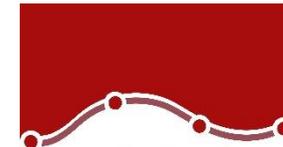




**SPSS**



**Statistics for  
Data Analysis**

## WHO WE ARE

SPS is an Italian center of statistical data analysis with more than 20 years of experience.

SPS was born in 1994 as SPSS Italia and it was the only reseller in Italy for SPSS software suite, authorised by SPSS inc.

Today SPS is an IBM Gold Business Partner, Software Support Provider and Expert Level in Data Science & Business Analytics.

## CONTACTS

Registered office:  
Via Antonio Zanolini, 36 A/B  
40126 Bologna (BO)

Operational headquarters:  
Via Isonzo, 55/2  
40033 Casalecchio di Reno (BO)  
P.I. 04222630370  
Tel. 051-252573  
[www.spss.it](http://www.spss.it)

## DATASHEET

---

**Statistics  
Neural Networks**



# Statistics for Data Analysis

Organizations can solve a wide array of business and research problems with the solution Statistics for Data Analysis.

Compared to other statistical software, the solution is easier to use, has a lower total cost of ownership and more comprehensively addresses the entire analytical process, from planning to data collection to analysis, reporting and deployment.

Organizations of all types rely on Statistics for Data Analysis to help increase revenue, outmaneuver competitors, conduct research and make better decisions. With decades of built-in expertise and innovation, it's a leading choice for reliable statistical analysis.

Statistics Base is part of the solution Statistics for Data Analysis, which consists of:

- Software license
- Add-On
- SPS Service Program

This comprehensive, easy-to-use solution includes many different procedures and tests to help users solve complex business and research challenges.

## Highlights Statistics for Data Analysis

- Get support through every step of the analytical process.
- Carry out essential analyses from an intuitive graphical interface.
- Select from more than a dozen integrated products to make specialized analyses faster and easier.



# Statistics for Data Analysis

The solution analytical capabilities to meet the analysis requirements of any type of organization, from basic tools for solving common problems to advanced analytical techniques that enable all type of organization to address complex challenges.

Statistics for Data Analysis can help you:

- Analyze your data with new and advanced statistics, including a variety of new features within UNIANOVA methods
- Integrate better with third-party applications, including stronger integration with Microsoft Office
- Save time and effort with productivity enhancements:
  - More attractive and modern-looking charts in Chartbuilder
  - New groundbreaking features in Statistics Amos 25
  - Data and syntax editor enhancements
  - Accessibility improvements for the visually impaired
  - Updated merge user interface
  - Simplified toolbars

Statistics for Data Analysis can access quickly, manage and analyze any kind of dataset, including survey data, corporate databases or data downloaded from the web.

In addition, the software can process Unicode data. This eliminates variability in data due to language-specific encoding and enables your organization to view, analyze and share data written in multiple languages.

## Business Benefit Statistics for Data Analysis

- Support business decisions with data-based analytics for improved outcomes.
- Be more confident in your results by incorporating data from many different sources, including geospatial information, in your analysis and using proven, tested techniques to perform your analysis.
- Save time and effort with capabilities that enable experienced analysts to develop procedures or dialogs that others can use to speed through repetitive tasks.
- Give results greater impact by using visualization capabilities that clearly show others the significance of your findings.



# Statistics Neural Networks

## Datasheet

### *New tools for building predictive models*

Your organization needs to find patterns and connections in the complex and fast-changing environment you work in so that you can make better decisions at every turn. You may be using Statistics Base and one or more of its modules to help you do this. If so, you know the power and versatility you have at your fingertips. But there's even more you can do.

You can explore subtle or hidden patterns in your data, using Statistics Neural Networks. This module offers you the ability to discover more complex relationships in your data and generate better performing predictive models.

The procedures in Statistics Neural Networks complement the more traditional statistics in Statistics Base and its modules. Find new associations in your data with Neural Networks and then confirm their significance with traditional statistical techniques.

### Highlights:

- Build better-performing models
- No programming needed
- Choice of MLP or RBF algorithms



# Statistics Neural Networks

## Datasheet

### Why use a neural network?

A computational neural network is a set of non-linear data modeling tools consisting of input and output layers plus one or two hidden layers. The connections between neurons in each layer have associated weights, which are iteratively adjusted by the training algorithm to minimize error and provide accurate predictions.

You set the conditions under which the network “learns” and can finely control the training stopping rules and network architecture, or let the procedure automatically choose the architecture for you.

