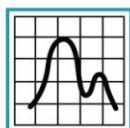


StatSoft White Paper

Opções de Integração e Recursos para Aproveitar a Funcionalidade Especializada R nas Soluções do STATISTICA e STATISTICA Enterprise Server

Sumário

1	Introdução	3
2	Visão Geral e Resumo	5
3	Arquitetura Básica e Recursos do Suporte R no STATISTICA	11
3.1	Interface COM para o Ambiente R	12
3.2	Macros de Suporte à Integração do R (R.svb / R.r)	13
3.3	Scripts do R como Macros Nativas do STATISTICA	14
3.4	Recuperação de Resultados	16
3.4.1	Capacidades Únicas do STATISTICA	18
4	Extensões da Linguagem R: Transmissão de Dados ao R e Recuperação de Resultados	18
4.1	ActiveDataSet	19
4.2	Spreadsheet (FilePathOrName)	19
4.3	RouteOutput (RObject, [SpreadsheetName], [SpreadsheetHeader])	21
4.4	Uses (RPackageName, [AttachImports])	22
5	Executando Scripts do R a partir do STATISTICA Visual Basic	23



5.1	Scripts do R são Macros do STATISTICA	24
5.2	Passagem de Parâmetros para Scripts do R: Objeto <code>Collection</code>	25
5.3	Gestão de Resultados de Script	28
5.3.1	<i>Macro.ExecuteNoRouteOutput([Parameters As Collection]) As StaDocCollection</i>	29
5.4	Mais Exemplos.....	30
6	Suporte do R no STATISTICA Enterprise Server	30
6.1	STATISTICA Enterprise Server é um Poderoso R Server.....	32
6.2	Carregamento-off de Scripts a partir do STATISTICA para o STATISTICA Enterprise Server.....	33
7	Criação de Nós do STATISTICA Data Miner Baseados no R	35
8	Integração do R no STATISTICA Enterprise e no STATISTICA Enterprise Server	37
8.1	Criação de Configurações de Análise Baseadas no R.....	37
8.2	Chamar Scripts do R a partir das Configurações de Análise do SVB	40
8.3	Resumo	41
9	Comentários Finais e Algumas Ressalvas.....	41
9.1	Tratamento de Erros	41
9.2	Pontos Fortes e Limitações	42

1 Introdução

R é uma linguagem de programação e ambiente para computação estatística (www.r-project.org). A maior parte do ambiente R e seu código-fonte estão disponíveis sob a licença GNU GPL (www.r-project.org/licenses). Observe que nenhum dos componentes do ambiente R constituem um “freeware irrestrito”, em vez disso ele está disponível apenas sob os termos das licenças específicas, os quais os usuários que pretendem baixar os aplicativos precisam aceitar antes de baixar e com as quais eles precisam cumprir. Além disso, as licenças podem mudar ao longo do tempo e, portanto, os usuários são aconselhados a se familiarizarem completamente com seus termos cada vez que baixarem componentes do ambiente R.

O STATISTICA pode interagir com aplicativos de terceiros (como o Microsoft Office) através de determinados canais compatíveis, nomeadamente COM, que é uma interface para a comunicação construída em sistemas operacionais da Microsoft.

Mesmo que o R em si não forneça uma interface COM, existem aplicativos de terceiros que permitem a comunicação com R através do COM. Um desses aplicativos é a biblioteca *statconnDCOM* (rcom.univie.ac.at ou www.statconn.com).

O STATISTICA pode interagir com o R via *statconnDCOM* (se esta biblioteca está instalada no sistema) e permitirá que os usuários abram os scripts R e os submetam à execução no ambiente do R de dentro do STATISTICA, facilitando a transferência bidirecional de dados e apresentação das saídas resultantes, através de macro ajustável pelo usuário, que é executada no STATISTICA no lugar do script R.

Essa interface permite que todos os produtos do STATISTICA, das soluções Desktop às soluções Enterprise e Web, forneçam um suporte abrangente para interação com a plataforma R (em sistemas onde componentes de terceiros e bibliotecas necessárias estão presentes), dando aos usuários uma forma de:

- * Rodar scripts R dentro do ambiente do STATISTICA, retornando resultados para relatórios, workbooks e gráficos do STATISTICA.
- * Processar conjuntos de dados no R e retornando resultados tabulados a partir do R para spreadsheets do STATISTICA.
- * Chamar o R a partir do STATISTICA Visual Basic (SVB) para criar nova funcionalidade que aproveitem as bibliotecas do R.
- * Utilizar o R no STATISTICA Enterprise, STATISTICA Server e STATISTICA Data Miner

NOTA: É de responsabilidade do usuário garantir a conformidade com os termos de todas as licenças aplicáveis ao R e a todos os componentes do ambiente do R. Sempre analise cuidadosamente todos os contratos de licença antes de aceitá-los, visto que eles podem mudar ao longo do tempo. Por favor, observe que esses produtos são externos ao ambiente STATISTICA e eles não estão cobertos por qualquer contrato de licença da StatSoft.

A fim de configurar a interface entre o STATISTICA e o R, você terá que seguir as instruções encontradas em <http://statsoft.com/r-integration>, que incluem:

- Faça o download e instale a biblioteca *statconnDCOM*, disponível em rcom.univie.ac.at (procurar a página de Download e lá – pela seção *statconnDCOM*)

Observe que esta biblioteca tem certos requisitos mínimos em relação às versões do R.

- Instale o pacote *rcom* do R (isto vai também instalar o pacote *rscproxy* que é necessário para o *statconnDCOM*)

Observe que este pacote tem que ser instalado enquanto roda o R como administrador. Para conseguir isso, clique com o botão direito sobre o ícone do aplicativo R e selecione *Executar como Administrador*.

Para instalar, rode os seguintes comandos no console do R:

```
install.packages('rcom')  
library(rcom)  
comRegisterRegistry()
```

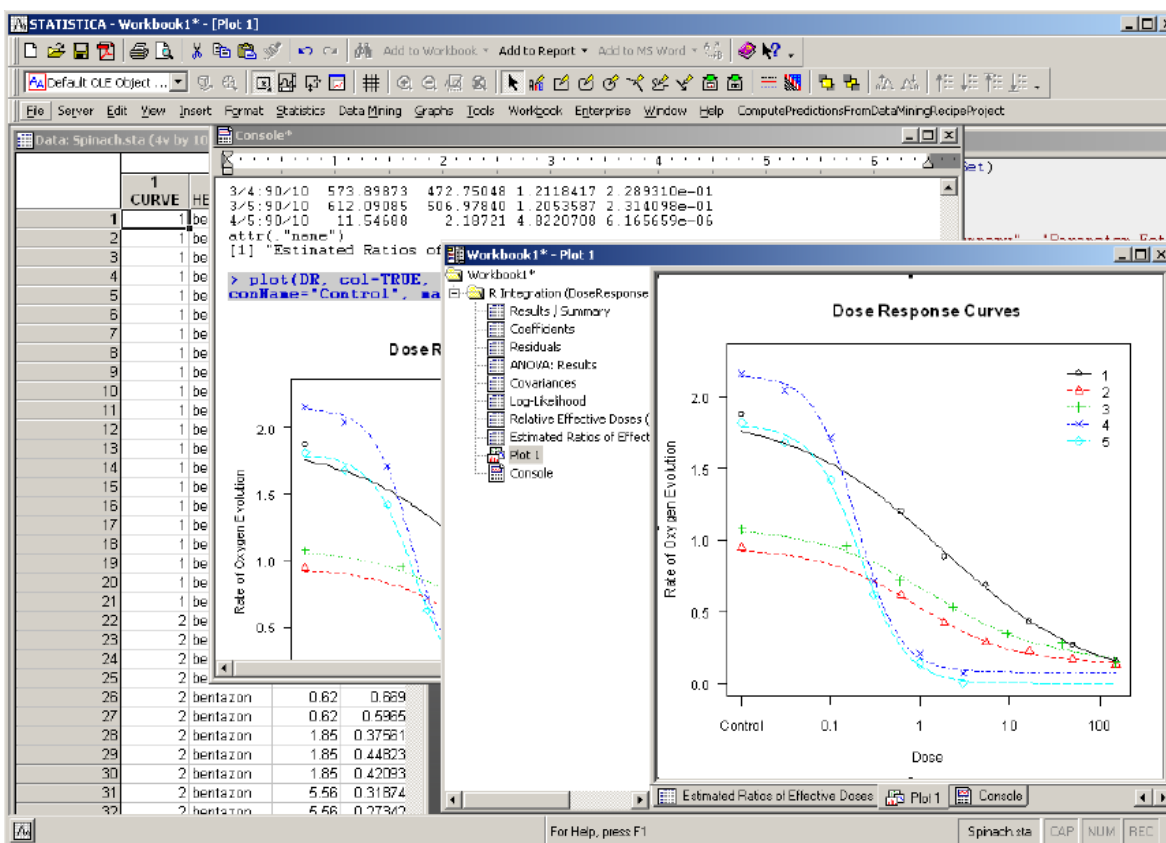
O R é altamente extensível: os usuários podem enviar as bibliotecas ("pacotes") implementando um conjunto de funções, geralmente para uma área específica de seu conhecimento/pesquisa. A comunidade do R mantém vários repositórios centralizados que compõem centenas de pacotes prontamente disponíveis para todos os usuários na Internet. Muitos destes pacotes atendem especificamente ao público altamente especializado com necessidades de análise de dados específicas.

2 Visão Geral e Resumo

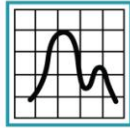
O objetivo deste documento é fornecer uma descrição detalhada dos novos recursos que tornam a diversidade e o poder do R totalmente disponíveis para usuários de todas as soluções da StatSoft. Além disso, esses recursos permitem ao usuário combinar as capacidades únicas de ambas plataformas STATISTICA e R.

Iniciando o STATISTICA:

- Scripts nativos do R podem ser executados diretamente no STATISTICA, STATISTICA Enterprise, e STATISTICA Enterprise Server;
- Saídas do R podem ser recuperadas como planilhas e gráficos nativos do STATISTICA, e gerenciados através de workbooks altamente flexíveis do STATISTICA.

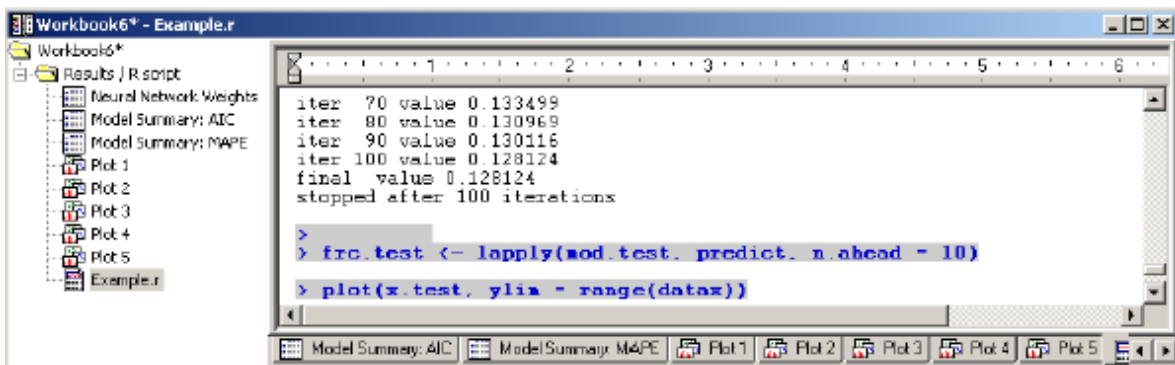
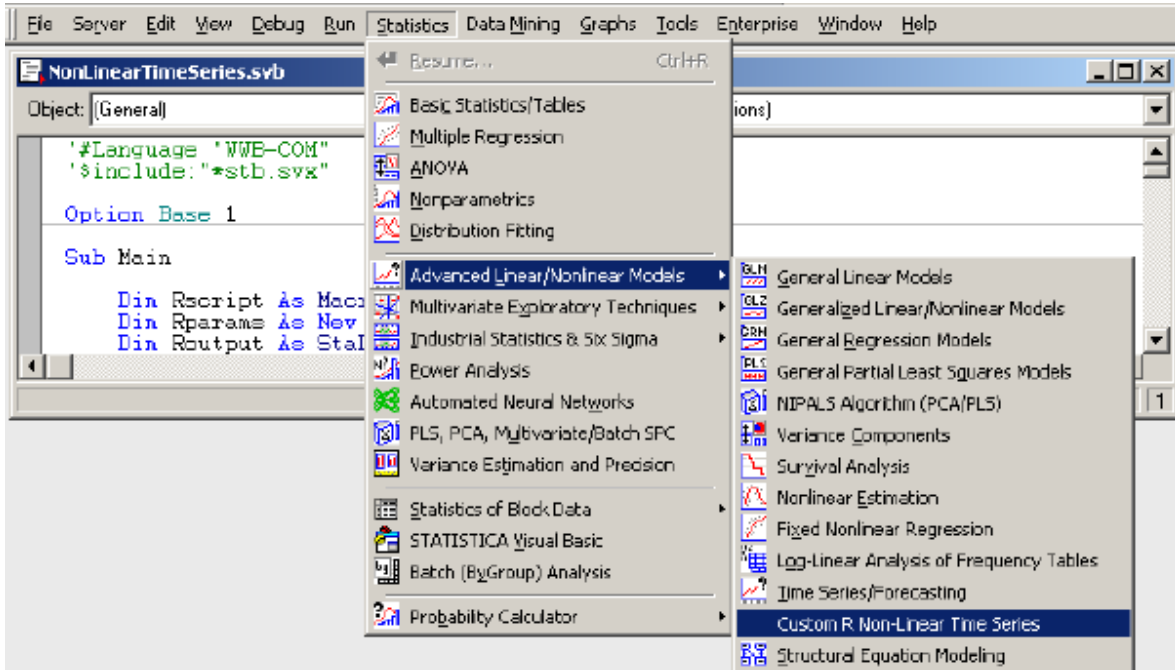


