

Grundlage dieser Gruppenarbeit ist der Datensatz "datensontbest2.RData", den Sie aus dem OC runterladen können. Eine Beschreibung dieses Datensatzes "Beschreibung des Datensatzes datensontbest2.pdf" finden Sie ebenfalls im OC. Der Datensatz beinhaltet 19 Variablen, von denen Sie aber nur drei Variablen benötigen.

Die beiden ersten Aufgaben sind per Hand zu berechnen. Die weiteren Aufgaben sind mittels R zu berechnen.

1. Die folgenden Berechnungen sollen nur für die ersten fünf Werte der Variable *note\_p*, d.h. der Werte für die ersten 5 Personen, per Hand vorgenommen werden. Beschreiben Sie dabei auch den Rechenweg. Berechnen Sie für diese Variable
  - a) den Mittelwert, die Varianz und die Standardabweichung
  - b) die z-standardisierten Werte (für die ersten 5 Personen)
  - c) Erklären Sie bitte kurz den wesentlichen Unterschied zwischen dem z-standardisierten Wert und dem ursprünglichen Wert (Note) für die erste Person.
2. Berechnen Sie die folgenden Statistiken ebenfalls nur für die ersten 5 Personen (per Hand)
  - a) Kovarianz zwischen zwischen den Variablen *note\_p* und *sk\_p*
  - b) Korrelation zwischen zwischen den Variablen *note\_p* und *sk\_p*

Nehmen Sie die folgenden Berechnungen mittels R für alle 100 Personen vor. Berichten Sie dabei jeweils den Output der Berechnungen mittels R. Alle notwendigen R-Befehle sind in den Dateien "R-Befehle für die deskriptive Statistik" und "R-Befehle für die Inferenzstatistik" enthalten. Beide Dateien können Sie aus dem OC runterladen.

3. Berechnen Sie bitte für die Variable *note\_p*
  - a) die Häufigkeitstabelle mit den absoluten Häufigkeiten, den relativen Häufigkeiten und der empirischen Verteilungsfunktion anhand der absoluten und relativen Häufigkeiten
  - b) die Fünf-Punkte-Zusammenfassung
  - c) den Mittelwert, die Varianz und Standardabweichung
  - d) das Histogramm
  - e) den Boxplot
    - i. Wieviele Ausreißer zeigt der Boxplot an?
    - ii. Wieviele Extremwerte zeigt der Boxplot an?

4. Berechnen Sie

- a) den Korrelationskoeffizienten zwischen den Variablen *note\_p* und *sk\_p*
- b) Wie ist das Ergebnis in inhaltlicher Hinsicht anhand von zwei, drei Sätzen zu interpretieren?
- c) Stellen Sie zu der inhaltlichen Hypothese, dass eine positive Korrelation zwischen diesen beiden Variablen besteht, die statistische Null- und Alternativhypothese auf.
- d) Überprüfen Sie diese statistische Nullhypothese anhand eines Signifikanzniveaus von  $\alpha = .05$ .
- e) Wie groß ist der *p*-Wert?
- f) Muss die Nullhypothese bei einem Signifikanzniveau von 5% verworfen werden? Begründen Sie bitte ihre Antwort.
- g) Geben Sie bitte das 95%ige Konfidenzintervall an.
- h) Kann man anhand dieses Konfidenzintervalls entscheiden, ob die Nullhypothese bei einem Signifikanzniveaus von  $\alpha = .05$  zu verwerfen ist? Begründen Sie bitte Ihre Meinung.

5. Überprüfen Sie anhand eines geeigneten Tests, ob sich die Erwartungswerte der Variable *note\_p* für Schülerinnen und Schüler signifikant unterscheiden.

- a) Stellen Sie zu diesem Testproblem die statistische Null- und Alternativhypothese auf.
- b) Wie groß ist der *p*-Wert?
- c) Muss die Nullhypothese bei einem Signifikanzniveau von 5% verworfen werden? Begründen Sie bitte ihre Antwort.
- d) Interpretieren Sie das Ergebnis dieses Tests aus inhaltlicher Sicht.