You can combine Statistics Neural Networks with other statistical procedures to gain clearer insight in a number of areas. In market research, for example, you can create customer profiles and discover customer preferences. In database marketing, you can segment your customer base and optimize marketing campaigns.

In financial analysis, you can use Statistics Neural Networks to analyze applicants’ creditworthiness and to detect possible fraud. In operational analysis, use this tool to manage cash flow and improve logistics planning. Scientific and healthcare applications include forecasting treatment costs, performing medical outcomes analysis, and predicting the length of a hospital stay.



# Statistics Neural Networks

## Datasheet

Control the process from start to finish

With Statistics Neural Networks, you select either the Multilayer Perceptron (MLP) or Radial Basis Function (RBF) procedure. Both of these are supervised learning techniques - that is, they map relationships implied by the data. Both use feedforward architectures, meaning that data moves in only one direction, from the input nodes through the hidden layer of nodes to the output nodes.

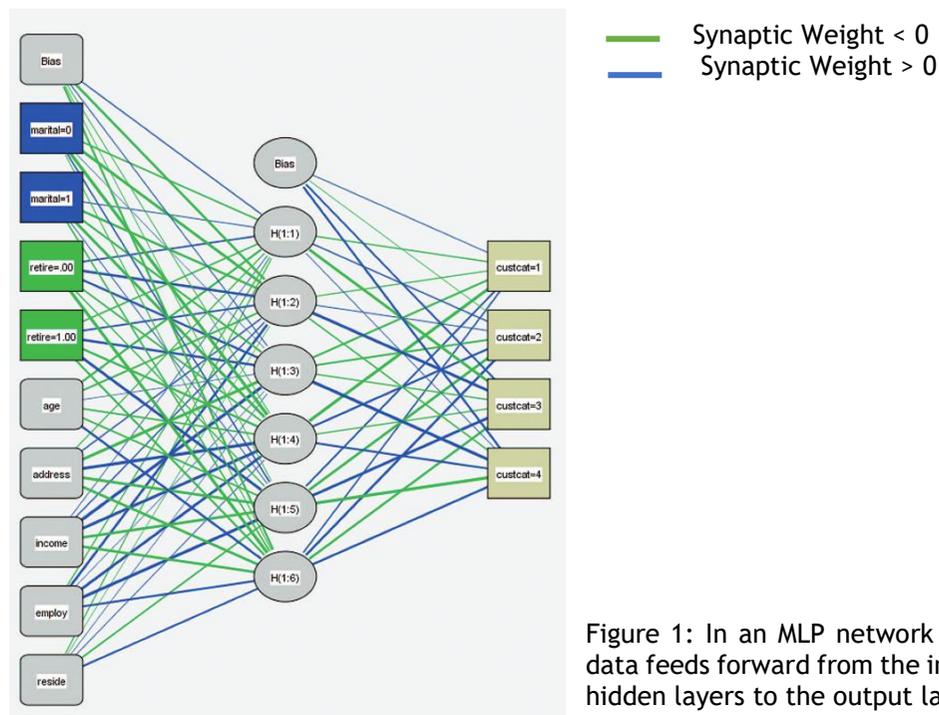


Figure 1: In an MLP network like the one shown here, the data feeds forward from the input layer through one or more hidden layers to the output layer



# Statistics Neural Networks

## Datasheet

Your choice of procedure will be influenced by the type of data you have and the level of complexity you seek to uncover. While the MLP procedure can find more complex relationships, the RBF procedure is generally faster.

With either of these approaches, you divide your data into training, testing, and holdout sets. The training set is used to estimate the network parameters. The testing set is used to prevent overtraining. The holdout set is used to independently assess the final network, which is applied to the entire dataset and to any new data.

You specify the dependent variables, which may be scale, categorical, or a combination of the two. If a dependent variable has scale measurement level, then the neural network predicts continuous values that approximate the “true” value of some continuous function of the input data. If a dependent variable is categorical, then the neural network is used to classify cases into the “best” category based on the input predictors.

You adjust the procedure by choosing how to partition the dataset, what sort of architecture you want, and what computation resources will be applied to the analysis. Finally, you choose to display results in tables or graphs, save optional temporary variables to the active dataset, and export models in XML-file formats to score future data.

