

# Data Mining - Aufgabenblatt 1

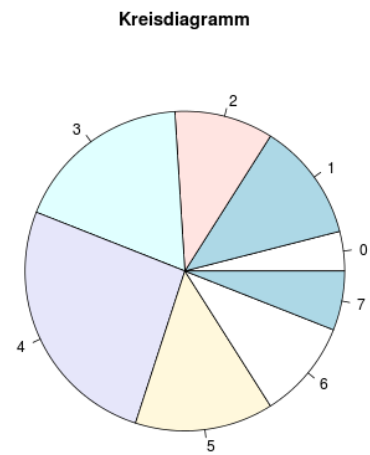
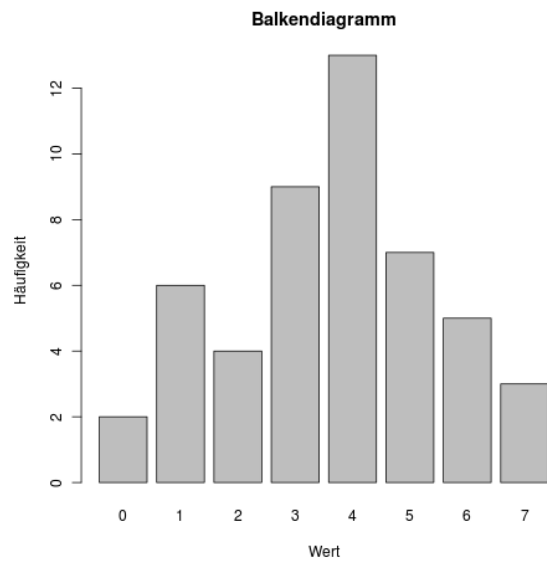
Andra Herta, Tobias Dreher

07.11.2011

## Aufgabe 1

Zahl	Absolute Häufigkeit	Relative Häufigkeit
0	2	0,04
1	6	0,12
2	5	0,10
3	9	0,18
4	13	0,26
5	7	0,14
6	5	0,10
7	3	0,06

Tabelle 1: Häufigkeitstabelle



## Aufgabe 2

Modalwert: 4

Median:  $\tilde{x} = 4$

1/5-Quantil:  $\tilde{x}_{\frac{1}{5}} = 2$

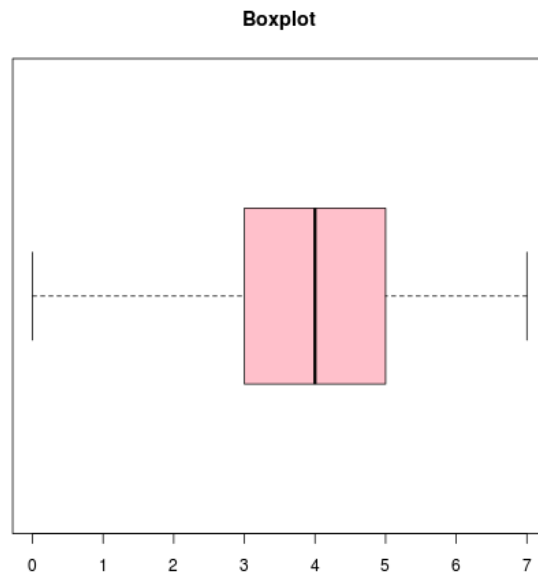
Arithmetischer Mittelwert:  $\bar{x} = 3,653061$

Spannweite:  $R = 7$

Interquartilbereich:  $Q := 2$

Varianz:  $s^2 = 3.314626$

Standardabweichung:  $s = 1.820611$



## Aufgabe 3

Korrelationskoeffizient  $r_{xy}$ : -0.860932

## Aufgabe 4.1

Beweis durch Cauchy-Schwarzsche Ungleichung

$$|a \cdot b| \leq |a| \cdot |b| \quad (1)$$

## Aufgabe 4.2

Wenn  $y_i = ax_i + b$  gilt, gilt auch  $\bar{y} = a\bar{x} + b$

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x}) (ax_i + b - (a\bar{x} + b))}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (ax_i + b - (a\bar{x} + b))^2}} \quad (3)$$

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x}) (ax_i - a\bar{x})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (ax_i - a\bar{x})^2}} \quad (4)$$

$$r_{XY} = \frac{a \sum_{i=1}^n (x_i - \bar{x}) (x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{a^2 \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (5)$$

$$r_{XY} = \frac{a \sum_{i=1}^n (x_i - \bar{x})^2}{|a| \sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

$$r_{XY} = \frac{a}{|a|} \quad (7)$$

$\Rightarrow r_{XY} = 1$ , für  $a > 0$

## Aufgabe 4.3

Nach Beweisführung von Aufgabe 4.2:

$$\Rightarrow r_{XY} = \frac{a}{|a|}$$

$\Rightarrow r_{XY} = -1$ , für  $a < 0$