

Bernd Klaus (bernd.klaus@imise.uni-leipzig.de)

Verena Zuber (verena.zuber@imise.uni-leipzig.de)

<http://uni-leipzig.de/~zuber/teaching/ws09/r-kurs/>

1 Aufgabe: Einfaktorielle ANOVA: Pisa Score in Abhängigkeit von der Ausbildung des Vaters

Die Leistung von Kindern in der Schule gilt als hochgradig abhängig vom Bildungsgrad der Eltern. Im Datensatz *PISAnova* finden Sie folgende Variablen:

- **Country**: Land
 - **Degree**: Bildungsabschluß des Vaters: (1) Universitätsabschluß bis (5) kein Schulabschluß
 - **M**: Mittlerer PISA-Score Mathematik
 - **R**: Mittlerer PISA-Score Lesen
 - **S**: Mittlerer PISA-Score Naturwissenschaften
- (a) Lesen Sie den Datensatz mit Hilfe des `read.table()`-Befehls ein und stellen Sie sicher, dass `Degree` als Faktor verwendet wird.
- (b) Untersuchen Sie mit einer Varianzanalyse, ob der mittlere PISA-Score in den Bereichen Mathematik, Lesen und Naturwissenschaften von der Ausbildung des Vaters abhängt. Welche Aussagen können Sie mittels einer Varianzanalyse untersuchen?
- (c) Betrachten Sie genauer den mittleren PISA-Score Mathematik. Erstellen Sie dazu einen Boxplot. Zwischen welchen Bildungsabschlüssen finden Sie signifikante Unterschiede?

2 Aufgabe: t-Tests auf Gruppenunterschiede

Der Datensatz *Singers.csv* enthält Körpergrößen (in inches) eines New Yorker Orchesters in Abhängigkeit von der Stimmlage.

- **Soprano**: Hohe Frauenstimme
 - **Alto**: Tiefe Frauenstimme
 - **Tenor**: Hohe Männerstimme
 - **Bass**: Tiefe Männerstimme
- (a) Lesen Sie den Datensatz mit Hilfe des `read.table()`-Befehls ein.
- (b) Erstellen Sie einen Boxplot der Daten.
- (c) Versuchen Sie, Vermutungen über die Unterschiede zwischen den einzelnen Gruppen, die sich aus dem Boxplot ergeben, mit Hilfe geeigneter *t*-Tests zu bestätigen.
- (d) Nutzen Sie Ihre Ergebnisse aus der vorangehenden Aufgabe, um die Sänger in zwei Gruppen einzuteilen. Gibt es zwischen diesen Gruppen einen signifikanten Unterschied? Welche Schlussfolgerung lässt sich aus der Antwort auf die vorangegangene Frage ziehen?

3 Aufgabe: Mehrfaktorielle ANOVA: Der Untergang der Titanic

“Frauen und Kinder zuerst!” lautet die Losung beim Untergang eines Schiffes. Eine zusätzliche interessante Frage ist, ob die Reiseklasse ebenfalls Einfluss auf die Überlebenswahrscheinlichkeit hat. Im Datensatz *titanic* finden Sie folgende Variablen:

- **Class:** Reiseklasse (1), (2) oder (3)
 - **Age:** (1) Erwachsene (0) Kinder
 - **Sex:** (1) Männer (0) Frauen
 - **Survival:** Mittlere Überlebenswahrscheinlichkeit
- (a) Lesen Sie den Datensatz mit Hilfe des `read.table()`-Befehls ein und stellen Sie sicher, dass `Class`, `Age` und `Sex` als Faktor verwendet werden.
- (b) Untersuchen Sie in je einer Varianzanalyse, ob die Überlebenswahrscheinlichkeit vom Alter, dem Geschlecht und der Reiseklasse abhängig ist.
- (c) Untersuchen Sie in einer gemeinsamen Varianzanalyse, ob die Überlebenswahrscheinlichkeit vom Alter, dem Geschlecht und der Reiseklasse abhängig ist.
- (d) Können Sie aus diesen Daten folgern, dass Alter, Geschlecht und Reiseklasse die Überlebenswahrscheinlichkeit beeinflussen? Rufen Sie mit `?Titanic` den Ursprungsdatensatz in **R** auf. Haben Sie eine Idee, wie die Fragestellung anders untersucht werden könnte?