Statistik für Wirtschaftsinformatiker
Univariate Verteilungsanalyse - Einführung

Dr. Wolfgang Böhme

Oktober 2010

2. Univariate Verteilungsanalyse - Gliederung

- **2.0 Einführung**
- 2.1 Häufigkeitsverteilung / Verteilungsfunktion
- 2.2 Maßzahlen der Verteilung
- 2.3 Konzentrationsanalyse

Erhobene Daten

Ausgangspunkt: **Urliste**

$$X_{(n,k)} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ \vdots & \vdots & \ddots & \\ x_{n1} & x_{n2} & & x_{nk} \end{pmatrix}$$

mit

x_{ij} – *Auspraegung des* $j-$ *ten Merkmals am* $i-$ *ten Merkmalstraeger*

j – *Index Merkmal* $(j = 1, 2, ..., k) : X_1, X_2, ..., X_k$

i – *Index Merkmalstraeger* $(i = 1, 2, ..., n)$

univariate Statistik

Systematik der Statistik nach Anzahl der Merkmale

$k = 1$ - univariate Statistik
$k > 1$ - multivariate Statistik

Urliste für univariate Statistik

$$X_{(n,1)} = X_{(n)} \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

mit

$x_{i1} = x_i$ – *Auspraegung am* $i-$*ten Merkmalstraeger* (*Beobachtung* i)
i – *Index Merkmalstraeger* ($i = 1, 2, ..., n$)
$X_1 = X$ – *Merkmal* (1)

geordnete Beobachtungen

gegeben sei ein mindestens ordinal skalierter Datensatz. Ein aufsteigend geordneter Datensatz genügt der Bedingung:

$$x_1 \leq x_2 \leq \ ..., \ x_n$$

Bezeichnung: **Rangwertreihe**

Beispiel: Urliste

$$X_{(n,k)} = \begin{pmatrix} 6861 & 73.0 & 6 & 21765.0 & 44.9 & 1.0 & 310.0 \\ 5150 & 72.5 & 7 & 28245.0 & 60.4 & 1.0 & 332.0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 3018 & 74.6 & 7 & 12201.0 & 37.1 & 4.2 & 104.0 \end{pmatrix}$$

mit

x_{ij} – *Auspraegung des* $j-$ *ten Merkmals am* $i-$ *ten Merkmalstraeger*

j – *Index Merkmal* $(j = 1, 2, ..., 8)$: *LAND*, *ENERGIE*, ..., *ZEITUNG*

i – *Index Merkmalstraeger* $(i = B, DK, ..., E)$

univariate Statistik - Urliste

$$
\begin{array}{ll}
B: & \left(\begin{array}{c} 21765.0 \\ 28245.0 \\ 25179.0 \\ 19048.0 \\ 24608.0 \\ 7465.0 \\ 17471.0 \\ 14735.0 \\ 17921.0 \\ 27611.0 \\ 21536.0 \\ 24823.0 \\ 8822.0 \\ 22499.0 \\ 12201.0 \end{array}\right. \\
DK: & \\
D: & \\
FIN: & \\
F: & \\
EL: & \\
UK: & \\
IRL: & \\
I: & \\
L: & \\
NL: & \\
A: & \\
P: & \\
S: & \\
E: & \left.\begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array}\right)
\end{array}
$$

$$
\begin{array}{rcl}
\text{mit } X_{i1} = X_i & - & \text{Auspraegung am } i-\text{ten Merkmalstraeger (Beobachtung } i) \\
i & - & \text{Index Merkmalstraeger } (i = B,\ DK,\ \ldots,\ E) \\
X_1 = X & - & \text{Merkmal : BIP}
\end{array}
$$

univariate Statistik - Rangwertreihe

$$
\begin{array}{ll}
EL: & 7465.0 \\
P: & 8822.0 \\
E: & 12201.0 \\
IRL: & 14735.0 \\
UK: & 17471.0 \\
I: & 17921.0 \\
FIN: & 19048.0 \\
NL: & 21536.0 \\
B: & 21765.0 \\
S: & 22499.0 \\
F: & 24608.0 \\
A: & 24823.0 \\
D: & 25179.0 \\
L: & 27611.0 \\
DK: & 28245.0
\end{array}
$$

$$x_1 \leq x_2 \leq \ldots, x_n$$

Vorlesungsbeispiel 1

Datensatz: Vorlesungsbeispiel1.xls

Urliste

Merkmalsträger	Anzahl
1	4
2	4
3	3
4	5
5	5
6	5
7	3
8	4
9	5
10	4
11	3
12	4
13	3
14	3
15	3
16	6

Merkmalsträger	Anzahl
27	1
19	2
20	2
25	2
35	2
3	3
7	3
11	3
13	3
14	3
15	3
34	3
38	3
41	3
44	3
49	3

Statische Gesamtheit	5-Raum Wohnungen in einer Wohnungsbaugesellschaft
Merkmal	Anzahl der gemeldeten Personen

univariate Statistik - Urliste / Rangwertreihe

Urliste	
Merkmalstraeger	*Anzahl*
1	4
2	4
3	3
4	5
5	5
6	5
7	3
8	4
9	5
10	4
⋮	⋮

$\Longrightarrow$

Rangwertreihe	
Merkmalstraeger	*Anzahl*
27	1
19	2
20	2
25	2
35	2
3	3
7	3
11	3
13	3
14	3
⋮	⋮