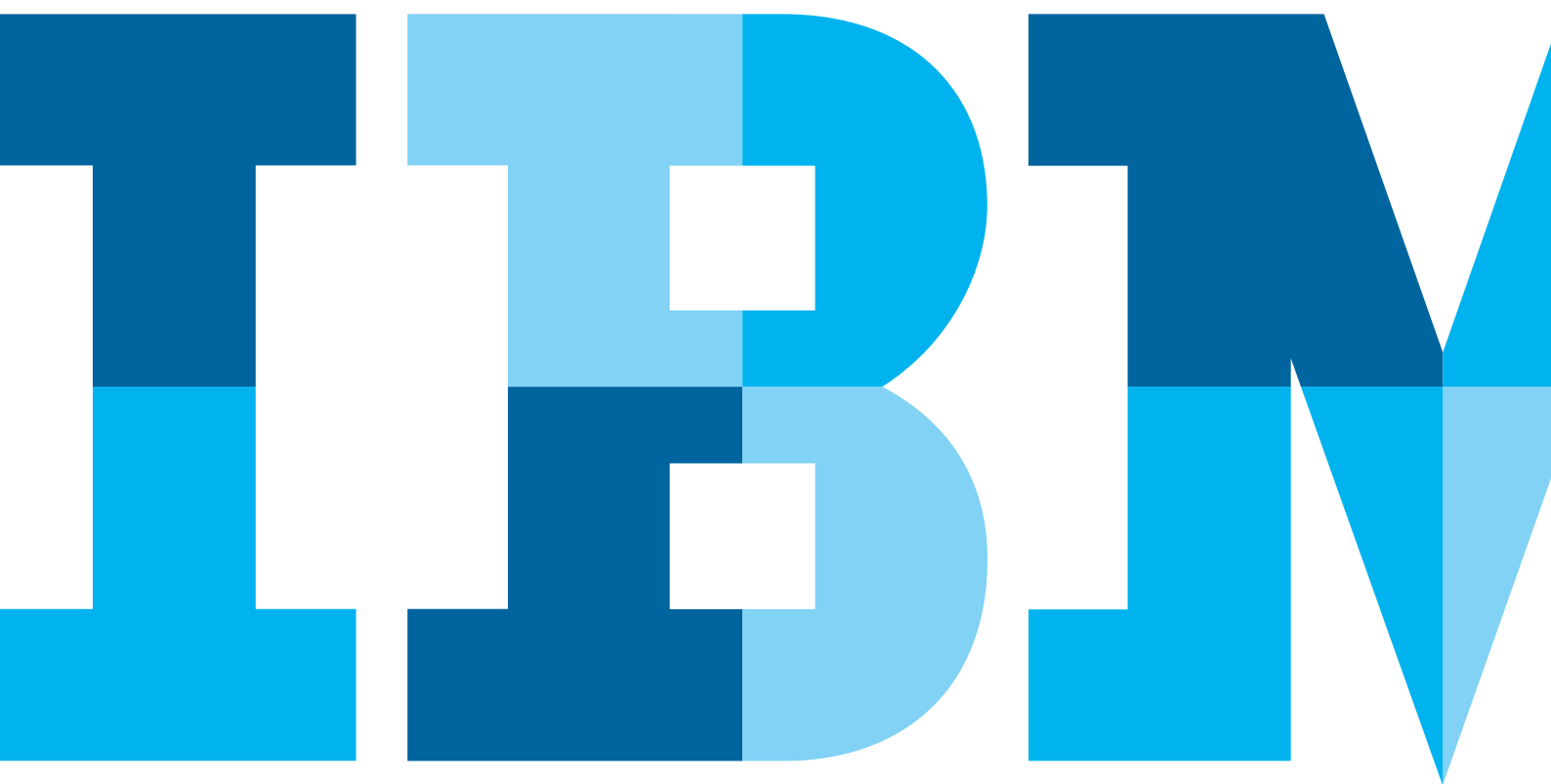


R und SPSS-Software: Eine Win-win-Situation



Inhalt

- 2 Überblick
- 3 Was ist R?
- 3 Die Grenzen von R
- 4 Vorteile der Integration von R in SPSS-Software
- 7 Fazit
- 7 Informationen zu IBM Business Analytics
- 7 Rückruf und Fragen