Assim, as empresas podem agora usar as rotinas e capacidades especializadas do R com o STATISTICA, STATISTICA Enterprise e STATISTICA Enterprise Server para:

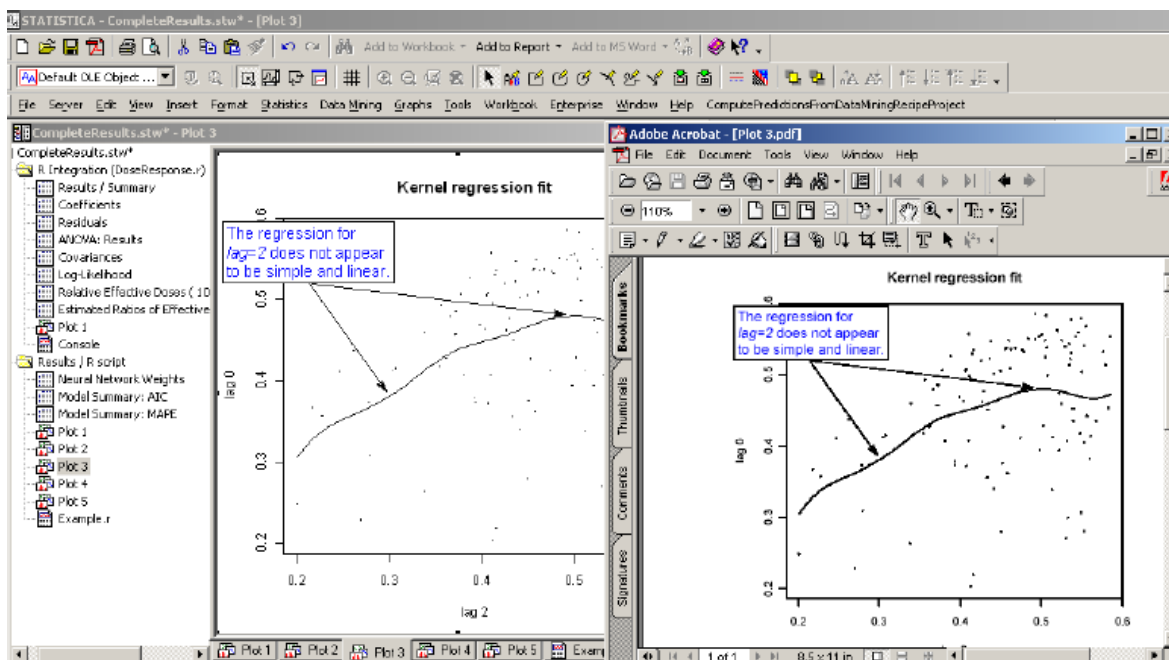


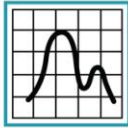
StatSoft
SOUTH AMERICA

- Adicionar novos “módulos” baseados no R;

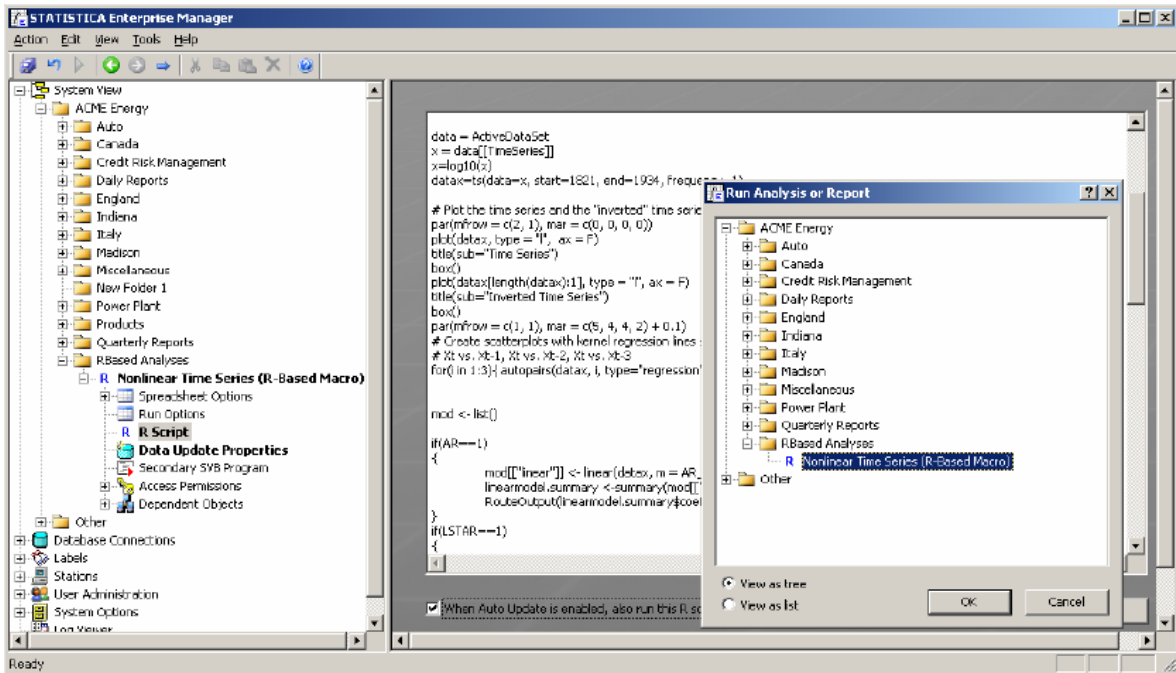


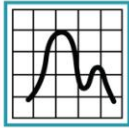
- Gráficos do STATISTICA de alta qualidade, planilhas flexíveis e práticos workbooks para vários tipos de documentos para manusear as saídas do R;



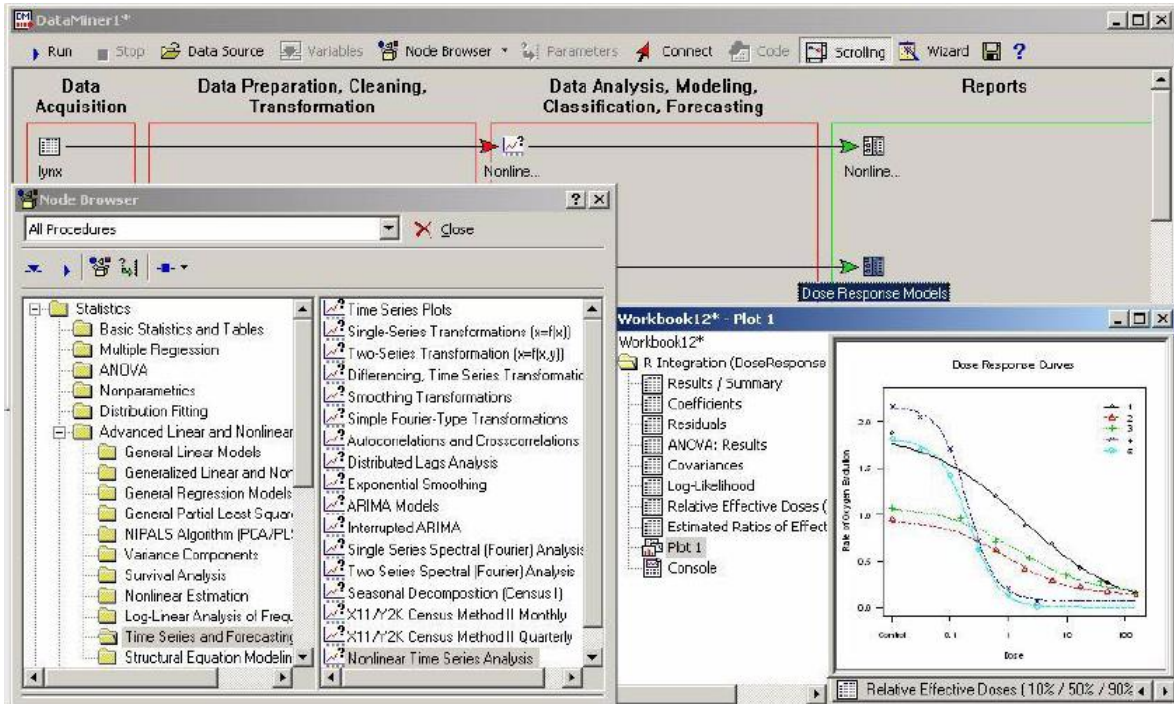


- Integrar o R no STATISTICA Enterprise para tornar a funcionalidade especializada R disponível como templates de análises reutilizáveis para usuários não familiarizados com a linguagem R de forma segura, baseado em funções do sistema de análise enterprise;

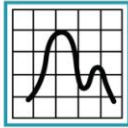




- Adicionar nós analíticos baseados no R para o STATISTICA Data Miner, aproveitando assim todas as capacidades do R dentro do STATISTICA e dos workspaces do Data Miner;



- Construir servidores do R dimensionáveis usando o STATISTICA Enterprise Server para lidar com a segurança e balanceamento de carga, e para tirar proveito de servidores de múltiplos processadores para rodar o R pela demanda e/ou aplicativos validados do enterprise.



StatSoft
SOUTH AMERICA

The screenshot displays the STATISTICA software environment with several components:

- STATISTICA - Workbook1* - [Console]:** Shows R code for fitting an ARCH(2) model. The code includes:

```
library("tseries")
n <- 1100
c <- c(0.1, 0.5, 0.2) # ARCH(2) coefficients
r <- rnorm(n)
d <- double(n)
[[1:2] <- rnorm(2, sd = sqrt(a[1]/(1.0-a[2]-a[3])))
for(i in 3:n) # Generate ARCH(2) process
{
  d[i] <- c[i]*sqrt(a[1]+a[2]*x[i-1]^2+a[3]*x[i-2]^2)
}
x <- ts(x[101:1100])
x.arch <- garch(x, order = c(0,2)) # Fit ARCH(2)
summary(x.arch) # Diagnostic tests
plot(x.arch)
```
- Task Status:** A table showing the execution status of various 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 5:38:10 PM	DMTest.adm	1m13s (1m10s)	Completed
11/13/2007 5:33:10 PM	Trees2	1m14s (1m9s)	Completed
- WebSTATISTICA Server - Microsoft Internet Explorer:** Displays the URL `/Datasets/Loan_applications.sta`.
- WebSTATISTICA DataMiner - Microsoft Internet Explorer:** Shows a workflow diagram for data analysis, including steps like "Data Acquisition", "Data Preparation, Cleaning, Transformation", "Data Classification", and "Model Evaluation".
- System Performance:** A taskbar window showing system metrics:
 - CPU Usage: 100%
 - MEM Usage: 231508
 - Totals: Handles (7678), Threads (407), Processes (30)
 - Physical Memory (K): Total (523828), Available (255572), System Cache (294080)
 - Commit Charge (K): Total (231508), Limit (1278692), Peak (392172)
 - Kernel Memory (K): Total (34408), Paged (24868), Nonpaged (9516)

3 Arquitetura Básica e Recursos do Suporte R no STATISTICA

O suporte do R no STATISTICA foi projetado para criar uma plataforma integrada STATISTICA-R que permita aos usuários executar programas do R (“scripts”) diretamente dentro do STATISTICA para que possam aproveitar ao máximo as capacidades especializadas disponíveis no R.

O ambiente de integração do R no STATISTICA foi projetado especificamente para:

- Permitir aos usuários executar scripts R “como são” e recuperar os resultados em relatórios do STATISTICA
 - Toda saída do console R é copiado para o relatório; comandos R são destacados;
 - Gráficos gerados pelo script são automaticamente incorporados no relatório como imagens dimensionáveis;
 - Estes gráficos são também replicados como gráficos do STATISTICA (imagens “metafiles” dimensionáveis são colocadas dentro destes gráficos), permitindo assim anotação usando poderosos recursos gráficos no STATISTICA. Os gráficos podem ser impressos ou exportados para uma variedade de formatos de imagem;
 - Os relatórios podem ser editados, impressos e salvos como arquivos PDF.
- Fornecer extensões da linguagem R (palavras-chave) para os scripts R executados a partir do ambiente STATISTICA que:
 - Transferir dados de planilhas do STATISTICA para *data frames* do R;
 - Extrair dados tabulados de variáveis do R em planilhas do STATISTICA. Estas planilhas de “resultados” (bem como todos os gráficos produzidos por script) são devolvidos de acordo com as configurações do *Output Manager*, colocadas, por exemplo, em workbooks do STATISTICA.
- Executar scripts R como macros nativas dentro de programas do STATISTICA Visual Basic (SVB)
 - Scripts podem ser parametrizados com uma coleção de objetos (números, strings, arrays, código de R adicionais ou funções do R substituídas, planilhas) que são mapeados para variáveis do R acessíveis ao script. Esta abordagem fornece controle sobre o comportamento de scripts em repetidos processamentos ou quando usados como *backend* para módulos personalizados do STATISTICA;
 - Por padrão, toda saída de script é roteada pelo *STATISTICA Output Manager*. Os scripts podem também ser executados usando um método que, como alternativa, retorna a saída como uma coleção de documentos, dando aos desenvolvedores uma maneira fácil de extrair resultados

específicos de análise que poderiam ser usados para processamentos adicionais, por exemplo, como dados de entrada para outras análises no STATISTICA ou no R, ou gráficas.

