Seagate Software

Seagate Holos Analytical Features

Find the Information You Need in the Data You Have

Overview

Holos provides wide variety of analytical tool sets and statistical functions in the standard product. These tools and functions are highly flexible and can be configured or customised to meet the specific analysis needs of each Holos application. Some of these tools and functions are available from the standard Holos worksheet (look for the “modelling” menu at the top of the worksheet) and can be applied to any Holos data available in the worksheet view. Other tools and functions are available from the Holos “Analysis Tools Desktop”. The tools and functions available from the Analysis Tools Desktop can also be provided to end users by adding buttons or menus to your Holos application. Other Analysis functions are implemented in the Holos language and can be easily included in any Holos application. Since all of the analytical and statistical functions are available in the Holos language, they can be customised or modified to meet your specific requirements.

Contents

General Numeric Functions 2
Built-in Arithmetic Functions 2
Number Generation "Holos_Support:random.hl" 2
Financial Functions 4
Holos Support Rule Functions ("Holos_Support:*.rfn") 4
Statistical Tools 6
General Analysis 6
Forecasting, Smoothing & Time-Series Analysis 7
Modelling/Optimisation 9
"What-if?" Analysis 9
"What-if?" and "What’s-best?" Analysis 9
Data Mining 11
Simple Data Mining using Pattern Matching 11
Advanced Data Mining 11
Neural Networks 13
Neural Network Analysis 13

General Numeric Functions

Built-in Arithmetic Functions

ABS (*x*)	absolute value
ACOS (*x*)	arc-cosine (returns angle in degrees: 0:180)
ASIN (*x*)	arc-sine (returns angle in degrees: -90:+90)
ATAN (*x*)	arc-tangent (returns angle in degrees: -90:+90)
AVG (*x*,...)	average value from list of values
COS (*x*)	cosine (angle in degrees)
EXP (*x*)	natural exponentiation (e**x)
FACTORIAL (*x*)	factorial of x (x!)
HCF(*value1*,*value2*)	highest common factor
INT (*x*)	integer value of x
LCM(*value1*,*value2*)	lowest common multiple
LN (*x*)	natural logarithm (base e)
LOG (*x*)	common logarithm (base 10)
LOGN (*base*,*value*)	calculates a logarithm to an arbitrary base
MAX (*x*,...)	maximum value from list of values
MIN (*x*,...)	minimum value from list of values
MOD (*x*,*y*)	modulo (gives remainder as x-INT(x/y)*y)
RANDOM ()	return random number 0.0 <= rnd < 1.0
ROUND	round to decimal places
SELECT(*true*,*false*,*condition*)	conditional calculation
SIN (*x*)	sine (angle in degrees)
SQRT(*x*)	square root
TAN (*x*)	tangent (angle in degrees)

A full range of text and date functions is also available.

Number Generation "Holos_Support:random.hl"

Random Number Generation

Loads the following functions to generate random numbers for a number of distributions:

rnd_normal (mean, standard deviation)

Returns a random number that conforms to a normal distribution with the specified *mean* and *standard deviation*.

rnd_uniform (*a*, *b*)

Returns a random number that conforms to a uniform distribution in the range a <= number <= b.

rnd_lognormal (*scale*, *shape*)

Returns a random number that conforms to a lognormal distribution with the specified *scale* and *shape* deviation.

`rnd_exp (mean)`

Returns a random number that conforms to an exponential distribution with the specified *mean*.

`rnd_tri (a, b, c)`

Returns a random number that conforms to a triangular distribution with the arguments *a*, *b*, and *c*.

`rnd_binomial (t, p)`

Returns a random number that conforms to a binomial distribution with a population of *t*, and probability *p*.

`rnd_poisson (mean)`

Returns a random value that conforms to a poisson distribution with the specified mean.

Series Generation

Loads the following functions to generate series of numbers:

`ap (n, a, d)`

Returns the n^{th} element of an Arithmetic Progression described by the first term *a*, and the common difference *d*.

`gp (n, a, r)`

Returns the n^{th} element of a Geometric Progression described by the first term *a*, and the common factor *r*.

