

Übung 03 - Datenanalyse und Statistik WS 2015/2016

Aufgabe 1: Skala, Grundgesamtheit und deskriptive Analyse

Laden Sie die Datei "Hydrochem.csv" von der Website herunter, wo sich auch die Übungsblätter befinden. Öffnen Sie es in R und führen Sie eine univariate Beschreibung jeder Variable dieses Datensatzes, graphisch (`plot`, `hist`, `barplot`) sowie numerisch (`summary`). Diskutieren Sie die Skala jeder Variable. Die Daten beinhalten hydrochemische Analyse Flusswasser Proben monatlich entlang 3 Jahre aus 30 Messstationen im einem Flussbecken Nord-Ost Spaniens, mit verschiedenen Unterbecken, Hauptflüsse und Nebenflüsse. Welche Grundgesamtheit(en) sind hier dargestellt?

Aufgabe 2: Boxplots

Führen Sie die folgende Zeilen aus und lesen die entsprechende Hilfeseiten.

```
byRiverData = split(Hydrochem, Hydrochem$River)
par(mfrow=c(2,2))
for(i in 1:4){ boxplot(-log10(H)~tributary, data=byRiverData[[i]]) }
```

Was ist die Klasse von `byRiverData`? Was ist ihr Inhalt? Was wird durch `split` genau erreicht? Warum wird die transformation `-log10` auf Variable `H` angewendet? Produzieren Sie äquivalente Diagramme für die Variablen `NO3` und `NH4` in ihren geeigneten Skalen. Vergleichen Sie Streuung und Lage zwischen Haupt- und Nebenflüssen.

Aufgabe 3: Zwei-Stichproben Testbatterie

Das Ziel dieser Aufgabe ist, die Eigenschaften von `pNO3` in den Nebenflüssen gegen ihren Hauptfluss zu vergleichen, und zwar für jedes Unterbecken. Deshalb werden wir folgende Tests führen. Zuerst nehmen wir an, dass die `pNO3` Messungen in allen Gruppen normalverteilt sind:

- für jeden Becken, testen wir ob die Messungen im Hauptfluss und in Nebenflüsse gleicher Varianz haben;
- für jeden Becken, testen wir ob die Messungen im Hauptfluss und in Nebenflüsse gleicher Mittelwert haben. Diese Tests sollten die Ergebnisse von (i) berücksichtigen;
- falls beide Tests positive Ergebnisse liefern, testen ob beide Verteilungen (Nebenflüsse-Hauptfluss) gleich sind.

Graphische Darstellungen von diesen Vergleichen können erzeugt werden, mit den Befehlen `boxplot` (evtl. notching aktiviert) und `qqplot`. Jetzt testen Sie ob die Annahme zu Normalverteilung angemessen ist oder nicht. Wenn nicht, führen Sie alternative Tests zu gleicher Varianz und gleichem Mittelwert durch. Beachten Sie, dass in einer Schleife Ergebnisse nicht automatisch ausgegeben werden. Benutzen Sie `print`. Fassen Sie tabularisch ihre Ergebnisse zusammen.

Aufgabe 4: Korrelationstests

Erzeugen Sie Streudiagramme von `Ba`, `TOC`, `SO4` und `Cl` gegen `Na`, in ihren geeigneten Skalen. Ordnen Sie die 4 Diagramme nach (scheinbar) zunehmender Korrelation. Berechnen Sie die Pearson und Spearman Korrelationskoeffizienten, und testen Sie, ob sie Null sind.

Aufgabe 5: Fishers exakter Test

Mithilfe des Befehls `fisher.test` testen Sie, ob die Überlebensrate in der verunglückten Titanic von Geschlecht oder von Alter abhängt (Seite 29/Overlay 36 von den Vorlesungsfolien "Vor3.pdf" könnte sehr inspirierend sein).