Tomadas em conjunto, essas melhorias não só permitem aos usuários executar scripts do R diretamente no ambiente de trabalho do STATISTICA, mas também proporcionam uma forma de incorporar a funcionalidade R especializada em seus programas do SVB, módulos interativos de análise padrões, nós do *Data Miner* e configurações de análise do *Enterprise*, ou para descarregar tais scripts para o *STATISTICA Enterprise Server* para processamento no servidor.

Empresas têm a oportunidade de integrar o R em soluções validadas do *STATISTICA Enterprise* (por exemplo, para indústrias regulamentadas pela FDA) ou para fornecer um *STATISTICA Enterprise Server* poderoso baseado no ambiente do Servidor R.

Desenvolvendo e depurando scripts do R: Um dos casos de uso que o suporte do R no STATISTICA não foi projetado para fornecer um completo ambiente de desenvolvimento e depuração do R. O aplicativo de console e ferramentas fornecidas com a instalação padrão do R executam essas funções muito bem, e já são familiares aos usuários e desenvolvedores do R.

3.1 Interface COM para o Ambiente R

Para que Integração do R no STATISTICA transfira dados entre o STATISTICA e o R, o ambiente do R deve ser instalado no mesmo computador (ou no *STATISTICA Enterprise Server* ou no servidor das soluções *Enterprise*). A última versão do ambiente do R pode ser obtida no site da CRAN (<http://cran.r-project.org>).

Para acessar o ambiente do R, o STATISTICA usa a biblioteca *statconnDCOM* (<http://rcom.univie.ac.at>) distribuída sob a licença *Statconn DCOM Public Licence* (esta biblioteca depende do pacote *rscproxy* do R). Esta biblioteca funciona como **R COM Server**. O pacote *rcom* do R é utilizado para chamadas de retorno do R para o STATISTICA.

Uma vez que o R é instalado, o STATISTICA automaticamente irá detectá-lo. A próxima vez que você executar o STATISTICA, uma caixa de diálogo será exibida, onde você pode selecionar permitir a integração com o R. Clique em *Yes* e o STATISTICA irá levá-lo para uma página Web da *statsoft.com* que irá explicar onde ir para baixar o conector *statconnDCOM* e também como instalar os pacotes do R necessários *rscproxy* e *rcom*. Observe que essas etapas exigem privilégios de administrador – dependendo da sua versão do sistema operacional e configurações, o sistema pode pedir que você confirme essas ações, possivelmente solicitando credenciais de administrador.

Complete a instalação da biblioteca R COM Server aceitando as configurações padrão do instalador. Agora você está pronto para usar o R dentro do ambiente do STATISTICA. Observe que se você não precisar dos recursos de integração do R ou decide cancelar a instalação dessas bibliotecas de apoio, a janela *Integrating R into STATISTICA* será exibida quando você iniciar o STATISTICA novamente – selecione a caixa de opção *Do not ask me again* para evitar que essa janela seja exibida novamente. Você pode habilitá-la novamente, marcando a caixa de seleção correspondente no painel de opções *Documents: Macros* da janela *Options*. Lembre-se que a biblioteca R COM Server é independente do STATISTICA e permanecerá em seu computador até que seja desinstalada manualmente.

A biblioteca R COM Server fornece uma simples, mas poderosa interface COM (*Component Object Model*) para o ambiente R. Esta interface pode ser usada diretamente por programas do SVB no STATISTICA – um exemplo de uso é incluído no STATISTICA, localizado na pasta `Examples\R\Dose Response` (abra *Direct Interface To R via COM.stw* e execute a macro do SVB embutida).

Mas o uso de tal interface diretamente pelos usuários finais é muito ineficaz, às vezes improdutivo e, normalmente inflexível. Também pode degradar significativamente o desempenho global das interações com o R se realizada de forma incorreta.

Portanto, as seguintes extensões de arquitetura tendo sido adicionadas à plataforma do STATISTICA para fornecer uma efetiva experiência de integração perfeita do R para usuários finais (o exemplo citado acima também demonstra a redução significativa dos esforços necessários para implementar a mesma análise utilizando os novos recursos incorporados: no STATISTICA, basta abrir e executar `DoseResponse.r`).

3.2 Macros de Suporte à Integração do R (R.svb / R.r)

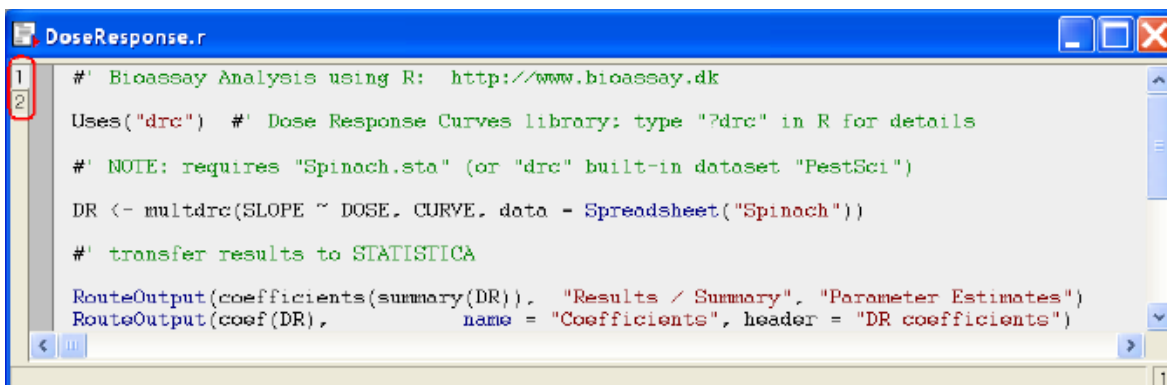
A instalação do STATISTICA inclui uma macro do STATISTICA Visual Basic chamada *R.svb* e um script do R chamado *R.r*. Esses arquivos contêm o código de suporte necessário para gerenciar as interações através do COM no STATISTICA e ao lado do R, respectivamente. Quando um script do R é executado no STATISTICA, ele é analisado pela macro de apoio, que então transfere dados e parâmetros de script, envia conteúdo de script para o ambiente do R, gerencia as condições de erro, e também lida com saídas de script, assegurando que elas sejam devidamente transferidas de volta para o STATISTICA. O script de apoio *R.r* implementa extensões específicas da linguagem R no STATISTICA (palavras-chave). Uma de suas principais responsabilidades é traduzir planilhas do STATISTICA para *data frames* do R (e vice-versa).

Embora, por padrão, o código de suporte esteja protegido contra gravação, está acessível para inspeção e pode ser modificado ou aperfeiçoado para suportar novas funcionalidades necessárias para casos de uso específico, embora os usuários devam fazê-lo “por sua conta e risco”. A macro *R.svb* suporta execução autônoma (*standalone*) para simplificar a depuração e testes de modificações.

3.3 Scripts do R como Macros Nativas do STATISTICA

O STATISTICA reconhece extensões de arquivo *.R* (e *.S*) como os scripts do R. Tais arquivos podem ser abertos selecionando *Open* no menu *File*. Ele também registra (no nível do sistema operacional) arquivos *.R/.S* como macros do STATISTICA e, portanto, estes arquivos podem ser abertos no STATISTICA de um navegador de arquivos dando um duplo-clique.

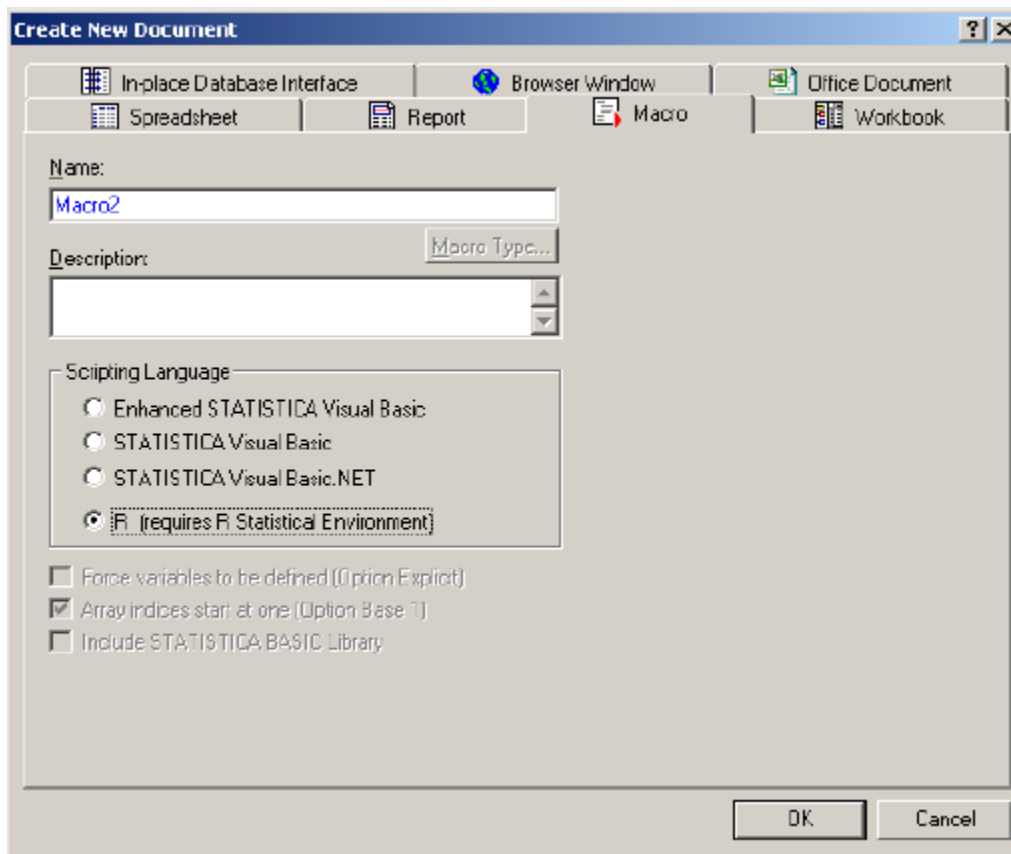
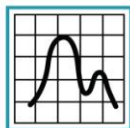
Os scripts do R são exibidos em janelas do Visual Basic do STATISTICA ligeiramente modificadas. Tais janelas na verdade contêm dois scripts: o script do R em si e a macro *R Integration Support Macro (R.svb)*, acessíveis através de duas abas no canto superior esquerdo (circulado na imagem abaixo).




```
#' Bioassay Analysis using R: http://www.bioassay.dk
Uses("drc") #' Dose Response Curves library: type "?drc" in R for details
#' NOTE: requires "Spinach.sta" (or "drc" built-in dataset "PestSci")
DR <- multdrc(SLOPE ~ DOSE, CURVE, data = Spreadsheet("Spinach"))
#' transfer results to STATISTICA
RouteOutput(coefficients(summary(DR)), "Results / Summary", "Parameter Estimates")
RouteOutput(coef(DR), name = "Coefficients", header = "DR coefficients")
```

Há código restrito ao R destacada (strings, extensões de linguagem, comentários estilo VB).

Criação de um script novo do R: Você também pode criar um script novo do R dentro do STATISTICA. Selecione *New* no menu *File* para exibir a janela *Create New Document*, clique na aba *Macro* e selecione o botão de opção *R* na área *Scripting Language* (esta opção só estará disponível se o suporte de integração ao R estiver instalado em sua máquina):



Será aberta uma janela de script do R vazia – agora você pode digitar ou colar um programa do R. A integração com o R também inclui um arquivo de texto opcional chamado *R.inc*, colocado no diretório padrão de instalação, juntamente com *R.svb*. O conteúdo do arquivo é copiado para o início de cada novo script do R criado desta forma.

Execução de scripts do R dentro STATISTICA: clique no botão  na barra de ferramentas ou selecione *Run > Run Macro* no menu de comando (ou pressione *F5*). Esta ação irá executar a macro *R.svb* para o script do R ativo.