Überblick

In diesem Dokument werden die Vorteile beschrieben, die sich ergeben, wenn die Programmiersprache R nicht allein, sondern in Kombination mit IBM SPSS Statistics- und Modeler-Software genutzt wird. Mit SPSS-Software erhalten R-Benutzer leistungsfähiges Datenmanagement, eine Point-and-click-Schnittstelle, Ausgaben in Präsentationsqualität und höhere Skalierbarkeit. SPSS-Softwarebenutzer können zudem auf eine umfangreiche, stetig wachsende Sammlung an Bibliotheken für statistische Analyse und Grafiken zugreifen, um aus ihren Daten aussagekräftige Informationen zu ziehen.

Wenn SPSS-Software und R zusammen verwendet werden, profitieren Sie von beiden Welten. SPSS-Software kann R-Syntax aus Ihrer SPSS-Schnittstelle ausführen. Mit angepasstem R-Code, der zusätzliche Leistung und Flexibilität verspricht, können Sie die leistungsfähige Datenbearbeitung, statistische Analyse und Vorhersagealgorithmen der SPSS-Software noch erweitern. Sie können benutzerdefinierte Analysen durchführen, Ausgaben erstellen und bearbeiten und neue Informationen in Ihre Analysepläne einfließen lassen. Darüber hinaus können Sie mit SPSS Custom Dialog Builder R-Code mit Personen gemeinsam nutzen und wiederverwenden, die von den neuen Analyseoptionen profitieren würden, aber keine Programmierung für die Durchführung von Analysen verwenden. Entwickler können sich auf das Schreiben von Code konzentrieren, während sich Benutzer auf die Durchführung von Analysen konzentrieren und R-basierte Funktionalität nutzen können, ohne die Programmiersprache R erlernen zu müssen.

Was ist R?

R ist eine Open-Source-Programmiersprache und Softwareumgebung für statistische Datenverarbeitung und Grafiken (www.r-project.org). Die Programmiersprache R wird inzwischen gern von Statistik- und Data-Mining-Experten für die Entwicklung statistischer Software verwendet. Darüber hinaus wird R häufig für die erweiterte Datenanalyse verwendet. R bietet eine breite Palette an statistischen und grafischen Verfahren, wie z. B. lineare und nicht lineare Modellierung, klassische statistische Tests, Zeitreihenanalyse, Klassifizierung und Clustering. R ist zu den Bedingungen der [Free Software Foundation](#) und der [GNU General Public License](#) erhältlich.

R bietet über 4.800 Pakete aus mehreren Repositories, die auf Ökonometrie, Data-Mining, Raumanalyse, Bioinformatik etc. spezialisiert sind. Die Programmiersprache R wurde von intern anerkannten Statistikern und IT-Wissenschaftlern geprüft. Da es sich jedoch um eine Open-Source-Lösung handelt, verfügen weder Releases noch Pakete über codifizierte Prozesse.

Die Grenzen von R

Da R kostenfrei erhältlich ist, hat man den Eindruck, dass die Lösung statt kostenpflichtiger Statistik- und Modellierungssoftware verwendet werden kann und damit für ein Unternehmen große Einsparungen zu erzielen sind. Doch dieser Eindruck täuscht. Trotz der beträchtlichen Vorteile kann R nicht alles bieten, was Sie benötigen, um aus Ihren Analysen größtmöglichen geschäftlichen Nutzen zu ziehen. R-Benutzer müssen die folgenden Einschränkungen hinnehmen:

- *Bereitstellung.* Die Verwendung von R, um Vorhersageausgaben in eine Betriebsumgebung zu integrieren, kann schwierig sein.
- *Schnittstelle.* R verfügt über keine moderne grafische Benutzerschnittstelle, was die Nutzung durch Personen erschwert, bei denen es sich nicht um R-Programmierer handelt.
- *Einarbeitungsaufwand.* R ist nicht für jeden einfach erlernbar. Nicht jeder ist ein Programmierer.
- *Daten.* R bietet keine einfache native Datenbankverbindung.
- *Ausgabe.* Die Erstellung einer zur Veröffentlichung geeigneten Ausgabe ist schwierig.
- *Leistung.* R kann sehr schnell den gesamten verfügbaren Hauptspeicher verbrauchen.
- *Collaboration.* R erschwert die Arbeitsteilung in einem Analytenteam, insbesondere wenn Teammitglieder nicht über denselben Wissensstand bezüglich R verfügen.
- *Unternehmenssicherheit.* Die Sicherheit der Pakete, die heruntergeladen werden, ist nicht gewährleistet.

