

# R e software SPSS: Vantagens para todos



## Sumário

- 2 Visão geral
- 3 O que é R?
- 3 As limitações do R
- 4 Benefícios da integração do R com o software SPSS
- 7 Conclusão
- 7 Sobre Business Analytics
- 7 Solicitar uma ligação

## Visão geral

A finalidade deste documento é demonstrar os benefícios do uso da linguagem de programação R com o software IBM® SPSS® Statistics e Modeler, em vez de tentar fazer tudo apenas com o R. Com o software SPSS, usuários de R têm acesso a um gerenciamento de dados superior, uma interface de apontar e clicar, um resultado com qualidade de apresentação e possibilidade de escalabilidade aprimorada. Usuários do software SPSS têm acesso a uma coleção farta e em expansão de análises estatísticas e bibliotecas de gráficos para ajudar a obter insights mais profundos a partir de seus dados.

O uso conjunto do software SPSS e R aproveita o melhor dos dois mundos. O software SPSS é capaz de executar sintaxe R a partir da sua interface SPSS. É possível ampliar a poderosa manipulação de dados, a análise estatística e os algoritmos de previsão que já existem no software SPSS com código R personalizado para fornecer mais poder e flexibilidade. É possível realizar análises personalizadas, criar e trabalhar com resultados e integrar novos insights aos seus planos de análise. Além disso, com o SPSS Custom Dialog Builder, é possível compartilhar e reutilizar código R com pessoas que se beneficiariam com novas opções de análise, mas que optam por não usar programação para análise. Os desenvolvedores podem se concentrar na criação de código, enquanto os usuários se concentram na análise e usam a funcionalidade baseada em R sem terem que aprender R.

## O que é R?

R é uma linguagem de programação e ambiente de software de código aberto para computação estatística e gráficos ([www.r-project.org](http://www.r-project.org)). A linguagem R tem se tornado muito popular com estatísticos e data miners que a usam para desenvolver softwares estatísticos. Além disso, R é amplamente usado para análises de dados avançadas. R fornece uma ampla variedade de técnicas estatísticas e gráficas, como modelagem linear e não linear, testes estatísticos clássicos, análise de série temporal, classificação e agrupamento. R está disponível de acordo com os termos da [Free Software Foundation](#) e da [GNU General Public License](#).

R tem mais de 4800 pacotes disponíveis em vários repositórios, especializados em métricas econômicas, *data mining*, análise espacial, bioinformática e vários outros. R foi analisado por estatísticos de renome internacional e cientistas da computação. Entretanto, como ele é de código aberto, lançamentos e pacotes não têm processos codificados.

## As limitações do R

Como R está disponível sem qualquer tipo de cobrança, uma percepção comum é a de que ele pode ser usado no lugar de software comercial de estatística e modelagem para economizar gastos a uma organização. Essa percepção é um tanto equivocada. Apesar de seus consideráveis méritos, R não tem como oferecer tudo de que você precisa para obter o máximo de valor comercial de suas análises. As seguintes limitações devem ser considerados pelos usuários de R:

- *Implementação*. Usar R para integrar resultados preditivos a um ambiente operacional pode ser difícil.
- *Interface*. R não tem uma interface de usuário gráfica moderna, o que torna difícil seu uso para quem não é programador de R.
- *Curva de aprendizado*. R não é de fácil aprendizado para todos. Nem todos são programadores.
- *Dados*. R não se conecta facilmente a bancos de dados de forma nativa.
- *Resultados*. A geração de resultados prontos para publicação é difícil.
- *Desempenho*. R pode consumir muito rapidamente toda a memória disponível.
- *Colaboração*. R dificulta o compartilhamento de trabalho entre uma equipe de analistas, especialmente quando os integrantes não têm o mesmo nível de conhecimento de R.
- *Segurança corporativa*. A segurança dos pacotes que você baixa não é garantida.

Além disso, os usuários precisam se certificar de terem os pacotes R e o código corretos para unir os pacotes e os resultados. Essas limitações costumam adicionar custos ao uso de R.