Financial Functions

Holos Support Rule Functions ("Holos_Support:*.rfn")

General Calculations

ACCUMULATE (*a*,*b*,*c*)

Cumulates row *a* into row *b*, resetting at the interval specified by *c*.

CUMULATE (*a*,*b*,*c*)

Returns the cumulative sum of row *a* to row *b* at the frequency specified by *c*.

OFFSET (*a*, *b*, *c*)

Shifts row *a* by *c* number of columns and outputs the result of the shift to row *b*.

SPREAD (*a*, *b*, *c*)

The data in row *a* is split up according to the percentages in row *b* and output to row *c*.

Averaging

MEDIAN (*a*)

Returns the median average for the values in row *a*. The median is the middle value after the row has been sorted.

MODAL (*a*)

Returns the modal average for the values in row *a*. The mode is defined as the most frequent value. Zero is returned if there is more than one potential modal class.

MOVING (source, destination, c)

Evaluates the moving average of *c* numbers in the *source* row, outputting to the *destination* row.

Interest Calculations

COMPOUND (*a*,*b*,*c*)

Applies the specified rates of a compound interest in row *b* to a row *a* and outputs to row *c*.

DISCOUNT (*a*,*b*,*c*)

Applies the specified rates of a compound discount in row *b* to a row *a* and outputs to row *c*.The first column is base and not discounted.

Loan Related Calculations

LOAN_BALANCE (*loan, rate, repay, balance*)

Returns the balance outstanding on a *loan* given the interest *rate* and the *repay*ments, outputting to row *balance*.

LOAN_INTEREST (loan, rate, repay, interest)

Returns the interest on a *loan* given the interest *rate* and the *repay*ments, outputting to row *interest*.

LOAN_PAYMENT (*loan, rate, repay, payment*)

Returns the sum of interest paid and the loan repayment for each period of a *loan* given the interest *rate* and the *repay*ments, outputting to row *payment*.

Cash Flow Calculations

DCF (*a*)

Evaluates the Discounted Cash Flow (or internal rate of return IRR%) of row *a*.

NPV_ROW (*a, b, discount*)

Returns the Net Present Value of a stream of cash flows in row *a* given a percentage *discount*, outputting to row *b*.

NPV_SINGLE (*a, discount*)

Returns the Net Present Value of a stream of cash flows in row *a* given a percentage *discount*.

Depreciation

REDUCE_ASSET (*a, b, residual, life*)

Depreciates assets in row *a* using the reducing balance method given the *life* of each asset and the *residual*, outputting to row *b*.

REDUCE_DEP (*a, b, residual, life*)

Returns the depreciation charges for assets in row *a*, calculated using the reducing balance method given the *life* of each asset and the *residual*, outputting to row *b*.

STRAIGHT_ASSET (*a, b, residual, life*)

Returns the depreciation charges for assets in row *a*, calculated using the straight line method given the *life* of each asset and the *residual*, outputting to row *b*.

STRAIGHT_DEP (*a, b, residual, life*)

Returns the depreciation charges for assets in row *a*, using the straight line method given the *life* of each asset and the *residual*, outputting to row *b*.

Statistical Tools

General Analysis

ANALYSE Block

Generates statistical information about a tuple. Returns the following statistics:

Number of uninitialised cells	Number of values processed
Maximum value	Minimum value
Range	Mean value
Population Deviation + Variance	Sample Deviation + Variance
Coefficient of variation	Tau
Skewness	Kurtosis
Median Value	Lower Quartile
Upper Quartile	Inter-Quartile Range

These statistical calculations can be applied to any Holos structures or data. These calculations can be added as optional calculation selections to Holos application screens or reports.

"Holos_Support:stat_util.hl"

Declares the following statistical functions:

`stt$_chidist_prob (degrees of freedom, chi-square value )`

Returns the Chi-Square Probability.

`stt$_tdist_prob (degrees of freedom, t value)`

Returns the two tailed p-value for the t-distribution.

`stt$_fdist_prob (degrees of freedom, p value)`

Returns the p-value for the F-distribution.