Darüber hinaus müssen Benutzer sicherstellen, dass sie über die richtigen R-Pakete und den richtigen R-Code verfügen, um die Pakete und die Ausgabe miteinander zu verknüpfen. Diese Einschränkungen bringen häufig zusätzliche Kosten mit sich, wenn die Programmiersprache R verwendet wird.

Vorteile der Integration von R in SPSS-Software

Es ist sinnvoll, die Einschränkungen der Programmiersprache R durch Verwendung von R in der SPSS-Software aufzuheben. Die kombinierten Stärken beider Lösungen werden den Anforderungen von Unternehmen gerecht, die nur über wenige Experten verfügen und die ohne großen Einarbeitungsaufwand von R profitieren möchten. SPSS-Software ist eine praktische Plattform, von der R-Benutzer große Datenmengen bearbeiten können und qualitativ hochwertige Grafiken und andere Ausgabearten erhalten. Andere Vorteile sind z. B. die Benutzerfreundlichkeit der SPSS-Software und die Möglichkeit, integrierte R-Pakete an eine große Anzahl an Benutzern weiterzugeben, die mit R nicht vertraut sind. Am wichtigsten ist jedoch, dass die Einschränkungen von R durch die Verwendung von R in der SPSS-Software aufgehoben werden können.

Bereitstellung

SPSS-Software kann die Einschränkungen von R hinsichtlich der Bereitstellung kosteneffizient aufheben. Mit IBM SPSS Modeler Gold können Benutzer beispielsweise die Auswirkungen von Geschäftsregeln und Vorhersagemodellen auf Verkäufe, Kundenservice, Wartung etc. optimieren. SPSS Modeler Gold unterstützt die Verwendung von Modellen, die auf R basieren. Ferner optimiert SPSS Modeler den Zugriff auf geschäftsrelevantes Wissen durch den kombinierten Einsatz mit IBM Cognos TM1. IBM kann Benutzern mehrere Arten der Verwendung von Vorhersageinformationen bieten. Benutzer können Vorhersagekenntnisse in ihre interaktiven, mobilen Dashboards neben Langzeit- und Echtzeitdatenansichten integrieren, ohne zusätzliche Software erstellen oder erwerben zu müssen.

Schnittstelle

SPSS Modeler und SPSS Statistics bieten eine einfache grafische Benutzerschnittstelle, die eine Vielzahl von Algorithmen für die Datenaufbereitung, statistische Analyse und Vorhersagemodellierung unterstützt. R-Code wird neben allen anderen Funktionen und Features, die bereits vorhanden sind, in derselben Schnittstelle ausgeführt. Darüber hinaus können Benutzer R ein GUI-Front-End hinzufügen. Mit diesem Front-End können Benutzer, die keine Programmierer sind, Kundenwerte, wie z. B. Variablenamen etc., eingeben (Abbildung 1). Sie können außerdem von erweiterten Analysen profitieren, ohne den zugrunde liegenden Code verstehen zu müssen.

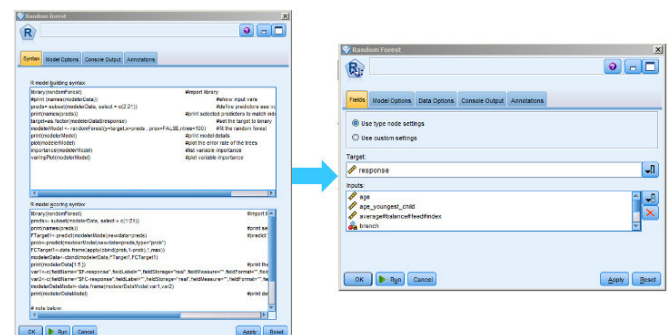


Abbildung 1: R-Code kann verwendet werden, um Kunden-Dialogfenster zu erstellen, damit bei der Analyse auch Personen vom Code profitieren können, die keine Programmierer sind.

