABLE

After selecting the “Use Memorized Model’s Equation to create a variable” button, memorized models will appear in the “Select Variable to Transform” window instead of variables. Select one, then, if desired, change elements of the model formula within the “Model Equation” window before providing a name and saving the new variable.

CREATE A CONTINUOUS VARIABLE

After selecting the "Create a continuous variable from text" button, select the text variable to transform, rename the variable in the "New Variable Name", and click "Create this Variable".

Frequency Analysis

GENERAL DESCRIPTION

The Frequency Analysis tab looks at the frequency of occurrence for binary and categorical variables.

SINGLE VARIABLE CATEGORY FREQUENCIES

By clicking the "Single Variable Category Frequencies" button, the counts and percentages for each variable selected are displayed. Click "Submit" to view the selected variables, values, counts, and percentages in the window below the variable lists. Users may also choose to view the data in report form by clicking "Frequency Report" or as an Excel spreadsheet by clicking "Export to Excel".

Variable	Value	Count	Percentage
Gender			
	Female	1007	40.0%
	Male	1512	60.0%

TWO WAY CROSS FREQUENCIES

Crosstab	Type	By Variable	By Value	Crosstab Results		
Gender by Athlete Flag						
	Raw Numbers					
				Gender:		
				Female	**Male**	**Total**
		Athlete Flag:	**0**	876	1328	**2204**
			1	131	184	**315**
			Total	**1007**	**1512**	**2519**
	Row Percentages					
	Column Percentages					
	Total Percentages					

"Two Way Cross Frequencies" further breaks down variable frequencies (for one or more variables) by including another variable. First, select one or more variables as your "Selected Variables" from the "Categorical and Binary Variables" window. To see the selected variables by another variable, select a "By" variable from the "Categorical and Binary Variables" window. Click the "Submit" button to view the frequencies in the window below the variables lists. There are four view for each comparison of two variables - raw numbers, row percentages, column percentages, and total percentages. Users may also choose to view the data in report form by clicking "Frequency Report" or as an Excel spreadsheet by clicking "Export to Excel".

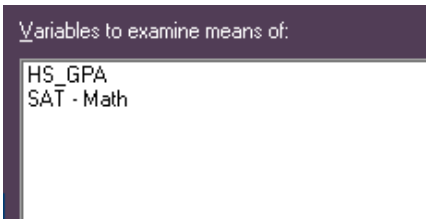
Means Analysis

GENERAL DESCRIPTION

The Means Analysis tab performs means analysis on any continuous variable, which includes a mean, count of the observations, max, and min for each category or subclass. The means for each category of

a categorical or binary variable can also be further sub-classed by a second categorical or binary variable. To see a report of this data, select the "Means Report" button, or to view as an Excel spreadsheet, select the "Export to Excel" button.

Variables to Examine Means



From the variable list, select a variable to examine the mean of by highlighting the variable and clicking the green arrow to move it to the "Variables to examine means of" window. You may select more than one variable to examine the means of in this step. The means that appear in the output table are the means (categorized by any other variables selected) for the variables selected in this field.

Take Means By



From the variable list, select a variable by highlighting it and clicking the green arrow to move it to the "Take means by" window. Next click "Submit" if you are ready to view the means analysis.

Sub-Class By



Further subclass the means analysis of variables by selecting a third variable to "View By". Do this by highlighting a variable from the variable list and clicking the green arrow to move it to the "Sub-class by" window. Next click "Submit" when ready to view the means analysis.

Correlation Analysis

GENERAL DESCRIPTION



	SAT - Math	Scholarship Flag
SAT - Math	1	0.0227
Scholarship Flag	0.0227	1

The Correlation Analysis tab displays positive and negative correlations between variables. The Correlation Matrix displays the correlation coefficients for each variable pair. The correlations range from -1 to +1 and indicate the degree of association between two variables. A positive value implies a positive association while a negative value implies an inverse association. The closer each value is to -1 or +1, the stronger the degree of association.

Select All / Clear All / Select Variables

Select All to see all of the variables and their statistics at once, Clear All to clear all variables from the table, or individually Select Variables by clicking the checkboxes next to each one.



Export to Excel



The Correlation Matrix can be exported as a spreadsheet to Excel by clicking the "Export to Excel" button below the list of variables.

Profiling Analysis

GENERAL DESCRIPTION

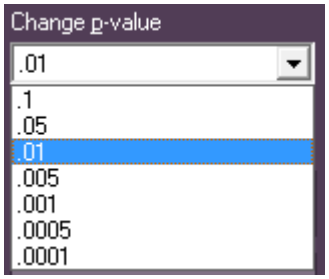
The Profiling Analysis tab compares two groups of a binary variable to determine how statistically different they are.