`stt$_normal_cum (x-value, mean, standard deviation)`

Returns the probability that a Normal Distribution is greater than the value specified.

`stt$_gammln (x)`

Returns the Natural Logarithm of the Gamma Function for the given integer *x*.

`stt$_binomial (n,k)`

Returns the Binomial Coefficient where 0 <= *k* <= *n*.

REGRESSION Block

Allows regression (best-fit) analysis to find and analyse the relationship between two sets of numbers. The following curve types can be fitted to the data:

Exponential	Geometric
Hyperbolic	Linear
Log Quadratic	Modified Exponential
Modified Hyperbolic	Polynomial
Quadratic	Rational
Semilog	

Also allows further best-fit analysis of a single curve-type, recording the information into Holos variables and record objects.

Can be accessed graphically from the Worksheet, Modelling menu.

REGRESSION MULTI Block

Uses Multilinear Regression to analyse relationships between a dependant variable and two or more independent variables.

The regression routines can be run using a recursive method that can be used to track evolving regression coefficients. Holos supports options for Least Square, Recursive and Choleski Decomposition.

Can be accessed graphically from the Worksheet, Modelling menu.

REGRESSION MULTI “Logistic” Block

Analyses the relationship between binary response models, such as where analogue values present a binary success/failure result.

Forecasting, Smoothing & Time-Series Analysis

SMOOTH Block

Used to smooth data traces and to forecast future values based on past trends.

Exponential Smoothing & Forecasting is for data with a continuous trend, such as linear or geometric.

Winters Smoothing & Forecasting is for data with a continuous trend, but also containing a recognised periodic (time-series) component.

FOURIER Block

Used to analyse data with a periodic component. Provides periodic component required by Winters smoothing & forecasting. Can also be used for forecasting directly.

Fourier Analysis can be applied to time series data with contributions of differing frequencies (day, week, month, etc.) The results indicate the amplitude of contribution at the fundamental frequency and successive harmonic frequencies. Once again, a graphical interface enables a user with no knowledge

of the Fourier block to rapidly calculate Spectral densities. These can be displayed in terms of a normalized frequency or directly in periods.

Can be accessed graphically from the Worksheet, Modelling menu.

BOX JENKINS Block

A more sophisticated method of time-series analysis. The Box Jenkins method of analysis predicts the future, based solely on past values. Handles all aspects of Identification, Estimation, Diagnosis and Forecasting of general ARIMA models

Difference Time Series at any required lags (seasonal and nonseasonal)
Generate Autocorrelation Sequence
Generate Partial Autocorrelation Sequence
Estimate (using Maximum Likelihood method) the parameters and their standard deviations in any desired model.
Generate Forecasts with lower and upper confidence intervals.
Generate various diagnostics such as Akaikes AIC and FPE.

The Box Jenkins analysis is also available from the Worksheet under the Modelling menu. The graphical interface to the Box Jenkins block enables the identification, estimation and diagnosis to be rapidly effected. For example, the auto correlation of various differencings of the original series and the auto correlations of prediction errors can be calculated and graphed immediately. There is also a direct interface to the Fourier block so that spectral densities can be viewed directly.

An extension of the Box Jenkins block to handle Box Tisc (ARIMAX) models, A STATE SPACE procedure based on identifying a state space model (in Kalman Filter innovations form) to handle general multi output multi input time series.

Modelling/Optimisation

"What-if?" Analysis

RISK Block

Used for Risk Analysis (Monte Carlo simulation).

Risk Analysis or Monte Carlo simulation is an important scenario analysis tool where the decision-maker is interested in the range of possible outcomes and the likelihood of those outcomes. It is essentially an extension of "What-if?"' analysis to assessing the probability of possible outcomes indicated by a Holos model.

Understanding Risk Analysis inevitably will involve the use of statistical and probability terminology. It is strongly recommended that the user understand the rudiments of distributions and probabilities, otherwise it is difficult to make an educated interpretation of the results - since the outputs will depend on assumptions made about the inputs.

"What-if?" and "What's-best?" Analysis

TARGET Block

The Target analysis can be used as a powerful optimization tool to answer "What if?" and "What's best" questions concerning Holos models.