Einarbeitungsaufwand

R ist eine Programmiersprache und muss daher erlernt werden. Doch das Erlernen von R ist nicht einfach. Während Sie sich die Zeit nehmen, R zu erlernen, können Sie den vollen Funktionsumfang der Programmiersprache nicht nutzen. Wenn R jedoch in SPSS-Software integriert ist, können Sie praktisch direkt von den Vorteilen profitieren. Sie können Ihren Lernaufwand auf die R-Routinen konzentrieren, die wirklich speziell sind, und die Daten, die statistischen Analysen und die Modellierung nutzen, die bereits in der SPSS-Software enthalten sind.

Daten

Wenn R verwendet wird, ist für den Zugriff auf die für die Analyse benötigten Daten ein hoher Zeit- und Arbeitsaufwand erforderlich. Es müssen große Mengen an Code geschrieben, Pakete implementiert und sogar Java verwendet werden. SPSS Statistics und SPSS Modeler lösen das Problem des Datenzugriffs. Beide Lösungen können mit kommerziellen SQL-, Oracle-, SAP-, IBM Netezza-, DB2- und weiteren Datenbanken verwendet werden. SPSS Statistics und SPSS Modeler können Texteingaben, Spreadsheets, SAS-Dateien und andere Formate lesen. SPSS Modeler kann direkt aus IBM Cognos Business Intelligence und Cognos TM11 lesen. Assistenten mit vordefinierten Konnektoren greifen auf die Daten zu, sodass die extrem zeitaufwendige Aufgabe des Extrahierens, Transformierens und Bearbeitens von Daten vor der Analyse entfällt. SPSS Statistics und SPSS Modeler bieten leistungsfähige Verfahren für die Datenbearbeitung, die in Point-and-click-Schnittstellen eingebunden sind. Benutzer können Daten transponieren, prüfen und neu formatieren. SPSS Modeler bietet zudem eine automatische Datenaufbereitung, die Daten für die Vorhersagemodellierung mit einem einzigen Klick optimiert.

Ausgabe

SPSS Statistics und Modeler umfassen mehrere Möglichkeiten zur Erstellung von Diagrammen und Grafiken, die für die Verwendung in Präsentationen bereit sind. SPSS Statistics-Software bietet R-Programmierern die Möglichkeit, R-Funktionen in die Syntax der SPSS-Software einzubinden. Auf diese Weise können Grafiken, Pivot-Tabellen und andere Ausgabearten (Abbildung 2) in Präsentationsqualität erstellt werden. Benutzer können darüber hinaus ihre Ergebnisse problemlos in gängigen Formaten wie PDF, Word, PowerPoint, Excel etc. veröffentlichen. Außerdem kann R mit erweiterter Funktionalität von SPSS Modeler wie Entitäts-, Social-Media-, Textanalyse etc. kombiniert werden.

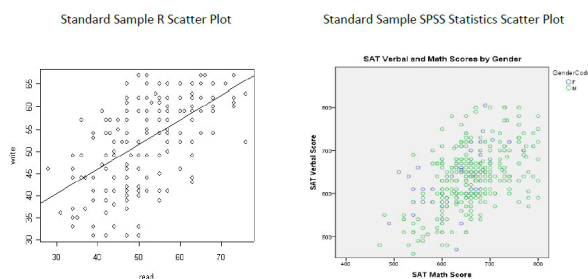


Abbildung 2: SPSS erstellt Diagramme und Grafiken, die für die Verwendung in Präsentationen bereit sind. Damit können Mitarbeiter aus dem Geschäftsbereich Zeiteinsparungen bei der Vorbereitung und Formatierung von Analyseausgaben erzielen.

