

# Funzionalità di R ed Integrazione in *STATISTICA*



- ✓ *Esecuzione di programmi nativi di R dall'interno di STATISTICA*
- ✓ *Miglioramento di STATISTICA con le funzionalità uniche di R*
- ✓ *Miglioramento di R con le funzionalità uniche di STATISTICA*
- ✓ *Creazione e supporto d'installazioni validate (FDA) sulla base di R*
- ✓ *Uso di WebSTATISTICA per creare un server R scalabile e sicuro*

**data analysis • data mining • quality control • web-based analytics**

**StatSoft Italia Srl • Via Parenzo 3 • 35010, Vigonza (PD) • Tel: 049 8932897 • Fax: 049 8934654 • [info@statsoft.it](mailto:info@statsoft.it) • [www.statsoft.it](http://www.statsoft.it)**

Australia: StatSoft Pacific Pty Ltd.

Brazil: StatSoft South America

Bulgaria: StatSoft Bulgaria Ltd.

Czech Rep.: StatSoft Czech Rep. s.r.o.

China: StatSoft China

France: StatSoft France

Germany: StatSoft GmbH

Hungary: StatSoft Hungary Ltd.

India: StatSoft India Pvt. Ltd.

Israel: StatSoft Israel Ltd.

Italy: StatSoft Italia srl

Japan: StatSoft Japan Inc.

Korea: StatSoft Korea

Netherlands: StatSoft Benelux BV

Norway: StatSoft Norway AS

Poland: StatSoft Polska Sp. z o.o.

Portugal: StatSoft Ibérica Lda

Russia: StatSoft Russia

Spain: StatSoft Ibérica Lda

S. Africa: StatSoft S. Africa (Pty) Ltd.

Sweden: StatSoft Scandinavia AB

Taiwan: StatSoft Taiwan

UK: StatSoft Ltd.

# Tabelle dei Contenuti: Supporto Completo di R in *STATISTICA*

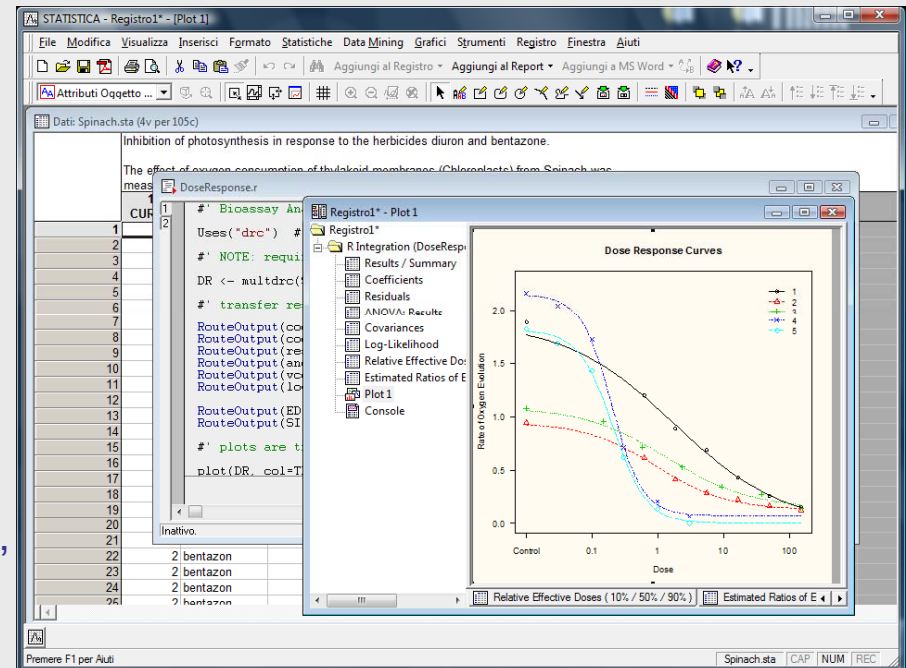
---

- Panoramica esecutiva
- Esecuzione di programmi in R come macro native di *STATISTICA*
- Caricamento su *WebSTATISTICA* Server
- Cattura di risultati dettagli da R in spreadsheet, report, grafici di *STATISTICA*
- Esecuzione di Script in R da *STATISTICA* tramite l'interfaccia flessibile di *STATISTICA*
- Costruzione di nuove funzioni di *STATISTICA* basate su librerie di R
- Integrazione delle librerie di R in *STATISTICA*: Dettagli tecnici
- Interfacciamento diretto con R attraverso l'interfaccia COM
- Creazione di nodi di *STATISTICA Data Miner* basati su R
- Integrazione di R in *STATISTICA Enterprise* (uso di R in reportistica analitica validata)
- Uso di *WebSTATISTICA* per la creazione di un server R multi-processore e multi-utente
- Riepilogo: Supporto completo ad R in (*Web*)*STATISTICA*
- Per maggiori informazioni si contatti:

StatSoft Italia srl.  
Via Parenzo 3, 35010, Vigonza (PD)  
Tel: 0498934654  
Fax: 0498932897  
Oppure si visiti [www.StatSoft.it](http://www.StatSoft.it)

