

Die beste Software für die Datenanalyse

Vorab: Es gibt nicht DIE beste Software. Für unterschiedliche Anwendungsbereiche, verschiedene Ziele und Anwender:innen sind unterschiedliche Software-Pakete gut geeignet. Ich arbeite je nach Projekt und Auftrag auch mit verschiedenen Programmen.

In den Tabellen sind die Software-Pakete für die Datenanalyse gelistet, die ich oft verwende. Die Kriterien sind subjektiv und erheben weder einen Anspruch auf Vollständigkeit noch auf Richtigkeit.

Vor der Entscheidung für oder gegen ein bestimmtes Software-Paket würde ich verschiedene Programme ausprobieren und prüfen, ob die Funktionen und das Arbeiten für meine Ziele geeignet sind und wie gut und schnell ich bei Fragen Unterstützung direkt vom Hersteller oder im Internet bekomme bzw. finde.

Legende: +++ sehr gut / ++ gut / + ok / - ausreichend / -- viel Luft nach oben / --- ungenügend / x fehlt

Allgemeines

	JMP	Minitab	R	KNIME
Art & aktuelle Version der Software	kommerziell JMP 16	kommerziell Minitab 20	OpenSource R-4.1.0	OpenSource KNIME Analytics Plattform 4.3
Kosten	1. Jahr ca. 1700 EUR, ab 2. Jahr ca. 850 EUR		0 EUR	0 EUR
Einarbeitungsaufwand	mittel	niedrig	hoch	mittel
Betriebssystem	Windows, Mac	Windows, (Mac)	Windows, Mac, Linux	Windows, Mac, Linux
Sprache	deutsch, englisch	deutsch, englisch	englisch	englisch
Benutzeroberfläche	Menü-basiert ++	Menü-basiert +++	Programmcode -- R alleine/ + mit RStudio	Operatoren/ Nodes ++

Analyse-Möglichkeiten

	JMP	Minitab	R	KNIME
einfache Kennzahlen & Grafiken	+++	+++	+++	+
komplexe Grafiken	+++	+	+++	--
Qualitäts-Statistik (Mess-System-Analyse, SPC, Prozessfähigkeit)	+++	+++	+	---
statistische Tests (t-Test, F-Test...)	+++	+++	+++	--
klassische Modellierung (Regression, ANOVA)	+++	+++	+++	+
erweiterte Modellierung (PLS, Lasso, Generalized Linear Model)	+++	-	+++	++
Data Mining Modelle (neuronale Netze, Entscheidungsbäume...)	++	-	+++	+++

Erweiterungen, Automatisierung & Schnittstellen

	JMP	Minitab	R	KNIME
Erweiterungen	Add-Ins ++	Makros +	>17.000 Packages	Extensions (H2O...)
Automatisierung der Analyse	+++	+	+++	+++
Schnittstellen	+++ R, Python, SAS, Matlab	- (Python, R, Datenbanken)	+ Python, Datenbanken	++ R, Python, PMML, Datenbanken
Datenmenge	durch Rechner / RAM begrenzt	max. 4000 Spalten	durch Rechner / RAM begrenzt	durch Rechner / RAM begrenzt
Performance bei großen Datenmengen / komplexen Auswertungen	+++	--	+++	+++

Support

	JMP	Minitab	R	KNIME
Anzahl Treffer bei Google	ca. 18 Mio	ca. 5 Mio	ca. 92 Mio	ca. 1 Mio
Dokumentation	+++	+	+++	++
Support durch Hersteller	++	-	x	++
Community (Forum usw.)	+++	x	+++	+++

Software-Validierung

	JMP	Minitab	R	KNIME
Strategie	Validierungstool	Validierungstool	Konzept für Validierung	x-
weitere Informationen	[1]	[2]	[3], [4], [5]	x

[1] JMP „A Commitment to Quality“ https://www.jmp.com/en_us/software/quality-statement.html, abgerufen 14.06.2021

[2] Minitab „Software-Validierung“ <https://www.minitab.com/de-de/support/software-validation/>, abgerufen 14.06.2021

[3] Schaefer, Peter; Fontana, Debra „Validating R – Part of the Uphill Battle in the Pharmaceutical Industry“, PharmaSUG 2018, <https://www.pharmasug.org/proceedings/2018/SI/PharmaSUG-2018-SI13.pdf>, abgerufen 14.06.2021

[4] PharmaR „Validation Overview“, <https://www.pharmar.org/overview/>, abgerufen 14.06.2021

[5] PharmaR „A risk-based approach for assessing R package accuracy within a validated infrastructure“ <https://www.pharmar.org/white-paper/>, abgerufen am 14.06.2021