Leistung

R-Befehle funktionieren in einem speicherinternen Arbeitsbereich. Der Hauptspeicher wird von R-Befehlen nicht berücksichtigt. Während einer R-Sitzung werden alle Objekte in einem temporären Arbeitsspeicher gespeichert. Aus diesem Grund kann der verfügbare Hauptspeicher recht schnell verbraucht sein. Die Kombination von R und SPSS Modeler bietet die Möglichkeit, die an R übertragenen Daten zu partitionieren oder eine Stichprobe zu nehmen (Abbildung 3). Darüber hinaus handelt es sich bei SPSS Modeler Server um eine speicherbasierende Technologie, die Analysen auf die Platte übertragen kann, damit Hauptspeicher verfügbar bleibt. Sie können ohne schwerwiegende Auswirkungen auf die Gesamtleistung von Data-Mining und Modellierung R-Befehle verwenden und R-Objekte erstellen. Darüber hinaus kann SPSS Modeler unter anderem R datenbankintern für IBM Netezza- und SAP Hana-Umgebungen skalieren. Mit SPSS Analytic Server kann R auch in Hadoop skaliert werden.

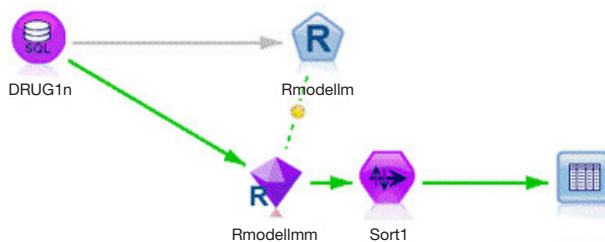


Abbildung 3: IBM SPSS Modeler wird verwendet, um über ein R-Modell zur Auswertung von Datensätzen nativen Zugriff auf Daten in einer SQL-Datenbank zu erhalten. Violett bedeutet, dass der Analyseschritt SQL-Pushback verwendet, wodurch die Leistung durch Ausführung speicherintensiver Aufgaben in der Datenbank selbst verbessert wird.

Collaboration

R ist eine Programmiersprache, die für Einzelpersonen vorgesehen ist. Beim Großteil der Analysearbeit handelt es sich jedoch um eine gemeinsame Aufgabe, in deren Rahmen mehrere Personen ihren Beitrag zu Modellen und statistischen Analysen leisten. Wenn SPSS-Software und R zusammen genutzt werden, verliert die Programmiersprache ihren „Einzelgänger“-Charakter und gewinnt erstklassige IBM Collaboration-Funktionalität hinzu. SPSS-Software bietet ein umfassendes Framework für die Zentralisierung, Sicherung und Automatisierung von Analyseassets, die mit SPSS Statistics und SPSS Modeler entwickelt wurden. Damit wird sichergestellt, dass entwickelte Modelle und statistische Analysen sicher gemeinsam genutzt werden können und Corporate Governance angewendet werden kann.

Unternehmenssicherheit

R hat keinen offiziellen Releaseprozess. Das heißt, eine offizielle Qualitätssicherung erfolgt nicht. Ohne Qualitätssicherung verwenden Sie R auf eigenes Risiko. Sie können nicht sicher sein, dass das Release oder Paket die versprochenen Funktionen auch wirklich bietet. Darüber hinaus werden R-Pakete aus dem Web und Benutzercommunitys heruntergeladen, sodass die Sicherheit der Pakete zweifelhaft sein kann. Ohne Ihr Wissen könnten Sie im Rahmen eines Downloads Malware, Trojaner, Data-Taps etc. herunterladen. SPSS Modeler und Statistics werden dagegen als Bestandteil des IBM Qualitätssicherungsprozesses für Software gründlich getestet. Da die Software von IBM ist, müssen Sie keine riskanten Verfahren nutzen, die die Sicherheit Ihrer Umgebung beeinträchtigen können.

Fazit

Sowohl SPSS als auch R verfügen jeweils über Stärken, die im Laufe der Zeit optimal getestet und in der Analyse-Community gut angenommen werden. Darüber hinaus ergänzen sich diese Stärken, sodass noch leistungsfähigere Funktionen und Features entstehen, von denen die Analyse-Community insgesamt profitiert. R-Benutzer können auf leistungsfähige Datenmanagementfunktionen zugreifen, mit denen sie deutlich größere Datenmengen bearbeiten können. Außerdem bietet SPSS-Software R-Benutzern eine größere Auswahl an Ausgabeoptionen in Form von Grafiken und Pivot-Tabellen, die die Attraktivität für den Benutzer steigern. Abschließend fungiert SPSS-Software als ideales Instrument für die Bereitstellung von R-Paketen für eine große Zahl an Benutzern.

