

AUFGABE 1: STATISTISCHE DATENANALYSE IN R

In Aufgabe werden Sie uni- und bivariate statistische Analysen eines geochemischen Datensatzes in der Software R durchführen. Dabei handelt es sich um den Datensatz `GeoChem_rev.dat`, der Daten zu Mergel-Proben enthält. Mergel sind kalkhaltige Sedimente, die unterschiedliche Anteile an Ton enthalten. In der Industrie werden Mergel hauptsächlich für die Produktion von Zement verwendet. Der Datensatz beinhaltet Werte für sieben Oxide, die in den Mergel-Proben enthalten sind.

Diese Teilaufgabe umfasst sowohl die Berechnung deskriptiver als auch inferentieller Statistiken. Im Detail:

- Berechnen von deskriptiven statistischen Maßen und explorative Datenanalyse für jedes Oxid;
- Berechnen von Korrelation und Regression;
- Interpretieren der Ergebnisse und schriftliche Dokumentation.

Der Datensatz **GeoChem_rev.dat** ist eine ASCII-Datei, die Daten zu 1269 Mergel-Proben enthält. Jede Zeile in der Tabelle enthält die Werte für eine Probe. Die Spalten enthalten die jeweiligen Oxid-Werte. Die Werte beziehen sich auf den Gewichtsanteil des jeweiligen Oxids in %. Zum Beispiel gibt ein `SiO2`-Wert von 17,2% an, dass eine 100 Gramm schwere Probe 17,2 Gramm `SiO2` enthält:

Spalte 1	2	3	4	5	6	7
SiO2	Al2O3	Fe2O3	CaO	MgO	Na2O	K2O

Anleitung:

Teil 1: Verschaffen Sie sich einen Überblick über den geochemischen Datensatz.

- Lesen Sie die Datei `GeoChem_rev.dat` in R ein.
- Berechnen Sie deskriptive statistische Maße für jedes der sieben Oxide (Minimum, Maximum, Wertebereich, Mittelwert, Standardabweichung, 1-2-

- 3 Quartil, Schiefe, Kurtosis) und fassen Sie die Ergebnisse in einer Übersichtstabelle zusammen.
- c. Erstellen Sie für jedes Oxid ein Histogramm.
 - d. Schreiben Sie einen kurzen Bericht (ca. 100 Wörter). Diskutieren Sie die deskriptiven statistischen Werte und die Histogramme für jedes Oxid in Hinblick auf zentrale Tendenz (mittels Lageparameter), Streuung und Gestalt der Verteilungen.

Teil 2: Bivariate Statistik

- e. Erstellen Sie 6 Scatterplots mit jeweils SiO₂ auf der x-Achse und je einem der anderen 6 Oxide auf der y-Achse.
 - f. Berechnen Sie den Korrelationskoeffizienten zwischen SiO₂ und den restlichen sechs Oxiden.
 - g. Führen Sie lineare bivariate Regressionen für SiO₂ (unabhängige Variable) und den restlichen sechs Oxiden (abhängige Variablen) durch. Integrieren Sie die Regressionsgeraden in die Scatterplots.
 - h. Kann anhand der explorativen Betrachtung angenommen werden, dass die Voraussetzungen für eine lineare Regressionsanalyse gegeben sind? Was beobachten Sie? Interpretieren Sie die Ergebnisse in einem kurzen Bericht (100-150 Wörter)!
2. Erstellen Sie ein Lösungsdokument, das Ihre Berichte und die erstellten Grafiken (Histogramme, Scatterplots etc.) beinhaltet.