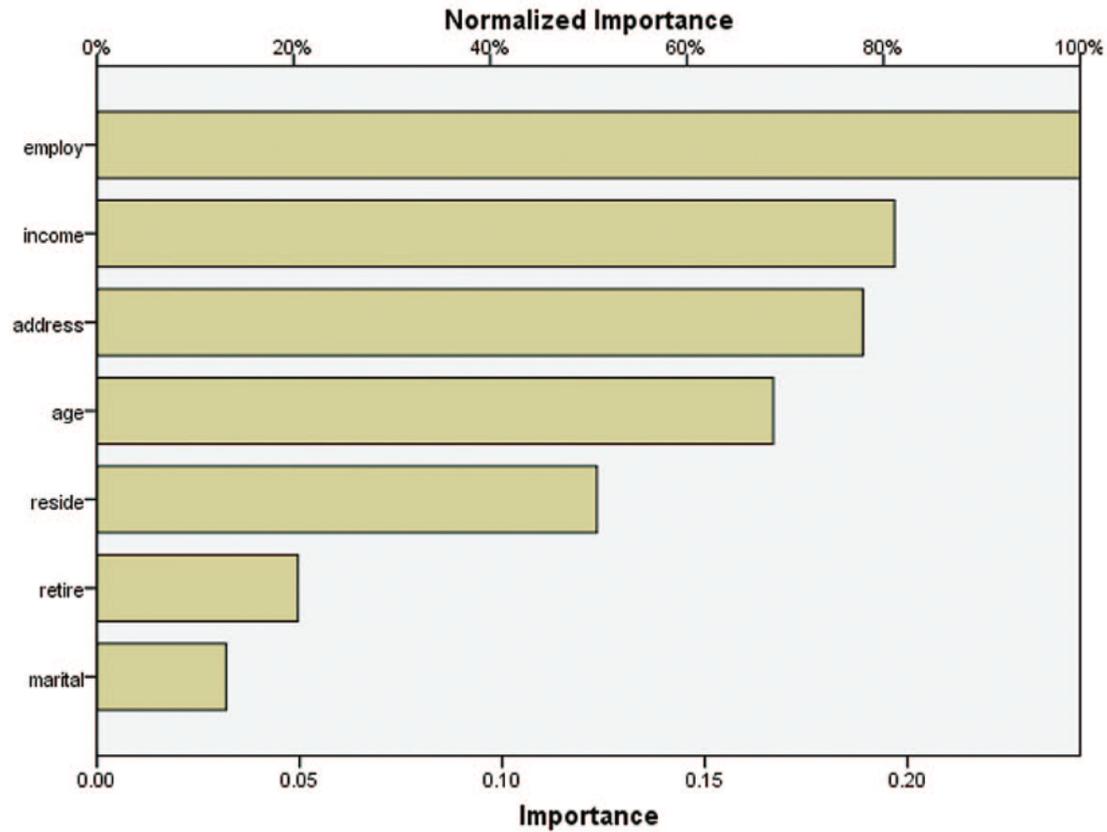


Figure 2: The results of exploring data with neural network techniques can be shown in a variety of graphic formats. This simple bar chart is one of many options.



# Statistics Neural Networks Features

## Multilayer Perceptron (MLP)

The MLP procedure fits a particular kind of neural network called a multilayer perceptron. The multilayer perceptron is a supervised method using feedforward architecture. It can have multiple hidden layers. One or more dependent variables may be specified, which may be scale, categorical, or a combination. If a dependent variable has scale measurement level, then the neural network predicts continuous values that approximate the “true” value of some continuous function of the input data. If a dependent variable is categorical, then the neural network is used to classify cases into the “best” category based on the input predictors.

- Predictors
  - Factors
  - Covariates
- The EXCEPT subcommand lists any variables that the MLP procedure should exclude from the factor or covariate lists on the command line. This subcommand is useful if the factor or covariate lists contain a large number of variables
- The RESCALE subcommand is used to rescale covariates or scale dependent variables
  - Dependent variable (if scale): standardized, normalized, adjusted normalized, or none
  - Covariates: standardized, normalized, adjusted normalized, or none
- The PARTITION subcommand specifies the method of partitioning the active dataset into training, testing, and holdout samples. The training sample comprises the data records used to train the neural network. The testing sample is an independent set of data records used to track prediction error during training in order to prevent overtraining. The holdout sample is another independent set of data records used to assess the final neural network. You can specify:
  - The relative number of cases in the active dataset to randomly assign to the training sample
  - The relative number of cases in the active dataset to randomly assign to the testing sample
  - The relative number of cases in the active dataset to randomly assign to the holdout sample
  - A variable that assigns each case in the active dataset to the training, testing, or holdout sample
- The ARCHITECTURE subcommand is used to specify the neural network architecture. You can specify:
  - Whether to use the automatic architecture or, if automatic is not used:
    - The number of hidden layers in the neural network
    - The activation function to use for all units in the hidden layers (Hyperbolic tangent or Sigmoid)
    - The activation function to use for all units in the output layer (Identity, Hyperbolic tangent, Sigmoid, or Softmax)



