

1. Beschreibende Statistik

1.1 Begriffe

- Beschreibende (deskriptive) Statistik

Zugrundeliegende Daten werden mit Hilfe von Grafiken und statistischen Kennzahlen ausgewertet

- Schließende (induktive) Statistik

Es werden statistische Modellannahmen für die zugrundeliegenden Daten getroffen und damit Fehlerwahrscheinlichkeiten für die berechneten Kennzahlen ermittelt.

—> Parameterschätzung, Hypothesentests

- Grundgesamtheit (N Elemente)

Bsp.: Untersuchung der Wirtschaftskraft von "kleinen" deutschen "IT-Unternehmen"

Die Objekte, für die eine Datenerhebung durchgeführt wird.

Wichtig: Präzise Definition!

z.B. von "kleinem" Unternehmen bzw. "IT-Unternehmen"

- Stichprobe (n Elemente)

(i) zufällig

(ii) repräsentativ

(iii) "proportional geschichtet", s. Bsp.

(iv) ausreichend groß

Bsp.: Ein Unternehmen bezieht Komponenten von drei verschied. Zulieferfirmen, die sich in der Qualität der Komponenten unterscheiden.
 $N_1 = 2000$ Komponenten von Firma 1, $N_2 = 1000$ Komponenten von Firma 2, $N_3 = 3000$ Komponenten von Firma 3.
 Es wird ein Umfang der Stichprobe von $n = 300$ festgelegt.

Welche Alternative würden Sie für eine Stichprobe wählen?

- (1) 300 rein zufällig
- (2) Je 100 von jeder der drei Firmen
- ✓ (3) 100 von Firma 1, 50 von Firma 2, 150 von Firma 3

• Merkmal

Bsp.: Geschlecht, Mitarbeiterzahl, Lebensdauer eines elektr. Bauteils

Welche "Klassen" von Merkmalen lassen sich an diesen Beispielen erkennen?

diskret	—	stetig	qualitativ	—	quantitativ
einige wenige Werte/Ausprägungen		≥ 30 verschiedene Ausprägungen	kein Zahlenwert z.B. Geschlecht		numerische Werte

• Univariate / multivariate Daten

Datenvektor (nur ein Merkmal) Datenmatrix

• Datenmatrix / Datenframe: s. Statistiksoftware R

n sei Stichprobenumfang, $p > 1$ sei Anzahl der Merkmale

$n \times 1$ - Matrix : univariate Daten

$n \times p$ - Matrix : multivariate Daten

1.2 Darstellung diskreter Merkmale

3

1.2.1 Häufigkeitstabellen

Definition 1.1:

Sei $\{x_1, x_2, \dots, x_n\}$ eine Stichprobe vom Umfang n .

Innerhalb der Stichprobe treten die Ausprägungen a_1, a_2, \dots, a_k auf.

Dann heißen die Häufigkeiten h_i ($i = 1, \dots, k$) der Ausprägungen a_i **absolute Häufigkeiten** und die Anteile $f_i = \frac{h_i}{n}$ **relative Häufigkeiten**.

