STATGRAPHICS Expressions

STATGRAPHICS expressions may be used to assign new values to columns that have been assigned the *Formula* type. For example, assume that a datasheet contains columns named *weight* and *horsepower*. Then new columns could be defined with data values determined by formulas such as

LOG(weight)

or

horsepower/weight

STATGRAPHICS expressions may consist of:

1. **Names of columns** in the datasheet.
2. **Numeric constants**.
3. **Character strings**
4. **Algebraic operators** such as + and -.
5. **Boolean operators** such as = and <.
6. **STATGRAPHICS functions** such as LOG and SQRT.
7. **Parentheses** to override the normal order of evaluation.

Examples of each are shown below, assuming that the datasheet contains the following:

weight	horsepower	make
2705	140	Acura
3560	200	Acura
3375	172	Audi
3405	172	Audi
3640	208	BMW

Column Names

The most common STATGRAPHICS expression is a simple column name, such as:

weight

The result of this expression is simply a vector containing the data values:

2705 3560 3375 3405 3640

Expressions are **not** case-sensitive with respect to column names.

Numeric Constants

Expressions can also consist of numeric values, such as

3.14159265

or

1.4e6

The latter example represents the number 1,400,000 since e6 indicates that 1.4 is multiplied by 10 to the 6^{th} power.

Character Strings

A character string in an expression is surrounded by double quotes, as in

"Chevrolet"

Expressions **are** case sensitive with respect to the **contents** of character strings.

Algebraic Operators

Objects within expressions may be combined algebraically using standard symbols, as in

100+10*horsepower/weight

The available operators are shown in the table below.

Symbol	Definition
+	addition
-	subtraction
*	multiplication
/	division
^	exponentiation

When evaluating expressions containing more than one algebraic operator, normal precedence rules apply: exponentiation is done first, then multiplication and division, then addition and subtraction.

Boolean Operators

Boolean operators result in a vector of 0's (for FALSE) and 1's (for TRUE). For example, the expression

weight < 3500

results in the following vector:

1 0 1 1 0

The available Boolean operators are:

Symbol	Definition
=	equal
<>	not equal
<	less than
<=	less than or equal
>	greater than
>=	greater than or equal
&	and
|	or

The operator & performs a logical "and", resulting in TRUE only if the conditions on both sides are TRUE. For example, the expression

weight < 3500 & horsepower < 150

results in

1 0 0 0 0

The operator | (Shift-'\' on most keyboards) performs a logical "or", resulting in TRUE if either of the conditions on both sides are TRUE. For example, the expression

weight < 3500 | horsepower > 200

results in

1 0 1 1 1

Note: use double-quotes when comparing variable names to character strings, as in

make = "Audi"

STATGRAPHICS Functions

STATGRAPHICS contains a large number of special functions for various types of operations. Functions take one or more argument and have the basic form

FunctionName(argument1, argument2, agrument3, …)

Note: function names are **not** case sensitive.

Functions fall into several basic classes:

- **Mathematical transformations** – for mathematical calculations performed on a column of data or a numeric constant, such as LOG(weight) to calculate the natural logarithms of weight.
- **Sequential operators** – operators which perform sequential operations on a column of data, such as DIFF(x) to calculate the first differences of a time series.
- **Random number generators** – for generating sets of random numbers, such as RNORMAL(100,10,3), which generates 100 random numbers from a normal distribution with mean = 10 and standard deviation = 3.
- **Statistical summaries** – for calculating statistics from a sample of data, such as AVG(horsepower), which calculates the average value of *horsepower*.
- **Distribution functions** – for determining probabilities or critical values for common probability distributions, such as NORMAL(12.5,10,3), which finds the probability that a random variable from a normal distribution with mean = 10 and standard deviation = 3 will be less than 12.5.
- **Boolean selector** – functions designed for the *Select* field that return 0's and 1's.
- **Data selectors** – functions designed for input directly in data fields that return a subset of the data, such as TAKE(3,weight), which returns the first 3 values of *weight*.
- **Pattern generators** – for generating data with simple patterns, such as COUNT(1,100,1), which generates the integers from 1 to 100.
- **Utility functions** – for manipulating data in useful ways, such as REPLACE(make,"Acura","A") which will replace all occurrences of the string "Acura" in the column named *make* with an "A".

