

## Schriftliche Ausarbeitung zum Erwerb des Leistungsnachweises

Nachfolgend finden Sie Aufgaben, die sich auf den Datensatz **leistungstest.dat** beziehen. Bitte senden Sie **bis 21.04.2017** eine E-Mail an [olga.kunina-habenicht@dipf.de](mailto:olga.kunina-habenicht@dipf.de), die folgende Dokumente als Anlage enthält:

- eine schriftliche Ausarbeitung zu den nachfolgenden Aufgaben (Namen und Immatrikulationsnummer nicht vergessen!) sowie
- das R-Skript zu den von Ihnen ausgeführten Analysen, die den gesamten Verlauf Ihrer Datenanalyse dokumentieren (Einlesen der Daten, ggf. Datenaufbereitung, Analysen).

Bitte notieren Sie die zentralen Ergebnisse und Antworten zu diesen Aufgaben in Ihrer schriftlichen Ausarbeitung und achten Sie darauf, die Ausarbeitung sowie das Skript entsprechend der Aufgaben durchnummerieren.

### Der Datensatz: **leistungstest.dat**

Ihnen liegt ein Datensatz vor, der Informationen von 1425 Schülerinnen und Schülern aus fünf Schulen (*schule*) in den Klassenstufen (*klasse*) 6, 7 und 8 enthält. Für alle Schüler liegt das Ergebnis dreier Leistungstests (d.h. die erreichte Punktzahl) in den Fächern Mathematik (max. 60 Punkte, *mathe*), Deutsch (max. 50 Punkte, *deutsch*) und Englisch (max. 60 Punkte, *englisch*) vor. Der Datensatz enthält zudem Informationen zum Geschlecht (*geschl*), zur letzten Zeugnisnote in Englisch (*englisch\_note*) sowie über die vom Schüler zu Beginn angegebene Motivation zur Testbearbeitung (*motivation*), wobei Letztere auf Ordinalskalenniveau (1 – sehr gering bis 6 – sehr hoch) erfasst wurde.

Viel Erfolg!

### **Hinweise**

*Notieren Sie in Ihrer Ausarbeitung und im Skript bitte Ihren Namen sowie Ihre Matrikelnummer.*

*Für die Aufgaben 1, 4.1 und 7 reicht es aus, wenn Sie diese im Skript abhandeln. Sie brauchen sie nicht in der schriftlichen Ausarbeitung zu beantworten.*

*Aufgaben, die durch ein \* gekennzeichnet sind, erfordern R-Wissen, dass im Workshop nicht explizit behandelt wurde. Die Bearbeitung dieser Aufgaben ist daher freiwillig. Versuchen Sie es trotzdem - nur dann erhalten Sie Rückmeldung!*

*Bei der Erstellung von Grafiken dürfen Sie sich gern austoben. ☺*

# Aufgaben

1. Lesen Sie den Datensatz „leistungsdaten.dat“. Verschaffen Sie sich einen Überblick über dessen Struktur. Sortieren Sie den Datensatz nach Schulen und Klassen.
  - 1.1 Wandeln Sie die Variable zum Geschlecht in einen Faktor um, wobei „0“ männliche und „1“ weibliche Testteilnehmer bezeichnet.
  - 1.2 Auf welchen Variablen gibt es fehlende Werte?
2. An welcher Schule nahmen die meisten Schüler teil? Bestimmen Sie das Skalenniveau der Variablen *schule*. Berichten Sie (angemessene) statistische Kennwerte.
3. Erstellen Sie eine übersichtliche Tabelle über die Geschlechterverteilung in den drei Klassenstufen. Wie viele Mädchen gibt es in der siebten Klassenstufe? Welchem prozentualen Anteil an allen Testteilnehmern entspricht diese Gruppe?
  - 1.1 Beschreiben Sie den Aufbau der von Ihnen erstellten Tabelle.
  - 1.2 Stellen Sie die prozentuale Auftretenshäufigkeit von Jungen und Mädchen in den verschiedenen Klassenstufen grafisch dar und beschreiben Sie den Aufbau Ihrer Grafik.
4. Vergleichen Sie deskriptiv die Deutschleistung der Achtklässler zwischen den fünf Schulen. Berichten Sie den Median und den Mittelwert für die Gesamtstichprobe. Erstellen Sie einen Boxplot der Verteilung der Deutschleistung pro Schule und zeichnen Sie zusätzlich (und gut erkennbar) den jeweiligen Mittelwert ein.
  - 4.1 Ein Schüler aus Schule 3 hat in Deutsch einen hohen Punktwert (= 47 Punkte) erhalten. Sie halten diesen Schüler für einen Ausreißer und möchten seinen Punktwert - natürlich nach langem und äußerst sorgfältigem Abwägen - aus dem Datensatz entfernen. Kennzeichnen Sie den Wert im Datensatz als fehlend (NA).
  - 4.2 Verändern sich der Median und der Mittelwert, nachdem Sie den Ausreißerwert entfernt haben? Wenn ja, warum? Wenn nein, warum nicht?
5. Bilden Sie einen Summenscore für das allgemeine schulische Leistungsniveau aus den gemessenen Leistungen im Deutsch-, Mathe- und Englischtest. Beschreiben Sie die Verteilung dieser neuen Variablen anhand geeigneter deskriptivstatistischer Kennwerte. Erstellen Sie ein Histogramm der Verteilung. Führen Sie mindestens zwei Veränderungen an dieser Grafik durch, die sie von der Standard-Ausgabe unterscheidet. (Hier dürfen Sie sich austoben 😊)
6. Betrachten Sie nun den Zusammenhang zwischen der Motivation zur Testbearbeitung und der Matheleistung. Berechnen Sie ein geeignetes Zusammenhangsmaß und begründen Sie Ihre Wahl. Was besagt der Koeffizient inhaltlich? Erstellen Sie zur grafischen Anschauung zusätzlich ein Streudiagramm.
7. Wussten Sie, dass R auch Tabellenblätter aus Excel (.csv Dateien) einlesen kann? Suchen Sie nach einem geeigneten Befehl und laden (und überprüfen) Sie den Datensatz *leistungsdaten.csv*! \*