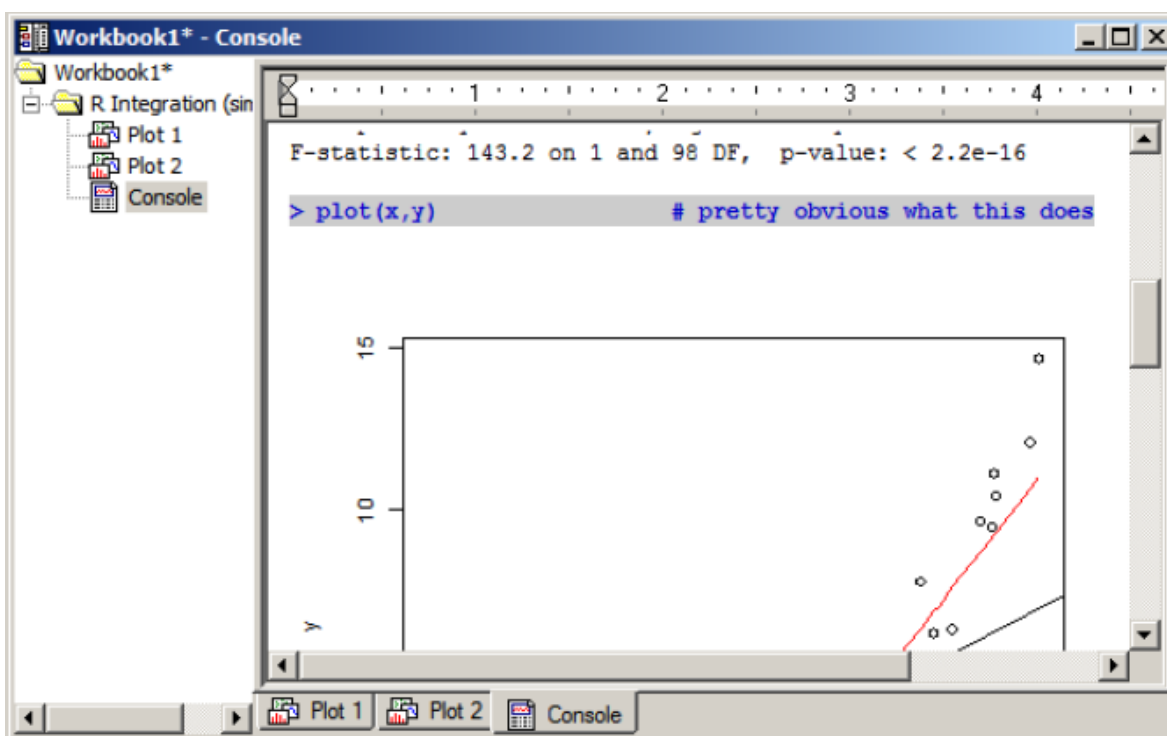
Embora os pontos de interrupção não sejam suportados para scripts do R em si, é possível definir pontos de interrupção e depurar a macro *R.svb* na segunda aba durante a execução do script do R.

Observe que, a fim de tirar vantagem dos recursos de integração do R descritos neste documento, os scripts do R devem ser executados a partir do STATISTICA. Embora seja possível desenvolver e depurar programas complexos do R dentro deste ambiente, ele não

foi projetado especificamente para esse fim. O ambiente do R por si só poderia ser mais adequado para tais atividades.

3.4 Recuperação de Resultados

Sessão de Console: A saída mínima produzida durante a execução de um script do R é um workbook do STATISTICA que representa uma sessão de console do R, incluindo comandos destacados e qualquer saída gerada pelo ambiente R. Esse relatório será produzido mesmo se o script estiver vazio. O conteúdo deste relatório pode ser editado e manipulado da mesma forma que você pode editar qualquer outro relatório do STATISTICA.



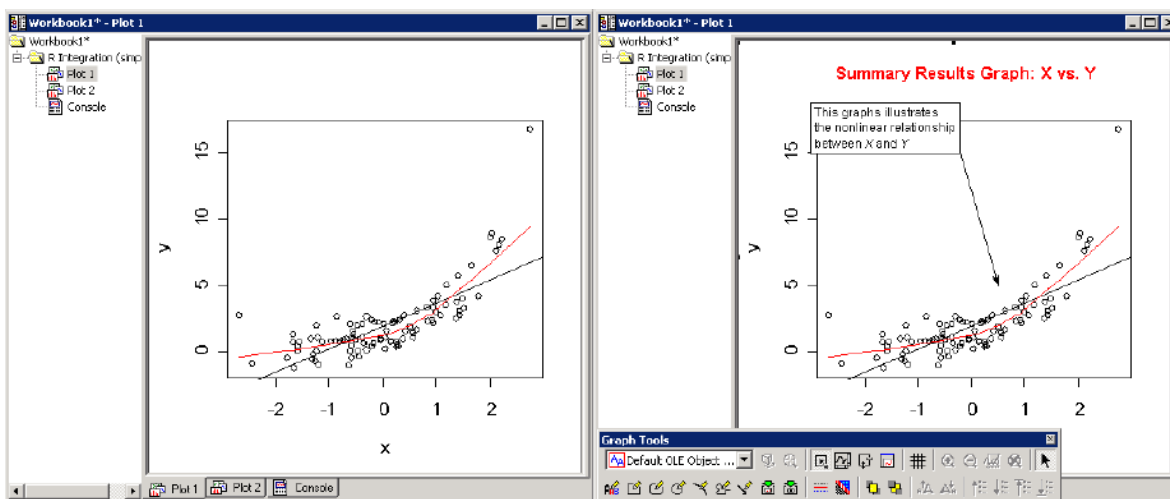
Saída gráfica: Todos os gráficos criados durante uma sessão do R são automaticamente transferidos para o ambiente do STATISTICA como *metafiles* do Windows (formato gráfico vetorial), o que significa que pode ser redimensionado sem perda de qualidade.

Estes gráficos são colocados no relatório da sessão do R console, criando um relatório da sessão do R com os gráficos embutidos ligados aos comandos gráficos (algo que às vezes não é tão facilmente alcançado no ambiente R em si).

Além disso, os gráficos também são replicados como gráficos do STATISTICA que se tornam parte da saída do script do R. As imagens *metafile* estão embutidas em objetos gráficos com “fundo” bloqueado – isso permite aos usuários fazer anotações em gráficos do R no STATISTICA usando uma interface familiar *point-and-click* com um conjunto de textos e objetos de desenho (tais como linhas e setas, retângulos e elipses, polígonos e padrões/áreas de cores de preenchimento, etc.). E uma vez que essas anotações estão ancoradas em posições relativas na área do gráfico, elas permanecerão corretamente “ligadas” se o gráfico é redimensionado.

Portanto, esses gráficos podem agora ser “personalizados” de forma flexível usando ferramentas gráficas do STATISTICA, salvos em outros formatos (por exemplo, JPG ou GIF), ou impressos (por exemplo, para arquivos PDF).

Os componentes individuais de gráficos do R (a estrutura do gráfico) não são acessíveis para a manipulação no STATISTICA, e, portanto, os recursos avançados do STATISTICA para a criação e edição de gráficos (marcadores de escala, o ponto de ajuste, linhas, etc.) não estão disponíveis. No entanto, a integração entre R e o STATISTICA oferece oportunidades para extrair dados do R e depois processar gráficos importantes dentro do ambiente STATISTICA (escrevendo macros do STATISTICA Visual Basic que vão executar scripts do R, extrair resultados e depois pós-processar os resultados, se necessário).



3.4.1 Capacidades Únicas do STATISTICA

Uma vez que saída do script do R consiste inteiramente de objetos nativos do STATISTICA:

- Esses objetos são devidamente “roteados” pelo STATISTICA de acordo com as configurações do *Output Manager*, o que significa que, dependendo da seleção do usuário, eles podem ser colocados em janelas individuais ou em um workbook;
- Todos e cada um deles podem ser ainda gerenciados usando os amplos recursos da plataforma STATISTICA. Por exemplo, eles podem ser anotados, armazenados em um workbook (comprimidos), exportados como documentos do Microsoft Office, impressos, salvos como um arquivo PDF, convertidos em um de muitos formatos populares, arquivados como versão-controlada, itens “auditáveis” no *STATISTICA Document Management System*, compartilhados com outros usuários em uma rede baseada em cliente-servidor do ambiente *STATISTICA Enterprise Server*, etc.

4 Extensões da Linguagem R: Transmissão de Dados ao R e Recuperação de Resultados

A macro *R Integration Support Macro (R.svb)* implementa várias extensões para a linguagem R – palavras-chave e funções que podem ser usadas dentro de scripts do R executadas dentro do ambiente do STATISTICA. Essas extensões permitem que os scripts passem dados ao R e recuperem os resultados do ambiente do R.

Importante: A linguagem R diferencia maiúsculas e minúsculas, portanto extensões de linguagem do R para o STATISTICA também diferenciam – que somente serão reconhecidos pelo ambiente STATISTICA quando digitados exatamente como mostrado.

Dados tabulados representados no STATISTICA em forma de planilhas são mapeados nas estruturas equivalentes do R – *data frames*. O mapeamento preserva informações, tanto quanto possível para os dois formatos: por exemplo, etiquetas de texto de variáveis (*text labels*) em uma planilha se tornam objetos de fatores em um *data frame*, dados faltantes são mapeados como indicadores *NA*. Tipos de dados e nomes de variáveis e casos são transferidos para os dois lados, etc.

4.1 ActiveDataSet

A palavra-chave `ActiveDataSet` foi adotada a partir da linguagem do STATISTICA Visual Basic e executa a mesma função em scripts R: ela faz referência à planilha de dados ativa do STATISTICA.

No ambiente do STATISTICA desktop, *active data set* normalmente significa a planilha visível que pode atuar como uma fonte de dados (que pode ser também uma planilha em um workbook selecionada como *Active Input*). Esta noção é redefinida e ampliada para ambientes baseados em servidor (*STATISTICA Enterprise Server, Enterprise*), mas a palavra-chave ainda é válida e refere-se ao mapeamento do servidor correspondente da fonte de dados ativa.

Se nenhum conjunto de dados ativo está definido/disponível, o script do R utilizado irá falhar (o mesmo vale para macros do SVB).

Se a macro *R Integration Support Macro* encontra a palavra-chave `ActiveDataSet` no script do R, ela transfere os conjunto de dados ativo do STATISTICA ao ambiente do R e atribui a ele uma variável com o mesmo nome. Portanto, esta palavra-chave representa uma variável *data frame* e pode ser tratada como tal no script.

Exemplo:

```
activeDataSet [1:5] # exibe um subconjunto do conjunto de dados ativo (variáveis de 1 a 5)
```

```
str(ActiveDataSet) # ver a estrutura do data frame
```

```
plot(ActiveDataSet$Measure01) # plota os valores da variável "MEASURE01"
```

4.2 Spreadsheet (FilePathOrName)

`FilePathOrName` – vetor de character (literal ou variável), por exemplo `"c:/data/sample.sta"` ou `"adstudy"`

A extensão `Spreadsheet()` permite carregar um arquivo de dados específico do STATISTICA no R e transferir os dados do arquivo para um *data frame* do R.

Semelhante à palavra-chave `ActiveDataSet`, o valor retornado da função `Spreadsheet()` deve ser tratado como uma variável *data frame* com o conteúdo da planilha do STATISTICA correspondente.

Um recurso útil suportado por esta função é a utilização de *default search paths* para arquivos spreadsheets que estão especificados apenas com o simples nome do arquivos. Isto significa que, se o parâmetro de função consiste apenas de um nome de arquivo, por exemplo, `Spreadsheet ("some.sta")`, o código *R Integration Support* vai olhar para este arquivo em vários locais: primeiro, ele irá verificar a pasta onde o próprio script do R está localizado (se ele foi salvo no disco), e em seguida ele irá verificar a pasta `Examples\Datasets` da atual instalação do STATISTICA. O código de suporte também adiciona a extensão padrão de arquivo `.sta` se não estiver presente. Portanto, as seguintes opções estão disponíveis:

- Scripts do R podem referenciar os conjuntos de dados de acompanhamento (colocados na mesma pasta) simplesmente pelo nome;
- Planilhas que estão incluídas em cada instalação STATISTICA como conjuntos de dados de exemplo podem ser referenciadas pelo nome da mesma forma como conjuntos de dados embutidos dentro no R.

Exemplo:

```
Spreadsheet("c:\myfiles\mydata.sta") # exibe o conjunto de
dados do console R
Spreadsheet("\\server\share\data.sta") # o mesmo, mas lê de
um compartilhamento de rede
Spreadsheet("thisdemo.sta") # arquivo na mesma pasta que este
script
...
# plota um histograma para a variável MEASURE05 do conjunto
de dados de exemplo do STATISTICA Adstudy:
hist(Spreadsheet("Adstudy")$MEASURE05)
...
advert = Spreadsheet("Adstudy")$ADVERT # recupera uma
variável do data frame (fator)
is.factor(advert) # => [1] TRUE
levels(advert) # => [1] "PEPSI" "COKE"
```

4.3 RouteOutput (RObject, [SpreadsheetName], [SpreadsheetHeader])

`RObject` - um data frame, matriz, array, número ou uma string

`SpreadsheetName`, `SpreadsheetHeader` – string literal, por exemplo, "Uma Tabela de Frequência"

Esta extensão transfere os vários tipos de dados do ambiente do R em planilhas do STATISTICA.

Embora a função `t` introduzida para recuperar dados tabulados (como *data frames*, matrizes ou arrays) em planilhas, dados de valor único, como números ou strings podem ser passados também e serão colocados em planilhas de única célula. O `RObject` pode ser uma variável do R ou um valor literal.

O nome dessa extensão, `RouteOutput()`, aponta para a semelhança de seu comportamento para a função equivalente do STATISTICA Visual Basic: as planilhas de resultados "resultados" recriadas pela função no ambiente do STATISTICA se tornam a saída padrão do script/análise do R e acompanha as configurações do *Output Manager* [selecione *Output Manager* no menu *File* no STATISTICA], ou seja, são "encaminhados" quer para janelas individuais ou para um workbook (ou múltiplos workbooks para cada análise, com opções de relatórios de saída, por exemplo, como um documento do Microsoft Word). A configuração mais usual, é um único workbook de resultados.

