

Aufgabe 1.

- a) Definieren Sie die Begriffe Varianz und Standardabweichung. Erläutern Sie, wie diese Streuungsmaße in der deskriptiven Statistik verwendet werden, um die Verteilung von Daten zu beschreiben. Diskutieren Sie die Unterschiede zwischen diesen beiden Maßen und geben Sie an, in welchen Situationen eines der Maße dem anderen vorzuziehen ist.
- b) Erläutern Sie die Bestandteile und die Bedeutung der Formel für die Varianz.

Aufgabe 2.

Ein Datensatz von Testergebnissen einer Klasse von 36 Schülern ergab eine durchschnittliche Punktzahl von 75 bei einem maximal möglichen Ergebnis von 100 Punkten. Die Varianz der Testergebnisse war 144.

- a) Interpretieren Sie diese Ergebnisse sowie die zugehörige Standardabweichung im Kontext der Leistung der Klasse.
- b) Wie groß ist der Standardfehler? Wie groß ist das 95%-Konfidenzintervall des Mittelwerts, wenn Sie die Näherung 2 für das benötigte Quantil der t-Verteilung verwenden?

Aufgabe 3.

Beschreiben Sie die Eigenschaften der Normalverteilung und erklären Sie, warum die Normalverteilungsannahme in vielen statistischen Tests (z.B. t-Test) und Modellen wichtig ist.

Aufgabe 4.

Erläutern Sie die Begriffe univariater, bivariater und multivariater Datensatz. Geben Sie für alle drei Typen von Datensätzen ein Beispiel an, sowie jeweils zwei mögliche graphische Darstellungen. Welche statistischen Modelle kennen Sie für bivariate Daten?

Aufgabe 5.

Diese Aufgabe befasst sich mit dem Datensatz *Einkommen.jmp*. Bearbeiten Sie die Aufgabe sowohl in JMP als auch mit Python.

- a) Führen Sie für alle Variablen im Datensatz eine explorative Datenanalyse durch.
- b) Prüfen Sie die Variable Einkommen auf Normalverteilung.
- c) Führen Sie mit diesem Datensatz eine einfaktorielle Anova sowie einen t-Test durch. Welche Schlussfolgerungen ziehen Sie?

Aufgabe 6.

Diese Aufgabe befasst sich mit dem Auto-Datensatz (auto.csv). Bearbeiten Sie die Aufgabe sowohl in JMP als auch mit Python. Stellen Sie sicher, dass die fehlenden Werte aus den Daten entfernt wurden.

- a) Welche möglichen Prädiktoren sind quantitativ und welche sind qualitativ?
- b) Wie groß ist der Bereich jedes quantitativen Prädiktors?
- c) Wie sind der Mittelwert und die Standardabweichung jedes quantitativen Prädiktors?
- d) Entfernen Sie nun die Beobachtungen 11 bis 80. Welche Reichweite, Mittelwert und Standardabweichung hat jeder Prädiktor im verbleibenden Datensubset?
- e) Untersuchen Sie mithilfe des vollständigen Datensatzes die Prädiktoren grafisch, indem Sie Streudiagramme oder andere Ihnen zur Verfügung stehende Werkzeuge verwenden. Erstellen Sie einige Diagramme, die die Beziehungen zwischen den Prädiktoren hervorheben. Kommentieren Sie Ihre Erkenntnisse.
- f) Angenommen, wir möchten den Benzinverbrauch (mpg) auf Basis der anderen Variablen vorhersagen. Deuten Ihre Diagramme darauf hin, dass eine der anderen Variablen nützlich sein könnte, um mpg vorherzusagen? Begründen Sie Ihre Antwort.