

Gemeinsames Potenzial von IBM SPSS Statistics und R

A large, stylized graphic of the letters 'IBM' in a bold, sans-serif font. The letters are composed of horizontal stripes in dark blue and light blue, with white spaces between the stripes, mimicking the classic IBM logo design.

Inhalt

- 2 Kurzübersicht
- 2 Warum IBM SPSS Statistics und R integrieren?
- 4 Integration von R und IBM SPSS Statistics
- 4 Erweiterungsbundles: Verwenden fremder R-Programme
- 7 Schreiben eigener R-Programme
- 10 Fazit

Kurzübersicht

Dieses White Paper soll die Funktionen und Leistungsmerkmale veranschaulichen, die durch die Integration von IBM SPSS Statistics und R zur Verfügung stehen. R-Benutzer profitieren von einem leistungsfähigem Datenmanagement, hohem Bedienungskomfort und der von IBM SPSS Statistics bereitgestellten Darstellungsqualität. SPSS Statistics-Benutzer können auf eine umfangreiche, stetig wachsende Sammlung an Bibliotheken für die statistische Analyse und für Grafiken zugreifen, um aus ihren Daten fundiertere Erkenntnisse zu gewinnen. IBM SPSS Statistics und R ergänzen sich in idealer Weise.

Warum IBM SPSS Statistics und R integrieren?

IBM SPSS Statistics ist eine der weltweit führenden Softwarelösungen im Bereich Statistik. Die Software bietet Vorhersagemodelle und hochwertige Analysen, mit denen Sie geschäftliche und wissenschaftliche Fragestellungen lösen können. Für viele Unternehmen, Forschungseinrichtungen und Statistikfachleute stellt sie den De-facto-Standard für die statistische Analyse dar. Unternehmen setzen IBM SPSS Statistics für Folgendes ein:

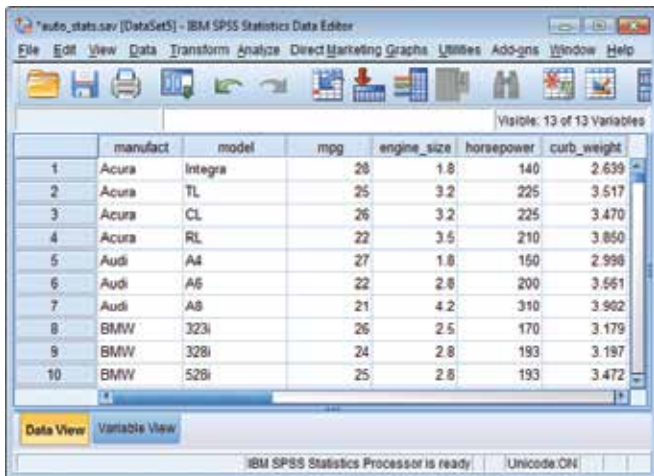
- Verstehen von Daten
- Analyse von Trends
- Vorhersage und Planung
- Überprüfung von Annahmen
- Unterstützung für präzise Schlussfolgerungen

IBM SPSS Statistics wurde seit 1968 fortlaufend weiterentwickelt und getestet. In diesem Zeitraum wurden zahlreiche Formen der statistischen Analyse in die Software integriert. Darüber hinaus wurden die Algorithmen, die die Gleichungen berechnen, von Entwicklern und akademischen Benutzern sowie in Laboratorien und in praktisch jeder Art von Unternehmen getestet. Somit können Benutzer darauf bauen, dass die Software gründlich geprüft ist und dass die Ergebnisse zuverlässig sind.

Die IBM SPSS Statistics-Umgebung vereinfacht es, auf Datasets zuzugreifen, sie zu verwalten und zu analysieren: z. B. Umfragedaten, Unternehmensdatenbanken oder Daten aus dem World Wide Web und vieles mehr. Intelligente statistische Prozeduren und Visualisierungen stellen eine leistungsfähige, benutzerfreundliche und integrierte Plattform für das Verständnis der Daten und das Lösen komplexer geschäftlicher und wissenschaftlicher Fragestellungen bereit.

IBM SPSS Statistics unterstützt Sie in *allen* Aspekten des Analyseprozesses, von der Datenvorbereitung und dem Datenmanagement bis hin zur Analyse und Berichterstellung. Das Produkt bietet angepasste Funktionalität und anpassungsfähige Schnittstellen für unterschiedliche Qualifikationsstufen und Funktionsbereiche. Außerdem können Benutzer damit Diagramme und präsentationsfertige Berichte mit hoher Auflösung erstellen, damit die Ergebnisse ohne großen Aufwand kommuniziert werden.