# Panoramica

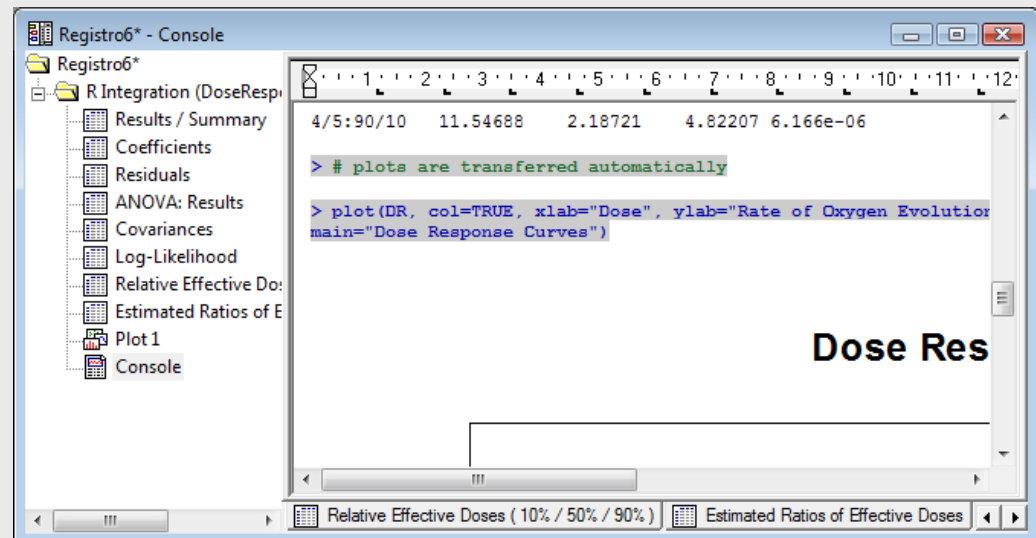
- R è un linguaggio ed un ambiente di programmazione per il calcolo statistico; R ed il suo codice sorgente è liberamente disponibile sotto licenza GNU GPL (si veda <http://cran.r-project.org>)
- Con *STATISTICA*, gli script nativi di R possono essere eseguiti direttamente dall'interno di *STATISTICA*; l'output di R può essere trasformato in spreadsheet e grafici nativi di *STATISTICA*, e gestito attraverso i *Registri* di *STATISTICA*
- Le aziende potranno adesso utilizzare le funzionalità specializzate di R con *STATISTICA*, *STATISTICA Enterprise*, e *WebSTATISTICA Server*:
  - Aggiungere nuovi “moduli” in R
  - Sfruttare le superiori capacità grafiche, gli Spreadsheet, ed i Registri di *STATISTICA* per gestire i diversi tipi di output proveniente da R
  - Costruire server scalabili in R sulla base di *WebSTATISTICA* per gestire servizi di sicurezza, caricamenti su server, e per beneficiare dei vantaggi derivanti da server multi-processori per l'esecuzione di R in ambienti regolamentati



# Esecuzione degli script di R in *STATISTICA*

## script di R come macro native di *STATISTICA*

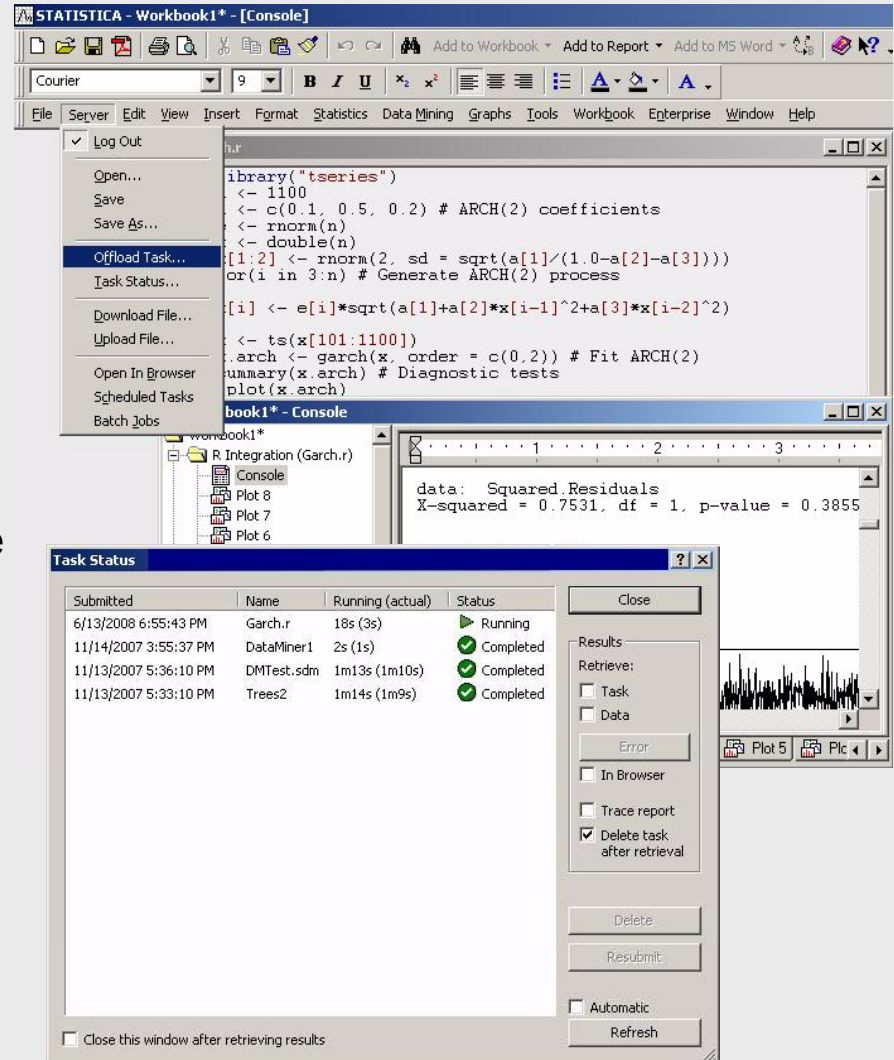
- Adesso è possibile eseguire uno script completo in R dall'interno di *STATISTICA* piuttosto che attraverso la console di R:
  - Creare nuovi script di R o caricarne di esistenti
    - Caricare file *.R* o *.S*;
    - *STATISTICA* li tratterà come macro native
  - Semplice esecuzione degli script
  - L'output della console di R verrà automaticamente catturato in un Report di *STATISTICA*
    - Comandi di R evidenziati
    - Tutti i grafici vengono catturati
  - Adesso è possibile
    - Creare file in PDF
    - Inserire report in *STATISTICA Document Management System* sotto forma di report validati
    - ....



