

Bernd Klaus (bernd.klaus@imise.uni-leipzig.de)
Verena Zuber (verena.zuber@imise.uni-leipzig.de)

<http://uni-leipzig.de/~zuber/teaching/ws10/r-kurs/>

1 Aufgabe: Google-Suche in Deutschland

Im Datensatz "google.csv" finden Sie getrennt nach dem Bundesland die absolute Suchhäufigkeit nach:

- "Krise" im Jahr 2008 (Krise08)
- "Krise" im Jahr 2009 (Krise09)
- "Arbeitsamt" im Jahr 2010 (Arbeitsamt)

Dazu enthält der Datensatz einen Ost-West Indikator (OW) (Berlin wird von dieser Kategorisierung ausgenommen).

- Lesen Sie den Datensatz ein und wandeln Sie die Variable "OW" in einen Faktor um, der fehlende Wert zulässt und mit 0 West und 1 Ost kodiert.
- Führen Sie den geeigneten t -Test durch, um zu untersuchen, ob nach dem Begriff "Krise" im Jahr 2008 stärker gesucht wurde als im Jahr 2009. Erstellen Sie zusätzlich einen Boxplot und führen Sie den entsprechenden nonparametrischen Test durch. Interpretieren Sie die Ergebnisse.
- Überprüfen Sie mittels des Ost-West Indikator, ob diese Abnahme der Suche nach dem Begriff "Krise" im Osten stärker ausgeprägt ist als im Westen.
- Untersuchen Sie die Hypothese, dass im Osten häufiger nach dem Begriff "Arbeitsamt" gesucht wird als im Westen. Begründen Sie Ihre Entscheidung.

2 Aufgabe: Der α -Fehler

- Ziehen Sie 100 normalverteilte Zufallszahlen mit den Parametern $\mu = 0$ und $\sigma = 1$ und speichern Sie diese in dem Objekt `X1`. Erstellen Sie analog ein Objekt `X2` mit 100 (neuen) normalverteilten Zufallszahlen mit den Parametern $\mu = 0$ und $\sigma = 1$.
- Führen Sie den t -Test durch, um zu untersuchen, ob die beiden Objekte unterschiedliche Mittelwerte besitzen.
- Schreiben Sie eine Funktion, die
 - die Teilaufgaben (a) und (b) n -mal durchführt
 - in jedem Schleifenschritt soll nur die t -Statistik in einem Vektor abgespeichert werden
- Testen Sie Ihre Funktion mit $n = 1000$. Plotten Sie die Verteilung der t -Statistiken mit einem Kerndichteschätzer. Wie viele t -Statistiken sind im Betrag fälschlicherweise größer als der kritische Wert von $\kappa = 1.972$?

3 Aufgabe: Blutdruckvergleich

Der Datensatz *Calcium.csv* enthält Daten über den Einfluss von Calcium auf den Blutdruck. 12 Wochen lang erhielt eine Gruppe von 10 Männern ein Calciumpräparat und eine andere Gruppe von 11 Männern ein Placebo. Dabei bedeuten:

- **Treatment:** Gruppenzuordnung (Calcium oder Placebo)
- **Begin:** Blutdruck zu Beginn der Studie
- **Ende:** Blutdruck nach der Behandlung
- **Decrease:** Blutdruckdifferenz (Anfang - Ende)

- Lesen Sie den Datensatz mit Hilfe des `read.table()`-Befehls ein.
- Erstellen Sie einen geeigneten Boxplot, um zu untersuchen, ob ein Unterschied zwischen den zwei Gruppen besteht.
- Untersuchen Sie mit Hilfe einer geeigneten Hypothese und eines geeigneten *t*-Tests, ob der Unterschied zwischen den Behandlungsgruppen statistisch signifikant ist. Überprüfen Sie auch, ob die Voraussetzungen des Tests erfüllt sind.

4 Aufgabe: Testen der Füllmenge

Die Kontrolleure des Leipziger Amt für Verbraucherschutz sind alarmiert worden, dass auf dem Weihnachtsmarkt zu wenig Glühwein in den Bechern ausgeschenkt wird. Sie entschließen sich zum Selbstversuch. An einem Abend messen sie die folgenden Werte in ihren 12 Glühweingläsern:

1	2	3	4	5	6	7	8	9	10	11	12
0.240	0.249	0.251	0.252	0.244	0.241	0.247	0.261	0.225	0.212	0.264	0.201

- Lesen Sie den Datensatz *Weihnachtsmarkt.csv* ein.
- Untersuchen Sie mit dem geeigneten *t*-Test, ob signifikant weniger als 0.25l in den Bechern zu finden war!

Die Kursleiter wünschen ein schöne Adventszeit!