SELECT A BINARY VARIABLE TO PROFILE



The "Select a Binary Variable to Profile" window lists all of the binary variables in a dataset. After highlighting a variable, create names for the binary variable's two values in the Profile Work Area.

CHANGE P-VALUE



The "Change P-Value" drop-down menu gives the ability to change the p-value according to how statistically significant the relationship between our X-variables and Y-variable should be. The lower the p-value, the more significant the relation is.

AUTO-PROFILE



Clicking the "Auto-Profile" button generates a chart that compares the categories of the binary variable chosen to variables for which the two categories have statistically significantly different means or proportions. The upper portion of the chart compares each category to each of the continuous variables in a dataset by showing the means of each variable for both categories of the binary variable. The lower portion of the chart compares each category to each of the binary and categorical variables in the dataset for which the two categories have statistically significantly different means or proportions. By checking the "Show Z Score" button, a Z-score (measure of how many standard deviations an observation is above or below a mean) is included. The larger the Z-score, the further away each observation is from the mean.

VIEW PROFILE GRAPHICALLY



Clicking the "View Profile Graphically" button displays graphs for each variable via the "Profile Against Variable" drop-down menu. Options to see the lift index, add count labels, angle the X-axis and Y-axis labels, and print each chart individually are also available. Users can display all values by means/proportions, percentages of sample, or both. Users can also click the "View Chart Data" button to view each chart's data.

PROFILE REPORT



Clicking the "Profile Report" button automatically generates a report of the values shown in the Auto-Profile chart. In this window we can print

the report or send a copy of it to a word processing program.

EXPORT TO EXCEL



These profiles can be exported as a spreadsheet to Excel by clicking the "Export to Excel" button below the list of variables.

Model

GENERAL DESCRIPTION

Regression models are powerful tools for predicting the likely response of one variable ("Y-variable") based upon on the known behavior of other variables. The Model tab allows users to set specifications about which variable to model on and how accurate it should be.

REGRESSION TYPE



Analytics uses regression analysis to develop its predictive models. Regression uses several techniques for modeling and analyzing data describing past behavior. More specifically, regression analysis helps explain how the typical value of the Y-variable changes when any one of the other, independent variables is changed, while the other independent variables are held fixed. This is known as a "step". These series of stepwise results are then combined into a formulaic whole – the predictive model. Analytics offers two types of regression analysis – Logistic and OLS (Ordinary Least Squares).

Logistic Regression

Logistic regression is used when the Y-variable is a binary. The resulting model looks to predict the likelihood of an event to take place.

OLS Regression

Ordinary Least Squares (OLS) Regression is used when the Y-variable is continuous. Users may select the Logistic regression method to model continuous variables, though this is not recommended. The resulting model looks to predict the rate or quantity of an event.

MISSING VALUES DEFAULT HANDLING



The Missing Handling column identifies any missing values within each variable. If missing values are found, an option must be selected from the drop-down menu. "Default" (which is defined on the "Control Panel" tab), "Fill with

Variable Mean", or "Exclude Entire Record". It is important to keep in mind that excluding multiple records can greatly reduce sample size for analysis; do so with caution.

Fill with Variable Means

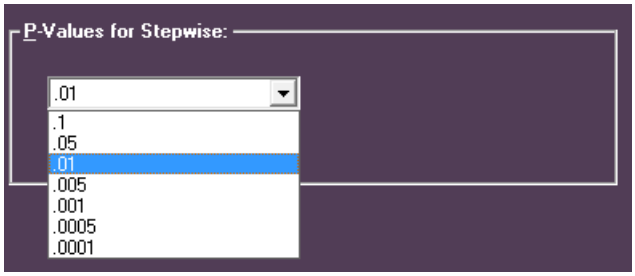
Clicking the "Fill with Variable Means" button will fill in the missing values for each variable with the average value for that variable.

Exclude Entire Record

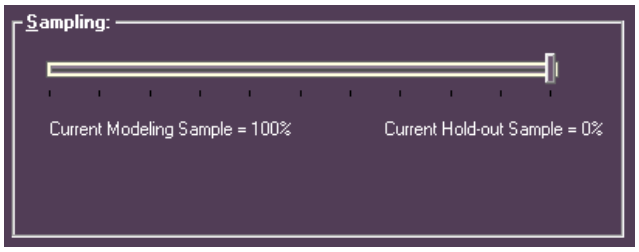
Clicking the "Exclude Entire Record" will not replace a missing value with any other value; rather, it will throw out that record for all variables so that it will not affect the analysis.