## Statistics Neural Networks Features

- The CRITERIA subcommand specifies the computational and resource settings for the MLP procedure. You can specify the training type, which determines how the neural network processes training data records: batch training, online training, mini-batch training. You can also specify:
  - The number of training records per mini-batch (if selected as the training method)
  - The maximum number of cases to store in memory when automatic architecture selection and/or mini-batch training is in effect
  - The optimization algorithm used to determine the synaptic weights: Gradient descent, Scaled conjugate gradient
  - The initial learning rate for the gradient descent optimization algorithm
  - The lower boundary for the learning rate when gradient descent is used with online or mini-batch training
  - The momentum rate for the gradient descent optimization algorithm
  - The initial lambda, for the scaled conjugate gradient optimization algorithm
  - The initial sigma, for the scaled conjugate gradient optimization algorithm
  - The interval [a0-a, a0+a] in which weight vectors are randomly generated when simulated annealing is used
- The STOPPINGRULES subcommand specifies the rules that determine when to stop training the neural network. You can specify:
  - The number of steps n to allow before checking for a decrease in prediction error
  - Whether the training timer is turned on or off and the maximum training time
  - The maximum number of epochs allowed
  - The relative change in training error criterion
  - The training error ratio criterion
- The MISSING subcommand is used to control whether user-missing values for categorical variables - that is, factors and categorical dependent variables - are treated as valid values.
- The PRINT subcommand indicates the tabular output to display and can be used to request a sensitivity analysis. You can choose to display:
  - The case processing summary table
  - Information about the neural network, including the dependent variables, number of input and output units, number of hidden layers and units, and activation functions
  - A summary of the neural network results, including the average overall error, the stopping rule used to stop training and the training time
  - A classification table for each categorical dependent variable
  - The synaptic weights; that is, the coefficient estimates, from layer i-1 unit j to layer i unit k
  - A sensitivity analysis, which computes the importance of each predictor in determining the neural network



## Statistics Neural Networks Features

- The PLOT subcommand indicates the chart output to display. You can display:
  - Network diagram
  - A predicted by observed value chart for each dependent variable
  - A residual by predicted value chart for each scale dependent variable
  - ROC (Receiver Operating Characteristic) curves for each categorical dependent variable. It also displays a table giving the area under each curve
  - Cumulative gains charts for each categorical dependent variable
  - Lift charts for each categorical dependent variable
- The SAVE subcommand writes optional temporary variables to the active dataset. You can save:
  - Predicted value or category
  - Predicted pseudo-probability
- The OUTFILE subcommand saves XML-format files containing the synaptic weights

### Radial Basis Function (RBF)

The RBF procedure fits a radial basis function neural network, which is a feedforward, supervised learning network with an input layer, a hidden layer called the radial basis function layer, and an output layer. The hidden layer transforms the input vectors into radial basis functions. Like the MLP procedure, the RBF procedure performs prediction and classification.

The RBF procedure trains the network in two stages:

1. The procedure determines the radial basis functions using clustering methods. The center and width of each radial basis function are determined.
2. The procedure estimates the synaptic weights given the radial basis functions. The sum-of-squares error function with identity activation function for the output layer is used for both prediction and classification. Ordinary Least Squares regression is used to minimize the sum-of-squares error.

Due to this two-stage training approach, the RBF network is in general trained much faster than MLP.

Subcommands listed for the MLP procedure perform similar functions for the RBF procedure, with the following exceptions:

- When using the ARCHITECTURE subcommand, users can specify the Gaussian radial basis function used in the hidden layer: either Normalized RBF or Ordinary RBF.
- When using the CRITERIA subcommand, users can specify the computation settings for the RBF procedures, specifying how much overlap occurs among the hidden units System requirements.



### **Statistics for Data Analysis solution**

Add more analytical power, as you need it, with optional modules and stand-alone software from the Statistics for Data Analysis family.

#### **Statistics Base**

Statistics Base includes the core capabilities to take the analytical process from start to finish. It is easy to use and includes a broad range of procedures and techniques to increase revenue, outperform competitors, conduct research and make better decisions.

