

## Aufgaben zur Vorlesung „Werkzeuge der empirischen Forschung“

**Aufg. 19)** (Sterblichkeit und Wasserhärte)

Hängen Wasserhärte und Sterblichkeit mit der Lage der Orte zusammen?

**Aufg. 20)** (Ägyptische Schädel)

- a) (2 P.) Welche Hypothesen lassen sich formulieren? Bitte als Kommentar angeben.
- b) (4 P.) Berechnen Sie, ausgehend von der Datei skull.dat für alle Variablen elementare statistische Maßzahlen und erstellen Sie Boxplots, wobei die Boxen für die vier Epochen nebeneinander stehen sollen.
- c) (2 P.) Führen Sie eine einfache Varianzanalyse durch. Die Daten können hier als approximativ normalverteilt angenommen werden (wie überprüft man das?) Die Varianzen seien als homogen angenommen (vgl. Box-Plot).
- d) (1 P.) Wie überprüft man die Varianzhomogenität?

**Aufg. 21)** (Nicht-Robustheit des t-Tests)

Diese Aufgabe soll bestimmte Probleme bei Durchführung des t-Tests illustrieren. Die Ergebnisse sollen später mit denen des Wilcoxon-Tests verglichen werden.

- a) (3 P.) Seien die folgenden fiktiven Beobachtungen gegeben:  $\frac{X : | c \ c \ c \ c \ c}{Y : | c \ c \ c \ c \ y}$  mit  $c = 1$  und  $y \neq c$ . Vergleichen Sie die beiden Stichproben mit Hilfe des (unverbundenen) Zweistichproben t-Tests, jeweils für die Werte  $y = 2$ ,  $y = 5$  und  $y = 10$ .
- b) (2 P.) Seien jeweils die folgenden unverbundenen fiktiven Beobachtungen gegeben:  $\frac{X : | 1 \ 2 \ 3}{Y : | 0 \ 0 \ 0}$  bzw.  $\frac{X : | 1 \ 2 \ 3 \ 10}{Y : | 0 \ 0 \ 0}$  Führen Sie jeweils einen Zweistichproben t-Tests durch.
- c) (+2 P.) Führen Sie für die in Aufgabe 20a und 20b gegebenen Beobachtungen jeweils einen Wilcoxon-Test durch.
- d) (+1 P.) Es bietet sich wegen des geringen Stichprobenumfangs an, exakte Tests zu verwenden. Versuchen Sie es!

**Aufg. 22)** (2 P.) Geben Sie (verbal) zwei Aufgabenstellungen aus der Praxis an, die zum Zweistichprobenproblem führen, je eine für verbundene Stichproben und eine für unverbundene Stichproben. Bitte verwenden Sie keine Beispiele aus Vorlesung und Übung! Es genügen jeweils zwei Zeilen.

**Aufg. 22+** (+3 P.) (Power von t-Test und Wilcoxon-Test bei Normalverteilung)

Generieren Sie jeweils 100 pseudozufällige Vektoren  $X_i$  bzw.  $Y_i$  der Länge 100, deren Komponenten wie folgt verteilt sind:

$$X_{i,j} \sim \mathcal{N}(0, 1) \text{ und } Y_{i,j} \sim \mathcal{N}(0.2, 1).$$

Führen Sie mit diesen 200 Vektoren je einen (zweiseitigen) t-Test und einen Wilcoxon-Test der Nullhypothese  $\mu = 0$  durch. Geben Sie die vier Anzahlen der signifikanten ( $\alpha = 0.05$ ) Ergebnisse jeweils für  $X$  bzw.  $Y$  und jeden der Tests als Kommentar an. Interpretieren Sie die Ergebnisse. Setzen sie zur Nachvollziehbarkeit vor dem Experiment den Pseudozufalls-generator mit `set.seed` (Woher bekommen Sie einen geeigneten Seed?).