Es gilt: $\sum_{i=1}^k h_i = n$ und $\sum_{i=1}^k f_i = 1$

Bsp.: Bewerberzahlen der Bachelorstudiengänge in Rosenheim im WS 16/17

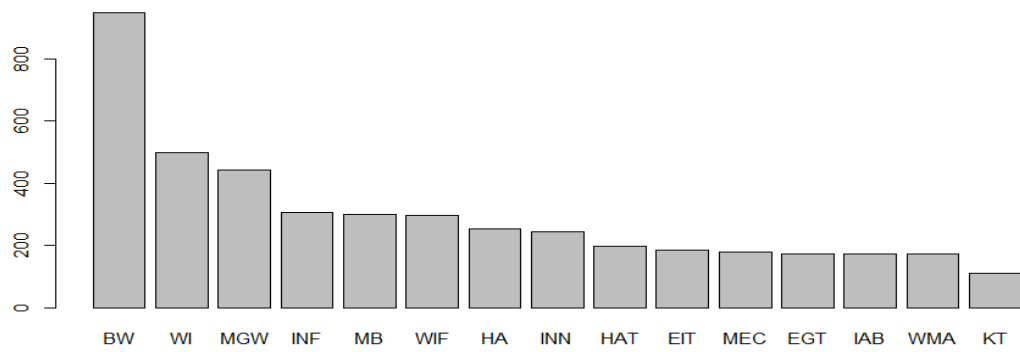
	STUDIENGANG	BEWERBER. abs.	BEWERBER. rel.
1	BW-B	947	21,19%
2	WI-B	497	11,12%
3	MGW-B	443	9,91%
4	INF-B	306	6,85%
5	MB-B	298	6,67%
6	WIF-B	297	6,65%
7	HA-B	252	5,64%
8	INN-B	245	5,48%
9	HT-B	197	4,41%
10	EIT-B	185	4,14%
11	MEC-B	177	3,96%
12	EGT-B	172	3,85%
13	IAB-B	171	3,83%
14	WMA-B	171	3,83%
15	KT-B	111	2,48%

1.2.2 Grafische Darstellung

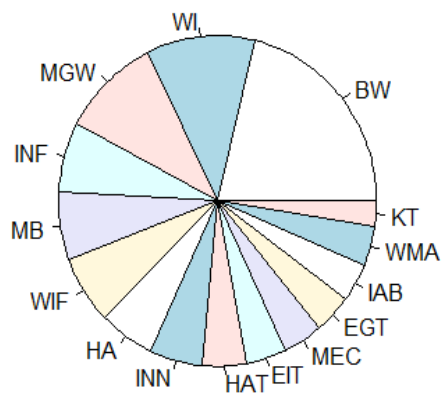
Welche grafische Darstellungen für diskrete Merkmale mit wenigen Ausprägungen kennen Sie?

- Balken - / Säulen - / Stabdiagramme
- Kreisdiagramme

R- Befehl: `barplot`



R- Befehl: `pie`



1.3 Darstellung stetiger Merkmale

5

Kennzeichen: Sehr viele verschiedene Beobachtungswerte für das Merkmal

Bsp.: Durchschnittsnoten von Studierenden in ihrem Abschlusszeugnis

1.97 ; 2.33 ; 3.51 ; 5.11 ; 1.12 ; 2.59 ; 4.18 ; 2.81 ; 3.27 ; 1.50

- (i) Sortierung nach Größe
- (ii) Klassenbildung
- (iii) Darstellung

1.3.1 Klassenbildung

Ungefähr für $k \geq 30$ Ausprägungen

Untersuchung der Häufigkeiten für Intervalle, sog. Klassen, des Wertebereichs

Problematisch bei Klassenbildung:

zu wenig Klassen \Rightarrow Informationsverlust

zu viele Klassen \Rightarrow unübersichtlich

Faustregeln bei der Klassenbildung:

- (1) Möglichst gleichbreite Klassen verwenden (Ausnahme: 1. und letzte Klasse)
- (2) Zwischen 5-20 Klassen, nicht größer als \sqrt{n}
- (3) Die linke Intervallgrenze gehört dazu, die rechte nicht

Bilden Sie Klassen für das Beispiel "Abschlussnoten" und ermitteln Sie die absoluten Häufigkeiten:

6

Klasse i	h_i	f_i	$\sum_{i: x_i \leq x} f_i = \hat{F}(x)$ empirische Verteilungsfkt.
$[1; 1.5[$	1	10%	10% $x = 1.5$
$[1.5; 2.5[$	3	30%	40% $x = 2.5$
$[2.5; 3.5[$	3	30%	70% $x = 3.5$
$[3.5; 4.5[$	2	20%	90% $x = 4.5$
$[4.5; 6.0]$	1	10%	100% $x = 6.0$

1.3.2 Histogramm (grafische Darstellung)

Histogramme sind Rechtecke mit folgenden Eigenschaften:

- (1) Die Breiten entsprechen den Intervalllängen Δx_i ;
- (2) Die Höhe entspricht $\frac{h_i}{\Delta x_i}$ bzw. $\frac{f_i}{\Delta x_i}$;
- (3) Die Rechtecksfläche entspricht h_i bzw. f_i . !

Zeichnen Sie die Histogramme bzw. die empirische Verteilungsfunktion für obiges Bsp.