

Prof. Dr. Katharina Morik,
Prof. Dr. Claus Weihs
Dipl.-Inform. Christian Bockermann,
Dipl.-Stat. Julia Schiffner

Dortmund, 19.04.11
Abgabe: bis Di, 26.04., 10.00 Uhr an
schiffner@statistik.tu-dortmund.de

Übungen zur Vorlesung
Wissensentdeckung in Datenbanken
Sommersemester 2011
Blatt 2

Aufgabe 2.1 (2 Punkte)

Ein handelsüblicher sechsseitiger Würfel mit den Zahlen 1 bis 6 wird zweimal geworfen.

a) Geben Sie die Grundgesamtheit Ω an.

b) Geben Sie die folgenden Ereignisse an:

A: Die Summe der Augenzahlen der beiden Würfe ist eine Primzahl.

B: Die Augenzahl des zweiten Wurfs ist um mindestens eins höher als die des ersten Wurfs.

C: Die Augenzahlen beider Würfe betragen mindestens fünf.

Aufgabe 2.2 (3 Punkte)

Die Eingänge eines Supermarkts sind mit einer Alarmanlage gegen Diebstahl gesichert. Wenn ein Dieb mit gestohlener Ware die Anlage passiert, wird mit einer Wahrscheinlichkeit von 0.995 Alarm ausgelöst. Bei einem unbescholtenen Kunden beträgt diese Wahrscheinlichkeit 0.006. Nehmen Sie an, dass jeder 500. Kunde einen Diebstahl probiert.

a) Wie groß ist die Wahrscheinlichkeit, dass ein beliebiger Kunde mit dem Alarm konfrontiert wird?

b) Wie groß ist die Wahrscheinlichkeit, dass die Anlage zu Recht alarmiert?

Hinweis: Der Satz von Bayes könnte nützlich sein!

Aufgabe 2.3 (5 Punkte)

In einer Spinnerei wurden für 6 Monate die Anzahl der Maschinenschäden (Y) sowie die durchschnittliche Außentemperatur (X) beobachtet. Es ergab sich folgende Tabelle:

Temperatur (x_i)	Schäden (y_i)
9	29
5	21
13	36
24	51
30	66
29	70

- Passen Sie ein lineares Regressionsmodell der Form $y_i = \beta_0 + \beta_1 \cdot x_i + \epsilon_i$, $i = 1, \dots, 6$, an die Daten an.
- Tragen Sie in einem Diagramm die Anzahl der Maschinenschäden gegen die mittlere Außentemperatur auf und zeichnen Sie die Regressionsgerade ein.
- Die Maschinen standen in einem Raum, dessen Innentemperatur durch eine Klimaanlage geregelt wurde. Wie würden Sie dem leitenden Ingenieur die Ergebnisse der Regressionsrechnung vermitteln?

Eine Einführung in R findet in der Übung am 29.04. statt.

Aufgabe 3 ist gut per Hand lösbar. Wer möchte, kann natürlich versuchen Aufgabe 3 in R zu lösen. Die in der obigen Tabelle angegebenen Daten finden Sie auf der Homepage als `.txt`-Datei. R können Sie entweder durch Eintippen von R in der Kommandozeile oder durch Doppelklick auf die Verknüpfung im Startmenü oder auf dem Desktop öffnen. Falls die Installation der R-extension bereits funktioniert hat, können Sie auch R innerhalb von RapidMiner benutzen.

Hilfreich zur Lösung von Aufgabe 3 sind die Befehle:

- `read.table` – zum Einlesen von Daten,
- `lm` – zum Erstellen eines linearen Regressionsmodells,
- `plot` – zum Plotten,
- `abline` – zum Einzeichnen der Regressionsgerade.

Mit `?<Befehl>` können Sie sich die Dokumentation zu den einzelnen Befehlen anschauen, z.B. `?read.table`.

Bitte geben Sie Ihren R-Code im plain text Format, z.B. als `.R` oder `.txt`-Datei, ab.