Os parâmetros opcionais `SpreadsheetName` e `SpreadsheetHeader` especificam o nome e o cabeçalho da spreadsheet resultante. Recomenda-se fornecer um valor para o nome da planilha para a distinção visual na árvore de resultados do workbook.

Observe que gráficos do R transferidos para o STATISTICA como gráficos nativos não requerem roteamento de saída explícito – todos os gráficos gerados durante a execução do script são automaticamente transferidos e roteados de acordo com as configurações do *Output Manager*.

Importante: *Muitas funções no R, especificamente as que realizam a modelagem estatística, representam os seus resultados como objetos estruturados, às vezes de grande complexidade. Estes objetos não podem ser reduzidos a uma única tabela, e, portanto, não podem ser tratados pela extensão RouteOutput() (podem ser automaticamente percorridos em busca de componentes tabulados, mas uma vez que as estruturas de objetos são específicas para um método particular, tal abordagem geraria uma quantidade significativa de "lixo" de saída). No entanto, como os resultados (os dados reais de interesse) são armazenados em objetos como componentes tabulados ou produzidos pela aplicação do método de um objeto de alguns dados de entrada, esta*

limitação não coloca quaisquer problemas – resultados particulares podem ser facilmente extraídos de um objeto de modelo estatístico e encaminhados de volta para o STATISTICA.

Exemplo:

```
# constrói um modelo estatístico (Curvas de Resposta de
Doses)

model <- multdrc(SLOPE ~ DOSE, CURVE, data = PestSci)

...

str(model) # apresenta a estrutura complexa deste
modelo/resultados no console do R

...

# transfere os resultados para o STATISTICA
RouteOutput(coefficients(summary(model)), "Summary",
"Parameter Estimates")

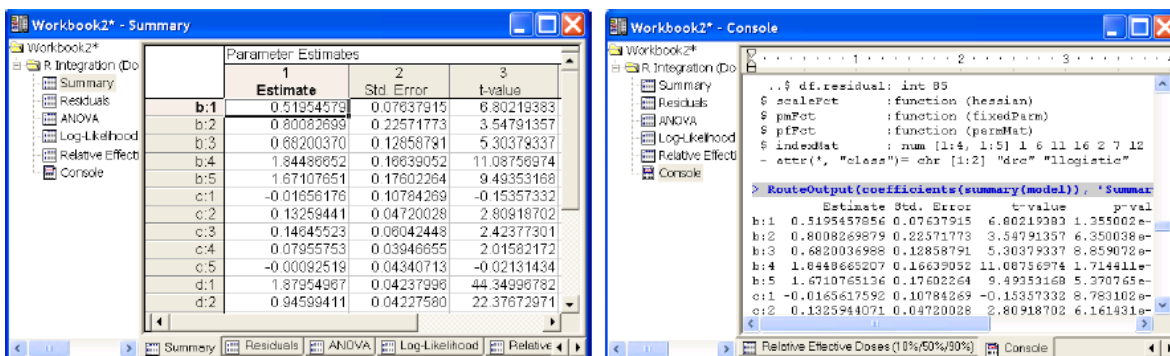
RouteOutput(residuals(model), name = "Residuals", header =
"Some Header")

RouteOutput(anova(model), "ANOVA") # herda o título padrão do
data frame

RouteOutput(logLik(model), "Log-Likelihood")

RouteOutput(ED(model, c(10, 50, 90)), "Relative Effective
Doses (10%/50%/90%)")
```

Os resultados para este exemplo podem parecer com:



4.4 Uses (RPackageName, [AttachImports])

`RPackageName` – string literal, por exemplo `"drc"`

`AttachImports` – valor booleano, por padrão = FALSE

A função `Uses(RPackageName)` pode ser usada para garantir que o respectivo pacote (biblioteca) chamado `RPackageName` e suas *dependências* (pacotes necessários para que este pacote execute) sejam instalados (no computador/servidor onde o script do R está sendo executado) e carregados no ambiente do R automaticamente. Se algumas dessas bibliotecas não estão presentes, elas serão silenciosamente instaladas a partir de repositório CRAN (<http://cran.r-project.org>) e depois carregados.

O R define vários níveis de especificação de dependências de pacotes. Uma delas é `Imports` – que lista os pacotes *with namespaces* que têm de ser carregados para o pacote rodar, mas os namespaces não precisam ser *anexados* ao ambiente atual/global, significando que o pacote acessa tais pacotes importados usando namespaces qualificados, por exemplo, `tseriesChaos::mutual(x)`, em vez de `mutual(x)`. Se você precisa usar algumas dessas funções de namespaces fechadas em seu próprio código R, você pode usar explicitamente qualificação de namespace [`tseriesChaos::mutual(x)`], anexar a biblioteca ao ambiente atual [`Uses("tseriesChaos")` ou `library(tseriesChaos)`], ou definir segundo parâmetro opcional `Uses()` `AttachImports` para TRUE (isso vai anexar todas as importações do pacote).

Observe que esta extensão não é necessária para a interação entre o STATISTICA e o R, e é possível aplicá-la dentro da linguagem R em si. Em vez disso *simplifica o processo de instalação e carregamento da biblioteca condicional, encapsulando-a em uma única chamada*.

Exemplo:

```
Uses("drc") # certifica-se de que o respectivo pacote está
instalado e carregado
...
DR <- multdrc(SLOPE ~ DOSE, CURVE, data = PestSci) # chama os
métodos do pacote
```

Este programa ajusta curvas de resposta de doses para as respectivas variáveis do conjunto de dados construído *PestSci* chamando a função *multdrc* definida no pacote "drc". `Uses("RDC")` assegura que a função está disponível através da instalação e carrega o pacote, se necessário.

5 Executando Scripts do R a partir do STATISTICA Visual Basic

Um caso de uso típico para aproveitar a funcionalidade especializada do R dentro do STATISTICA é chamar os scripts R de dentro de uma macro do STATISTICA Visual Basic (SVB). Dessa forma, podemos construir novos “módulos” usando a biblioteca de interface de usuário SVB e métodos implementados em scripts R. Da mesma forma, essa funcionalidade é necessária para criar configurações de análise do *STATISTICA Enterprise SVB*, ou de nós do *STATISTICA Data Miner* (para STATISTICA ou STATISTICA Enterprise Server) que aproveitam o R.

Mas, a fim de fornecer quaisquer funcionalidades não-triviais em um script do R em casos de uso dessas, precisamos ser capazes de “parametrizar” o script com parâmetros selecionados pelo usuário, listas de variáveis de entrada, planilhas, etc. O STATISTICA fornece uma simples e poderosa forma de passar tais parâmetros para scripts do R.

5.1 Scripts do R são Macros do STATISTICA

Em (quase) todos os casos, o STATISTICA trata os scripts do R da mesma forma que as macros nativas do SVB. Isto se aplica ao *STATISTICA Object Model* também: objetos `Macro` em programas do SVB podem agora representar scripts do R. Portanto, scripts do R podem ser criados, abertos, editados, salvos e executados a partir de scripts do SVB.

Isto por sua vez significa que a funcionalidade do R está disponível nas configurações de análise do *STATISTICA Enterprise* e nos nós do *STATISTICA Data Miner* uma vez que são baseados em SVB.

Arquivos de script existentes do R podem ser abertos com `Macros.Open` ("`path\to\some.r`") ou criados em tempo real com `Macros.New()` e `Macro.Code`. Observe que, em último caso, o STATISTICA precisa de ajuda para distinguir scripts do R distintas de macros do SVB – isso pode ser alcançado tanto especificando o nome de uma nova macro com a extensão `.R` (mesmo se você não está salvando no disco), ou definindo explicitamente `Macro.Scripting` para 5 (tipo de macro do R). Execute os scripts chamando `Macro.Execute`.

Importante: O tipo `Macro.Scripting` para scripts do R é 5 (mais tarde será mapeado por uma constante simbólica).

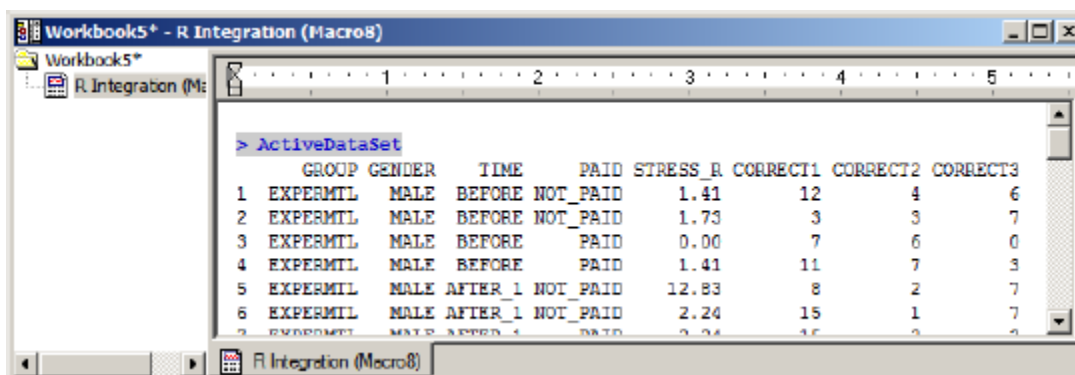
Exemplo:


```

Sub Main
Dim R As New Macro
R.Code = "ActiveDataSet" ' simple R script created on-the-fly
R.Scripting = 5 ' R Macro Type = 5
R.Execute
End Sub

```

Esta macro do *STATISTICA Visual Basic* executa um script do R simples, contendo apenas um único comando *ActiveDataSet* que, como descrito na seção anterior, é uma extensão da linguagem R para o *STATISTICA* que irá transferir (e, neste caso, exibir) o arquivo de dados atualmente ativo do *STATISTICA* no R. Por exemplo, se você executar esta macro após a abertura do arquivo de dados de exemplo *Exp.sta*, uma lista do que há no arquivo será exibida em uma janela de relatório que representa a sessão de console do R:



5.2 Passagem de Parâmetros para Scripts do R: Objeto Collection

Collection é uma biblioteca COM introduzida com a versão 8 MR3 do *STATISTICA* que implementa objetos genéricos de Coleção. Tais objetos permitem aos usuários armazenar palavras-chave marcadas de valores Variant. Variants são estruturas COM que podem armazenar um dos muitos tipos de dados suportados, incluindo arrays e referências a outros objetos COM, por exemplo, spreadsheets.

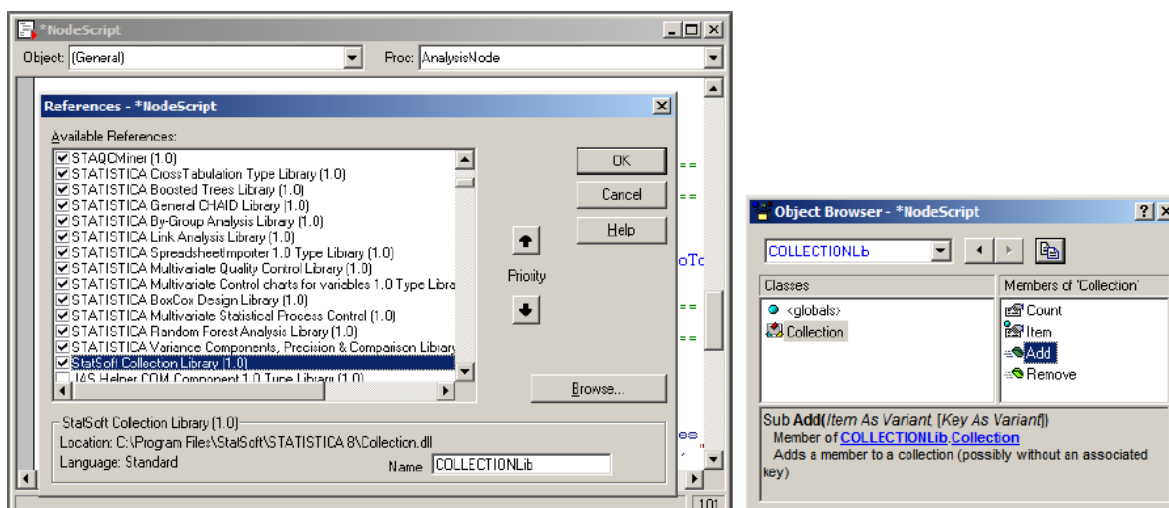
O que isto significa para a Integração do R é que agora podemos armazenar pares de {nome do parâmetro = valor do parâmetro} na coleção, em seguida, passar este parâmetro definido para o *STATISTICA* antes da execução do script do R, alcançando assim a parametrização do script. A macro *R Integration Support* vai transferir todos os parâmetros da Coleção para o meio ambiente do R como variáveis do R nomeadas com

respectivas tags de parâmetros de palavras-chave, permitindo que o script que é executado nesse ambiente faça referência a todas essas variáveis (parâmetros de script) diretamente pelo nome.

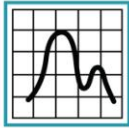
A integração do R suporta os seguintes tipos de parâmetros nomeados: strings, números, arrays e spreadsheets. O suporte para spreadsheets do STATISTICA é particularmente útil: por exemplo, tal parâmetro pode substituir a palavra-chave `ActiveDataSet`, permitindo que um único script sirva como um “motor” do R para um módulo interativo, configuração de análise do *Enterprise* e implementações de nó do *Data Miner*.

Observe também que os parâmetros string sem uma tag de palavra-chave são tratados como código do R que devem ser executados antes da execução do script em si. Isto é análogo ao “código escondido” do SVB e pode ser usado para, por exemplo, definir um conjunto comum de novas funções ou variáveis/constantes globais.