Beispiel: Ein Benutzer ist daran interessiert, die Verbrauchsdaten (mpg = Miles per Gallon) für unterschiedliche Fahrzeugarten zu analysieren. Abbildung 1 stellt dar, wie ein Teil der Daten im IBM SPSS Statistics-Dateneditor aussieht.



	manufact	model	mpg	engine_size	horsepower	curb_weight
1	Acura	Integra	28	1.8	140	2.639
2	Acura	TL	25	3.2	225	3.517
3	Acura	CL	26	3.2	225	3.470
4	Acura	RL	22	3.5	210	3.850
5	Audi	A4	27	1.8	150	2.998
6	Audi	A6	22	2.8	200	3.561
7	Audi	A8	21	4.2	310	3.902
8	BMW	323i	26	2.5	170	3.179
9	BMW	328i	24	2.8	193	3.197
10	BMW	528i	25	2.8	193	3.472

Abbildung 1: Der IBM SPSS Statistics-Dateneditor zeigt Informationen zum Verbrauch (mpg) für unterschiedliche Marken und Modelle an.

Für diese Analyse kann der Benutzer z. B. zuerst die Prozedur für deskriptive Statistik (Descriptives) ausführen, um eine Vorstellung von der Verteilung der Verbrauchsdaten (mpg) zu bekommen. Dazu kann der Benutzer das Dialogfeld für deskriptive Statistik (Descriptives) verwenden (Abbildung 2).

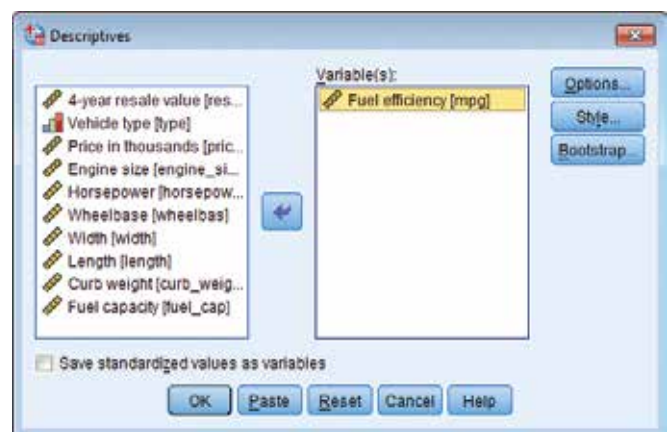


Abbildung 2: Dialogfeld für deskriptive Statistik (Descriptives).

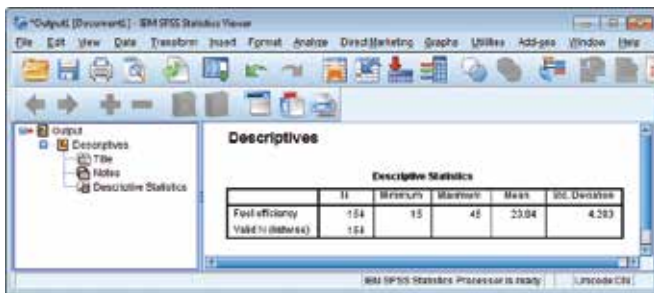


Abbildung 3: Ausgabe des Dialogfelds für deskriptive Statistik (Descriptives) im IBM SPSS Statistics Viewer.

Die Ausgabe (hier tabellarisch) wird im IBM SPSS Statistics Viewer angezeigt (Abbildung 3).

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 und für die professionelle Datenanalyse verwendet. R stellt eine große Vielfalt intelligenter statistischer und grafischer Verfahren bereit und ist hochgradig erweiterbar. R ist als „Free Software“ unter den Lizenzbedingungen der [Free Software Foundation GNU General Public License](http://www.fsf.org/) verfügbar. R kann unter Windows und Mac OS, auf einer großen Vielfalt von UNIX-Plattformen und ähnlichen Systemen (einschließlich FreeBSD und Linux) ausgeführt werden. Außerdem kann R ohne großen Aufwand mit Paketen erweitert werden.

Zu den Themen, die für R-Benutzer allgemein relevant sind, gehören Fragen zur Skalierbarkeit und zum grundlegenden Daten- und Ausgabemanagement: z. B. das Herstellen von Verbindungen zu Datenbanken, das Verbessern der Ausgabequalität und die gemeinsame Nutzung von R-Algorithmen mit anderen Personen, die nicht mit R vertraut sind. Andererseits können IBM SPSS Statistics-Benutzer einige R-Funktionen, die in IBM SPSS Statistics nicht verfügbar sind, nutzen, ohne R erlernen zu müssen. Schließlich ist die Programmierung in R nicht jedermanns Sache.

