

Statistik Grundlagen und Programmieren mit R







Statistik Schulung und
R Einführung mit Rstudio



Statistik Grundlagen und Programmieren mit R

Statistik Schulung und R Einführung mit Rstudio

Über den Kurs

-  **Dauer:** 3 Tage
-  **Gruppengröße:** 3-10
-  **Level:** Anfänger ohne Programmiererfahrung*
-  **Anteil Coding:** 50%
-  **Sprache:** R
-  **Bibliotheken:** data.table, ggplot2, stats, lm,...

Auf einen Blick

- ✓ Einführung in R und in wichtige Datenstrukturen
- ✓ Grundlagen eines data.table zur Datenanalyse
- ✓ Statistiken (Lagemaße, Streuungsmaße)
- ✓ Kreuztabelle & bedingte Wahrscheinlichkeiten
- ✓ Statistische Verteilungen und Zufallszahlen
- ✓ Daten einlesen und schreiben (csv, .RData)
- ✓ Korrelationsanalysen
- ✓ Lineare Regression
- ✓ Mittelwertsvergleiche und t-Tests

Überblick über das Seminar

Der dreitägige Einführungskurs in R erläutert Aspekte in Statistik und deren Umsetzung in R. Zudem wird die Datenstruktur data.table erläutert, um wichtige Grundlagen in der Datenanalyse umsetzen zu können.

Die frei verfügbare Programmiersprache R hat sich zu einer der meistverwendeten Sprachen bei statistischen Fragestellungen und Datenanalysen in Wissenschaft und Wirtschaft entwickelt. R wird in vielen Industrien verwendet, vor allem in der Versicherungsbranche, im Finanzsektor und in der medizinischen Statistik. Die Anwender schätzen die Leistungsstärke bei statistischen Berechnungen, die Einfachheit beim Erzeugen hochwertiger Grafiken sowie die umfangreichen Funktionen und Verfahren für Statistik und Datenanalyse. Auch im Machine Learning bietet R durch seine Fülle an Algorithmen und leichte Bedienbarkeit einen großen Vorteil. Schätzungsweise mehr als zwei Millionen Anwender benutzen R weltweit.

Der Schwerpunkt der Schulung ist auf anschaulichen Erklärungen, welche nur das Notwendige an mathematischen Formeln enthalten. Der Fokus liegt auf der direkten Umsetzung mit R und dem eigenen Programmieren.

Am Ende des Seminars haben Sie einen Überblick über wichtige Grundlagen in Statistik, kennen das data.table und können selbstständig erste Daten analysieren und Statistiken berechnen.

Inhalte des Seminars

Sie erhalten einen Überblick über R mit der Entwicklungsumgebung/DIE RStudio und lernen die Grundlagen zur Benutzung kennen, z.B. Variablen zuweisen, Pakete installieren und laden. Sie kennen wichtige Datentypen in R (vector, factor array) und können schnell grundlegende deskriptive Statistiken der Daten berechnen. Wichtige Lagemaße (Mittelwert, Median, Quantile) und Streuungsmaße (Varianz, Standard Abweichung) der Statistik und die Umsetzung in R sind Ihnen bekannt. Desweiteren werden Häufigkeitstabellen und Kreuztabellen und bedingte Wahrscheinlichkeiten behandelt.

Nach dem Seminar kennen Sie bekannte statistische Verteilungen, wissen was Dichtefunktion und Verteilungsfunktion ist und können Zufallszahlen aus einer Verteilung ziehen.

* Programmierkenntnisse in R oder einer anderen Sprache sind nicht notwendig. Erste Erfahrung in der einfachen Analyse von Daten (z.B. mit Excel) sind Voraussetzung.

Für die Grundlagen der Datenanalyse wird das Paket `data.table` verwendet. Es ist neben dem tidyverse-Paketen (u.a. mit `dplyr`) das beliebteste Paket zur Analyse von Daten. Sie können Daten einlesen und schreiben, Statistiken auf den Daten berechnen und fehlende Daten ergänzen. `Data.table` ist im Vergleich zu tidyverse insbesondere für größere Datenmengen effizienter. Die `data.tables` entsprechen den häufig in der statistischen Programmiersprache R verwendeten `data.frames` und ermöglichen eine elegante Analyse der Daten. In diesem Kurs wird nicht `dplyr` bzw. tidyverse geschult.

In der deskriptiven, bivariaten Statistik werden Korrelationsanalysen durchgeführt und die bekannten Korrelationsmaße Korrelationskoeffizient nach Pearson und Rangkorrelation nach Spearman behandelt.

Das beliebteste Paket zur Erstellung von Grafiken, `ggplot2`, und die dahinterliegende `grammar of graphics` wird durchgenommen. Sie können eigene Visualisierungen erstellen, anpassen, abspeichern und kennen die Grundlagen, um Plots individuell anzupassen. Sie beherrschen die Grundlagen von statistischen Tests (Stichprobe, Nullhypothese, Signifikanz, Konfidenzintervall) und können Mittelwertvergleiche mit dem t-Test durchführen. Häufige Pre-Tests für die Überprüfung der Voraussetzungen vom t-Test und nicht-parametrische Tests sind Ihnen geläufig.