# Caricamento su WebSTATISTICA Server

## WebSTATISTICA è un potente Server per R

- WebSTATISTICA Server è una potente architettura client-server basata sulle librerie di STATISTICA; si veda anche Usare WebSTATISTICA per creare un server scalabile multi-processore e multi-utente per R
- Gli script di R (così come gli script in SVB, gli spazi di lavoro di Data Miner, ecc.) possono essere caricati su WebSTATISTICA per una loro esecuzione e per beneficiare dei vantaggi derivanti da un ambiente server multi-processore
- Con STATISTICA e WebSTATISTICA, gli utenti di R avranno a disposizione potenti strumenti multi-utente, multi-processore, e per l'elaborazione a lotti, per eseguire i propri "R-jobs",...



The screenshot displays the STATISTICA software interface. The main window shows a console window with R code for generating ARCH(2) process data and fitting an ARCH(2) model. The code includes comments and diagnostic tests. Below the console, a 'Task Status' window is open, showing a table of submitted tasks.

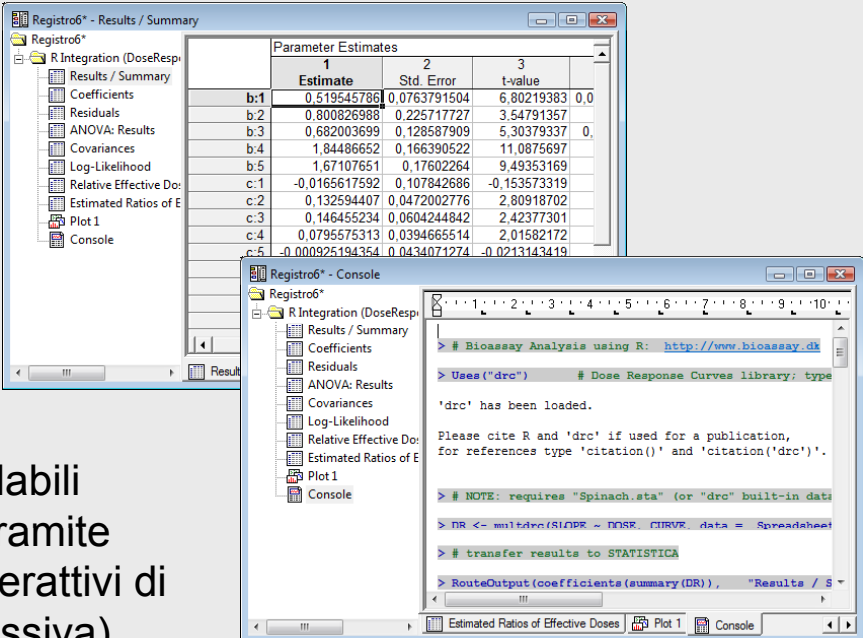
Submitted	Name	Running (actual)	Status
6/13/2008 6:55:43 PM	Garch.r	18s (3s)	Running
11/14/2007 3:55:37 PM	DataMiner1	2s (1s)	Completed
11/13/2007 5:36:10 PM	DMTest.sdm	1m13s (1m10s)	Completed
11/13/2007 5:33:10 PM	Trees2	1m14s (1m9s)	Completed

The 'Task Status' window also includes a 'Results' section with options for retrieving task or data, and buttons for 'Error', 'Delete', 'Resubmit', and 'Refresh'. A checkbox at the bottom indicates 'Close this window after retrieving results'.

# Esecuzione degli script di R in **STATISTICA**

## Cattura di Risultati Dettagliati

- Con solo poche modifiche applicate allo script di R, è possibile
  - Passare i dati/risultati in un file dati di **STATISTICA**
  - Inviare le tabelle dei risultati in spreadsheet “reali” di **STATISTICA**
  - Estrarre i grafici dei risultati in oggetti grafici di **STATISTICA**
  - Inserire tutti i risultati in Registri di **STATISTICA**, proprio come per l’output nativo di **STATISTICA**
- Usare estensioni del linguaggio:
  - *ActiveDataSet* e *Spreadsheet(filename)* per trasferire gli spreadsheet in R sotto forma di “data frame”
  - *RouteOutput (array/matrix-object)* per recuperare vettori, matrici, data frame in tabelle di **STATISTICA**
- Tutti i *grafici di R* vengono automaticamente copiati su grafici di **STATISTICA** in Metafile
  - Questi grafici sono immagini vettoriali scalabili a cui è possibile aggiungere annotazioni tramite testo, frecce, ecc. usando gli strumenti interattivi di **STATISTICA** (si veda la diapositiva successiva)



The screenshot displays two windows from the StatSoft STATISTICA software. The top window, titled 'Registro6\* - Results / Summary', shows a table of 'Parameter Estimates' for an R integration model. The bottom window, titled 'Registro6\* - Console', shows the R script being executed, including commands for loading the 'drc' library and routing results to the STATISTICA environment.

	1	2	3	
	Estimate	Std. Error	t-value	
b:1	0.519545786	0.0763791504	6.80219383	0.0
b:2	0.800826988	0.225717727	3.54791357	
b:3	0.682003699	0.128587909	5.30379337	0.
b:4	1.84486652	0.166390522	11.0875697	
b:5	1.67107651	0.17602264	9.49353169	
c:1	-0.0165617592	0.107842686	-0.153573319	
c:2	0.132594407	0.0472002776	2.80918702	
c:3	0.146455234	0.0604244842	2.42377301	
c:4	0.0795575313	0.0394665514	2.01582172	
c:5	-0.000925194354	0.0434071274	-0.0213143419	