P-VALUES FOR STEPWISE

The "P-Values for Stepwise" drop-down menu gives the ability to change the p-value according to how statistically significant the relationship between the X and Y-variables should be. The lower the p-value, the more significant the relation is.



SAMPLING



The "Sampling" slider adjusts how many records to use to create the model, and how many to hold out from development to evaluate a completed model's accuracy. Records are chosen at random.

MODELING OPTIONS

Modeling options control how a model is built and what is displayed in the outcome.

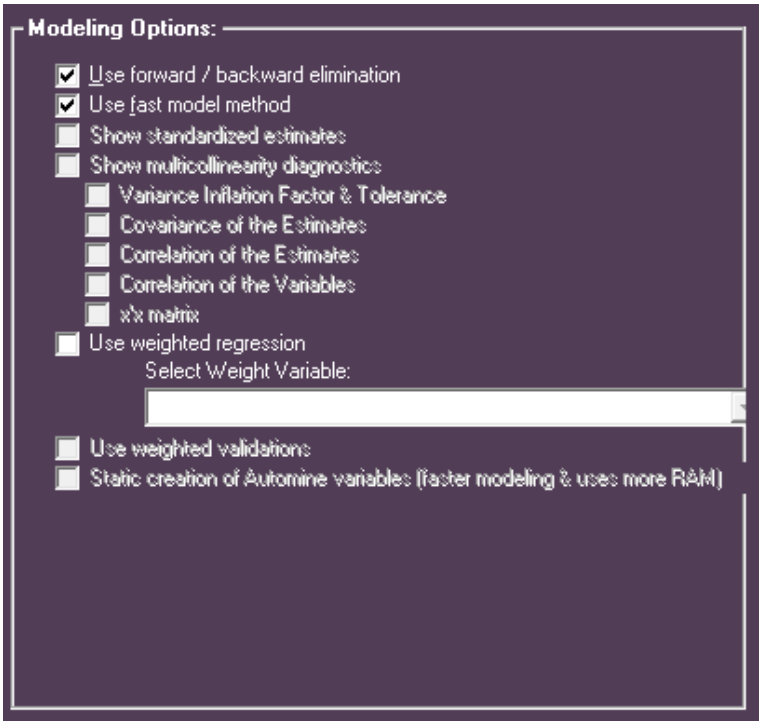
Use Forward/Backward Elimination

By checking "Use forward / backward elimination", Analytics will check to see if any of the variables should be dropped from the model during every model-building step. If a variable that has already entered the model is no longer statistically significant at the chosen p-value, that variable will be removed from the model.

Use Fast Model Method

Checking the "Use fast model method" box speeds up model development time by eliminating any variables once their significance level falls below the chosen threshold. All variables will be considered during each step if this option is not checked. Automatically-developed models may also take significantly longer to build on large datasets.

Show Standard Estimates



By checking the "Show standardized estimates" box, the output will include a column labeled "Std Est". This column represents the estimates of the regression coefficients from the "Coef" column standardized to a normal distribution. This option is only available when working with an OLS Regression model.

Show Multicollinearity Diagnostics

Checking the "Show Multicollinearity diagnostics" box displays any highly correlated relationships between predictor variables. Although correlated variables can exist in a successful predictive model, the values assigned to their coefficients may not accurately estimate each collinear variable's impact on the Y-variable if the other correlated variables were to be removed. Diagnostics we can include in our output are:

- Variance Inflation Factor & Tolerance: The Variance Inflation Factor (VIF) is equal to the reciprocal of tolerance. In general, the VIF represents the severity of Multicollinearity by measuring how much the variance of a coefficient is increased because of collinearity.
- Covariance of the Estimates: The matrix for the covariance of the estimates can be found in the "Diagnostics" section of the output. In general, a large positive value of covariance can mean that two variables are linearly related; a negative value can indicate that the two variables tend to move in opposite directions.
- Correlation of the Estimates: The correlation matrix can be found in the "Diagnostics" section of the output. The correlation between two estimates is equal to the slope of the regression line between the variables when they have both been normalized. Thus, a value of one indicates a direct correlation, and the closer to + 1 or -1 the value is, the more correlated the two variables are. The sign of the correlation represents a positive or inverse relationship.
- Correlation of the Variables: The correlation matrix can be found in the "Diagnostics" section of the output. The correlation between two variables is equal to the slope of the regression line between the two variables. Thus, a value of one indicates a direct correlation, and the closer to + 1 or -1 the value is, the more correlated the two variables are. The sign of the correlation represents a positive or inverse relationship.
- x'x Matrix: The x'x matrix (also called the variance-covariance matrix) can be found in the "Diagnostics" section of the output. This matrix displays the variances of each variable on the diagonal and the covariances between variables as the off-diagonal values.