## Benefícios da integração do R com software SPSS

Faz sentido abordar as limitações de R usando a linguagem dentro do software SPSS. O poder combinado de ambos atende às necessidades de uma organização que tenha poucos especialistas e queira se beneficiar de R sem uma acentuada curva de aprendizado. O software SPSS é uma plataforma prática a partir da qual os usuários de R podem lidar com grandes conjuntos de dados e obter gráficos e outras formas de resultado de alta qualidade. Alguns dos benefícios são a facilidade de uso do software SPSS e a capacidade de distribuir pacotes R integrados a uma variedade de usuários que não têm familiaridade com R. Acima de tudo, quando você usa R a partir do software SPSS, as limitações de R podem ser trabalhadas.

### Implementação

O software SPSS é capaz de lidar, com boa relação custo-benefício, com as limitações de implementação de R. Por exemplo, o IBM SPSS Modeler Gold permite que os usuários otimizem a maneira como regras corporativas e modelos preditivos afetam vendas, atendimento ao cliente e outros. O SPSS Modeler Gold é compatível com o uso de modelos que incorporam R. O SPSS Modeler também simplifica o acesso à inteligência corporativa usando integração com o IBM Cognos® TM1®. A IBM é capaz de fornecer aos usuários diversas maneiras de usar inteligência preditiva. Os usuários podem integrar o *insight* preditivo em seus painéis interativos móveis juntamente com visualizações de dados históricos e em tempo real sem a necessidade de criar ou comprar software adicional.

### Interface

O SPSS Modeler e o SPSS Statistics fornecem uma interface de usuário gráfica simples que é compatível com uma variedade de algoritmos de preparação de dados, análise estatística e modelagem preditiva.

O código R é executado na mesma interface, juntamente com todas as outras funções e recursos que já são fornecidos. Os usuários também podem adicionar um *front-end* de GUI ao R. Com esse *front-end*, usuários que não sejam programadores podem incluir valores do cliente (Figura 1), como nomes de variáveis e outros. Eles são capazes de aproveitar a análise avançada sem ter que entender o código subjacente.

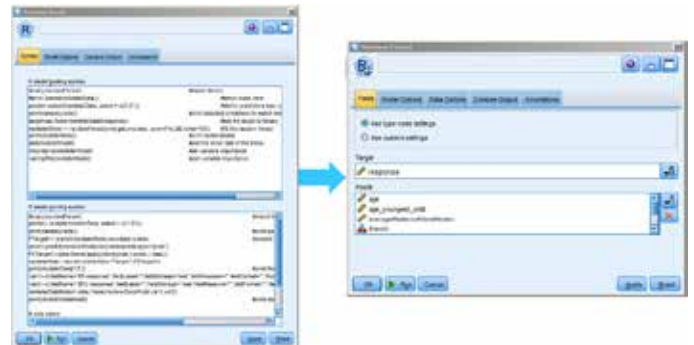


Figura 1: O código R pode ser usado para criar caixas de diálogo para o cliente, facilitando que não programadores usem o código durante a análise.

### Curva de aprendizado

R é uma linguagem de programação e, portanto, é algo que precisa ser aprendido. No entanto, aprender R não é uma tarefa fácil. E, enquanto você estiver usando seu tempo para aprender R, não será capaz de aproveitar todas as capacidades dele. Contudo, quando R é integrado ao software SPSS, você pode explorar seus benefícios de maneira quase imediata. Você pode focar seu aprendizado nas rotinas de R que sejam verdadeiramente singulares e usar os dados, análises estatísticas e modelagem que já estão inclusos no software SPSS.

### Dados

Com R, acessar os dados necessários à análise exige bastante tempo e esforço. É necessário gravar volumes de código, implementar pacotes e até mesmo utilizar Java. SPSS Statistics e SPSS Modeler removem o problema do acesso aos dados. Ambos podem ser usados com bancos de dados comerciais SQL, Oracle, SAP, IBM Netezza®, DB2® e outros. SPSS Statistics e SPSS Modeler são capazes de ler entrada em texto, planilhas, arquivos SAS e outros formatos. O SPSS Modeler é capaz de ler diretamente do IBM Cognos Business Intelligence e Cognos TM11. Assistentes com conectores pré-integrados acessam os dados, removendo o excessivamente demorado fardo de extrair, transformar e manipular dados antes da análise. SPSS Statistics e SPSS Modeler fornecem poderosas técnicas de manipulação de dados, encapsuladas em interfaces de apontar e clicar. Os usuários podem transpor, verificar e reformatar dados. O SPSS Modeler também fornece preparação automática de dados, que otimiza dados para modelagem preditiva com um único clique.