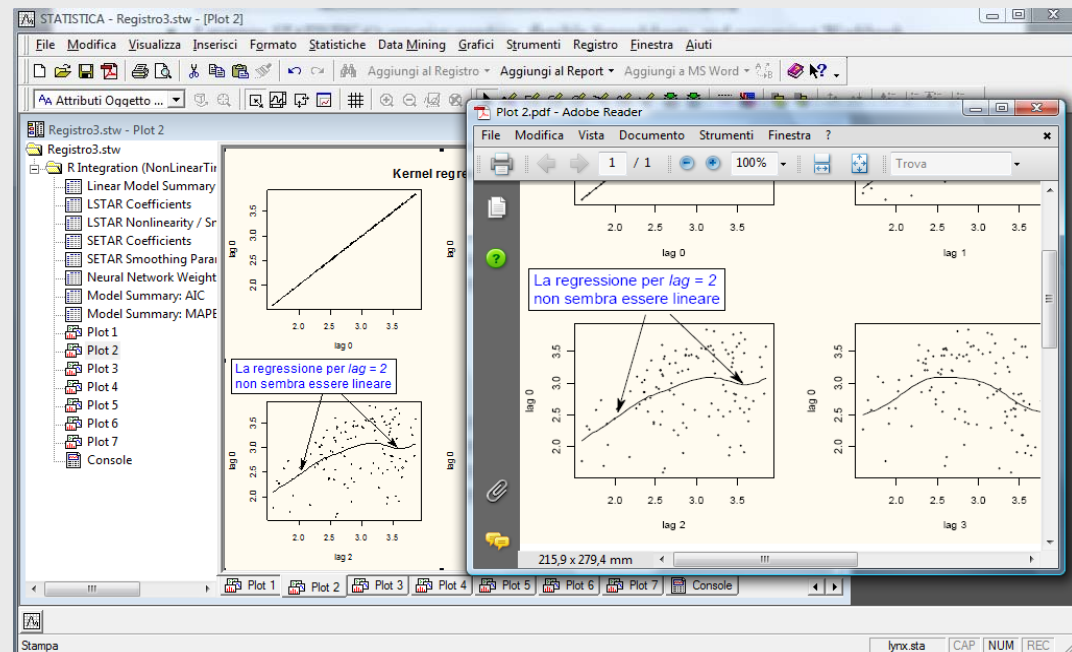
```

> # Bioassay Analysis using R: http://www.bioassay.dj
> Uses("drc") # Dose Response Curves library: type
'drc' has been loaded.
Please cite R and 'drc' if used for a publication,
for references type 'citation()' and 'citation("drc")'.
> # NOTE: requires "Spinach.ata" (or "drc" built-in data)
> DR <- multdrc(SLOPP ~ DOSE, CURVE_data = Spreadsheet
> # transfer results to STATISTICA
> RouteOutput(coefficients(summary(DR)), "Results / S
  
```

# Esecuzione degli script in R da *STATISTICA*

## Uso dell'IU flessibile di *STATISTICA*

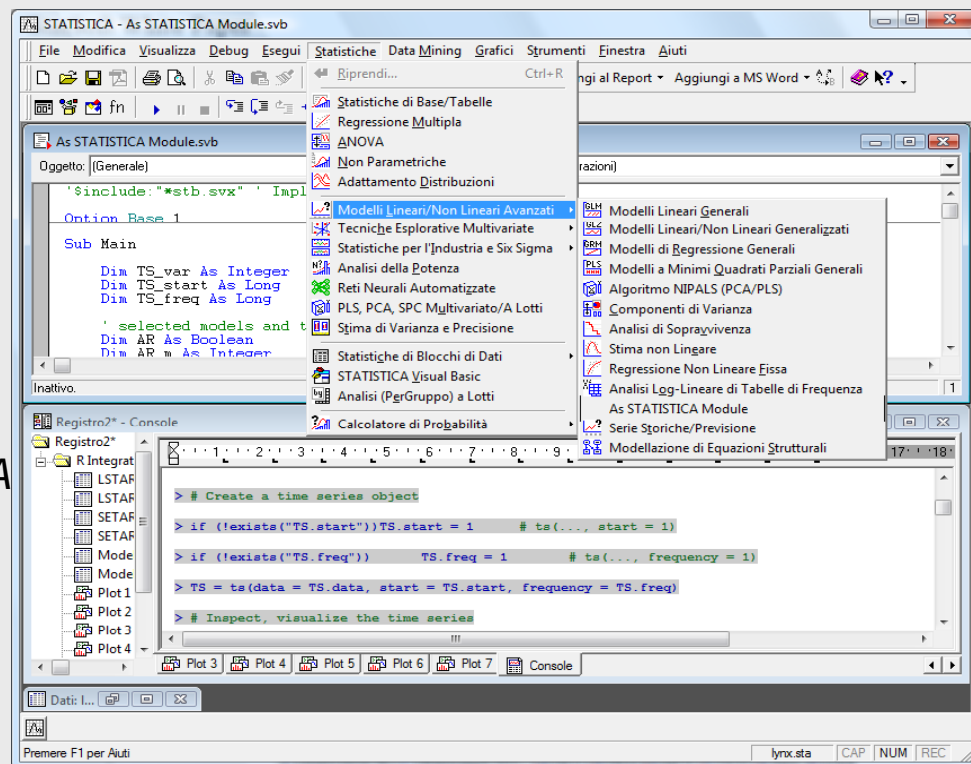
- Una volta trasferiti i risultati di *R* in *STATISTICA*, la potenza degli strumenti interattivi di *STATISTICA* desktop saranno a completa disposizione dell'utente per:
  - Stampare tabelle e report in forma di file in PDF
  - Eseguire analisi seguenti usando l'insieme di strumenti disponibili in *STATISTICA*
  - Modificare, migliorare, annotare interattivamente i grafici
  - Gestire insiemi di risultati in registri convenienti
  - Archiviare risultati versionati attraverso *STATISTICA Document Management*



# Costruzione di Nuove Funzioni di STATISTICA

- Il codice di R può essere eseguito direttamente dall'interno di *STATISTICA Visual Basic*
- I parametri (numeri, stringhe, array, spreadsheet, o codice aggiuntivo di R) possono essere passati ad R attraverso l'oggetto *STATISTICA Collection* – possono divenire variabili di R
- In questa maniera, nuove funzioni possono essere costruite in *STATISTICA* ricorrendo interamente o parzialmente ad R, o ad un “mix” di funzionalità di R e di *STATISTICA*; è possibile ad esempio:

- Creare una macro d'analisi e allegarla al menu in modo da farla divenire un nuovo “modulo di *STATISTICA*”
- La macro può presentarsi con un'Interfaccia Grafica attraverso la quale selezionare o definire i parametri per l'analisi basata su R
- I risultati saranno realizzati all'interno dei registri di *STATISTICA*
- Quindi, ad esempio sarà possibile costruire un nuovo modulo di *Serie Storiche Non-Lineari* di *STATISTICA*