Die Schulung beinhaltet ein einfaches Verfahren des Machine Learnings, welches ein Standardverfahren der Statistik ist: die lineare Regression. Sie lernen die theoretischen Grundlagen kennen, können den Algorithmus in R umsetzen und die Ergebnisse interpretieren.

Die Einstiegshürde für die Benutzung von R für Statistik und Datenanalyse ist genommen, so dass Sie eigenständig Ihr Wissen nach dem Seminar erweitern können. Durch die kleine Gruppengrößen und die eigene Umsetzung auf Ihrem Laptop mit Unterstützung des Dozenten wird eine hohe Qualität und ein bestmöglicher Lernerfolg sichergestellt.

Wer sollte teilnehmen?

Dieser Kurs mit R richtet sich an Teilnehmer mit keiner oder geringer Erfahrung in Statistik, der Analyse von Daten und der Bedienung der Statistik-Software R und der Entwicklungsumgebung Rstudio. Ziel sind statistische Grundlagen und Grundlagen der Datenanalyse kennen zu lernen. Teilnehmer haben in Ihrem Unternehmen mit der statistischen Auswertung von Daten zu tun oder wollen sich in diese Richtung entwickeln.

Methode des Seminars



Dieses Seminar ist sehr praxisorientiert. Die Teilnehmer arbeiten direkt und selbstständig mit der Programmiersprache R in der Entwicklungsumgebung RStudio, so dass das Erlernte direkt geübt und vertieft werden kann. Der Trainer moderiert dabei verschiedene Aufgaben und begleitet die Teilnehmer durch die einzelnen Lehreinheiten.

Voraussetzungen

Dieser R Kurs setzt keine Kenntnisse in R oder einer anderen Programmiersprache voraus. Notwendig sind jedoch Vorerfahrung mit dem Umgang von Daten, z.B. in Excel oder einer BI-Software, damit Konzepte einer spaltenweisen Berechnung bzw. einfache Statistiken (Mittelwert, Varianz) bekannt sind.

Allgemeine Computerkenntnisse und mathematische Grundkenntnisse (Was ist eine Funktion, Summenzeichen, Integral, Ableitung, Grundrechenarten, Koordinatensystem zum Plotten, ...) werden vorausgesetzt.

Das Seminar wird auf Deutsch gehalten. Englischkenntnisse (lediglich im Verstehen von englischen Texten) sind sehr empfehlenswert, da die Programmiersprache, verwendete Fachbegriffe in der Programmiersprache und die Dokumentationen im Internet auf Englisch sind. Aus diesem Grund sind auch die erstellten Folien in der Schulung auf Englisch, wobei mündlich auch die deutschen Terme verwendet werden.

Technische Voraussetzungen der Teilnehmer (Laptop, etc.)

- ❑ Die Teilnehmer benötigen für die Übungsaufgaben Laptops. Wir empfehlen, Ihren eigenen Laptop mit der vorab installierten Software mitzubringen. Eine genaue Installationsanleitung für die Software wird Ihnen vor dem Seminar per E-mail zugesandt. Auf Anfrage stellen wir auch Schulungslaptops zur Verfügung.
- ❑ Bitte prüfen Sie, ob Ihr Firmenlaptop Zugangsbeschränkungen im Internet hat. Die digitalen Unterlagen (Skript, Code, Dateien) werden im Seminar online zum Download zur Verfügung gestellt. Sie erhalten vor dem Seminar per E-Mail den Link zu einer Testdatei zum Download, um dies überprüfen zu können.
- ❑ Sie sollten sich in firmenfremde WLAN-Netze registrieren können.
- ❑ Als Backup Lösung ist es möglich, dass der USB Port bei Ihrem Laptop freigeschaltet ist, um damit verwendete Dateien oder sonstige Unterlagen übertragen zu können.
- ❑ Im Seminar wird das Betriebssystem Windows verwendet. Der Umgang mit Ihrem verwendeten Betriebssystem und Laptop sollte bekannt sein. Insbesondere sollten Sie ohne Schwierigkeiten Sonderzeichen auf der Tastatur finden (insbesondere bei Apple Geräten werden auf manchen Tastaturen nicht immer runde, eckige bzw. geschweifte Klammern dargestellt).

10.00-10.15

Begrüßung und Organisatorisches

- Vorstellungsrunde
- Erwartungen der Teilnehmer

10.15-11.45

Einführung in R

- Überblick über R und R-Studio
- Bedienung von R und R-Studio (Editor, Konsole, Hilfe usw.)
- Pakete installieren und laden
- Zuweisung von Variablen
- Skripte schreiben und ausführen

11.45-12.00

Kaffeepause

12.00-13.30

Wichtigste Datenstrukturen

- Vector, Factor und Array in R
- Elemente extrahieren
- Überblick über die Daten erhalten mit einfachen Funktionen
- Daten in andere Datenstrukturen umwandeln

13.30-14.30

Mittagspause

14.30-16.00

Statistiken in R

- Lagemaße (Mittelwert, Median, Quantile)
- Streuungsmaße (Varianz, Standard Abweichung)
- Weitere mathematische Funktionen (Minimum, Maximum, Summe, Wurzel, ...)