Target analysis enables a user to invoke well established optimization methods to answer business decisions. The business problem will be typically formulated within a Holos model, where the objective depends on a number of decision variables. The Target analysis "solves" the problem by an iterative procedure varying inputs values and observing the corresponding results and

Target Analysis will handle:

Linear Programming problems
Mixed Integer Linear programming problems
Quadratic Programming problems
General Nonlinear Unconstrained problems.
General Nonlinear, nonlinearly constrained problems, (solution is based on sequential quadratic programming algorithm).

IMPACT Block

Used to assess the changes to a cell value when changes are made to one or more other cells.

Values of particular interest within the model, such as net profit, are likely to depend on a number of other variables, like raw materials costs, fuel charges and many other factors. It is often useful to know which of these variables has most influence on the value of interest. If it turns out that a particular variable has a

very significant impact upon profit for instance, then it is probably worthwhile analysing it further. If the variable has an element of uncertainty, then risk analysis may well be useful, and target analysis could be used, targeting profit and using the variable as an altered value.

Impact analysis allows you to specify a value to watch, and lists of values to change. When the analysis is complete, a ranking list is produced, in an order you select, showing the effect of altering the changed values. The amount of alteration can be specified, and can include ranges of values.

Data Mining

Simple Data Mining using Pattern Matching

"Holos_Support:dmin.hl"

Provides data mining capability using pattern-matching of data values. Can search for the following patterns:

Outliers	Slopes
Spikes	Steps
Trends	Periodicity
Noise	

For each pattern search you have the ability to select the data set to search, the type of pattern, and specific parameters for each type of pattern search. Pattern Searches could be combined with Holos Agents to intelligently look for patterns in data and automatically detect and notify end users. Data Pattern Matching is provided in the Holos Analysis Tools desktop and is driven with a simple dialog based point and click interface. Pattern matching functions such as Outliers and Trends can be added to any Holos application.

Advanced Data Mining

CLUSTER Block

Used for identifying clusters of related data within a structure and dimension reduction using Principal Component Analysis.

The simplest approach to discovering distinct groups or clusters is by the examination of scattergrams. These may be obtained by plotting the first two or three principal components or from the results of multidimensional scaling. Other exploratory methods useful in the search for clustering are the use of Andrews plots and Chernoff 'faces'. Once evidence of clustering has been found it will often be useful to provide some sort of explicit classification using one or more cluster analysis algorithms. The type of cluster analysis that is implemented is hierarchical agglomerative clustering.

CHAID Block

Used for tree-based modelling, similar to building up a large IF statement surrounding various business factors.

CHAID stands for Chi-squared Automatic Interaction Detector. It is a tree-based modelling technique. Tree-based modelling is an exploratory technique for uncovering structure (or relationship) in data. Specifically, the technique is useful for classification and regression problems where one has a set of classification or predictor variables (often called independent variables in statistics), and a single response variable (also called dependent variable by statisticians). CHAID can help solve a variety of business problems, in which all

the variables are discrete and you are interested in finding out the relationship between one dependent variable, such as profit of the company, and other independent variables, such as weather conditions (good or bad), type of products, etc.. CHAID can produce explicit decision rules.

Neural Networks

Neural Network Analysis

NEURAL Block

A specialised data-mining tool used to analyse complex non-linear relationships between dependent and independent variables.

Creates a Neural Network that attempts to provide a “black-box” mathematical model of the relationship between dependent and independent variables. Previous understanding of the relationship is not required to create a successful Neural Network.

Neural Nets are a further step towards data mining. Neural nets can be trained by experience to perform a certain task. Currently Holos builds Radial Basis Function Neural Networks. This is one of several neural net approaches. Others are likely to be made available in the future.

To make Neural Networks accessible to non-specialist, Holos provides a wizard that guides users through a set of dialogs to create a Neural Network. The Neural Net wizard is available from the Holos Analysis Tools desktop. It can be added to any Holos application. The Neural Network is also available from the Holos language.

NEURAL INFORMATION Block

A subset of the NEURAL Block, used to sort independent variables in order of magnitude of their effect on the value of a dependent variable.