### Resultados

SPSS Statistics e Modeler incluem diversos meios de gerar tabelas e gráficos prontos para apresentação. O software SPSS Statistics permite que programadores de R envolvam funções de R em sintaxe do software SPSS. Por causa disso, é possível gerar gráficos, tabelas dinâmicas e outras formas de resultado com qualidade de apresentação (Figura 2). Os usuários também podem publicar facilmente seus resultados em formatos populares, como PDF, Word, PowerPoint, Excel e outros. Além disso, é possível integrar R às capacidades avançadas do SPSS Modeler, como análises de entidade, análises de rede social, análises de texto e outros.

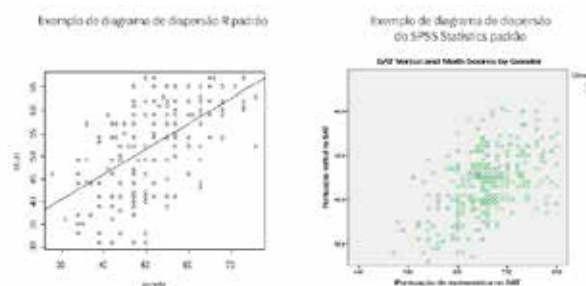


Figura 2: SPSS gera tabelas e gráficos prontos para apresentação, poupando o tempo que o público corporativo costuma gastar elaborando e formatando resultados analíticos.

## Desempenho

Os comandos de R operam em um espaço de trabalho interno à memória.

Os comandos de R não são projetados para levar a memória em consideração e, durante uma sessão de R, todos os objetos são armazenados em uma memória funcional temporária. Por causa disso, a memória disponível pode ser ocupada de forma bem rápida. A combinação de R com o SPSS Modeler permite particionar ou amostrar os dados passados para R (Figura 3). Além disso, o SPSS Modeler Server é uma tecnologia de exploração de memória que pode transferir a análise para o disco, para que a memória permaneça disponível. É possível usar comandos de R e criar objetos de R sem um grande impacto sobre o desempenho geral de sua extração de dados e modelagem. Além disso, o SPSS Modeler pode dimensionar seu R dentro do banco de dados para ambientes IBM Netezza e SAP Hana, entre outros. E, com o SPSS Analytic Server, ele pode dimensionar R também em Hadoop.

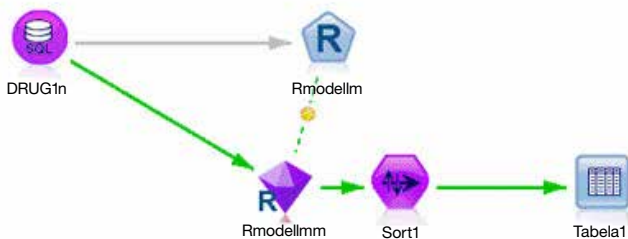


Figura 3: O IBM SPSS Modeler é usado para acessar nativamente dados em um banco de dados SQL com um modelo R usado para fazer a pontuação dos registros. O roxo indica que a etapa analítica está usando SQL *pushback*, que aprimora o desempenho ao executar tarefas que utilizam muita memória dentro do próprio banco de dados.

## Colaboração

R é uma linguagem projetada para o indivíduo. No entanto, a maior parte do trabalho de análise é um esforço colaborativo, com diversas pessoas contribuindo com modelos e análises estatísticas.

Ao usar o software SPSS juntamente com R, perde-se o aspecto de lobo solitário da linguagem e ganham-se os recursos de colaboração de nível internacional da IBM. O software SPSS fornece uma estrutura completa para centralizar, proteger e automatizar ativos analíticos desenvolvidos com SPSS Statistics e SPSS Modeler.

Isso garante que modelos e análises estatísticas desenvolvidos possam ser compartilhados com segurança e que a governança corporativa possa ser aplicada.

## Segurança corporativa

R não tem um processo de lançamento formal, o que significa que a garantia de qualidade formal não faz parte dele. Sem essa garantia, você usa R por sua própria conta e risco. Não há como ter certeza de que a versão ou o pacote fornecerá as funções que afirma fornecer. Além disso, pacotes de R são baixados da Internet e de comunidades de usuários; por isso, a segurança dos pacotes pode ser questionável. Na realidade, sem que você saiba, um download pode introduzir *malware*, um vírus do tipo Trojan, vazamento de dados e outros. SPSS Modeler e Statistics, por outro lado, são testados rigorosamente como parte do processo de garantia de qualidade de software da IBM. Como o software vem da IBM, não é necessário usar práticas arriscadas que podem pôr em risco a segurança do seu ambiente.