Examples of each function are given below. In the descriptions, a function shown with a list of values separated by spaces indicates a single argument consisting of the contents of a column. For example, ABS(1 –2 3) indicates the application of the absolute value function to a column of containing the values 1, 2, and 3. A function with multiple arguments shows the arguments separated by commas, as in NORMAL(100,10,3).

Notes:
Missing values in the examples below are shown as –32768, the internal placeholder STATGRAPHICS uses for numeric missing values (they appear as empty cells in the datasheet).

ABS(x)
Purpose: absolute value
Type: mathematical transformation
Argument: numeric values
Example: ABS(1 –2 3)
Result: 1 2 3

ACOS(x)
Purpose: inverse cosine in degrees
Type: mathematical transformation
Argument: cosine of angles
Example: ACOS(0 1 -1)
Result: 90 0 180

ACOSR(x)
Purpose: inverse cosine in radians
Type: mathematical transformation
Argument: cosine of angles
Example: ACOSR(0 1 -1)
Result: 1.57080 0 3.14159

ASIN(x)
Purpose: inverse sine in degrees
Type: mathematical transformation
Argument: sine of angles
Example: ASIN(0 1 -1)
Result: 0 90 -90

ASINR(x)
Purpose: inverse sine in radians
Type: mathematical transformation
Argument: sine of angles
Example: ASINR(0 1 -1)
Result: 0 1.57080 –1.57080

ATAN(x)
Purpose: inverse tangent in degrees
Type: mathematical transformation
Argument: tangent of angles
Example: ATAN(0 1 -1)
Result: 0 45 -45

ATANR(x)
Purpose: inverse tangent in radians
Type: mathematical transformation
Argument: tangent of angles
Example: ATANR(0 1 -1)
Result: 0 0.785398 –0.785398

AVG(x)
Purpose: average
Type: statistical summary
Argument: numeric data column
Example: AVG(1 2 4)
Result: 2.33333

BETA(x,alpha1,alpha2)
Purpose: cumulative beta distribution
Type: distribution function
Arguments: value of random variable, shape parameter 1, shape parameter 2
Example: BETA(0.5,1,2)
Result: 0.75

CELL(x,row)
Purpose: value of specified cell in data column
Type: data selector
Argument: data column, row number
Example: CELL(x,4)
Result: value in 4^{th} row of column x

CHISQUARE(x,df)
Purpose: cumulative chi-squared distribution
Type: distribution function
Arguments: value of random variable, degrees of freedom
Example: CHISQUARE(5,3)
Result: 0.828206

COMPRESS(x,c)
Purpose: removal of missing values from a data column and shortening of the result
Type: data selector
Arguments: data column, selection criterion
Example: COMPRESS(x,x<3) for x={1,2,3}
Result: 1 2

COS(x)
Purpose: cosine of angles measured in degrees
Type: mathematical transformation
Argument: angles in degrees
Example: COS(0 90 180)
Result: 1 0 -1

COSR(x)
Purpose: cosine of angles measured in radians
Type: mathematical transformation
Argument: angles in radians
Example: COSR(0 1.57081 3.14159)
Result: 1 0 -1

COUNT(from,to,by)
Purpose: generation of equally spaced data values
Type: pattern generator
Argument: starting value, ending value, increment
Example: COUNT(1 5 2)
Result: 1 3 5

CV(x)

Purpose: coefficient of variation for numeric data as a percentage
Type: statistical summary
Argument: numeric column
Example: CV(1 2 4)
Result: 64.4654%

DATENUM(x)
Purpose: converts character dates to numeric sequence values, where January 1, 1950 is day 1 and January, 1950 is month 1.
Type: utility function
Argument: column of dates or months.
Example: DATENUM(1/1/50 1/2/50 1/1/05)
Result: 1 2 20090