Daher ist die Integration von R und IBM SPSS Statistics sinnvoll. Die Stärken in Kombination miteinander erfüllen die Bedürfnisse der beiden Gruppen von Benutzern. IBM SPSS Statistics-Software ist eine praktische Plattform, von der R-Benutzer große Datenmengen bearbeiten können und qualitativ hochwertige Grafiken und andere Ausgabearten erhalten. Weitere Vorteile sind z. B. die Benutzerfreundlichkeit der IBM SPSS Statistics-Software und die Möglichkeit, integrierte R-Pakete an eine große Anzahl an Benutzern weiterzugeben, die mit R nicht vertraut sind. Durch diese Integration erhalten IBM SPSS Statistics-Benutzer einen einfachen Zugriff auf nahezu 4000 Open-Source-Statistikfunktionen.

Integration von R und IBM SPSS Statistics

Damit Sie R-Programmierungsfunktionen mit IBM SPSS Statistics verwenden können, benötigen Sie das IBM SPSS Statistics-Integrations-Plug-in für R. Dieses Plug-in ist kostenlos erhältlich und wird als Teil von IBM SPSS Statistics Essentials für R installiert. Dieses Plug-in benötigen Sie, wenn Sie selbst geschriebene oder fremde Erweiterungen verwenden möchten. Das IBM SPSS Statistics-Integrations-Plug-in für R ist Teil einer Familie von Integrations-Plug-ins, die kostenlos erhältlich sind. Zu dieser Familie gehören Plug-ins für Python, Java und .NET. R kommuniziert mit IBM SPSS Statistics über APIs im Plug-in, und die Integration setzt das Schreiben von R-Wrapper-Code voraus. Das IBM SPSS Statistics-Integrations-Plug-in für R erweitert die Sprache der IBM SPSS Statistics-Befehlsyntax um alle Leistungsmerkmale der Programmiersprache R. Das Plug-in bietet darüber hinaus Zugriff auf eine integrierte R-Entwicklungsumgebung, die Benutzern das Entwickeln, Testen und debuggen von R-Programmen für die Verwendung mit IBM SPSS Statistics vereinfacht. Dies ist für Windows, Linux, Mac OS und IBM SPSS Statistics Server verfügbar.

Nach dem Installieren des IBM SPSS Statistics-Integrations-Plug-ins für R können Sie R-Programme verwenden, die bereits geschrieben wurden, oder Sie können diese selbst schreiben.

Erweiterungsbundles: Verwenden fremder R-Programme

IBM SPSS Statistics und die R-Integration ermöglichen es Ihnen, die R-Programme zu verwenden, die von anderen geschrieben wurden und als Erweiterungsbundles in Paketen zusammengefasst sind. Die bereits codierten Algorithmen machen eine intensive R-Programmierung überflüssig, besonders dann, wenn Sie unter Zeitdruck sind oder das Fachwissen in R-Programmierung gering ist. Diese R-Programme werden als Erweiterungsbundles implementiert.

Nach der Installation eines Erweiterungsbundles können Sie auf das zugehörige Dialogfeld in den IBM SPSS Statistics-Menüs zugreifen. Außerdem kann der Erweiterungsbefehl wie ein beliebiger integrierter Befehl ausgeführt werden. Das R-Programm funktioniert wie ein natives Dialogfeld und wie ein Syntaxbefehl.

Zum besseren Verständnis eines Erweiterungsbundles betrachten wir erneut das Beispiel, in dem der Benutzer Daten zum Verbrauch (mpg) analysiert. Der Benutzer möchte nun z. B. den Verbrauch als Funktion der Motorgröße, der Pferdestärken und des Leergewichts des Fahrzeugs analysieren. Er möchte jedoch mehr als nur die übliche lineare Regression untersuchen und eine Quantil-Regression durchführen. Die Quantil-Regression wird als Erweiterungsbundle im Integrations-Plug-in für R bereitgestellt. Der Benutzer kann sie verwenden, um die Verteilung des Verbrauchs als Funktion der Einflussvariablen zu verstehen. Für den technisch Interessierten: Die Quantil-Regression ist eine Schätzung mit mindestens einem bedingten Quantil ($0 \leq q < 1$) für ein lineares Modell. Im Gegensatz dazu wird bei der normalen Regression lediglich der bedingte Erwartungswert geschätzt.

Das Dialogfeld für die Quantil-Regression (Quantile Regression) sieht genauso aus wie jedes andere native Dialogfeld in IBM SPSS Statistics (Abbildung 4).