A fim de usar objetos *Collection*, os scripts do SVB devem incluir uma referência à *StatSoft Collection Library* (adicioná-la usando as ferramentas do menu *Tools > References* ao editar a macro):



O objeto *Collection* tem várias propriedades e métodos genéricos que são necessários para manipular o conteúdo de uma coleção: *Count*, *Add(Item, [Key]), [Key]), Remove(KeyOrIndex)* e *Item(KeyOrIndex)* que retorna um objeto *Item* com as propriedades *Key* e *Value*. No entanto, devido ao uso das chamadas propriedades de objeto padrão (*Item* é a propriedade padrão de um *Collection* e retorna sua propriedade *Value* por padrão) a interação com um objeto de coleção é reduzida para limpar intuitivamente operações de atribuição:



Dim param As New Collection

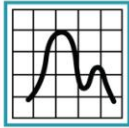
```
param("number") = 57
```

```
param("string") = "A string sample..."
```

Depois de ter reunida a coleção de parâmetros, execute o script do R parametrizado chamando Macro.ExecuteWithArgument (Parameters As Collection).

Exemplo:

```
Dim s1 As New Spreadsheet, s2 As New Spreadsheet ' ...
preencher s1, s2 com dados
var1 = Array("CASE 1", "CASE 2")
var2 = Array(1, 2, 3, 4, 5)
' * não utilize espaços em nomes de parâmetros
' * alguns nomes estão "bloqueados" e não podem ser usados
[por exemplo, 'text', 'str', 'sample']
Dim param As New Collection
param("number") = 57
param("string") = "A STRING sample..."
param("string_array") = var1 ' adiciona itens com um operador
de atribuição
param.Add(var2, "number_array") ' OU usando método Add()
explícito
param("SomeSpreadsheet") = s1
param("ActiveDataSet") = s2 ' substituir o valor de
parâmetros de palavra-chave da string 'ActiveDataSet', sem
chaves associadas será tratado como código R e será executado
antes do script - análogo ao "código oculto" do SVB
' *Definir uma função que estará disponível para o script R
param.Add("func <- function(x) { cat('Called func(x) with x
=', x) }")
' * uma outra forma de definir uma constante ou variável
global
param.Add("STATISTICA.Version = '" & Version & "'")
```



```
' agora executar o script do R com esta coleção de parâmetros  
' (parâmetros tornam-se variáveis do R - O script pode  
referenciá-los pelo nome)
```

```
Dim m As Macro
```

```
m = Macros.Open(MacroDir & "\parameterized.r")
```

```
m.ExecuteWithArgument(param)
```

```
Workbook6* - parameterized.r  
R Script Results  
parameterized.r  
This script has input parameters:  
  
number = 57  
string = A string sample...  
string_array = CASE 1 CASE 2  
number_array = 1 2 3 4 5  
SomeSpreadsheet =  
    First Second  
Case1      5    NA  
Case2      NA   25  
  
=====
```

```
> cat("Interesting fact: ", number, "- 1 =", number - 1)  
Interesting fact: 57 - 1 = 56  
  
> str(string_array) # display variable structure  
chr [1:2] "CASE 1" "CASE 2"  
  
> cat("STATISTICA Version: ", STATISTICA.Version)  
STATISTICA Version: 8.0  
  
> if (exists("func")) func(string)  
Called func(x) with x = A string sample...
```

5.3 Gestão de Resultados de Script

Ambos os métodos de `Macro.Execute()` e `Macro.ExecuteWithArgument()` lidam com saídas de scripts do R (planilhas,

gráficos, relatórios) exatamente da mesma forma como as saídas de macros SVB – serão encaminhadas ao STATISTICA de acordo com as configurações do *Output Manager*, por exemplo, colocado em um workbook.

Esta abordagem pode ser suficiente para muitos casos de utilização, mas não é suficientemente flexível em situações em que a saída de um script do R é tratada como um resultado intermediário para o processamento subsequente no STATISTICA.

Por exemplo, um script do SVB implementando um módulo personalizado interativo pode implementar um fluxo de trabalho de análise de múltiplos estágios, pondo no R, analisando os resultados, pedindo entradas adicionais do usuário, e pôr no R repetidamente com os resultados anteriores como entradas. Ou uma macro do SVB poderia usar resultados tabulares retornados do R para produzir gráficos que são mais expressivos e flexíveis do que o que R é capaz de gerar.

Uma vez que os itens processados pelo *STATISTICA Output Manager* destinam-se a ser uma representação final dos resultados da análise, não é trivial acessá-los individualmente para processamento adicional – seria necessário conhecer o tipo de representação selecionada pelo utilizador (um ou múltiplos workbooks, janelas individuais) e programaticamente percorrer o respectivo conjunto de itens de saída (janelas, conteúdo de workbooks representados como uma árvore), buscando o item de interesse pelo nome.

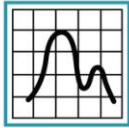
5.3.1 Macro.ExecuteNoRouteOutput([Parameters As Collection]) As StaDocCollection

Este método foi adicionado para acomodar esses casos de uso especial: saída de scripts do R não é “roteada” absolutamente, por isso o script é executado silenciosamente. Toda saída do script é coletada em um objeto *StaDocCollection* (recipiente padrão do STATISTICA para documentos como spreadsheets, gráficos, relatórios ou workbooks) que são retornados como resultados da execução do script. Ele também aceita uma *Collection* opcional para parametrização.

Agora, os desenvolvedores de macro do SVB podem facilmente acessar os componentes individuais de saída de script do R, por exemplo, para extrair dados de células individuais de planilhas ou para criar um gráfico complexo baseado em várias colunas de várias spreadsheets.

Exemplo:

```
Dim m As Macro  
m = Macros.Open(MacroDir & "\some.r")
```



```
Dim Routput As StaDocCollection
Routput = m.ExecuteNoRouteOutput() ' silenciosa, sem saída
exibida
' faz algo com "Routput": extrai e opera em documentos
individuais
' Caso trivial: resultados são encaminhados (exibidos), como
se a partir de Macro.Execute()
RouteOutput(Routput, "R Script Results").Visible = True
```

5.4 Mais Exemplos

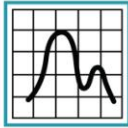
Neste ponto temos demonstrado todos os componentes funcionais necessários para construir aplicações personalizadas dentro da plataforma do STATISTICA que podem tirar proveito da funcionalidade especializado disponível no R.

Todas as instalações do STATISTICA têm agora um conjunto de exemplos que fornecem uma demonstração mais detalhada das características descritas, você vai encontrar esses exemplos na pasta [STATISTICA]\Examples\R. Estes exemplos podem também ser usados como modelos para o seu desenvolvimento.

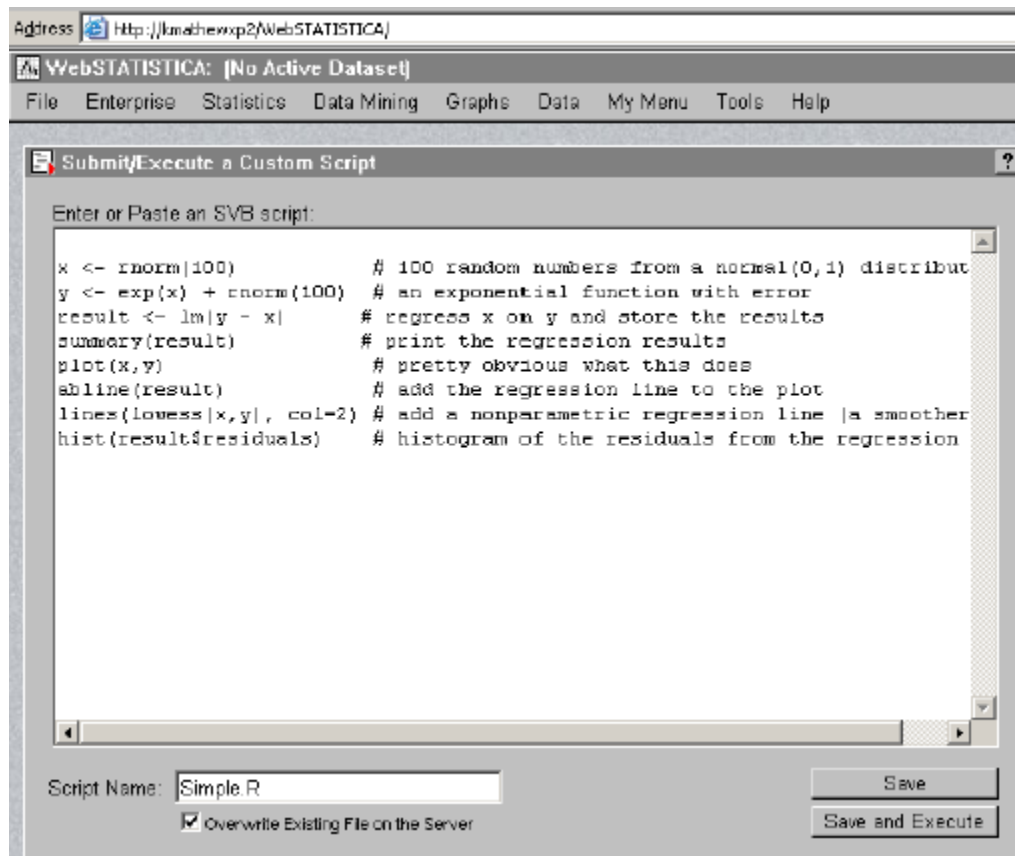
6 Suporte do R no STATISTICA Enterprise Server

O STATISTICA e o *STATISTICA Enterprise Server* são baseados em bibliotecas idênticas do STATISTICA e suportam idêntica funcionalidade. Isto é verdade para o suporte do R no STATISTICA também: você pode executar o R no *STATISTICA Enterprise Server* em grande parte da mesma forma como STATISTICA desktop.

Por exemplo, entre no *STATISTICA Enterprise Server* e selecione *Tools > Submit/Execute a Custom Script*, então digite a macro abaixo (disponível como *standalone.r* na pasta de exemplos exemplos /R).

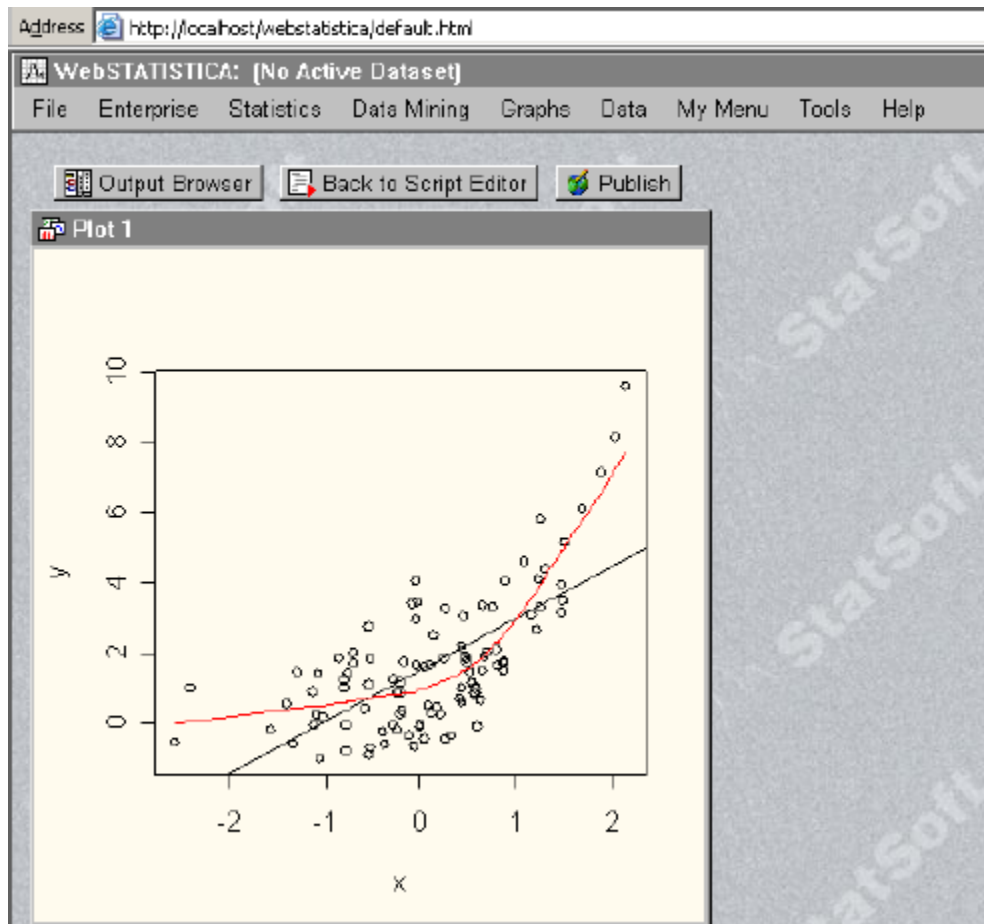
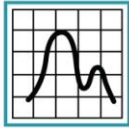


StatSoft[®]
SOUTH AMERICA



Importante: Certifique-se de salvar o script com a extensão *.R*.

Para executar o script, clique em *Save and Executer* ou use o menu *File* para abri-lo e executá-lo.