DIFF(x)
Purpose: first differences of a column of sequential data
Type: sequential operator
Argument: number data columns
Example: DIFF(1 2 4 6 10)
Result: -32768 1 2 2 4

DROP(x,k)
Purpose: drop from front and shorten column
Type: data selector
Argument: numeric data column, number of rows to drop
Example: DROP(1 2 4 6 10,2)
Result: 4 6 10

DROPLAST(x,k)
Purpose: drop from end and shorten column
Type: data selector
Argument: numeric data column, number of rows to drop
Example: DROPLAST(1 2 4 6 10,2)
Result: 1 2 4

EXCLUDE(row)
Purpose: Boolean exclusion of a single row
Type: Boolean selector
Argument: row number
Example: EXCLUDE(3) for data column with 5 values
Result: 1 1 0 1 1

EXP(x)
Purpose: exponentiation (e to the power x)
Type: mathematical transformation
Argument: numeric data column
Example: EXP(1 2 4)
Result: 2.71828 7.38906 54.5982

EXP10(x)

Purpose: exponentiation (10 to the power x)
Type: mathematical transformation
Argument: numeric data column
Example: EXP10(1 2 4)
Result: 10 100 10000

FACT(x)
Purpose: factorial
Type: mathematical transformation
Argument: numeric data column
Example: FACT(1 2 4)
Result: 1 2 24

FIRST(x)
Purpose: Boolean selection of first k rows
Type: Boolean selector
Argument: number of rows
Example: FIRST(2) for column of 5 values
Result: 1 1 0 0 0

FIRSTROWS(x)
Purpose: take from front and set others to missing
Type: data selector
Argument: numeric data column
Example: FIRSTROWS(1 2 4,2)
Result: 1 2 -32768

FISHERZ(x)
Purpose: Fisher's Z transformation for correlation coefficients
Type: mathematical transformation
Argument: numeric data column (all values between –1 and 1)
Example: FISHERZ(-.5 0 .5)
Result: -0.549306 0 0.549306

GEOMEAN(x)
Purpose: geometric mean
Type: statistical summary
Argument: numeric data column
Example: GEOMEAN(1 2 4)
Result: 2

INVBETA(p,alpha1,alpha2)
Purpose: inverse cumulative beta distribution
Type: distribution function
Argument: lower tail area, shape parameter 1, shape parameter 2
Example: INVBETA(0.75,1,2)
Result: 0.5

INVCHISQUARE(p,df)
Purpose: inverse cumulative chi-squared distribution

Type: distribution function
Arguments: lower tail area, degrees of freedom
Example: INVCHISQUARE(0.8,3)
Result: 4.62224

INVNORMAL(p,mu,sigma)
Purpose: inverse cumulative normal distribution
Type: distribution function
Arguments: lower tail area, mean, standard deviation
Example: INVCHISQUARE(0.8,3)
Result: 4.62224

INVSNEDECOR(p,df1,df2)
Purpose: inverse cumulative F distribution
Type: distribution function
Arguments: lower tail area, numerator degrees of freedom, denominator degrees of freedom
Example: INVSNEDECOR(0.8,3,10)
Result: 1.86146

INVSTUDENT(p,df)
Purpose: inverse cumulative t distribution
Type: distribution function
Arguments: lower tail area, degrees of freedom
Example: INVSTUDENT(0.8,3)
Result: 0.978476

IQR(x)
Purpose: interquartile range
Type: statistical summary
Argument: numeric data column
Example: IQR(1 2 4)
Result: 3

JOIN(x1,x2)
Purpose: joins two numeric or character columns end to end
Type: utility function
Argument: data column, data column
Example: JOIN(1 2 3,4 5 6)
Result: 1 2 3 4 5 6

JOIN3(x1,x2,x3)
Purpose: joins three numeric or character columns end to end
Type: utility function
Argument: data column, data column, data column
Example: JOIN3(1 2 3,4 5 6,7 8 9)
Result: 1 2 3 4 5 6 7 8 9