#### **Statistics Advanced**

Statistics Advanced includes these powerful multivariate techniques: generalized linear models (GENLIN), generalized estimating equations (GEE), mixed level models, general linear mixed models (GLMM), variance component estimation, MANOVA, Kaplan-Meier estimation, Cox regression, hiloglinear, loglinear and survival analysis.

#### **Statistics Bootstrapping**

Statistics Bootstrapping enables researchers and analysts to use bootstrapping techniques on a number of tests contained in Statistics for Data Analysis modules. This provides an efficient way to ensure that your models are stable and reliable. With Statistics Bootstrapping, you can reliably estimate the standard errors and confidence intervals of a population parameter like a mean, median, proportion, odds ratio, correlation coefficient, regression coefficient and numerous.

#### **Statistics Categories**

Unleash the full potential of your categorical data through perceptual maps with optimal scaling and dimension reduction techniques. This add-on module provides you with everything you need to analyze and interpret multivariate data and their relationships more completely.

#### **Statistics Complex Samples**

Incorporate complex sample designs into data analysis for more accurate analysis of complex sample data. Statistics Complex Samples, with specialized planning tools and statistics, reduces the risk of reaching incorrect or misleading inferences for stratified, clustered or multistage sampling.

#### **Statistics Conjoint**

Statistics Conjoint helps market researchers develop successful products. By performing conjoint analysis, you learn what product attributes are important in the consumer's mind and what the most preferred attribute levels are, and can perform pricing studies and brand equity studies.

#### **Statistics Tables**

Use Statistics Tables to present survey, customer satisfaction, polling and compliance reporting results. Features such as a table builder preview, included inferential statistics and data management capabilities make it easy to clearly communicate your results.



### Statistics Preparation

With Statistics Preparation, you gain several procedures that facilitate the data preparation process. This add-on module enables you to easily identify suspicious and invalid cases, variables and data values; view patterns of missing data; summarize variable distributions to get your data ready for analysis; and more accurately work with algorithms designed for nominal attributes.

### Statistics Decision Trees

Create highly visual classification and decision trees directly within Statistics for Data Analysis for segmentation, stratification, prediction, data reduction and variable screening, interaction identification, category merging and discretizing continuous variables. Highly visual trees enable you to present results in an intuitive manner.

### Statistics Direct Marketing

Statistics Direct Marketing helps marketers perform various kinds of analyses easily and confidently, without requiring a detailed understanding of statistics. They can conduct recency, frequency and monetary value (RFM) analysis, cluster analysis, and prospect profiling. They can also improve marketing campaigns through postal code analysis, propensity scoring, and control package testing. And they can easily score new customer data and access pre-built models.

### Statistics Exact Tests

Statistics Exact Tests always provides you with correct p values, regardless of your data structure, even if you have a

small number of cases, have subset your data into fine breakdowns or have variables where 80 percent or more of the responses are in one category.

### Statistics Forecasting

Improve forecasting with complete time-series analyses, including multiple curve-fitting, smoothing models, methods for estimating autoregressive functions and temporal causal modeling. Use the Expert Modeler to automatically determine

which ARIMA (autoregressive integrated moving average) process or exponential smoothing model best fits your time-series and independent variables, eliminating selection through trial and error.

### Statistics Missing Values

If values are missing from your data, this module may find some relationships between the missing values and other variables. In addition, the missing values module can estimate what the value would be if data weren't missing.

### Statistics Neural Networks

Use the Statistics Neural Networks module to model complex relationships between inputs and outputs or to discover patterns in your data. Choose from algorithms that can be used for classification (categorical outcomes) and prediction (numerical outcomes). The two available algorithms are Multilayer Perceptron and Radial Basis Function.



### **Statistics Regression**

Predict behavior or events when your data go beyond the assumptions of linear regression techniques. Perform multinomial or binary logistic regression and nonlinear regression, weighted least squares, two-stage least squares and probit analysis.

### **Complementary product**

Use these products with Statistics for Data Analysis to enhance your analytical results.

### **Statistics Amos**

Support your research and theories by extending standard multivariate analysis methods when using this stand-alone software package for structural equation modeling (SEM). Build attitudinal and behavioral models that more realistically reflect complex relationships, because any numeric variable, whether observed or latent, can be used to predict any other numeric variable. The latest release includes a new nongraphical method of model specification that improves accessibility for users who need scripting capabilities and enables large, complicated models to be run more quickly.