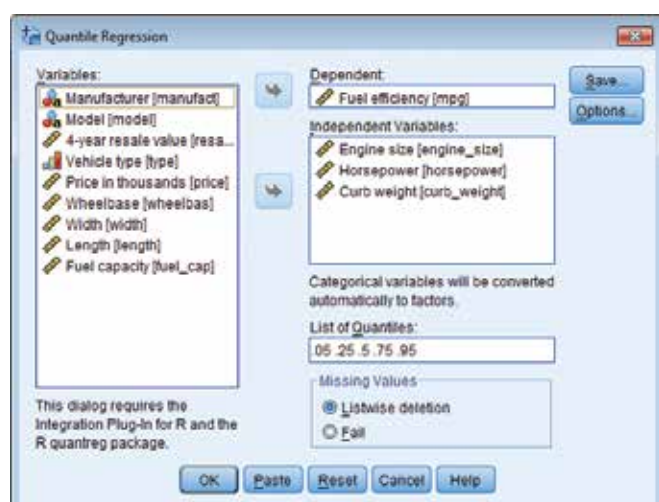


Abbildung 4: Dialogfeld für die Quantil-Regression (Quantile Regression).

Der Benutzer kann einfach die abhängigen und unabhängigen Variablen sowie die Liste der Quantile angeben und anschließend auf „OK“ klicken, um die Analyse auszuführen. IBM SPSS Statistics ruft R auf und führt die Analyse mithilfe des R-Pakets „quantreg“ aus. Der Benutzer sieht jedoch davon nichts. Die Ergebnisse der Analyse aus R werden tabellarisch oder als Diagramm im IBM SPSS Statistics Viewer ausgegeben. Wie in einer normalen Regressionsanalyse umfassen die Ergebnisse eine tabellarische Ausgabe der Regressionskoeffizienten. In diesem Fall wird eine separate Tabelle mit Regressionskoeffizienten für die einzelnen angegebenen Quantile erstellt (Abbildung 5). Die Tabelle für das 0,05-Quantil wird angezeigt. Die in der Tabelle angezeigten Ergebnisse werden bei der Ausführung der Analyse in R berechnet, aus R abgerufen und dann im IBM SPSS Statistics Viewer angezeigt.

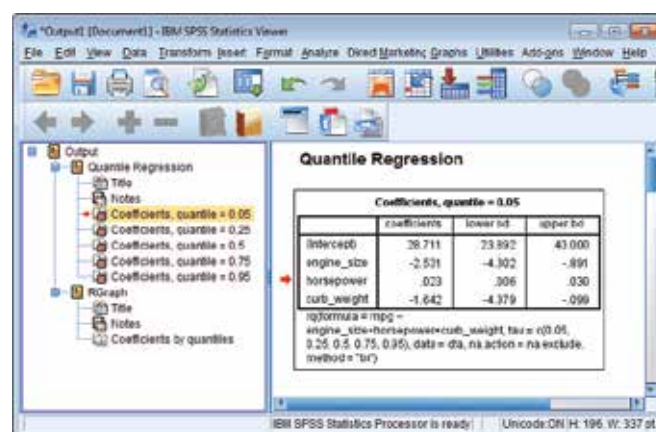


Abbildung 5: Tabelle der Regressionskoeffizienten.

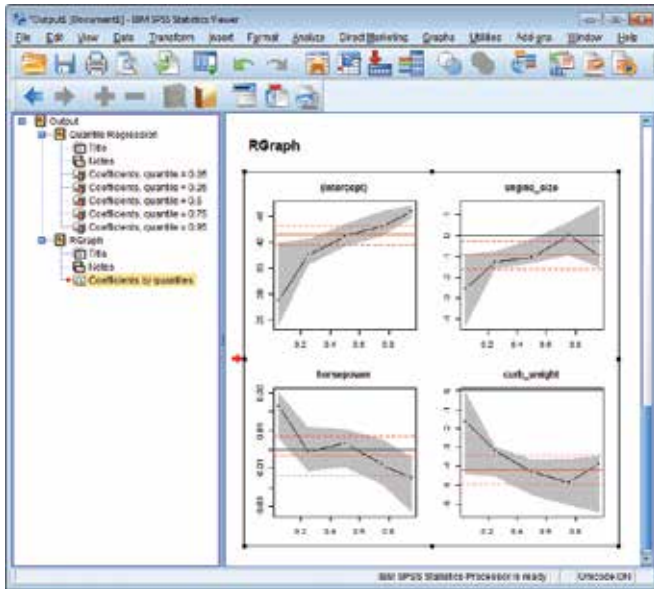


Abbildung 6: Diagrammanzeige der Koeffizienten für die einzelnen Einflussvariablen als Funktion der angegebenen Quantile.

Die Prozedur für die Quantil-Regression erzeugt darüber hinaus eine Diagrammausgabe. Darin werden die Koeffizienten für die einzelnen Einflussvariablen als Funktion der angegebenen Quantile dargestellt (Abbildung 6). Das Diagramm wird in Wirklichkeit von R generiert, jedoch automatisch im IBM SPSS Statistics Viewer wiedergegeben. Auch hierbei bleibt die gesamte Komplexität vor dem Benutzer verborgen. Der Benutzer sieht lediglich die Diagrammausgabe.