JOIN4(x1,x2,x3,x4)
Purpose: joins four numeric or character columns end to end
Type: utility function

Argument: data column, data column, data column, data column
Example: JOIN4(1 2 3,4 5 6,7 8 9,10 11 12)
Result: 1 2 3 4 5 6 7 8 9 10 11 12

JUXTAPOSE(x1,x2)
Purpose: joins two character columns side by side
Type: utility function
Argument: data column, data column,
Example: JUXTAPOSE("John" "Bob","Smith" "Jones")
Result: "John Smith", "Bob Jones"
Note: to remove excess blanks within the resulting strings, use STRIPBLANKS together with JUXTAPOSE.
Example: STRIPBLANKS(JUXTAPOSE("John" "Bob","Smith" "Jones"))

KURTOSIS(x)
Purpose: kurtosis
Type: statistical summary
Argument: numeric data column of length 4 or greater
Example: KURTOSIS(1 2 4 6 10)
Result: 0.147705

LAG(x)
Purpose: lags the data by the indicated amount. Useful for plotting time series data versus lagged values.
Type: sequential operator
Argument: numeric data column
Example: LAG(1 2 4 6 10,2)
Result: -32768 –32768 1 2 4

LAST(k)
Purpose: Boolean selection of last k rows
Type: Boolean selector
Argument: number of rows
Example: LAST(2) for column of 5 values
Result: 0 0 0 1 1

LASTROWS(x)
Purpose: take from end and set others to missing
Type: data selector
Argument: numeric data column
Example: LASTROWS(1 2 4,2)
Result: -32768 2 4

LOG(x)
Purpose: natural logarithm
Type: mathematical transformation
Argument: numeric data column
Example: LOG(1 2 4)
Result: 0 0.693147 1.38629

LOG10(x)
Purpose: log base 10
Type: mathematical transformation
Argument: numeric data column
Example: LOG10(1 10 100)
Result: 1 2 3

MAX(x)
Purpose: maximum value
Type: statistical summary
Argument: numeric data column
Example: MAX(1 2 4)
Result: 4

MDIFF(x,k)
Purpose: multiple differences of a column of sequential data
Type: sequential operator
Argument: numeric data column, order of differencing
Example: MDIFF(1 2 4 6 10,2)
Result: -32768 –32768 1 0 2 (second differences)

MEDIAN(x)
Purpose: median value
Type: statistical summary
Argument: numeric data column
Example: MEDIAN(1 2 4)
Result: 2

MIN(x)
Purpose: minimum value
Type: statistical summary
Argument: numeric data column
Example: MIN(1 2 4)
Result: 1

MODE(x)
Purpose: mode (most frequent value) or missing value is not unique
Type: statistical summary
Argument: numeric data column
Example: MODE(1 2 4 4)
Result: 4

NORMAL(x,mu,sigma)
Purpose: cumulative normal distribution
Type: distribution function
Arguments: value of random variable, mean, standard deviation
Example: NORMAL(2,0,1)
Result: 0.977250

PERCENTILE(x,p)

Purpose: sample percentile
Type: statistical summary
Argument: numeric data column, percentile
Example: PERCENTILE(1 2 4 6 10,90)
Result: 10

Q25(x)
Purpose: lower quartile
Type: statistical summary
Argument: numeric data column
Example: Q25(1 2 4)
Result: 1

Q75(x)
Purpose: upper quartile
Type: statistical summary
Argument: numeric data column
Example: Q75(1 2 4)
Result: 4

RANDOM(k)
Purpose: Boolean selection of random k rows
Type: Boolean selector
Argument: number of rows
Example: RANDOM(2) for column of 5 values
Result: 0 1 0 0 1

RANGE(x)
Purpose: range (maximum – minimum)
Type: statistical summary
Argument: numeric data column
Example: RANGE(1 2 4)
Result: 3

