Decision makers often need to consider the relationship between variables when analyzing a problem. Sensitivity analysis provides a tool for such an analysis.

@RISK uses two methods to calculate the sensitivity between each output and its input variables: multivariate stepwise regression and rank order correlation. This article provides more information on these techniques.

# Multivariate Stepwise Regression

## What is Linear Regression?

Regression is simply another term for fitting data to a theoretical equation. In the case of linear regression, the input data is fit to a line. You may have heard of the "Least-Squares" method, which is a type of linear regression.

Multiple regression tries to fit multiple input data sets to a planar equation that could produce the output data set. It uses an equation of the form:

 

where *y* is the output variable and *xi* are the input variables. The sensitivity values returned by @RISK are normalized variations of *ai*.

## What is Multivariate Stepwise Regression?

Stepwise regression is a technique for calculating regression values with multiple input values. Other techniques exist for calculating multiple regressions, but the stepwise regression technique is preferable for large numbers of inputs since it removes all variables that provide an insignificant contribution from the model.

The following steps are used in the stepwise regression method:

1. *The input with the highest correlation to the output is found and entered into regression.*
2. *The partial correlation coefficient of every input that is not in regression with the response is calculated, and the variable with the largest correlation value is entered into regression.*
3. *A regression equation is created for every input that has been entered into regression. Each input is tested for "significance" using a measurement called the F-test, and insignificant variables are removed from regression.*
4. *Steps 2 and 3 are repeated until the only remaining variables have been rejected from the regression.*
5. *A final regression equation is calculated from the inputs that have not been rejected from the regression. These normalized values are reported in @RISK.*
6. *The percentage of variation in the output that is explained by the linear regression equation is calculated and reported in @RISK as R2.*

## Interpreting the Results

The coefficients listed in the @RISK sensitivity report are normalized regression coefficients (*pi*) associated with each input. An *pi* value of 0 indicates that there is no significant relationship between the input and the output.

If you are interested in converting the normalized regression coefficient to an absolute regression coefficient, use the following equation:

 

where

 , *Yj* are the points in the output data set

 , *Xij* are the points in the input data set being analyzed.

The *ai* value represents the average change in the output corresponding to a 1 unit change in the input. The unit could be dollars, feet, pounds, etc. depending on the default units of your model. A positive *ai* indicates an increase in the output value when the input increases, while a negative value indicated a decrease.

The *R2* value listed at the top of the column is simply a measurement of the percentage of variation that is explained by the linear relationship. If this number is less than ~ 60% then the linear regression does not sufficiently explain the relationship between the inputs and outputs and another method of analysis should be used.

Even if your sensitivity analysis produces a relationship with a large value of *R2*, examine the results to verify that they are reasonable. Do any of the coefficients have an unexpected magnitude or sign? Please refer to the section titled *Comparison of Methods* for more information on this topic.

# Rank Order Correlation

## What is Correlation?

Correlation is a quantitative measurement of the strength of a relationship between two variables. The most common type of correlation is linear correlation, which measures the linear relationship between two variables.

The rank order correlation value returned by @RISK can vary between -1 and 1. A value of 0 indicates there is no correlation between variables, they are independent. A value of 1 indicates a complete positive correlation between the two variables, when the input value samples "high," the output value will sample "high." A value of -1 indicates a complete inverse correlation between the two variables, when the input value samples "high," the output value will sample "low." Other correlation values indicate a partial correlation, the output is affected by changes in the selected input, but may be affected by other variables as well.

## What is Rank Order Correlation?

Rank order correlation calculates the relationship between two data sets by comparing the *rank* of each value in a data set. To calculate rank, the data is ordered from lowest to highest and assigned numbers (the ranks) that correspond to their position in the order.

This method is preferable to linear correlation when we do not necessarily know the probability distribution functions from which the data were drawn. For example, if data set A was normally distributed and data set B was lognormally distributed, rank order correlation would produce a better representation of the relationship between the two data sets.

Given the following information:

 *x* = the input data set

 *y* = the output data set

 *Ri* = the rank of *xi* among the other *xi*’s

 *Si* = the rank of *yi* among the other *yi*’s

The Spearman Rank-Order Correlation Coefficient is simply the linear correlation coefficient of the ranks:

 

## Interpreting the Results

@RISK reports the *rs* value in the sensitivity analysis. Obviously, a correlation coefficient of 1 or -1 can be considered significant, but what about those in-between values? The *Null Hypothesis Test* can be used to determine whether the correlation coefficient is significant given a reasonable margin of error. The test statistic *t* can be computed from the correlation coefficient *rs* and the number of data points *n* as follows:

 

The test statistic *t* indicates how many standard deviations *rs* is from 0. To determine if you should consider *rs* significant, consult a t-distribution table (see *Groebner and Shannon*), which reports "critical values" for given levels of confidence. If *t* is greater than the critical value, the null hypothesis is rejected and you can conclude that *the correlation is significant* for the given level of confidence.

# Comparison of Methods

So, which measurement of sensitivity should you use? If the *R2* value reported by the Stepwise Regression is low, you can conclude that the relationship between the input and output variables in not linear. In this case, you should use the Rank-Order Correlation analysis to determine the sensitivity in your model.

If the *R2* value reported by the Stepwise Regression is high, it is easy to conclude that the relationship is linear. But, as mentioned above, you should always verify that the regression variables are reasonable. For example, @RISK might report a significant positive relationship between two variables in the regression analysis and a significant negative correlation in the rank-order analysis. This effect is called *multicollinearity*.

Multicollinearity occurs when independent variables in a model are correlated to each other as well as to the output. Unfortunately, reducing the impact of multicollinearity is a complicated problem to deal with, but you may want to consider removing the variable that causes the multicollinearity from your sensitivity analysis.

# References

The following texts include detailed, mathematically rigorous information on calculation regression between variables:

Davies and Goldsmith. *Statistical Methods in Research and Production*. **Chapter 8**.

Draper, N. R., and H. Smith. 1966. *Applied Regression Analysis*. New York: John Wiley & Sons, Inc. **Chapter 6**.

Efroymson, M. A. 1960. “Multiple Regression Analysis.” In *Mathematical Methods for Digital Computers*, eds. Ralston and Wiff. New York: John Wiley & Sons. **pp. 191-203**.

More information on calculating the Spearman Rank Order Correlation Coefficient may be found in the following text:

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1988. *Numerical Recipes in C: The Art of Scientific Computing*. New York: Cambridge University Press.

Any comprehensive, introductory text on statistics should contain information on correlations, including the following book, which can be purchased from Palisade Corporation:

Groebner, David F., and Patrick W. Shannon. 1993. *Business Statistics, A Decision-Making Approach*. Fourth Edition. New York: Macmillan Publishing Company.