

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

Statistik I

Übungsblatt 4

Abgabe: *Dienstag 15. Mai 2012, bis spätestens 12.00 Uhr*; Briefkasten: Statistik I oder per email an die Übungsleiter

Die Aufgaben können auch in 2er-Gruppen bearbeitet und abgegeben werden!

- Erstellen Sie aus "cars" einen Datensatz, der nur die Autos mit Vorderradantrieb beinhaltet. **(1P)**
 - Schätzen Sie μ und σ^2 des Verbrauchs für diesen Datensatz! **(1P)**
 - Plotten Sie die empirische Verteilungsfunktion der Daten (`ecdf`) und die der entsprechenden Normalverteilung in einen gemeinsamen Plot! **(1P)**
 - Bestimmen Sie für die Normalverteilung sowie auch für die Daten das 20%-Quantil. **(1P)**
- Schätzer (5P)** Seien \hat{p} und \hat{p}_B der Maximum-Likelihood-Schätzer bzw. Bayes-Schätzer (apriori Betaverteilung mit Parametern $a > 0$ und $b > 0$) für den Proportionsparameter p einer Binomialverteilung $B(N, p)$.
 - Bestimmen Sie die mittleren quadratischen Fehler dieser Schätzer.
 - Zeigen Sie, dass für die Wahl $a = b = \sqrt{N/4}$ der apriori Betaverteilung der mittlere quadratische Fehler des Bayes-Schätzers konstant ist, d.h. nicht von p abhängt. Bestimmen Sie hierbei explizit den Bayes-Schätzer und dessen mittleren quadratischen Fehler.
- Umfrage (5P)** Betrachtet wird die Umfrage aus Aufgabe 3 von Übungsblatt 3. Bestimmen Sie die Bayes-Schätzer für die Befürwortungsrate p unter den apriori Verteilungen
 - Gleichverteilung auf $[0, 1]$;
 - Betaverteilung mit $a = 1$ und Erwartungswert identisch dem Maximum-Likelihood-Schätzer \hat{p} für p ;
 - 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 $p = \hat{p}$?
- Mietspiegel (5P)**

Laden Sie den Datensatz zum Münchner Mietspiegel 2003 von der Webseite der Vorlesung in R.

 - Verschaffen Sie sich einen Überblick über den Datensatz. Welche Variablen gibt es und wie sehen die Verteilungen aus? Geben Sie eine kurze Zusammenfassung der wichtigsten Eckpunkte des Datensatzes an.
 - Verwenden Sie die Funktion `tapply` um herauszufinden, welche Bezirke
 - am günstigsten sind.
 - die größten Preisunterschiede aufweisen.

- (c) Verwenden Sie die Funktion `by` um sich summaries aller Variablen nach Bezirken ausgeben zu lassen.
- (d) Finden Sie mit den Funktionen `sapply` und `class` die Klasse jeder Variable heraus. Standardisieren Sie mit Hilfe des ersteren Befehls die ersten drei Variablen und berechnen Sie deren Korrelationen. Unterscheiden sie sich von den Korrelationen der nicht-standardisierten Variablen?
- (e) Berechnen Sie eine Kreuztabelle des Bezirks und der Anzahl der Wohnräume. Berechnen Sie Zeilen- und Spaltensummen der Tabelle mittels `apply`!
- (f) **(Zusatz für Interessierte)** Was gibt `xtabs(nm~bez+rooms, data=miete03)` aus? Wie kann man mittels `apply` alle Werte unter dem Wert 1000 auf Null setzen?

5. Smartphones (5P)

Betrachten Sie folgende Grafik zum Thema "Marktanteile von Handys".

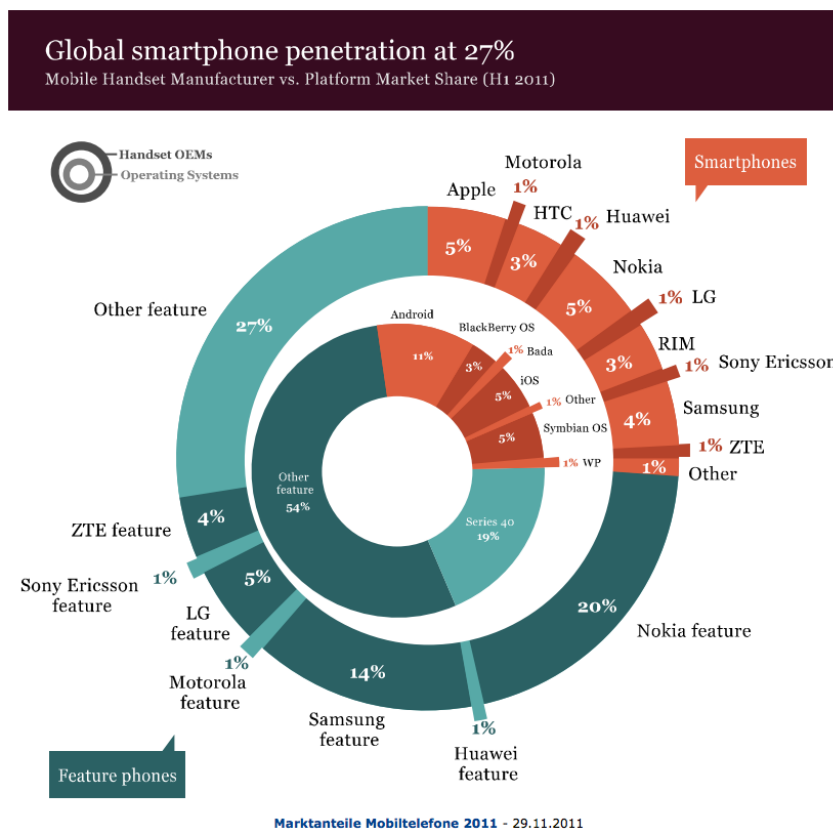


Abbildung 1: Quelle: <http://winfuture.de/news,66866.html>

- (a) Was genau wird dargestellt? Welche Größen fließen ein und wie sind sie visualisiert?
- (b) Beurteilen Sie die Grafik im Hinblick auf Übersichtlichkeit, Suggestivität und Ästhetik.
- (c) Welche Aussagen können Ihrer Meinung nach auf Basis der Grafik getroffen werden?
- (d) Ist die Visualisierung gelungen? Welche Verbesserungen würden Sie vornehmen?
- (e) Wie würden Sie die Daten visualisieren, wenn Sie
 - etwas über die Anteile der smartphones je Firma wissen wollten?
 - etwas über den Zusammenhang zwischen Betriebssystem und Firma erfahren möchten?