RANK(x)
Purpose: sample ranks
Type: mathematical transformation
Argument: numeric data column
Example: RANK(8 4 9 4 2)
Result: 4 2.5 5 2.5 1

RECODE(x)
Purpose: sorts data and returns indices
Type: utility function
Argument: numeric or character data column
Example: RECODE(NJ VA MA VA NJ)
Result: 2 3 1 3 2

REP(x,k)
Purpose: repeats each data value k times in groups

Type: pattern generator
Argument: numeric data column, replication factor
Example: REP(1 2 4,3)
Result: 1 1 1 2 2 2 4 4 4

REPLACE(x,oldvalue,newvalue)
Purpose: replaces all occurrences of the old value with the new value.
Type: utility function
Argument: numeric or character data column.
Example: REPLACE(1 2 3 4 1 2 3 4,3,7)
Result: 1 2 7 4 1 2 7 4

RESHAPE(x,length)
Purpose: reshapes a data column into the indicated length. If length if longer than original, repeats values in a circular fashion.
Type: pattern generator
Argument: numeric data column, desired column length
Example: RESHAPE(1 2 4,9)
Result: 1 2 4 1 2 4 1 2 4

REXPONENTIAL(n,mean)
Purpose: generates random numbers from an exponential distribution
Type: random number generator
Argument: sample size, distribution mean
Example: REXPONENTIAL(3,10)
Result: 1.67014 10.0075 20.4921

RGAMMA(n,shape,scale)
Purpose: generates random numbers from a gamma distribution
Type: random number generator
Argument: sample size, shape parameter, scale parameter
Example: RGAMMA(3,2,0.1)
Result: 49.4944 12.8622 4.70051

RINTEGER(n,min,max)
Purpose: generates random numbers from a discrete uniform distribution
Type: random number generator
Argument: sample size, lower limit, upper limit
Example: RINTEGER(3,1,0.1)
Result: 10 4 2

RLOGNORMAL(n,mu,sigma)
Purpose: generates random numbers from a lognormal distribution
Type: random number generator
Argument: sample size, mean, standard deviation
Example: RLOGNORMAL(3,10,3)
Result: 19.6869 5.29878 6.55710

RNORMAL(n,mu,sigma)
Purpose: generates random numbers from a normal distribution

Type: random number generator
Argument: sample size, mean, standard deviation
Example: RNORMAL(3,10,3)
Result: 13.4892 9.85616 11.9911

ROUND(x)
Purpose: round to integer
Type: mathematical transformation
Argument: numeric data column
Example: ROUND(1.11 2.22 5.55)
Result: 1 2 6

ROUNDTO(x,decimals)
Purpose: round to specified number of decimal places
Type: mathematical transformation
Argument: numeric data column, number of decimal places
Example: ROUNDTO(1.11 2.22 5.55,1)
Result: 1.1 2.2 5.6

ROWS(start,end)
Purpose: Boolean selection of range of rows
Type: Boolean selector
Arguments: start row, end row
Example: ROW(2,4) for column of 5 values
Result: 0 1 1 1 0

RUNIFORM(n,min,max)
Purpose: generates random numbers from a continuous uniform distribution
Type: random number generator
Argument: sample size, lower limit, upper limit
Example: RUNIFORM(3,0.0,1.0)
Result: 0.884547 0.745398 0.168033

RUNTOT(x)
Purpose: running total
Type: sequential operator
Argument: numeric data column
Example: RUNTOT(1 2 4)
Result: 1 3 7

RWEIBULL(n,shape,scale)
Purpose: generates random numbers from a Weibull distribution
Type: random number generator
Argument: sample size, shape parameter, scale parameter
Example: RWEIBULL (3,5,10)
Result: 11.2118 9.21719 7.87040

SD(x)
Purpose: standard deviation
Type: statistical summary

Argument: numeric data column
Example: SD(1 2 4)
Result: 1.52753

SDIFF(x,k)
Purpose: seasonal differences of a column of sequential data
Type: sequential operator
Argument: numeric data column, lag
Example: SDIFF(1 2 4 6 10,3)
Result: -32768 –32768 –32768 5 8 (differences between data 3 rows apart)