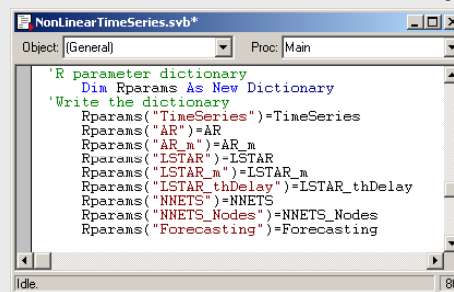
# Integrazione delle librerie di R in *STATISTICA*: Dettagli Tecnici (1)

- Per realizzare un modulo basato sulle funzionalità di R, occorre creare i file seguenti:
  - Un programma in R che esegue i calcoli (in R), ed utilizza speciali “estensioni” (ad es., *RouteOutput*, *ActiveDataset*) in modo che i dati (risultati) ed i grafici possano essere scambiati tra i contenuti di *STATISTICA* ed R
    - Questo programma in R verrà eseguito dall’interno di *STATISTICA* attraverso il ricorso ad una macro di supporto in *SVB*, che gestirà le parole chiave speciali per lo scambio dei dati (risultati, grafici) con R
    - Nota: per eseguire gli script di R in *STATISTICA* è necessario installare l’ambiente R (per maggiori dettagli si veda <http://cran.r-project.org/>)
  - Una macro in *STATISTICA Visual Basic (SVB)* che gestisce l’Interfaccia Utente (accetta i parametri, le selezioni di variabili, ...), ed esegue la “macro” in R
    - Il programma *SVB* caricherà ed eseguirà il programma in R ed invierà automaticamente tutti i risultati in *STATISTICA RouteOutput*, cioè, in registri, singoli stand-alone, o singoli oggetti di *STATISTICA* (spreadsheet, grafici).
- Le seguenti diapositive forniscono dettagli sugli
  - Esempi messi a disposizione con l’installazione di *STATISTICA*

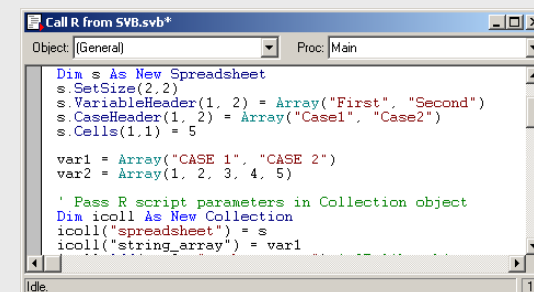
# Integrazione delle librerie di R in *STATISTICA*: Dettagli Tecnici (2)

- Per realizzare un modulo basato sulle funzionalità di R, si seguano i seguenti passaggi:
  - Si scriva il codice R come avviene per i tradizionali strumenti di R, e poi si esegua il debug dello script
    - Oppure si utilizzi una soluzione esistente creata dalla comunità di R
  - Si scriva lo script in *STATISTICA Visual Basic* per la creazione dell'IU, per la definizione delle selezioni di variabili, per l'accettazione dei parametri di script R, ecc.
    - *STATISTICA SVB* consente di costruire finestre complesse dotate di tutti i controlli standard di Window

- Si aggiunga nel codice SVB un oggetto *Collection* al fine di passare i parametri allo script R (numeri, stringhe, array, ecc.)
- Si apra lo script in R dentro



```
'R parameter dictionary
Dim Rparams As New Dictionary
'Write the dictionary
Rparams("TimeSeries")=TimeSeries
Rparams("AR")=AR
Rparams("AR_m")=AR_m
Rparams("LSTAR")=LSTAR
Rparams("LSTAR_m")=LSTAR_m
Rparams("LSTAR_thDelay")=LSTAR_thDelay
Rparams("NNETS")=NNETS
Rparams("NNETS_Nodes")=NNETS_Nodes
Rparams("Forecasting")=Forecasting
```



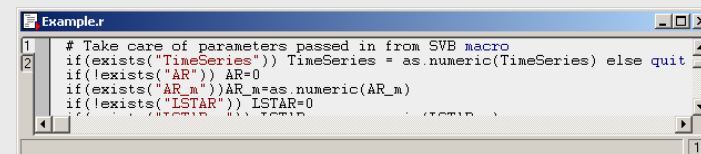
```
Dim s As New Spreadsheet
s.SetSize(2,2)
s.VariableHeader(1, 2) = Array("First", "Second")
s.CaseHeader(1, 2) = Array("Case1", "Case2")
s.Cells(1,1) = 5

var1 = Array("CASE 1", "CASE 2")
var2 = Array(1, 2, 3, 4, 5)

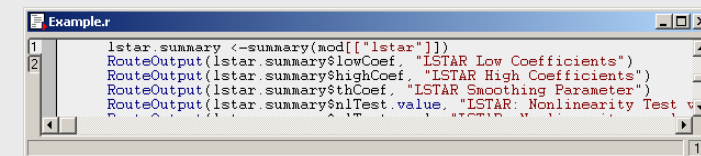
' Pass R script parameters in Collection object
Dim icoll As New Collection
icoll("spreadsheet") = s
icoll("string_array") = var1
```

*STATISTICA* e

- Controllare e definire i parametri
- Usare *ActiveDataSet* o *Spreadsheet(filename)* per trasferire i dati su un Data Frame di R
- Usare *RouteOutput()* per indirizzare l'output in registri di *STATISTICA*



