

Aufgabenblatt: Einführung in die statistische Datenanalyse mit R

Allgemeine Informationen

Die Bearbeitung der nachfolgenden Aufgaben ist Teil Ihrer Prüfungsleistung im Seminar “Einführung in die statistische Datenanalyse mit R”. Sie können insgesamt 25 Punkte erzielen, die zur Gesamtpunktzahl Ihrer Klausur am Semesterende addiert werden. Die Punktzahl jeder Aufgabe ist zur Orientierung jeweils in Klammern angegeben. Insgesamt macht dieses Aufgabenblatt 25% Ihrer Endnote aus.

Lösen Sie die Aufgaben, indem Sie gestellte Fragen beantworten und den dazu notwendigen R-Code in ein R-Skript schreiben. Eine Skriptvorlage können Sie im [GitHub-Repository](#) zur Veranstaltung einsehen und kopieren.¹

Das fertige R-Skript reichen Sie per E-Mail an david.bencek@ifw-kiel.de ein.

Abgabefrist: 10. Januar 2016, 23:59 Uhr

Aufgabe 1 (3 Punkte):

Lesen Sie den Aufsatz

Buhaug, H., Gates, S. (2002): The Geography of Civil War. Journal of Peace Research 39(4), 417–433.

und beantworten Sie folgende Fragen:²

- a) Was ist die Forschungsfrage des Aufsatzes?
- b) Was ist (sind) die abhängige(n) Variable(n), die die Autoren untersuchen?
- c) Erklären Sie die Operationalisierung der abhängigen Variable(n).
- d) Listen Sie kurz die relevanten unabhängigen Variablen auf.
- e) Was ist die Analyseeinheit der Studie?
- f) Vergleichen Sie die Fragestellung und den Analyseansatz mit dem im Rahmen des Seminars nachvollzogenen Aufsatz von Bethany Lacina (Lacina, B. (2006): Explaining the Severity of Civil Wars, Journal of Conflict Resolution 50(2), 276-289).

¹https://raw.githubusercontent.com/davben/stats-with-r/master/R/exercise_template.R

²Die pdf-Version des Aufsatzes ist über OLAT im Materialien-Ordner (im Unterordner “Aufgabenblatt”) verfügbar. Sollten Sie darauf keinen Zugriff haben, bringt Sie auch eine Suche bei scholar.google.com zum Ziel.

Aufgabe 2 (3 Punkte):

Verschaffen Sie sich einen Überblick zu den im Aufsatz analysierten Daten **mithilfe von R**. Die dazu notwendige csv-Datei finden Sie bei OLAT im Materialien-Ordner sowie im [GitHub-Repository](#).

- Wieviele Beobachtungen umfasst der Datensatz?
- Welcher Zeitraum wird von den Daten abgedeckt?
- Erstellen Sie einen Scatterplot für jede der abhängigen Variablen und zeigen Sie damit ihre Entwicklung in Abhängigkeit vom Konfliktbeginn. Können Sie einen Trend erkennen?
- Plotten Sie Histogramme der abhängigen Variablen.

Aufgabe 3 (4 Punkte):

- Vollziehen Sie die deskriptiven Statistiken der Tabelle I im Aufsatz von Buhaug/Gates nach und berechnen Sie alle angegebenen Werte.
- Erstellen Sie einen *data frame*, der exakt Tabelle I entspricht.

Aufgabe 4 (5 Punkte):

Tabelle II zeigt eine Matrix aller paarweisen Korrelationskoeffizienten der betrachteten Variablen.

- Was sagt Ihnen ein einzelner Wert in Tabelle II?
- Was sagt der Korrelationskoeffizient von “Identity” und “Incompatibility” aus? Interpretieren Sie kurz.
- Der Befehl `cor()` berechnet die Korrelation zweier Variablen. Wenn Sie ihn auf einen *data frame* anwenden, erhalten Sie alle paarweisen Korrelationskoeffizienten. Rechnen Sie damit Tabelle II nach. (Tips: (i) Erstellen Sie zunächst einen *data frame*, der nur die benötigten Variablen enthält. (ii) Das Argument `use` ist hilfreich, lesen Sie mehr darüber auf der Hilfeseite von `cor()`)

Aufgabe 5 (5 Punkte):

Replizieren Sie die Regressionsmodelle 1 bis 6. Betrachten Sie jeweils die Modellzusammenfassungen und geschätzten Koeffizienten. Beschreiben Sie kurz, was Sie über den Einfluss der einzelnen unabhängigen Variablen erfahren.

Aufgabe 6 (5 Punkte)

Hypothese: *Eine bessere Infrastruktur begünstigt die Ausweitung von Konflikten innerhalb eines Landes.*

Verwenden Sie als Approximation der Infrastrukturqualität das BIP pro Kopf zu Beginn jedes Konflikts im Datensatz. Finden Sie die notwendigen Daten in einer Datenbank Ihrer Wahl (mögliche Quellen sind: Weltbank, IWF, OECD), verknüpfen Sie die Informationen zum BIP pro Kopf mit dem Datensatz und schätzen Sie die Modelle aus Aufgabe 5 nochmals

mit der neuen Variable. Beschreiben Sie kurz ihre Erkenntnis zum Einfluss vom BIP pro Kopf und treffen Sie eine Aussage zur obenstehenden Hypothese.

Hinweise: (i) Das im Seminar gezeigte Paket `countrycode` ist hilfreich, um die einzelnen Länder zu identifizieren und aus unterschiedlichen Quellen zusammenzuführen. (ii) Reichen Sie bei Abgabe Ihrer Lösung auch Ihre Rohdaten aus der Datenbank ein.