SELECT(x,condition)
Purpose: Boolean selection of rows meeting a condition
Type: data selector
Argument: data column, Boolean condition
Example: SELECT(x,x>2) for data column x={1,2,4,6,10}
Result: -32768 -32768 4 6 10

SERROR(x)
Purpose: standard error of the mean
Type: statistical summary
Argument: numeric data column
Example: ASERROR(1 2 4)
Result: 0.881917

SIN(X)
Purpose: sine of angles measured in degrees
Type: mathematical transformation
Argument: angles in degrees
Example: SIN(0 90 180)
Result: 0 1 0

SINR(x)
Purpose: sine of angles measured in radians
Type: mathematical transformation
Argument: angles in radians
Example: SINR(0 1.57081 3.14159)
Result: 0 1 0

SIZE(x)
Purpose: number of nonmissing values
Type: statistical summary
Argument: numeric data column
Example: SIZE(1 2 4)
Result: 3

SKEWNESS(x)
Purpose: skewness
Type: statistical summary
Argument: numeric data column of length 3 or greater

Example: SKEWNESS(1 2 4)
Result: 0.935220

SKURT(x)
Purpose: standardized kurtosis
Type: statistical summary
Argument: numeric data column of length 4 or greater
Example: SKURT(1 2 4 6 10)
Result: 0.0674178

SNEDECOR(x,df1,df2)
Purpose: cumulative F distribution
Type: distribution function
Arguments: value of random variable, numerator degrees of freedom, denominator degrees of freedom
Example: SNEDECOR(2,3,10)
Result: 0.821993

SQRT(x)
Purpose: square root
Type: mathematical transformation
Argument: numeric data column or single number
Example: SQRT(2)
Result: 1.41421

SSKEW(x)
Purpose: standardize skewness
Type: statistical summary
Argument: numeric data column of length 3 or greater
Example: SSKEW(1 2 4)
Result: 0.661300

STANDARDIZE(x)
Purpose: standardized values; (x - mean)/standard deviation
Type: mathematical transformation
Argument: numeric data column
Example: STANDARDIZE(1 2 4)
Result: -0.872872 -0.218218 1.09109

STRIPBLANKS(x)
Purpose: remove consecutive blanks from contents of cells
Type: utility function
Argument: character data column
Example: STRIPBLANKS("John Smith" "Bob Jones")
Result: "John Smith" "Bob Jones"

STUDENT(x,df)
Purpose: cumulative t distribution
Type: distribution function
Arguments: value of random variable, degrees of freedom

Example: STUDENT(2,3)
Result: 0.930337

SUM(x)
Purpose: sum
Type: statistical summary
Argument: numeric data column
Example: SUM(1 2 4)
Result: 7

TAKE(x,k)
Purpose: take from front and shorten column
Type: data selector
Argument: numeric data column, number of rows to take
Example: TAKE(1 2 4 6 10,2)
Result: 1 2

TAKELAST(x,k)
Purpose: take from end and shorten column
Type: data selector
Argument: numeric data column, number of rows to take
Example: TAKELAST(1 2 4 6 10,2)
Result: 6 10

TAN(x)
Purpose: tangent of angles measured in degrees
Type: mathematical transformation
Argument: angles in degrees
Example: TAN(0 30 60)
Result: 0 0.577350 1.73205

TANR(x)
Purpose: tangent of angles measured in radians
Type: mathematical transformation
Argument: angles in radians
Example: TANR(0 0.523599 1.04720)
Result: 0 0.577350 1.73205

TRUNCATE(x)
Purpose: truncation to integer
Type: mathematical transformation
Argument: numeric data column
Example: TRUNCATE(1.11 2.22 5.55)
Result: 1 2 5

VARIANCE(x)
Purpose: variance
Type: statistical summary
Argument: numeric data column
Example: VARIANCE(1 2 4)

Result: 2.33333