Use Weighted Regression

By checking the "Use weighted regression" box and filling in the "Select Weight Variable" field with a continuous variable, we opt to assign a weight to each row equal to the value of the weight variable for

that row. A weight value of zero would, therefore, cause a row to be ignored, and a weight value of two would cause that row to be weighted twice – as though that row appeared twice in the dataset.

Static Creation of Automine Variables

This option is currently not available.

BEGIN REGRESSION ANALYSIS



When the regression type and options have been chosen, click "Begin Regression Analysis". The program will build a matrix of variables in preparation for the regression analysis.

Building a Model to Predict Y-Variable

GENERAL DESCRIPTION

This screen is where models to predict the Y-variable are built. Rapid Insight Analytics can build a model automatically or users can manually create their own custom models.

REGRESSION OPTIONS

By using the Regression Options, users can customize their regression model.

Build Model Automatically



The easiest and fastest way to build a model is by clicking the "Build Model Automatically" button. This option allows Rapid Insight Analytics to build what it feels is the best model that can be built on the current dataset. It takes all non-linearities in the data into account and makes all of the appropriate transformations on the candidate variables automatically.

Run Regression

Users may choose to build their own models by manually adding and/or removing variables or use the same method to change a "Build Model Automatically" model. Clicking "Run Regression" will then regressively analyze the customizations and produce the model.

Suggest Next Variable

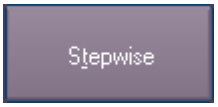


Clicking the "Suggest Next Variable" button allows Analytics to decide which variable from the "Variable List" window would best improve the model and moves it automatically to the "Include these" window.

Explain Model Graphically

To view graphs of the relationships between the X and Y-variables, click the "Explain Model Graphically" button. This action opens the Variable Relationships window (see the section Variable Relationships).

Stepwise



By clicking the "Stepwise" button, Rapid Insight Analytics will check to see if any of the variables should be dropped from the model following each step of the regression analysis,. If a variable that has already entered the model is no longer statistically significant at the chosen p-value, that variable will be removed from the model.

Exit Regression

Return to the main program screen by clicking "Exit Regression" or closing the window.

VARIABLE LIST

The "Variable List" window contains all of the variables available when building a model. Expand the list of available options by checking the "View new variable suggestions" box.

MEMORIZE MODELS

To capture a model, enter a name in the "Model Name" field and click the "Memorize Model" button. This retains the model for recall later. It also makes it available in other areas in the program, such as the Comparing Models and Create Variables tabs.

REGRESSION RESULTS

The "Regression Results" window is located below the variable lists. It displays the model's variables, their coefficients, standard errors, t-values, and p-values. All of the regression analysis steps taken to obtain the current model are also available by expanding the Model Steps section.

The "Diagnostics" section, when expanded, provides objective measures evaluating the quality of the model.

MODEL DECILE CHART

A decile column chart is automatically generated when a model is produced. It's a graphical representation of the results of the model having been tested by being applied to the Holdout Sample records. The red bars depict the test results. A set blue bars depicting the predicted outcome will appear when the "Show Actual and Predicted" option box that appears above the graph is checked. There are three separate decile views - "Decile Analysis", "Decile Lift", and "Cumulative Lift".

Decile Analysis

The decile analysis divides the total population into ten equal bins (deciles). The bars on this chart represent the average Y-value for each bin.

Decile Lift

The decile lift shows the percentage increase in accurate prediction by using the given model.

Cumulative Lift

The cumulative lift shows the cumulative percentage increase in accurate prediction for the given decile and lower.

Compare Models

GENERAL DESCRIPTION

The Compare Models tab provides the ability to evaluate the models memorized on the Model tab by comparing them to each other. Click on each model to view them individually or select multiple models to compare by holding down the "CTRL" or "SHIFT" key and clicking on each model to compare. Control what to view about each model by checking the boxes ("Model Data", "Decile Analysis", and "Cumulative Lift") on the right side of the screen. To display labels on a graph, simply click on it. Graphs are viewable by decile, twentile, and percentile by selecting an option from the drop-down menu.