Auffinden von Erweiterungsbundles

Erweiterungsbundles, die eine Vielzahl von Statistikalgorithmen in R implementieren, werden mit Essentials for R installiert. Eine Liste dieser Erweiterungsbundles finden Sie im Hilfesystem von IBM SPSS Statistics unter: *Integration Plug-in for R Help > R Extension Commands for SPSS Statistics*.

Viele weitere Erweiterungsbundles, die statistische Algorithmen in R implementieren, sind von der IBM SPSS-Community auf der Website von IBM developerWorks unter folgender Adresse erhältlich:

ibm.com/developerworks/spssdevcentral

Ab IBM SPSS Statistics 22 können Sie innerhalb von IBM SPSS Statistics nach Erweiterungsbundles, die sich in der SPSS-Community befinden, suchen und diese herunterladen. Diese Funktion finden Sie unter folgenden Optionen: *Utilities > Extension Bundles > Download and Install Extension Bundles*. Bereits installierte Bundles können auf dieselbe Weise aktualisiert werden.

Name	Summary	Latest version	Installed version	Parameters	Select
COV0	COV0 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV1	COV1 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV2	COV2 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV3	COV3 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV4	COV4 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV5	COV5 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV6	COV6 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV7	COV7 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV8	COV8 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV9	COV9 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV10	COV10 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV11	COV11 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV12	COV12 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV13	COV13 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV14	COV14 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV15	COV15 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV16	COV16 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV17	COV17 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV18	COV18 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV19	COV19 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV20	COV20 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV21	COV21 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV22	COV22 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV23	COV23 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV24	COV24 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV25	COV25 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV26	COV26 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV27	COV27 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV28	COV28 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV29	COV29 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV30	COV30 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV31	COV31 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV32	COV32 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV33	COV33 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV34	COV34 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV35	COV35 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV36	COV36 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV37	COV37 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV38	COV38 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV39	COV39 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV40	COV40 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV41	COV41 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV42	COV42 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV43	COV43 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV44	COV44 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV45	COV45 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV46	COV46 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV47	COV47 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV48	COV48 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV49	COV49 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV50	COV50 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV51	COV51 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV52	COV52 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV53	COV53 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV54	COV54 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV55	COV55 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV56	COV56 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV57	COV57 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV58	COV58 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV59	COV59 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV60	COV60 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV61	COV61 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV62	COV62 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV63	COV63 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV64	COV64 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV65	COV65 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV66	COV66 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV67	COV67 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV68	COV68 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV69	COV69 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV70	COV70 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV71	COV71 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV72	COV72 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV73	COV73 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV74	COV74 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV75	COV75 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV76	COV76 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV77	COV77 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV78	COV78 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV79	COV79 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV80	COV80 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV81	COV81 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV82	COV82 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV83	COV83 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV84	COV84 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV85	COV85 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV86	COV86 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV87	COV87 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV88	COV88 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV89	COV89 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV90	COV90 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV91	COV91 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV92	COV92 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV93	COV93 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV94	COV94 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV95	COV95 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV96	COV96 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV97	COV97 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV98	COV98 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV99	COV99 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select
COV100	COV100 - Current Working Directory Extension command to	1.0.0	1.0.0	Yes	Select

Abbildung 7: Liste der verfügbaren R-Erweiterungsbefehle in IBM SPSS Statistics.

Schreiben eigener R-Programme

Die Verwendung von Erweiterungsbundles ist lediglich eine von vielen Möglichkeiten, R in IBM SPSS Statistics zu nutzen. Sie können ein eigenes R-Programm schreiben und es in IBM SPSS Statistics wie folgt auf verschiedenen Ebenen integrieren:

- Sie können einen benutzerdefinierten Dialog erstellen, der die Syntax für einen R-Erweiterungsbefehl oder für expliziten R-Code generiert.
- Sie können einen in R implementierten Erweiterungsbefehl erstellen.
- Sie können R-Code direkt von IBM SPSS Statistics aus ausführen.

In diesen Strukturen haben Sie Zugriff sowohl auf die Programmiersprache R als auch auf die Funktionen, die für IBM SPSS Statistics spezifisch sind und die im R Integration Package für SPSS Statistics bereitgestellt werden. Außerdem können Sie R-Funktionen schreiben, die IBM SPSS Statistics-Funktionalität von R aus verwenden, jedoch Ergebnisse an R zurückgeben.