Note que este script não requer um arquivo de dados. Se você estiver usando extensões `ActiveDataSet` ou `Spreadsheet()`, tenha em mente que as fontes de dados têm de ser localizadas no servidor. `Spreadsheet()` deve referenciar conjuntos de dados das suas pastas de URL do *STATISTICA Enterprise Server*. `ActiveDataSet` será mapeada para a planilha que você abrir no *STATISTICA Enterprise Server*, mas isso pode ser substituído por um parâmetro com o mesmo nome passado para o script do R a partir do SVB. Além disso, a fim de executar todos os scripts no *STATISTICA Enterprise Server*, o usuário deve ter as permissões adequadas no servidor.

6.1 STATISTICA Enterprise Server é um Poderoso R Server

O *STATISTICA Enterprise Server* foi projetado como uma potente plataforma analítica servidor baseado na web, contando com o motor do *STATISTICA Visual Basic* para a diversidade de sua funcionalidade, bem como extensibilidade.

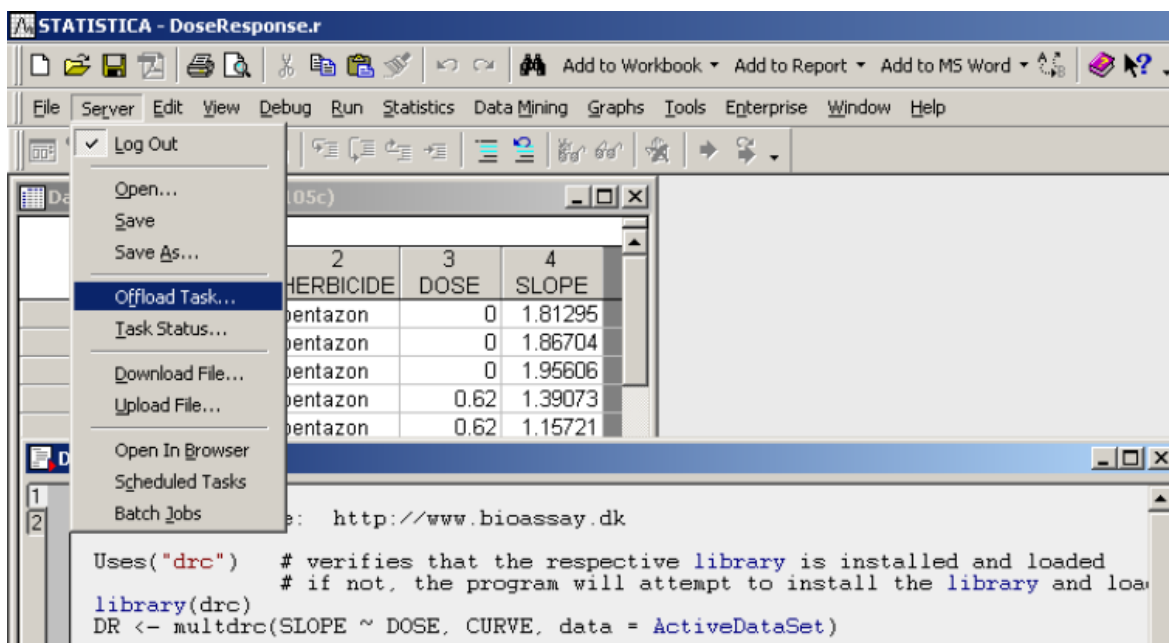
Scripts do R são tratados pelo *STATISTICA Enterprise Server* em grande parte da mesma forma como as macros padrão do SVB. Os componentes de terceiros que a plataforma *STATISTICA* depende para fornecer o seu ambiente de execução do R (como biblioteca R e biblioteca R COM Server) também são adequadas para o tratamento de várias sessões simultâneas do R.

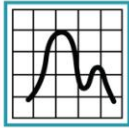
Assim, o *STATISTICA Enterprise Server* representa uma plataforma ideal para um poderoso e flexível servidor de R multi-processador que pode lidar com um grande número de usuários, fornecendo programação, segurança, balanceamento de carga, etc.

6.2 Carregamento-off de Scripts a partir do *STATISTICA* para o *STATISTICA Enterprise Server*

Como os scripts do SVB ou workspaces do Data Miner, scripts do R podem ser descarregados a partir de um ambiente de desktop para o processamento no *STATISTICA Enterprise Server*. Isto é particularmente útil quando um script do R deverá exigir recursos computacionais significativos – os usuários podem enviar tais scripts para o servidor (por exemplo, durante a noite) e recuperar os resultados em um momento posterior.

Se a sua instalação está configurada para trabalhar com o *STATISTICA Enterprise Server* (*Tools > Options: Server/Web*), você pode descarregar qualquer “tarefa” do R selecionando *Server > Offload Task...*:





Opções são fornecidas para também transferir o arquivo de dados em paralelo ao servidor, estas opções são descritas em detalhes em *STATISTICA documentation*.

Offload a task [?] [X]

Task

Current script or Data Miner project

Select a macro or a Data Miner project ...

This task does not require a data source

Data Source

Current spreadsheet or embedded dataset

Select data file stored on the server ...

Specify database connection Database...

Open Task Status dialog after offloading

Task Name (optional)

DoseResponse.r

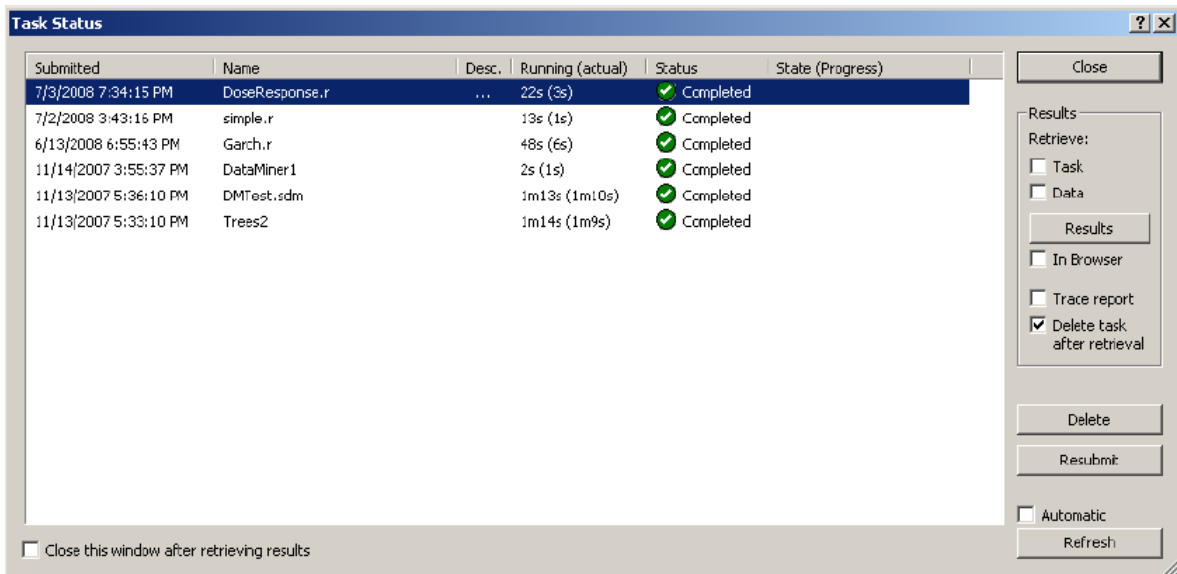
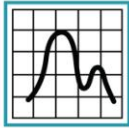
Task Description (optional)

OK

Cancel

Após clicar em *OK*, o respectivo script do R será executado no *STATISTICA Enterprise Server* (ou será programado para ser executado, dependendo da carga do servidor).


O progresso da análise pode ser monitorado através da janela *Server > Task Status...*

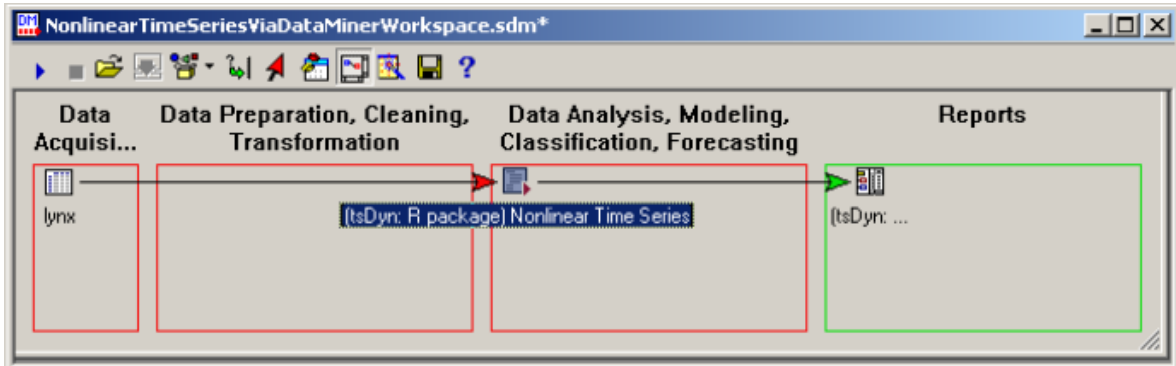
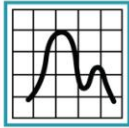


Os resultados podem ser recuperados do servidor através do botão *Results* (ou com um duplo-clique sobre a tarefa), a representação dos resultados de uma tarefa descarregada é equivalente à mesma tarefa rodando localmente.

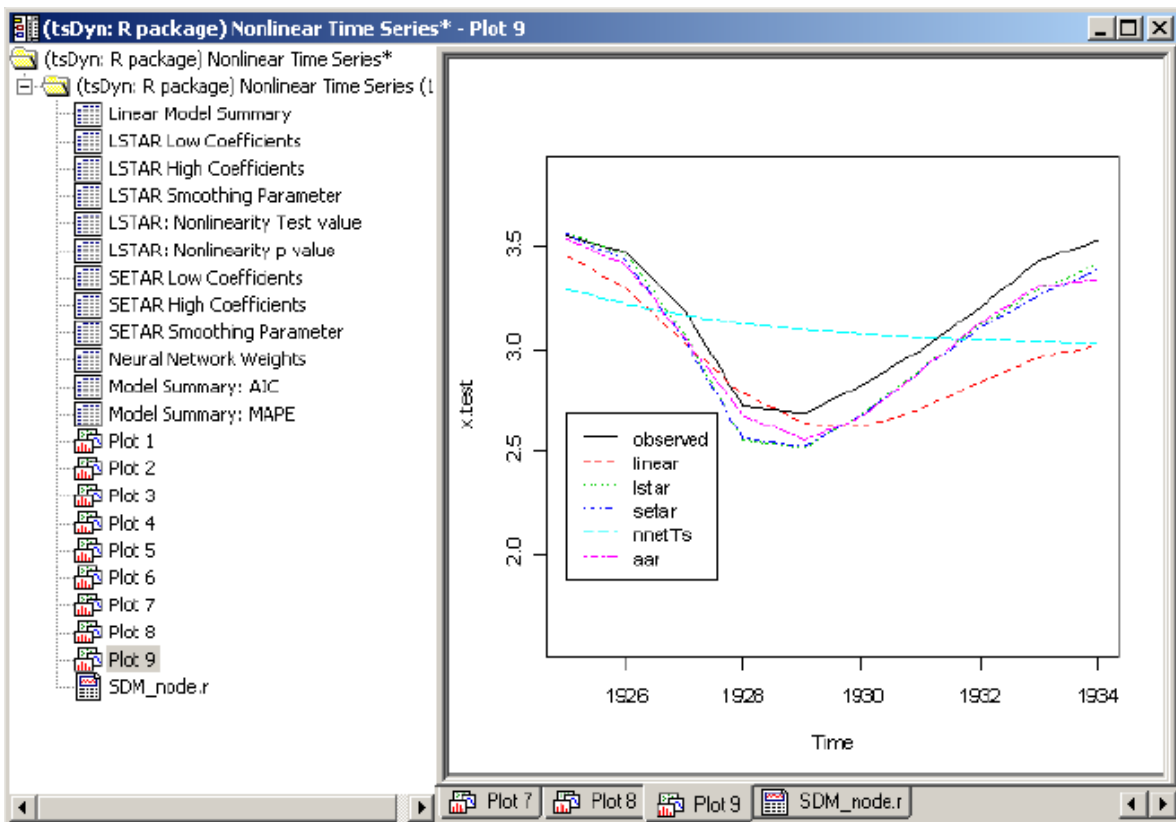
7 Criação de Nós do STATISTICA Data Miner Baseados no R

Podemos usar as técnicas descritas em seções anteriores para implementar um nó do Data Miner baseado no R: Edite a fonte do nó do SVB para gerar uma *Collection* com base em parâmetros do nó de entrada, use-a para executar um ou mais scripts parametrizados do R, e recupere o resultados como uma *StaDataCollection* que se torna a saída do nó.

Para um exemplo, veja `[STATISTICA]\Examples\R\NonLinear Time Series\As Data Miner Node.sdm`: Selecione o nó de análise e clique no ícone  *Edit Code* para rever o código SVB subjacente (`NonLinearTimeSeries.svx`). observe como o arquivo `.dmi` correspondente descreve os parâmetros do nó de entrada. O arquivo de script do R (`NonLinearTimeSeries.r`) realiza análise de séries temporais não-lineares usando o pacote do `tsDyn` do R. Este script pode servir como um modelo para os seus próprios nós do Data Miner baseados no R.