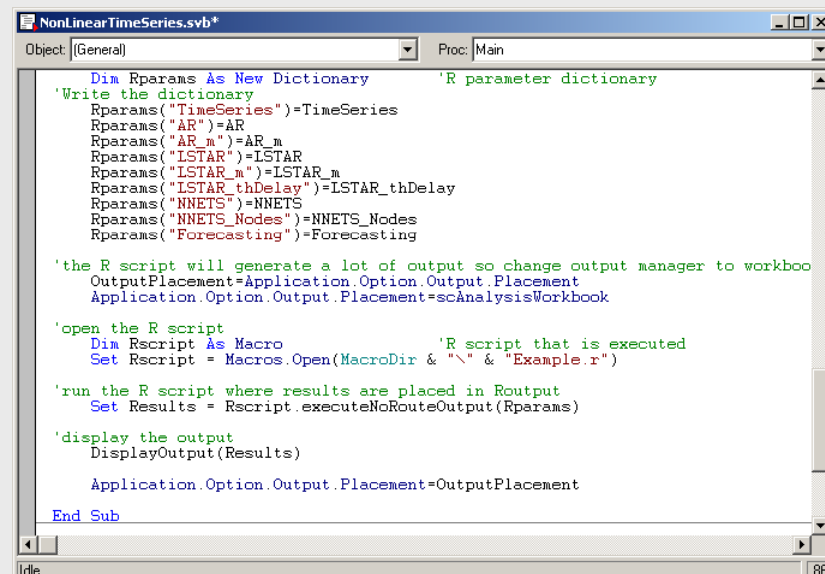
```
1 # Take care of parameters passed in from SVB macro
2 if(exists("TimeSeries")) TimeSeries = as.numeric(TimeSeries) else quit
if(exists("AR")) AR=0
if(exists("AR_m"))AR_m=as.numeric(AR_m)
if(exists("LSTAR")) LSTAR=0
```



```
1 lstar.summary <-summary(mod[["lstar"]])
2 RouteOutput(lstar.summary$lowCoef, "LSTAR Low Coefficients")
RouteOutput(lstar.summary$highCoef, "LSTAR High Coefficients")
RouteOutput(lstar.summary$thCoef, "LSTAR Smoothing Parameter")
RouteOutput(lstar.summary$nlTest.value, "LSTAR: Nonlinearity Test v
```

# Integrazione delle librerie di R in *STATISTICA*: Dettagli Tecnici (3)

- Il codice *SVB* può quindi richiamare o “eseguire” il codice in R; dall’interno del codice *SVB* è possibile:
  - Aprire/creare lo script in R all’interno della macro di *SVB*; *Macros.Open(filename), Macros.New()*
  - Eseguire lo script in R dalla macro in *SVB*; ad es.:
    - *Results = Macro.ExecuteNoRouteOutput([Parametri])*
    - *Results* è un oggetto di *StaDataCollection*
  - Visualizzare i *Risultati* attraverso la funzione *RouteOutput()* per inviarli a registri/report/..., o per iterare attraverso i contenuti per l’estrazione di dati specifici



```
NonLinearTimeSeries.svb*
Object: [General] Proc: Main

Dim Rparams As New Dictionary 'R parameter dictionary
Write the dictionary
Rparams("TimeSeries")=TimeSeries
Rparams("AR")=AR
Rparams("AR_m")=AR_m
Rparams("ISTAR")=ISTAR
Rparams("ISTAR_m")=ISTAR_m
Rparams("ISTAR_thDelay")=ISTAR_thDelay
Rparams("NNETS")=NNETS
Rparams("NNETS_Nodes")=NNETS_Nodes
Rparams("Forecasting")=Forecasting

'the R script will generate a lot of output so change output manager to workbook
OutputPlacement=Application.Option.Output.Placement
Application.Option.Output.Placement=scAnalysisWorkbook

'open the R script
Dim Rscript As Macro 'R script that is executed
Set Rscript = Macros.Open(MacroDir & "\ " & "Example.r")

'run the R script where results are placed in Routput
Set Results = Rscript.executeNoRouteOutput(Rparams)

'display the output
DisplayOutput(Results)

Application.Option.Output.Placement=OutputPlacement

End Sub

Idle. 86
```

# Interfacciamento diretto con R attraverso l'Interfaccia COM

- In generale, i programmi in R possono essere eseguiti da *STATISTICA Visual Basic (SVB)*
- Con *STATISTICA*, i dettagli dell'interfacciamento con R vengono automaticamente *gestiti* dall'interno di *STATISTICA* quando viene aperto un programma in R
- Tuttavia, è possibile avere accesso ad R direttamente attraverso "R COM Server library"
  - Si veda: <http://sunsite.univie.ac.at/rcom/>; la libreria è distribuita sotto licenza GNU LGPL
- Con "R Com Server" (ed R) installati, si aggiungano questi riferimenti nello script in SVB:

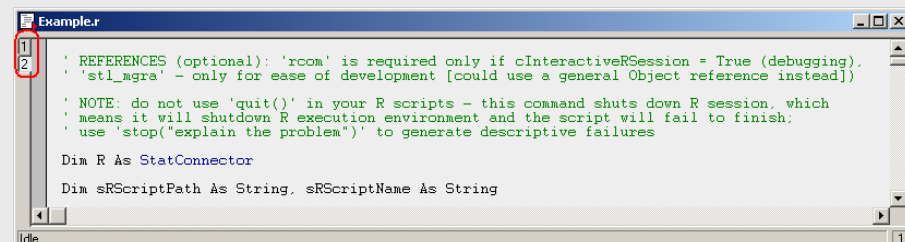
```
StatConnectorClnt 1.0 Type Library (1.0)
StatConnectorCommon 1.1 Type Library (1.1)
StatConnectorSvr 1.1 Type Library (1.1)
StatConnTools (10.0)
```

- In SVB, si instanzi un oggetto *StatConnector*.

```
Dim s As New STATCONNECTORSRVLib.StatConnector
S.Init("R")
s.EvaluateNoReturn( " R script command ")
Dim retval As Variant
retval = S.Evaluate( " R script command ")
```