Model Report

Click "Model Report" to generate a report that can be printed or exported as a text document.

Export Model to Excel

These model comparisons can be exported as a spreadsheet to Excel by clicking the "Export Model to Excel" button below the "Model Report" button.

Saving Scoring Model

In order to score a model later one, save it here by selecting just one model from the model diagnostics table and clicking the "Save Scoring Model" button. A file with an ".rism" file type is produced.

What If?

GENERAL DESCRIPTION

The "What If?" tab allows users to explore how the Y-variable would be affected if one or more of the related variables were increased or decreased.

Note: '*What If*' functionality makes assumptions regarding causalities in the data and for this reason results could turn out to be much different than projected. Rapid Insight® makes no guarantees regarding accuracy of these projections either implied or explicit and this tool is to be used at your own risk.

SELECT A MODEL / MODEL VARIABLES

Explore "what if?" scenarios using any saved model by selecting it from the "Select a Model" window. The "Model Equation" window displays the established formula for the selected model. To perform a "what if?" analysis, select any of the variables in the "Model Variables" window and use the slider control to change its value. The start value for any variable is its mean. The graph of the variable to the right will change to reflect the variable's "new" mean as one or more a slider controls is repositioned.

Reports

GENERAL DESCRIPTION

The Reports tab allows us to bring all of the information we've gathered into one report to be shared. Here we can import previously saved reports, or generate a report by using content saved to the Report Bar.

DOCUMENT EDITING CONTROLS

Here we can import previously saved reports, or generate a report by using content saved to the Report Bar.

New

Clicking the "New" button generates a new blank report.

Open

To open a saved report, click "Open" and select the report.

Save

Clicking "Save" will save the current report.

Print

Clicking "Print" will print the current report.

Cut

To cut, select the text or object you'd like to move, and click "Cut".

Copy

To copy, select the text or object you'd like copied, and click "Copy".

Paste

Click "Paste" to paste the last text or object you have cut or copied.

Bold

Highlighting text and selecting the "Bold" button will embolden that text.

Italics

Highlighting text and selecting the "Italics" button will italicize that text.

Underline

Highlighting text and selecting the "Underline" button will underline that text.

Left/Center/Right Justified

To change the justification or alignment of text or objects, highlight them and click on the justification you'd like.

Font Controls

Change the font color, size, and type by using the drop-down menus. Preview any combination of these options by clicking the "…" button next to the drop down menus and selecting the color, size, and type.

Report Bar

Clicking the "Report Bar" button allows you to drag and drop charts, graphs, etc. from the Report Bar onto a report.

Notes

GENERAL DESCRIPTION

The Notes tab creates and saves notes. Users can access them via this tab at any time. To create a note, first fill in the "Short Desc" field. Add more information via the "Note" window if desired. The program automatically numbers each note; click on each note number to view the "Note" window.

SCORING MODULE

Select and Load a Dataset

GENERAL DESCRIPTION

The Scoring module can provide a predicted Y-variable for each X-variable given any model. It also evaluates the model based on how closely each predicted Y-variable correlates to the actual values of the Y-variables.

SCORE TO

We have the option of scoring to a flatfile or database column by selecting the appropriate button option in the "Score to" box.

SELECT DATASET

Click "Select Dataset" and choose the type of file you'd like to open from the tabs on the left side of the screen. Find the correct file path and click "OK" to open.

SELECT SCORING MODEL

When your dataset has loaded, click "Select Scoring Model" and choose a .rism (Rapid Insight Scoring Model) file to evaluate.

Just Score the Dataset / Score the Dataset and Validate the Model

Choosing the "Just score the dataset" option presents only a column with a predicted Y-value appended to your original table or dataset as output. Choosing "Score the dataset and validate the model" will generate a chart below the drop-down menu displaying the decile analysis or cumulative list (decided by the toggle buttons to the right of the drop-down menu) of the model for predicting the given Y-variable.

Output to Flatfile

Here we have the options of "Output Decile", "Output Percentile", or "Don't Output Decile Info" in the drop-down menu. Choosing to output decile information will allow the program to display a column in your output (next to the scoring column) that displays the decile or percentile that each observation falls under.

Start Scoring

When you are ready to score, click "Start Scoring" to view a list of all variables in the dataset for selection. When you have selected the variables you would like to include, click "Ok".

Contact Us

For any issues/questions about Rapid Insight® Analytics, please send an email to RI Support at support@rapidinsightinc.com