

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

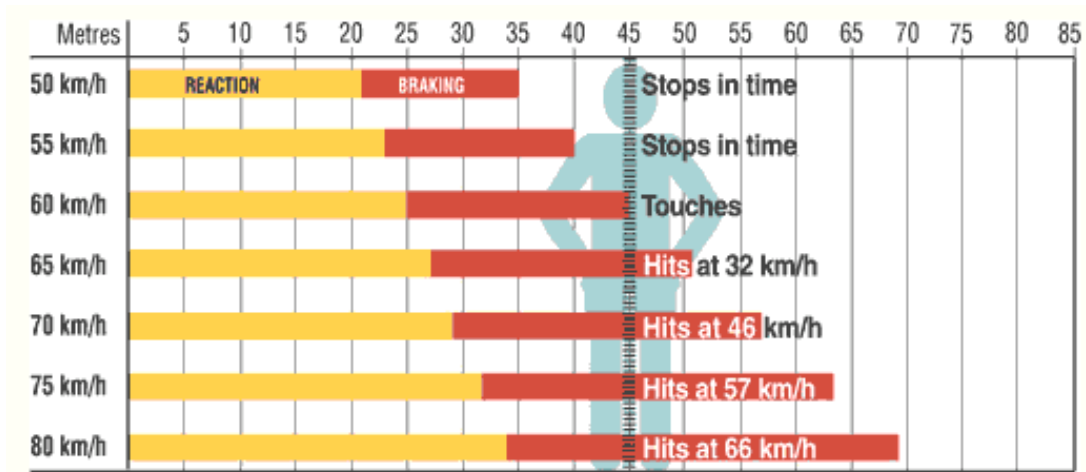
Stochastik II — Statistik I (Statistik und Datenanalyse, SS 2009)

Übungsblatt 5

Abgabe: Dienstag 26. Mai 2009, bis spätestens 12.00 Uhr; Briefkasten: Stochastik II

1. Betrachtet wird die Online-Umfrage zum Thema Abtreibungspille aus Aufgabe 3 von Übungsblatt 3. Bestimmen Sie die Bayes-Schätzer für die Befürwortungsrate p unter den apriori Verteilungen
 - (a) Gleichverteilung auf $[0, 1]$;
 - (b) Betaverteilung mit $a = 1$ und Erwartungswert identisch dem Maximum-Likelihood-Schätzer \hat{p} für p ;
 - (c) Betaverteilung mit $a = b$ und Varianz identisch der Varianz des Maximum-Likelihood-Schätzers (für p). Erhalten Sie einen einzigen Zahlenwert? Welcher Bayes-Schätzer ergibt sich für $p = \hat{p}$?
2. Betrachtet wird der Datensatz `crabs` im Paket `MASS` in R.
 - (a) Was beschreibt dieser Datensatz? Was sind hier die Fälle und Variablen? Wieviele Fälle und Variablen sind im Datensatz enthalten? Welche Variablen sind kategoriell, welche stetig?
 - (b) Vergleichen Sie die morphologischen Messungen mittels Boxplots. Bei welcher Variablen sieht man hierbei einen Ausreisser? Bestimmen Sie den konkreten Zahlenwert für diesen Ausreisser und färben Sie ihn im Boxplot ein.
 - (c) Untersuchen Sie graphisch die morphologischen Messgrößen auf Normalität, mittels QQ-Plots, und Histogrammen und überlagerten Normalverteilungen mit jeweils geeignet geschätzten Erwartungswerten und Varianzen. Erzeugen Sie das Ganze, versehen mit entsprechender Legende, in einem einzigen Graphikfenster.
 - (d) Berechnen Sie die jeweiligen Korrelationen zwischen den morphologischen Messgrößen. Betrachten Sie diese Messgrößen auch in einer Streudiagramm-Matrix. Was fällt auf?
3. Simulieren Sie in R 100-mal 1,000 Realisationen einer Normalverteilung mit den Parametern $\mu = 2$ und $\sigma = 5$. Bestimmen Sie in jedem dieser 100 Fälle das Konfidenzintervall für μ zum Konfidenzniveau 0.95 bei bekannt vorausgesetztem σ^2 . Bestätigt sich die Aussage *‘Wenn wir viele Stichproben ziehen und Konfidenzintervalle aus jeder berechnen, werden wir finden, dass ungefähr 95% der Intervalle den wahren Wert enthalten werden.’*? Stellen Sie das Ganze in geeigneter Weise graphisch dar.

4. Kommentieren Sie die folgende Graphik:



Von Edward Tufte Internetseite

http://www.edwardtufte.com/boardq-and-a-fetch-msg?msg_id=0000Yn
(abgerufen 2. Mai 2008)

- Was ist das Thema der Graphik und welche Informationen stellt sie bereit?
 - Finden Sie das Fehlen einer Überschrift schlecht? Wie könnte eine Überschrift lauten?
 - Was könnte an der Graphik verbessert werden und wie würden Sie das machen? Sollten weitere Informationen in der Graphik gezeigt werden, wenn ja welche?
 - Beurteilen Sie die Graphik in Bezug auf Suggestivität, Übersichtlichkeit und Ästhetik.
5. Betrachtet werden die Rochdale und Zehnkampf2006 Datensätze (verfügbar über die Internetseite der Lehrveranstaltung).

- (a) Erstellen Sie einen Mosaicplot für die 8 Variablen der Rochdale Daten.
- Welche Zelle hat die meisten Werte, und wie viel Prozent der gesamten Daten macht dies aus? Was können Sie über die Verteilung der Zellengrößen sagen?
 - Wie viele leere Zellen treten auf? In wie weit spielt die Stichprobengröße dabei eine Rolle?
 - In welcher Reihenfolge wird die Zahl der dargestellten leeren Zellen minimiert? Welchen Vorteil hat diese Darstellung, und wie gehen Sie vor, um sie zu finden?
- (b) Welche Disziplinen (Zehnkampf Daten) waren wichtig, um insgesamt eine hohe Punktezahl zu erreichen, und in welcher Disziplin würden Sie vermehrt trainieren, um sich einen Vorteil zu verschaffen (begründen Sie)? Erstellen Sie geeignete Parallelkoordinatenplots für die Zehnkampf Daten.

6. Einzeiler-Aufgaben.

- (a) Was gilt für den Bias aus Aufgabe 5 von Übungsblatt 3 in Theorie?
- (b) Was macht der R-Befehl
- ```
parcoord(crabs[1:100, 4:6], col=rep(1:2, rep(50, 2)), var.label=TRUE)
```
- (falls nötig, bitte MASS laden)?