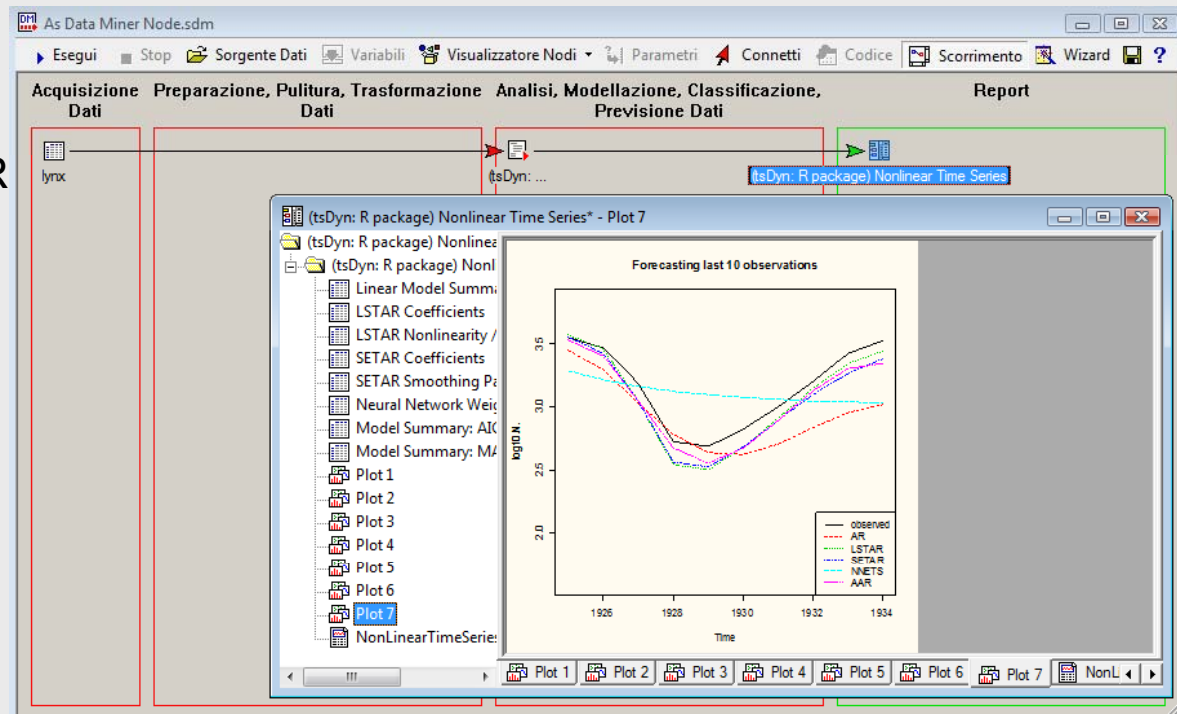
- Quando l'utente apre un programma in R (file con estensione *.R* o *.S*), verrà aperta automaticamente una macro in *SVB* che offre tutto il supporto necessario per l'esecuzione dello script in R *STATISTICA*

- Gli utenti possono espandere o personalizzare la macro di supporto in R (*R.SVB*)



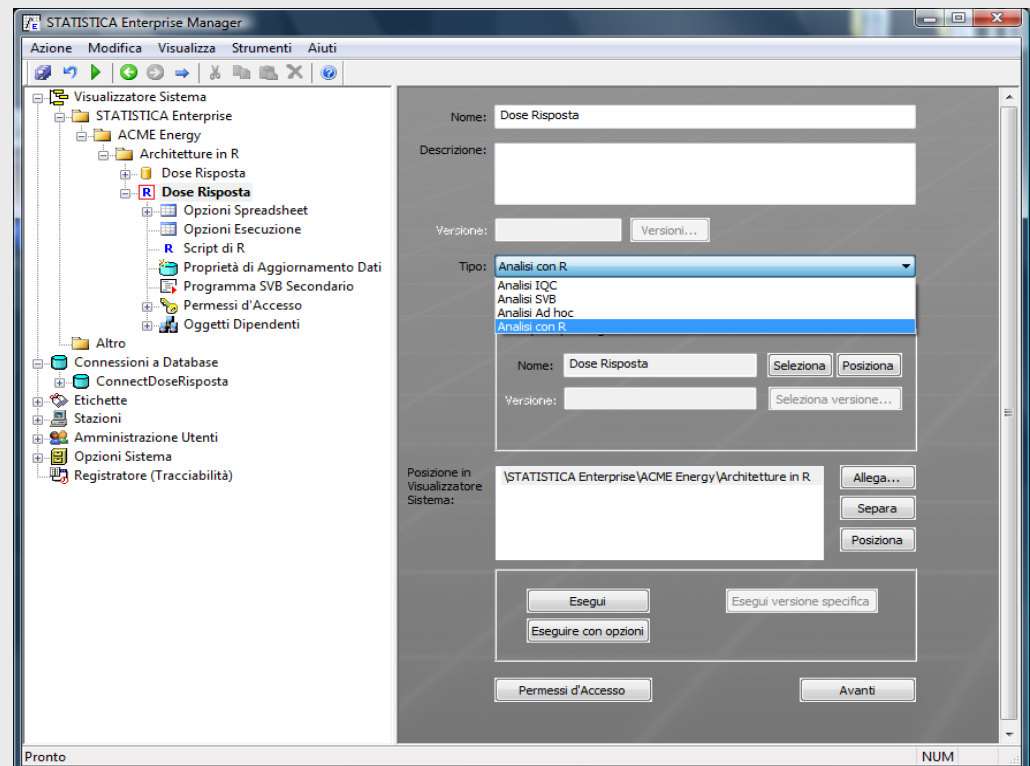
# Creazione Nodi di Data Miner basati su R

- *STATISTICA Data Miner* può essere espanso tramite script personalizzati di *STATISTICA Visual Basic (SVB)*
- Si segua semplicemente la procedura di realizzazione dei nodi di *STATISTICA Data Miner*, e si utilizzino le funzionalità di R nello stesso modo descritto nelle diapositive precedenti
- Considerata la semplicità con cui è possibile costruire interfacce in *STATISTICA Data Miner* (nodi), sarà semplice creare spazi di lavoro di *STATISTICA Data Miner* che incorporano al loro interno sia funzionalità di *STATISTICA Data Miner* che funzioni specializzate di R



