



## Statistik I

### Übungsblatt 1

**Abgabe:** Donnerstag, 04.05.06, 12:00 Uhr, Briefkasten: Statistik I.  
Insgesamt können maximal 25 Punkte erreicht werden.

1. Laden Sie die 2004-Auto-Daten von der Internetseite der Vorlesung, und bestimmen Sie für jede Variable den Datentyp.

Wie kann Ihnen die Software

- (a) R
- (b) Mondrian

dabei helfen?

2. Bestimmen Sie für jede Variable aus den 2004-Auto-Daten

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

Diskutieren und vergleichen Sie die Ergebnisse!

3. Erstellen Sie für die Variablen "Drive" und "Type" eine gemeinsame Tabelle, d.h. zweidimensionale Tabelle mit einer Variablen in den Zeilen und der anderen in den Spalten.

- (a) Welche Zelle ist am häufigsten besetzt?
- (b) Gibt es leere Zellen? Sind diese systematisch motiviert?
- (c) Fassen Sie geeignete Klassen zusammen, um die Tabelle zu vereinfachen!
- (d) Wie könnte eine geeignete graphische Darstellung der Tabelle aussehen?

4. Vergleichen Sie für eine Normalverteilung die Wahrscheinlichkeit, dafür, dass ein Wert innerhalb des Interquartilsabstands liegt, mit der, dass ein Wert im Intervall  $[\mu - \sigma, \mu + \sigma]$  liegt.

Wie ändert sich das Verhältnis bei einer Log-Normalverteilung?

5. Betrachtet wird wiederum der 2004-Auto-Datensatz:

- (a) Zeichnen Sie für die Variable "Horsepower" die empirische Verteilungsfunktion.
- (b) Bestimmen Sie graphisch und rechnerisch Median und Quartile.
- (c) Zeichnen Sie weiter die theoretische Verteilungsfunktion einer "passenden" Normalverteilung in die Graphik aus (a). Wie unterscheiden sich die Kurven und die Quantile aus (b)?