O nó de resultados no grupo *Reports* é um workbook.



Este exemplo demonstra que a capacidade de facilmente incorporar a funcionalidade especializada do R no *STATISTICA Data Miner* oferece oportunidades muito poderosas para criar fluxos de trabalho analíticos ou coleções de nós do Data Miner e configurações do *node-browser* dedicadas a análises altamente especializadas (como séries temporais não-lineares, ajuste curva de resposta de dose e assim por diante).

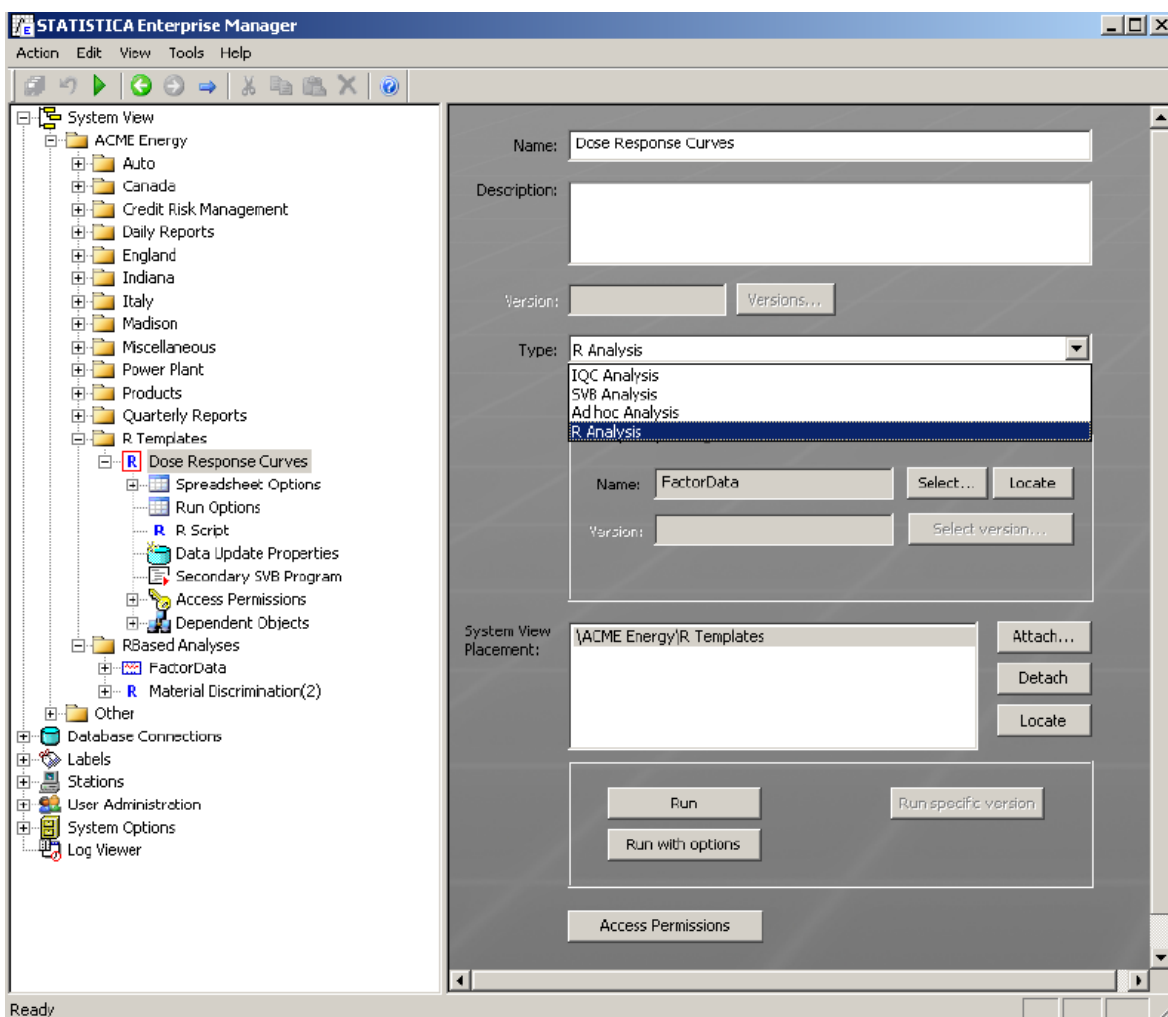
8 Integração do R no STATISTICA Enterprise e no STATISTICA Enterprise Server

A integração do R no STATISTICA seria incompleta sem o suporte para o *STATISTICA Enterprise* e soluções do *STATISTICA Enterprise Server*: Agora programas do R podem ser utilizados dentro ou como modelos de análise. Isso significa que análises do R podem ser distribuídas para usuários finais que não estão familiarizados com o VB ou com a programação do R. Além disso, pelo fato da plataforma *Enterprise* oferecer inúmeras opções e as recursos para especificamente permitir a aplicação de análises de *template* para aplicações validadas, tais como controle de versão e caminhos de auditoria, essas características tornam o R disponível para implantação em aplicações de manufatura de missão crítica.

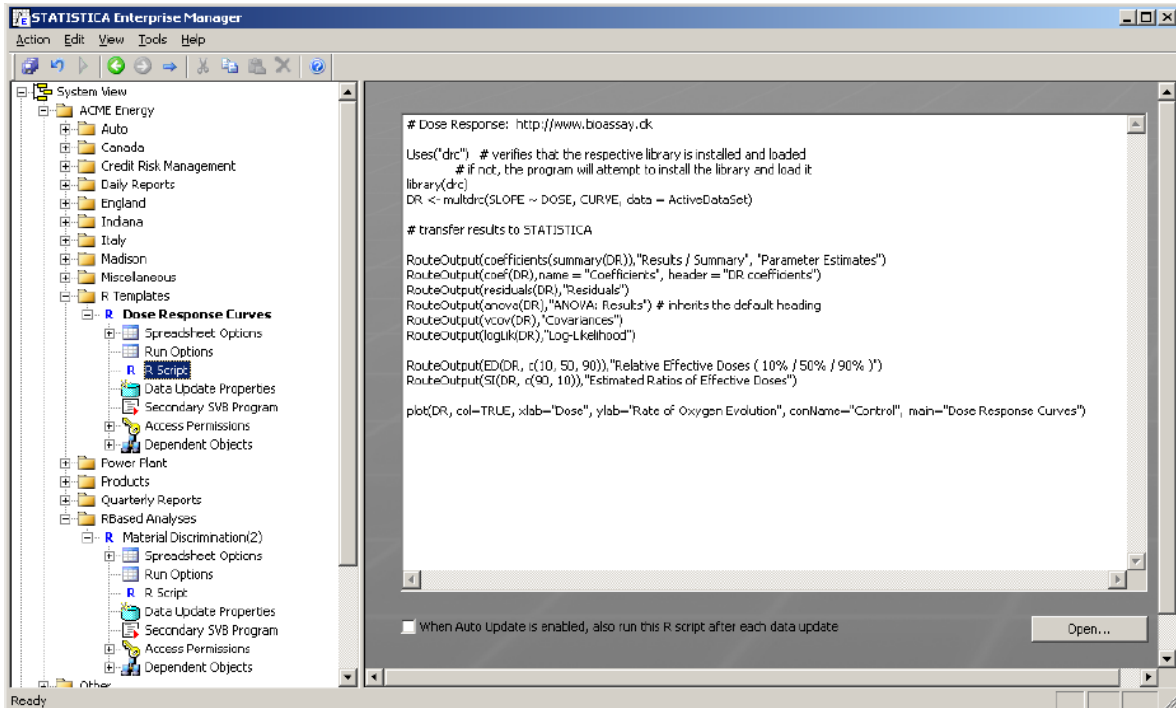
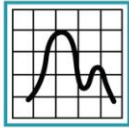
8.1 Criação de Configurações de Análise Baseadas no R

O sistema *STATISTICA Enterprise* implementa gráficos padrão de controle e análises SPC out-of-the-box. Modelos de análise não-triviais são usualmente criados usando o *STATISTICA Visual Basic*. Além disso, o *STATISTICA Enterprise* e o *STATISTICA Enterprise Server* agora reconhecem os scripts do R, e podem armazená-los e processá-los da mesma forma que os scripts do SVB.

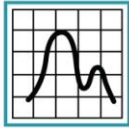
Ao criar uma nova *Analysis Configuration* no *STATISTICA Enterprise*, selecione a opção *R Analysis*:



Então, digite, cole ou carregue o respectivo script do R. Todas as outras opções são idênticas às disponíveis para os modelos baseados em análise do SVB:



Dependendo da configuração do sistema *STATISTICA Enterprise*, este modelo de análise reutilizável já está disponível para os usuários de ambos ambientes *STATISTICA* e *STATISTICA Enterprise Server*, e os resultados deste modelo podem ser combinados em relatórios padrão.



Address http://localhost/webstatistica/default.html

WebSTATISTICA: [No Active Dataset]

File Enterprise Statistics Data Mining Graphs Data My Menu Tools Help

Analysis or Report Selection

- ACME Energy
 - Indiana
 - Madison
 - Canada
 - Italy
 - England
 - Daily Reports
 - Quarterly Reports
 - Miscellaneous
 - Auto
 - Power Plant
 - Credit Risk Management
 - R Templates
 - RBased Analyses

name
Dose Response Curves

Display causes and actions for all samples (if applicable).
 Make resulting data set the active Data Source

Component Manager Explore Data Run Analysis or Report

Além disso, se a respectiva instalação do *STATISTICA Enterprise* está integrada com o *STATISTICA Document Management System*, então estes scripts do R podem ser “bloqueados”.

8.2 Chamar Scripts do R a partir das Configurações de Análise do SVB

Da mesma forma, a funcionalidade do R também pode ser utilizada a partir de dentro de modelos de análise SVB, que podem obter os resultados a partir do R para posterior processamento ou exibição. Neste caso, os scripts do R que são carregados para

as respectivas análises de template no *STATISTICA Enterprise* podem ser colocados em uma pasta de repositório segura a partir da qual eles podem ser carregados (e onde eles podem ser gerenciados através do *STATISTICA Document Management System* integrado com o *STATISTICA Enterprise*) ou eles podem ser incorporados no código SVB (atribuir o script do R à propriedade *Code* da macro SVB).

8.3 Resumo

A integração total do ambiente R no *STATISTICA Enterprise* permite que as organizações utilizem funcionalidade ampla e altamente especializada do R e alavanquem dentro de um sistema de análise de segurança toda a empresa, com base na função que isola os usuários finais a partir de detalhes de implementação e que pode ser implantado em validados ambientes de produção. Em conjunto com a capacidade de executar templates baseados no R na plataforma do *STATISTICA Enterprise Server* e extrair resultados em relatórios padrão pré-formatados, esses recursos fazem do *STATISTICA Enterprise*, sem dúvida o sistema de análise enterprise mais poderoso disponível até o momento.

9 Comentários Finais e Algumas Ressalvas

Os recursos fornecidos no *STATISTICA* para integração com R são bastante flexíveis e tornam as milhares de funções altamente especializadas do R e recursos disponíveis para todas as soluções do *STATISTICA*. No entanto, os usuários que estão planejando explorar esses recursos são aconselhados considerar os seguintes aspectos possíveis, em particular, certas limitações do sistema da plataforma R.

9.1 Tratamento de Erros

A maior parte das condições de erro geradas dentro do ambiente R (por exemplo, erros de sintaxe e de execução causados por um script do R) ou pelas bibliotecas de integração de apoio (por exemplo, instalação R quebrada ou falta de componentes) são interceptadas e manipuladas pelo *STATISTICA*. Os desenvolvedores podem utilizar as facilidades de manipulação de erros disponíveis em qualquer ambiente, por exemplo, em manipuladores *On Error* nas macros do SVB chamando scripts do R.

No entanto, ocasionalmente programas do R podem falhar ou travar (no caso, o controle do programa não retorna ao *STATISTICA*). Portanto, uma validação cuidadosa dos respectivos scripts do R é crucial para empresa em nível de implantação de modelos de análise do R.

9.2 Pontos Fortes e Limitações

Uma palavra de cautela em relação à qualidade de algoritmos do R: O R vem sem garantia ou garantias. Na prática, muitos (a maioria) dos algoritmos disponíveis em R são o resultado de um trabalho diligente longo de muitos anos por um ou alguns indivíduos que são especialistas na respectiva metodologia ou área. No entanto, isso não significa que o software foi criado seguindo rigorosa metodologia de desenvolvimento de software de ciclo de vida, ou rigorosos procedimentos operacionais padrão para a coleta de requisitos de software, projeto, implementação e testes. Portanto, a fim de construir uma aplicação de missão crítica ou validada em torno de um componente que depende do R, é absolutamente importante que você valide cuidadosamente todos os resultados para os casos de uso para o qual o software será aplicado.

Uma palavra de cautela em relação a escalabilidade, grandes conjuntos de dados, etc: Outra ressalva diz respeito ao R que precisa ser considerada antes de construir soluções em torno do R diz respeito à sua arquitetura básica. Ao contrário do STATISTICA, dados no R devem ser (em praticamente todos os casos) residentes na memória do computador. Esta restrição, em combinação com o nível de hardware e limitações operacionais de nível de sistema de memória, pode ou não representar um obstáculo para qualquer usuário individual, mas terá de ser analisado com cuidado ao criar aplicativos de servidor baseados no R acessível a vários usuários.