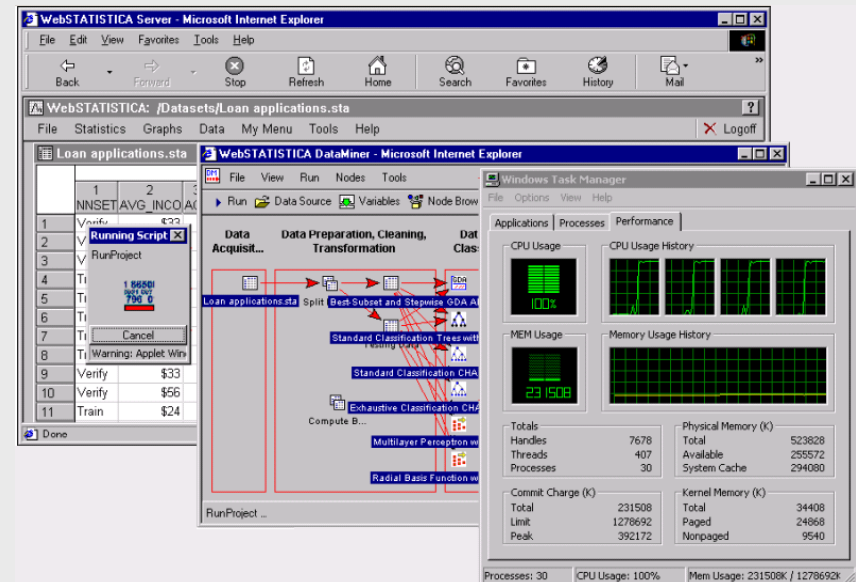
# Integrazione delle Funzionalità di R in *STATISTICA Enterprise*

- *STATISTICA Enterprise* è una piattaforma aziendale per l'analisi dei dati, dotata di un sistema di differenziazione dei ruoli e di distribuzione di reportistica standard, necessario negli ambienti regolamentati da normative adottate a livello internazionale (FDA)
- *STATISTICA Enterprise* consente agli utenti con permessi di amministrazione di creare configurazione di dati (query riutilizzabili, meta-dati) e configurazioni di dati (architetture di analisi riutilizzabili, report analitici)
- Usando i metodi descritti nelle diapositive precedenti, sarà possibile integrare in *STATISTICA Enterprise* analisi e report basati su funzionalità di R
- Sfrutta la potenza specializzata di R all'interno di una potente piattaforma aziendale d'analisi, attraverso la quale gli utenti finali non avranno bisogno di conoscere R, o qualsiasi altro linguaggio di programmazione!



# WebSTATISTICA crea un server di R scalabile multi-processore, e multi-utente

- WebSTATISTICA è un'architettura client-server in cui STATISTICA viene eseguito come servizio
  - Istanze multiple di STATISTICA possono essere eseguite contemporaneamente, consentendo in questo modo la gestione di più lavori simultanei
  - I singoli "lavori" di STATISTICA possono essere distribuiti su più processori
  - WebSTATISTICA sopporta grossi carichi di lavoro in modo molto efficace
  - Gli utenti (client) possono lavorare interattivamente con WebSTATISTICA, oppure inviando lavori a lotti programmati automaticamente o manualmente per l'esecuzione
- Con (Web)STATISTICA, è adesso possibile eseguire su server:
  - Programmi nativi in R
  - Script in STATISTICA Visual Basic (SVB) che possono richiamare i programmi in R
  - Progetti di (Web)STATISTICA Data Miner basati su R
  - Configurazioni d'analisi di (Web)STATISTICA Enterprise basate sulle funzionalità di R
- WebSTATISTICA è infatti un potente server multi-processore d'analisi



The screenshot displays the WebSTATISTICA server interface. On the left, a table titled 'Loan applications.sta' shows data for 11 rows, including variables like 'NNSETAVG\_INCOA', 'Verify', and 'Train'. A 'Running Script' dialog box is open over the table. The main window shows a data mining workflow with steps: 'Data Acquisition', 'Data Preparation, Cleaning, Transformation', 'Split', 'Best Subset and Stepwise ODA', 'Standard Classification Teaset', 'Standard Classification CHA', 'Exhaustive Classification CH2', 'Compute B...', 'Multilayer Perceptron', and 'Radial Basis Function'. On the right, the Windows Task Manager Performance tab is visible, showing system metrics: CPU Usage (100%), MEM Usage (231508K), Physical Memory (523828K Total, 255572K Available, 294080K System Cache), and Kernel Memory (34408K Total, 24868K Paged, 9540K Nonpaged).



# Riepilogo: Supporto Completo per l'integrazione di R in (Web)STATISTICA

- Con *STATISTICA*, gli utenti potranno adesso beneficiare della potenza specializzata di R, sfruttando nel frattempo tutte le funzionalità di *STATISTICA* e di *WebSTATISTICA (Enterprise)* (analisi, grafici, gestione flessibile delle tabelle dei risultati, supporto di stampa in PDF...)
- Con *STATISTICA*, saranno disponibili diversi modi d'impiegare R:
  - Accedendo ad interfacce R COM per un'interazione di basso-livello
  - Eseguendo programmi in R direttamente da *STATISTICA*, e recuperando i risultati in report, registri e grafici di *STATISTICA*
  - Usando i dataset di *STATISTICA* nell'ambiente di R e recuperando i risultati tabulati da programmi in R all'interno di spreadsheet di *STATISTICA*
  - Richiamando R attraverso *STATISTICA Visual Basic (SVB)*, per creare funzionalità di *STATISTICA* basate su librerie di R
  - Eseguendo R attraverso *STATISTICA Enterprise* (creando configurazioni d'analisi basate su R, e consentendo agli utenti che hanno *poca familiarità* con R di sfruttare la potenza di R)
  - Creando ed eseguendo nodi di *STATISTICA Data Miner* basati su R, per integrare routine specializzate in R all'interno di *STATISTICA Data Miner*
  - Eseguendo R da *WebSTATISTICA Server* (usando uno qualsiasi dei metodi descritti in precedenza), per creare server R potenti, sicuri e multi-utente con funzionalità di esecuzione a lotti, ecc.

# Per Maggiori Informazioni

- **Contattare StatSoft Italia**  
**via Parenzo 3 – 35010, Vigonza (PD)**  
**Tel: 0498934654**  
**Fax: 0498932897**
- **Oppure si visiti StatSoft ([www.StatSoft.it](http://www.StatSoft.it))**



- **Grazie per il tempo dedicato alla lettura di questo documento!**