## Conclusão

Independentemente, tanto SPSS quanto R têm pontos fortes que foram testados ao longo do tempo e são fortemente aceitos na comunidade analítica. Além disso, esses pontos fortes se complementam para criar um conjunto de funções e recursos ainda mais poderoso, que beneficia a comunidade analítica em geral. Usuários de R podem acessar capacidades superiores de gerenciamento de dados, permitindo que lidem com conjuntos de dados muito maiores. Além disso, o software SPSS fornece aos usuários de R um conjunto mais farto de opções de resultados gráficos e tabelas dinâmicas, o que pode levar a uma melhor experiência do usuário. Por fim, o software SPSS atua como o veículo de implementação ideal para distribuir pacotes de R a uma variedade maior de usuários.

Os usuários do software SPSS têm acesso a muitas outras funções estatísticas. Eles podem realizar análises complexas sem o incômodo de aprender uma linguagem de programação complexa como R. Vale a pena pensar nas vantagens do uso conjunto de R e do software SPSS.

## Sobre o Business Analytics

O software IBM Business Analytics fornece informações baseadas em dados que ajudam as organizações a trabalharem de maneira mais inteligente e superarem a concorrência. Esse portfólio abrangente inclui soluções para inteligência comercial, *analytics* preditiva e gerenciamento de decisões, de desempenho e de riscos. As soluções Business Analytics permitem que empresas identifiquem e visualizem tendências e padrões em áreas como *analytics* de clientes, as quais podem ter um efeito profundo no desempenho comercial. Elas podem comparar cenários; antever possíveis ameaças e oportunidades; planejar, fazer orçamentos e prever da melhor maneira os recursos; equilibrar os riscos contra devoluções inesperadas; e trabalhar no sentido de atender aos requisitos regulatórios. Ao disponibilizarem *analytics* amplamente, as organizações podem alinhar tomadas de decisões táticas e estratégicas para alcançar metas comerciais. Para obter mais informações, acesse [ibm.com/business-analytics](http://ibm.com/business-analytics).

## Solicitar uma ligação

Para solicitar uma ligação ou fazer uma pergunta, acesse [ibm.com/business-analytics/contactus](http://ibm.com/business-analytics/contactus). Um representante da IBM responderá à sua consulta no prazo de dois dias úteis.



---

© Copyright IBM Corporation 2014

IBM Corporation  
Software Group  
Route 100  
Somers, NY 10589

Produzido nos Estados Unidos da América  
Março de 2014

IBM, o logotipo IBM, **ibm.com**, Cognos, DB2 e SPSS são marcas comerciais da International Business Machines Corp., registradas em muitas jurisdições em todo o mundo. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atual das marcas comerciais da IBM está disponível na *web* em “Copyright and trademark information” (Informações de *copyright* e marcas comerciais), no *link* [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Netezza é uma marca registrada da IBM International Group B.V., uma empresa da IBM.

Linux é marca comercial registrada de Linus Torvalds nos Estados Unidos e/ou em outros países.

Microsoft, Windows, Windows NT e o logotipo do Windows são marcas comerciais da Microsoft Corporation nos Estados Unidos e/ou em outros países.

Java e todas as marcas comerciais baseadas em Java, bem como os respectivos logotipos, são marcas comerciais ou registradas da Oracle e/ou suas afiliadas.

UNIX é uma marca comercial registrada do The Open Group nos Estados Unidos e em outros países.

Este documento encontra-se atualizado na data inicial de sua publicação e pode ser alterado pela IBM a qualquer momento. Nem todas as ofertas estão disponíveis em todos os países em que a IBM opera.

AS INFORMAÇÕES CONTIDAS NESTE DOCUMENTO SÃO FORNECIDAS “NA FORMA EM QUE SE ENCONTRAM”, SEM QUALQUER GARANTIA, EXPRESSA OU IMPLÍCITA, INCLUINDO NENHUMA GARANTIA DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UMA DETERMINADA FINALIDADE E NENHUMA GARANTIA OU CONDIÇÃO DE NÃO-VIOLAÇÃO. Os produtos da IBM são garantidos de acordo com os termos e condições dos acordos sob os quais eles são fornecidos.



Recycle