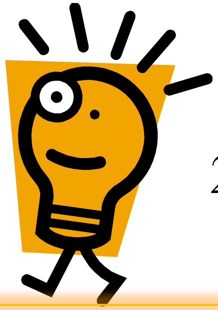
Beschreibende Statistik

Zusammengefasst von

Michaela Walla und Djurdica Filipovic

Die beschreibende Statistik erhebt Daten einer interessierenden Gesamtheit (z.B. Schüler des ibc), stellt die Daten in geeigneter Form dar, beschreibt sie und macht Aussagen über die untersuchte Gesamtheit.

Zur Untersuchung von Massenerscheinungen



1. Datenerhebung
2. Aufbereitung der Daten
3. Auswertungen

Beispiel: Datenerhebung: Schuhgrößen

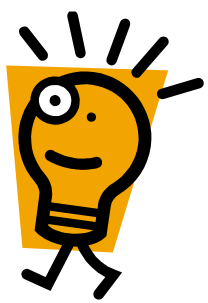
Urliste

x_i
$x_1 = 37$
$x_2 = 39$
$x_3 = 39$
$x_4 = 37$
$x_5 = 37$
$x_6 = 39$
$x_7 = 40$
$x_8 = 40$

Häufigkeitsdarstellung

x_i	H_i	h_i	%
$x_1 = 37$	$H_1 = 3$	$h_1 = 3/8 = 0,375$	37,50
$x_2 = 39$	$H_2 = 3$	$h2 = 3/8 = 0,375$	37,50
$x_3 = 40$	$H_4 = 2$	$h3 = 2/8 = 0,25$	25,00
	$n = 8$		100

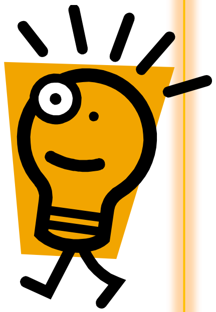
H_i = absolute Häufigkeit
h_i = relative Häufigkeit (Hi/n)
x_i = Daten, die die Schuhgrößen annehmen.



Zur Beschreibung der erhobenen Schuhgrößen gibt es

- Zentral- und
- Streuungsmaße

Welche Größen kann man hier am besten berechnen?



Zentralmaße

Mittelwert (arithmetisches Mittel)

Der Mittelwert ist verzerrt, wenn es Ausreißer gibt. Abhilfe wäre, wenn man die Ausreißer weg gibt oder den Median verwendet. (EXCEL **=MITTELWERT**)

Ergebnis: **38,5**

Definition: Es sei $x_1, x_2, \ldots, x_n$ eine Liste von Daten
Das **arithmetische Mittel** ist

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots\ldots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Definition: Es sei $x_1, x_2, \ldots, x_k$ eine Liste von Daten, die mit der Häufigkeit $H_1, H_2, \ldots.\ H_k$ auftreten. Dann ist das **arithmetische Mittel**

$$\bar{x} = \frac{H_1.x_1 + H_2.x_2 + H_3.x_3 + \ldots\ldots.. + H_k.x_k}{n} = \frac{1}{n} \sum_{i=1}^{k} H_i \cdot x_i \quad \text{für} \quad n = \sum_{i=1}^{k} H_i$$

Zentralmaße

Median

Der Median ist jene Zahl, die genau in der Mitte steht, wenn man die Daten ordnet (EXCEL **=MEDIAN**)

37 – 37 – 37 – **39** – **39** **–** 39 – 40 – 40

Bei gerader Anzahl der Daten :

37 – 37 – 37 – **39 – 39** – 39 – 40 – 40 berechnet man das arithmetische Mittel der beiden mittleren Zahlen:

Median = (39+39)/2 = 39

Bedeutung:
50% der Schuhgrößen sind kleiner oder gleich dem Median,
50% der Schuhgrößen sind größer oder gleich dem Median.

Zentralmaße

Modus

Der Modus ist der häufigste Wert.

Er ist auch für qualitative Daten geeignet.

Gibt es gleichviele häufigste Werte, existiert kein Modus

(EXCEL **=MODALWERT**)

In diesem Beispiel existiert kein Modus, da die Schuhgröße 37 und 39 gleich oft vorkommen!

