

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)

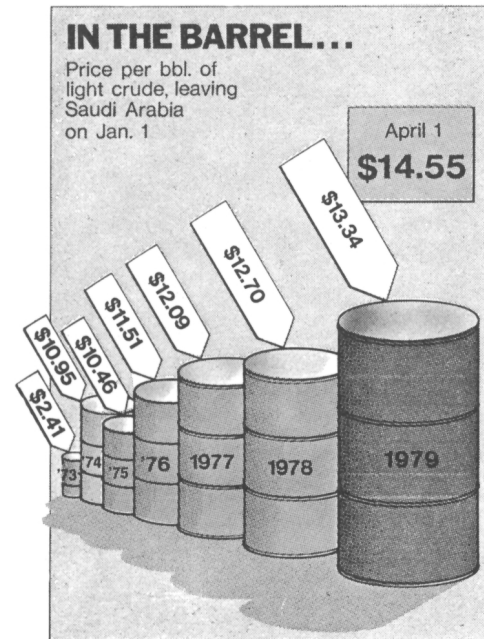
Übungsblatt 2

Abgabe: Dienstag 5. Mai 2009, bis spätestens 12.00 Uhr; Briefkasten: Stochastik II

1. Eine sehr gute Software für interaktive Graphiken ist `Mondrian`. (Sollten Sie Graphiken mit den Übungen abgeben, dann bitte als Screenshots.)
 - (a) Laden Sie diese Software unter
<http://stats.math.uni-augsburg.de/Mondrian/>
herunter und installieren Sie sie.
 - (b) Laden Sie den Datensatz `Titanic` (von der Internetseite der Vorlesung) in `Mondrian`. Erstellen Sie Barcharts (Säulendiagramme), um die darin enthaltenen Variablen zu erkunden. Bestimmen Sie für jede Variable den Datentyp, die Anzahl und Kodierungen der Kategorien. Was beschreiben die einzelnen Variablen?
 - (c) Welche interaktiven Optionen sind in den Barcharts aus vorheriger Teilaufgabe (b) anwendbar? (Die "Reference Card", unter Menü "Help", kann sicherlich behilflich sein; ist sie aber vollständig?) Wie viele weibliche Passagiere aus der 1. Klasse haben überlebt, wieviele nicht? Warum ist hier die Rede von "weiblichen Passagieren" an Stelle von "Frauen" (beschreiben Sie anhand konkreter Zahlen aus dem Datensatz)?
2. Laden Sie den Datensatz `Nikotin` (von der Internetseite der Vorlesung) in `Mondrian`.
 - (a) Erstellen Sie ein Histogramm der Messungen.
 - (b) Wieviele Klassen hat das anfängliche Histogramm, und wird damit die Verteilung der Nikotin-Werte zufriedenstellend wiedergegeben? Variieren Sie Ankerpunkt oder Klassenbreite so wie Sie es für sinnvoll halten. Zu welcher Wahl gelangen Sie (begründen Sie)? Wieviele Klassen hat das Histogramm jetzt? Bestimmen Sie die genaue Klasseneinteilung für die Messungen und die Anzahl der Nikotin-Messungen, die in die jeweiligen Klassen fallen.
 - (c) Welche interaktiven Optionen sind im Histogramm aus vorheriger Teilaufgabe (b) anwendbar?
3. Laden Sie den Datensatz `Cars` (von der Internetseite der Vorlesung) in `Mondrian`.
 - (a) Erstellen Sie für die Variablen "Horsepower" und "City Miles Per Gallon" einen Scatterplot (Streudiagramm), sowie einen Barchart für die Variable "Number of Cylinders".
 - (b) Färben Sie im Barchart die einzelnen Zylindergruppen in verschiedenen Farben. Zeigt der Scatterplot ein konsistentes Muster? Gibt es Ausreißer; wenn ja, welche (beschreiben Sie)?
 - (c) Löschen Sie die bisherige Färbung. Visualisieren Sie durch eine geeignete Färbung die proportionalen Verläufe der 4- und 8-Zylindergruppen in einem Spinogramm der Variablen "Highway Miles Per Gallon". Interpretieren Sie diesen Plot.

4. Analysieren Sie die nebenstehende Graphik.

- Was wird hier Ihrer Meinung nach dargestellt?
- Ist die Visualisierung der Zahlen gelungen? Schätzen Sie den (wahrgenommenen) Unterschied zwischen 1973 und 1979.
- Messen Sie die Höhen der Fässer. Berechnen Sie die mit Farbe ausgefüllten Frontflächen sowie die Volumina der Zylinder. Wozu sollten die Preise proportional sein?
- Was fällt Ihnen sonst noch auf?



Ölpreise im Zeitraum von 1973 bis 1979 (entnommen aus Tufte, S. 62).

5. (a) Berechnen Sie, wie groß der Prozentsatz der Werte einer Normalverteilung ist, die in einem Boxplot als Ausreißer bzw. krasse Ausreißer klassifiziert werden.
- (b) Erzeugen Sie in R eine exponentialverteilte Stichprobe zum Parameter $\lambda = 2$, und zeichnen Sie für diese Stichprobe einen Boxplot, in den Sie innere und äußere Zäune eintragen. Was fällt auf?

6. Einzeiler-Aufgaben.

- (a) Sei $X \sim N(8, 10^2)$. Berechnen Sie die Wahrscheinlichkeit $P(X \geq 12)$. Was gilt für $P(X = 5)$; warum?
- (b) Was macht die folgende R Befehlsfolge (beschreiben Sie):

```
> j <- 0; for (i in 1: 10) j <- j + i; j
[1] 55
```

An welche Geschichte über einen der wichtigsten Mathematiker erinnert Sie das?