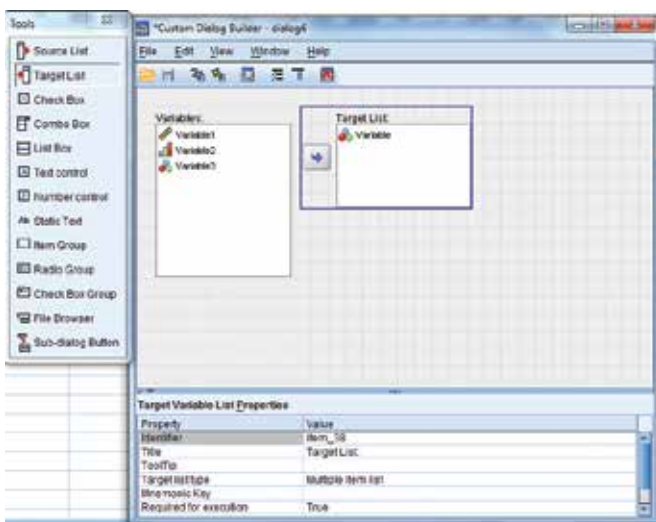


Abbildung 8: Beispiel für das Layout eines benutzerdefinierten Dialogs.

Mit diesen Funktionen können Sie Folgendes tun:

- Falldaten aus dem aktiven Dataset in R einlesen
- Informationen zu Daten im aktiven Dataset abrufen
- Ausgabeergebnisse aus Syntaxbefehlen abrufen
- Ergebnisse aus R schreiben (zurück in ein neues Dataset, in eine Pivot-Tabelle und in eine Grafik), um diese in IBM SPSS Statistics anzuzeigen

Erstellen eines benutzerdefinierten Dialogs, der die Syntax für einen R-Erweiterungsbefehl oder für expliziten R-Code generiert

Mit dem Programm für Dialogfeldaufbau (Custom Dialog Builder) können Sie eine Benutzerschnittstelle erstellen, die Befehlsyntax für einen in R implementierten Erweiterungsbefehl generiert. Anschließend können Sie die Ausgabe anzeigen (Abbildungen 7-9), die bei der Ausführung des Dialogs im IBM SPSS Statistics Viewer zurückgegeben wird. Ein R-Programm kann außerdem direkt in einen benutzerdefinierten Dialog eingebettet werden.

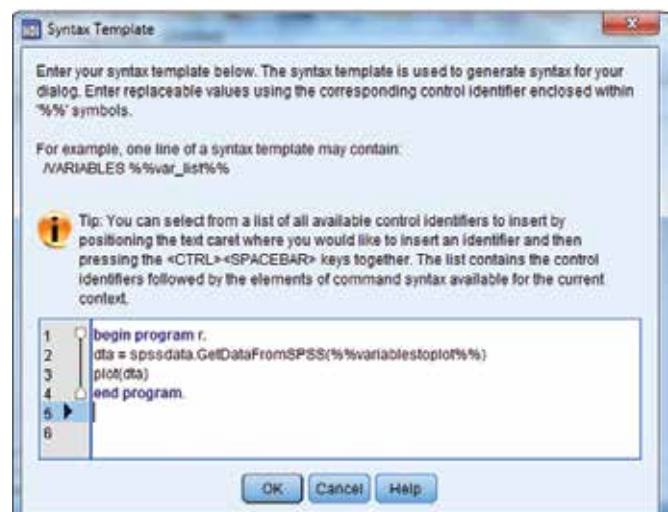


Abbildung 9: Syntaxvorlage für expliziten R-Code.

Erstellen eines in R implementierten Erweiterungsbefehls

Ein Erweiterungsbefehl ist ein benutzerdefinierter IBM SPSS Statistics-Befehl, der in R, in Python oder in Java implementiert ist. Sie können ein R-Programm in IBM SPSS Statistics integrieren, indem Sie einen Erweiterungsbefehl erstellen, der das R-Programm implementiert. Die Integration eines R-Algorithmus in IBM SPSS Statistics ist besonders nützlich, wenn ein Benutzer eine erweiterte statistische Funktion benötigt, jedoch nicht über das Fachwissen oder über die Zeit verfügt, ein derartiges Programm zu erstellen. In diesem Szenario, kann eine Methodikgruppe, die R-Algorithmen für dringend benötigte Statistikfunktionen erstellt, den Code schreiben und ihn als Erweiterungsbefehl verteilen. Anschließend kann der Benutzer diesen wie einen integrierten IBM SPSS Statistics-Befehl verwenden.