16.00-16.15

Kaffeepause

16.15-18.00

Kreuztabelle & bedingte Wahrscheinlichkeiten

- Eine Häufigkeitstabelle berechnen
- Auswertung einer Kreuztabelle
- Zeilen- und Spaltenwahrscheinlichkeiten berechnen
- Bedingte Wahrscheinlichkeiten interpretieren

18.00

Ende

09.00-09.15

Rückblick und offene Fragen von Tag 1

09.15-10.45

Statistische Verteilungen und Zufallszahlen

- Dichtefunktion und Verteilungsfunktion
- Bekannte statistische Verteilungen
- Zufallszahlen aus einer Verteilung ziehen
- Tests, ob Daten einer Normalverteilung folgen

10.45-11.00

Kaffeepause

11.00-12.30

Grundlagen eines DataTables

- Effiziente Berechnung mit Daten
- Reihen und Spalten filtern / auswählen
- Statistiken auf Spalten berechnen
- Berechnungen nach einer Variablen gruppieren

12.30-13.30

Mittagspause

13.30-15.15

Weitere Datenanalysen

- Spalten anlegen / modifizieren / löschen
- Fehlende Werte ergänzen
- Daten einlesen (mit der Funktion fread) und schreiben (als .csv bzw. .RData)

15.15-15.30

Kaffeepause

15.30-17.00

Bivariate Deskriptivstatistik

- Korrelationsanalysen
- Streudiagramme
- Bekannte Korrelationsmaße: Korrelationskoeffizient nach Pearson, Rangkorrelation nach Spearman

17.00

Ende

09.00-09.15

Rückblick und offene Fragen von Tag 2

09.15-10.45

Datenvisualisierung mit ggplot2

- Das Konzept hinter Grammar of Graphics
- Die Layer von ggplot2 zur Erstellung erster Plots und zum Plotten von Statistiken
- Darstellungen (Größe der Datenpunkte, Farbe, Gruppierung) mit einer Variable variieren oder festsetzen
- Mehrere Subplots erstellen, Plots anpassen und speichern
- Kerndichteschätzer
- Wichtige Plots: Scatterplot, Histogramm, Boxplot

10.45-11.00

Kaffeepause

11.00-12.30

Statistische Tests

- Grundlagen von statistischen Tests (Stichprobe, Nullhypothese, Signifikanz, Konfidenzintervall)
- Mittelwertsvergleiche und t-Tests
- Pre-Tests für die Überprüfung der Voraussetzungen vom t-Test. (Anderson-Darling, Shapiro-Wilk, Levene-Test)
- Nicht-parametrische Tests: U-Test und Vorzeichenrangtest bei Medianen

12.30-13.30

Mittagspause

13.30-15.15

Lineare Regression

- Algorithmische Grundlagen
- Umsetzung in R
- Interpretation und Signifikanz der Koeffizienten
- Bewertung der Gütemaße: R^2 , AIC/BIC
- Überprüfung der Normalverteilung der Residuen

15.15-15.30

Kaffeepause

15.30-17.00

Wiederholung mit eigenen oder neuen Daten

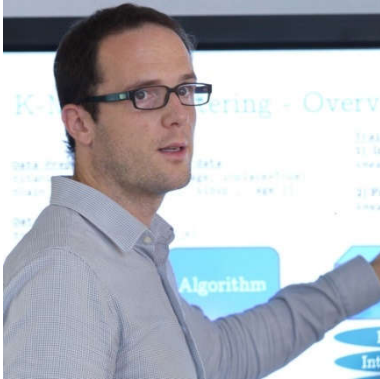
- Möglichkeit, das Gelernte am mitgebrachten Datensatz anzuwenden und direkt Fragen zu klären.
- Alternativ: Wiederholung am neuen Datensatz, um das Wissen zu vertiefen.

17.00

Ende

Ihre Dozenten

Unser folgender Experte leitet das Seminar



Jan Köhler

Vom Hintergrund Statistiker und Wirtschaftsingenieur hat er über 7 Jahre an den neuesten Technologien in Machine Learning, Deep Learning und Data Science im Bosch Center for Artificial Intelligence (BCAI) gearbeitet, hat in der Praxis bei über 25 Patentanmeldungen (meist als Haupterfinder) beigetragen und ist Mitautor bei Veröffentlichungen auf Machine Learning Konferenzen neben Veröffentlichungen im Bereich der Medizinstatistik bzw. des Operations Research. In vielen Praxisprojekten unterstützte er bisher als Data Scientist und hat verschiedene Teilnehmer, vom Projektmitarbeiter bis zum Konzern-Vorstand geschult.

Zusammenfassung

€ Preise

1420 € zzgl. MwSt.

📅 Termin und Ort

Termine und Orte finden Sie unter <https://enable-ai.de>

Haben Sie Fragen? Wir helfen Ihnen. Versprochen.

📍 Enable AI, Stuttgart

☎ 0711 96881553

✉ info@enable-ai.de