SPSS-Softwarebenutzer können auf eine größere Anzahl statistischer Funktionen zugreifen. Auf diese Weise können sie komplizierte Analysen durchführen, ohne sich zuerst eine komplexe Programmiersprache wie R aneignen zu müssen. Es lohnt sich also, die Vorteile einer gemeinsamen Nutzung von R und SPSS-Software zu bedenken.

Informationen zu IBM Business Analytics

IBM Business Analytics-Software bietet datenbasiertes Wissen, mit dem Unternehmen intelligenter arbeiten und Mitbewerber übertreffen können. Dieses umfassende Portfolio beinhaltet Lösungen für Business Intelligence, Vorhersageanalyse und Entscheidungsmanagement, Performance Management und Risikomanagement. Mit Business Analytics-Lösungen können Unternehmen Trends und Muster in bestimmten Bereichen (z. B. Kundenanalyse) identifizieren und visualisieren, die sich umfassend auf die betriebliche Leistung auswirken können. Mit diesen Lösungen können Szenarien verglichen, potenzielle Risiken und Geschäftschancen vorausgesehen, Ressourcen besser geplant, budgetiert und prognostiziert, Risiken gegen zu erwartende Erträge abgewogen und gesetzliche Bestimmungen eingehalten werden. Indem sie Analysefunktionen in großem Umfang verfügbar machen, können sie die taktische und strategische Entscheidungsfindung so ausrichten, dass die Unternehmensziele erreicht werden. Weitere Informationen finden Sie hier: ibm.com/software/de/analytics

Rückruf und Fragen

Wenn Sie ein Beratungsgespräch wünschen oder eine Frage haben, besuchen Sie uns unter:

ibm.com/software/de/analytics/contactus.

Ein IBM Ansprechpartner wird Ihre Anfrage innerhalb von zwei Arbeitstagen beantworten.



IBM Deutschland GmbH

IBM-Allee 1
71139 Ehningen
ibm.com/de

IBM Österreich

Obere Donaustraße 95
1020 Wien
ibm.com/at

IBM Schweiz

Vulkanstrasse 106
8010 Zürich
ibm.com/ch

Die IBM Homepage finden Sie unter:

ibm.com

IBM, das IBM Logo, ibm.com, Cognos, DB2 und SPSS sind eingetragene Marken oder Marken der IBM Corporation in den USA und/oder anderen Ländern. Weitere Produkt- und Servicenamen können Marken von IBM oder anderen Unternehmen sein. Eine aktuelle Liste der IBM Marken finden Sie auf der Webseite „Copyright and trademark information“ unter ibm.com/legal/copytrade.shtml

Netezza ist eine eingetragene Marke der IBM International Group B.V., einem IBM Unternehmen.

Linux ist eine eingetragene Marke von Linus Torvalds in den USA und/oder anderen Ländern.

Microsoft, Windows, Windows NT und das Windows-Logo sind Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

Java und alle auf Java basierenden Marken und Logos sind Marken oder eingetragene Marken der Oracle Corporation und/oder ihrer verbundenen Unternehmen.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Dieses Dokument ist zum Datum seiner Erstveröffentlichung aktuell und kann jederzeit von IBM geändert werden. Nicht alle IBM Angebote sind in jedem Land, in welchem IBM tätig ist, verfügbar.

Die Informationen in diesem Dokument werden auf der Grundlage des gegenwärtigen Zustands (auf „as-is“-Basis) ohne jegliche ausdrückliche oder stillschweigende Gewährleistung zur Verfügung gestellt, einschließlich, aber nicht beschränkt auf die Gewährleistungen für die Handelsüblichkeit, die Verwendungsfähigkeit für einen bestimmten Zweck oder die Freiheit von Rechten Dritter. Für IBM Produkte gelten die Gewährleistungen, die in den Vereinbarungen vorgesehen sind, unter denen sie erworben werden.

© Copyright IBM Corporation 2014



Bitte der Wiederverwertung zuführen