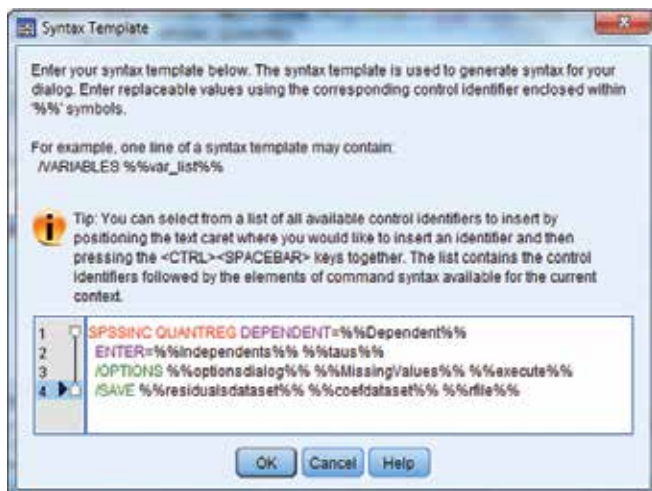


Abbildung 10: Syntaxvorlage für einen Erweiterungsbefehl.

Sie können einen Erweiterungsbefehl wie folgt erstellen:

- Schreiben Sie das Programm wie eine R-Funktion.
- Definieren Sie die IBM SPSS Statistics-Syntax für den Erweiterungsbefehl in einer XML-Datei, die den Befehlsnamen, die Unterbefehle und die Schlüsselwörter angibt (Abbildung 11).
- Deklarieren Sie die Syntax in einer „R-Run-Function“ und rufen Sie die Funktion auf (Abbildung 12).

Die Benutzereingabe wird automatisch überprüft, R-Variablen zugeordnet und der implementierenden Funktion übergeben.

Der R-Code ruft APIs im Integrations-Plug-in für R auf, die mit IBM SPSS Statistics interagieren. Text im R-Code, der für die Ausgabe bestimmt ist, z. B. Pivot-Tabellenbezeichnungen und Fehlernachrichten, kann für das Übersetzen aktiviert werden.



Abbildung 11: Syntax für den Erweiterungsbefehl.


```

52 Run<-function(args){
53   args <- args[1:(length(args)-1)]
54   oobj<-spsplg Syntax(temo=lst(
55     sspkg Template("DEPENDENT", subo="", Hype="existingvarist", var="y", idist=FALSE),
56     sspkg Template("ENTER", subo="", Hype="existingvarist", var="x", idist=TRUE),
57     sspkg Template("MISSING", subo="OPTIONS", Hype="str", var="missing",
58     valist = list("str", "1a1")
59   ))
60
61   if (HELP=="%in%" attr(args,"names"))
62     writeLines(helbed)
63   else
64     res <- sspkg processmo(oobj,args,"cbam")
65 }

```

Abbildung 12: Deklarieren der Syntax und Aufrufen der Funktion.

Direkte Ausführung von R-Code innerhalb IBM SPSS Statistics

Wenn Sie mit R vertraut sind, können Sie in IBM SPSS Statistics eigenen R-Code ausführen. Zum Ausführen des R-Codes schließen Sie diesen in einen Block mit BEGIN PROGRAM R – END PROGRAM ein (IBM SPSS Statistics-Befehlssyntax). Wenn die Syntax übergeben wird, wird der Code innerhalb des Blocks in R ausgeführt. Der Code ist in der Regel eine Kombination aus regulärem R-Code und Aufrufen SPSS-spezifischer R-Funktionen (aus dem Integrations-Plug-in für R), die es R ermöglichen, mit IBM SPSS Statistics zu interagieren.

Abbildung 13 zeigt ein Beispiel für ein R-Programm zum Ausführen einer Regression.

Es gibt folgende Funktionen, die das Verwenden von R-Programmfunktionen mithilfe von IBM SPSS-Befehlssyntax ermöglichen:

- `spsdata.GetDataFromSPSS` (Ruft Daten aus dem aktiven Dataset ab.)
- `spsdictionary.GetDictionaryFromSPSS` (Ruft Wörterbuchinformationen zu Variablen aus dem aktiven Dataset ab.)
- `spspivottable.Display` (Gibt eine tabellarische Ausgabe aus R als Pivot-Tabelle aus, die im IBM SPSS Statistics Viewer angezeigt oder mit dem IBM SPSS Statistics-Ausgabemanagementsystem in eine externe Datei geschrieben werden kann. Pivot-Tabellen, die mithilfe dieser Funktion erstellt werden, verhalten sich wie Pivot-Tabellen, die von nativem SPSS-Code erstellt wurden.)

1	begin program r.	2,3	y, x are dependent and independent variables
2	y = "salary"	5	combine into one list and get data, converting categorical variables to factors
3	x = "salbegin jobcat prevexp"	7	equation needs to look like y~x1+x2+x3 - convert blank(s) to + - "~ +" is a regular expression
4	data = spsdata.GetDataFromSPSS(c(y, x),	8,9	add dep var and combine into string
5	missingValueToNA=TRUE, factorMode="levels")	11	run the equation -convert string to formula object -identify data source -omit missing values -create result object "res" -notice the factor
6	plus = paste(" +", x)		
7	eqn = c(y, "~", plus)		
8	eqn = paste(eqn, collapse="")		
9	print(eqn)		
10	res = lm(as.formula(eqn), data=data, na.action=na.omit)		
11	print(res)		
12	end program.		