Streuungsmaße

Spannweite

$sp = x_{max} - x_{min}$

$sp = 40 – 37 = 3$

(EXCEL **=MAX** - **=MIN**)

Interquartil-Spannweite

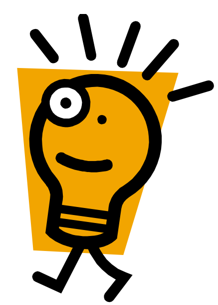
$isp = x_{0,75} – x_{0,25}$

$x_{0,75} = 37$

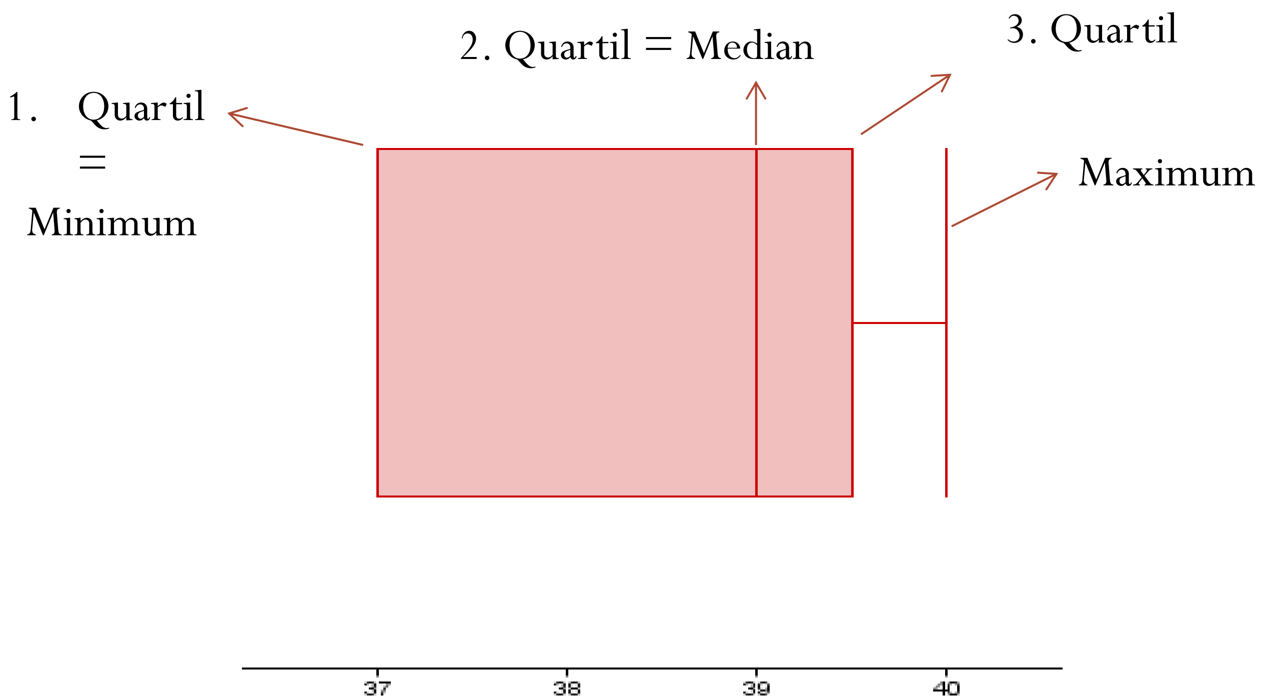
$x_{0,25} = 39,5$

$isp = \mathbf{1,5}$

$x_{0,75}$ ist das obere Quartil und
$x_{0,25}$ das untere Quartil.



Boxplot



Da das 1. Quartil mit dem Minimum zusammenfällt, liegen in der Box nicht die mittleren 50% der Daten, sondern sogar die unteren 75%.

Streuungsmaße

Empirische Standardabweichung

(EXCEL **=STABWN**)

Es ist nur sinnvoll gleichzeitig mit dem Mittelwert auch die empirische Standardabweichung zu berechnen.

Ergebnis: **1,2**

Definition: Es sei $x_1, x_2,\ldots,x_n$ eine Liste von Daten mit dem Zentralmaß $\bar{x}$.

Die (empirische) **quadratische Standardabweichung** vom Mittelwert $\bar{x}$ ist:

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Streuungsmaße

Empirische Standardabweichung

(EXCEL **=STABWN**)

Für gehäufte Daten:

Definition: Es sei $x_1, x_2, \ldots, x_k$ eine Liste von Daten, die mit der absoluten Häufigkeit $H_1, H_2, \ldots H_k$ auftreten. Das arithmetischen Mittel ist $\bar{x}$.
Die (empirische) **quadratische Standardabweichung** vom Mittelwert $\bar{x}$ ist:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^{k} H_i \cdot (x_i - \bar{x})^2} \qquad \text{für} \quad n = \sum_{i=1}^{k} H_i$$

Und nun viel Spaß beim Üben!