



Prof. Dr. Antony Unwin, Dr. Ali Ünlü
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg
<http://stats.math.uni-augsburg.de/>

Stochastik II — Statistik I (Statistik und Datenanalyse, SS 2009)

Übungsblatt 1

Abgabe: Dienstag 28. April 2009, bis spätestens 12.00 Uhr; Briefkasten: Stochastik II

1. Laden Sie den Datensatz **Zehnkampf2006** von der Internetseite der Vorlesung

<http://stats.math.uni-augsburg.de/lehre/SS09-f/S1.shtml>

Beschreiben Sie kurz die Bedeutung der Variablen des Datensatzes und bestimmen Sie für jede Variable den Datentyp. Wie kann Ihnen die Software R dabei helfen?

2. Bestimmen Sie für jede stetige Variable aus diesem Datensatz

- (a) Mittelwert,
- (b) Median,
- (c) Varianz und
- (d) Interquartilabstand.

Da es relativ viele Variablen gibt, sollten Sie sich überlegen wie Sie geschickt in R vorgehen können. Welche Kennzahlen/Zusammenfassungen würden Sie für die kategoriellen Variablen verwenden? Bestimmen Sie diese auch. Diskutieren und vergleichen Sie Ihre Ergebnisse.

3. Untersuchen Sie die stetigen Variablen auf mögliche Zusammenhänge mittels Streudiagrammen. Berechnen Sie Korrelationen zwischen den stetigen Variablen. Gehen Sie auch auf Ausreisser ein. Beschreiben Sie hierbei zwei Ihnen als wichtig erscheinende Ergebnisse.
4. Betrachtet wird wiederum der **Zehnkampf2006** Datensatz:
 - (a) Zeichnen Sie für die Variable P_{100m} ein Histogramm und die empirische Verteilungsfunktion.
 - (b) Bestimmen Sie graphisch und rechnerisch Median und untere und obere Quartile.
 - (c) Fügen Sie in das Histogramm und in den Plot der empirischen Verteilungsfunktion die theoretische Dichte bzw. Verteilungsfunktion einer "geeigneten" Normalverteilung hinzu. (Wie wird μ und σ^2 zu bestimmen sein?) Wie unterscheiden sich die Kurven und die Kennzahlen Median und Quartile?
5. Bearbeiten Sie die nachfolgenden Fragen über den Datensatz **Zehnkampf2006**.
 - (a) Bestimmen Sie den Zehnkämpfer mit der sechst-langsamsten Zeit im 100m-Lauf. Wie heisst dieser und für welchen Staat tritt er an? Bestimmen Sie dessen Sprunghöhe und erzielte Punkte im Hochsprung.
 - (b) Bestimmen Sie für alle Zehnkämpfer das Verhältnis der Zeiten für den 100m- und 400m-Lauf. Für welchen Athleten liegt das kleinste bzw. größte Verhältnis vor, und was sind die konkreten Zahlenwerte? Wie groß ist das Verhältnis beim Zehnkämpfer "Zhu Hengjun" und welchen Rang nimmt er ein?