Abbildung 13: R-Programm für die Ausführung einer Regression.

Alle Artefakte können ohne großen Aufwand so gepackt werden, dass ein Erweiterungsbundle zur einfachen Verteilung und Installation erstellt wird. Das Erweiterungsbundle enthält Folgendes:

- Eine Paketdatei (.spd), die den benutzerdefinierten Dialog angibt
- Eine XML-Datei, die die Syntax des Erweiterungsbefehls angibt
- Die Datei bzw. die Dateien mit dem Implementierungscode in R

Andere Benutzer, die Essentials for R installiert haben, können über die Menüs von IBM SPSS Statistics das Erweiterungsbundle ohne großen Aufwand installieren. Anschließend können Sie den Dialog und den Erweiterungsbefehl auf dieselbe Art und Weise verwenden wie einen nativen Dialog oder Befehl. Damit die Zusammenarbeit und der Wissensaustausch ermöglicht werden, können Beiträge mit R-basierten Erweiterungen gegebenenfalls auf der Website der IBM SPSS-Community gehostet werden.

Fazit

Sowohl IBM SPSS Statistics als auch R zeichnen sich durch Stärken aus, die im Laufe der Zeit optimal getestet und in der Statistik-Community gut angenommen werden. Darüber hinaus ergänzen sich diese Stärken, sodass noch leistungsfähigere statistische Funktionen und Features entstehen, von denen die Statistik-Community insgesamt profitiert.

R-Benutzer können auf leistungsfähige Datenmanagementfunktionen zugreifen, mit denen sie deutlich größere Datenmengen bearbeiten können. Außerdem bietet das IBM SPSS Statistics-Ausgabemanagementsystem R-Benutzern eine größere Auswahl an Ausgabeoptionen in Form von Grafiken und Pivot-Tabellen, die die Attraktivität für den Benutzer steigern. Und schließlich fungiert IBM SPSS Statistics als ideales Instrument für die Bereitstellung von R-Paketen für eine große Zahl an Benutzern.

Auf diese Weise können IBM SPSS Statistics-Benutzer komplizierte Analysen durchführen, ohne sich zuerst eine komplexe Programmiersprache wie R aneignen zu müssen. Aufgrund der zahlreichen Vorteile lohnt es sich also, eine gemeinsame Verwendung von R und IBM SPSS Statistics zu erwägen.

Weitere Informationen

Für Benutzer, die R in IBM SPSS Statistics verwenden möchten, stehen die folgenden Informationsquellen zur Verfügung:

- Anweisungen dazu, wie Sie Essentials for R erhalten, finden Sie im Hilfesystem von IBM SPSS Statistics unter folgendem Eintrag: *Integration Plug-in for R Help > How to get the IBM SPSS Statistics – Integration Plug-in for R.*
- Lerntexte finden Sie unter folgendem Eintrag: *Help > Working with R.*
- Eine vollständige Dokumentation zum Integrations-Plug-in für R finden Sie im Hilfesystem von IBM SPSS Statistics unter „Integration Plug-in for R Help“.
- Fragen zur Verwendung von R in IBM SPSS Statistics können an das Forum zur R-Programmierung gesendet werden. Das Forum wird von der SPSS-Community auf der Website von developerWorks gehostet (ibm.com/developerworks/spssdevcentral).
- Ausführliche Informationen zur Erstellung von Erweiterungsbefehlen finden Sie im Artikel „Writing IBM SPSS Statistics Extension Commands“ (Schreiben von IBM SPSS Statistics Erweiterungsbefehlen) und im Kapitel zu Erweiterungsbefehlen in der Programmierung und im Datenmanagement für IBM SPSS Statistics. Beides ist bei der SPSS-Community erhältlich.
- Eine Liste der Erweiterungsbundles, die bei der SPSS-Community zum Herunterladen verfügbar sind, finden Sie im Artikel „Extension Bundles from IBM SPSS“ (Erweiterungsbundles von IBM SPSS Statistics) in der SPSS-Community. Die Liste umfasst Erweiterungsbundles, die neben den in R implementierten Erweiterungsbefehlen auch Python-Erweiterungsbefehle enthalten.

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

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.

Der Benutzer ist dafür verantwortlich, den Betrieb von Produkten oder Programmen anderer Anbieter in Verbindung mit IBM Produkten und Programmen zu prüfen und zu verifizieren.

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 2015



Bitte der Wiederverwertung zuführen