Deskriptive Statistik

Deskriptive oder angewandte Statistik stellt statistische Methoden zur Betrachtung und Auswertung messbarer bzw. gezählter Beobachtungen oder Messergebnisse (Werte) zur Verfügung. Zentral für die deskriptive Statistik ist die Auswertung einer Zufallsstichprobe einer zunächst nicht näher bekannten Grundgesamtheit von Daten auf der Grundlage eines mathematisch-wahrscheinlichkeitstheoretischen Modells. Auf der Grundlage der Wahrscheinlichkeitsrechnung (Stochastik) untersucht also die Statistik die hinter Zufallsexperimenten liegenden mathematischen Modelle. Die deskriptive Statistik stellt Methoden zur Verfügung, um Datenmaterial (Stichproben) zu charakterisieren und zu analysieren.

Einfache Stichproben I

Eine geordnete Stichprobe ist eine (geordnete) Menge von n Werten $x_1, x_2, \ldots x_n$, $n\varepsilon\mathbf{N}$. Jeder Wert x_i besitzt die absolute Häufigkeit $a_i = 1$, die relative Häufigkeit $f_i = 1/n$, i=1, 2, …n. Als Lageparameter heißt

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

das arithmetische Mittel (Mittelwert, Durchschnitt) der Stichprobe. Als Streuungsmaß heißt

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

als Mittelwert der Quadrate der Abweichungen vom arithmetischen Mittel Varianz der Stichprobe. Die Wurzel aus der Varianz heißt Standardabweichung

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

Offensichtlich ist eine einfache Stichprobe durch arithmetisches Mittel und Standardabweichung charakterisierbar, was Lage und Streuung anbetrifft. Es folgen die Beziehungen:

a) $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow \sum_{i=1}^{n} x_i = n\bar{x} \Rightarrow \sum_{i=1}^{n}(x_i - \bar{x}) = 0$

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

$$\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{2\bar{x}}{n}\sum_{i=1}^{n}x_i + \frac{1}{n}\sum_{i=1}^{n}\bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\bar{x}\cdot\bar{x} + \frac{1}{n}\cdot n\bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\bar{x}^2 + \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2 =$$

$$\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n}x_i\right)^2 = \frac{1}{n}\left(\sum_{i=1}^{n}x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2\right).$$

Beispiel (einfache Stichprobe)

Gemäß den oben genannten Beziehungen a), b) lässt sich also eine einfache Stichprobe hinsichtlich arithmetischem Mittel und Standardabweichung wie folgt auswerten:

<u>Stichprobe</u>: Stichprobenumfang: n = 16; Stichprobe (geordnet): 1; 3; 4; 4; 5; 5; 6; 6; 7; 7; 8; 9; 9; 10; 10; 12

i	**Wert** x_i	**Relative Häufigkeit** $f_i = 1/n$	**Produkt** $x_i \cdot f_i$	**Differenz** $x_i - \bar{x}$	**Quadrat der Differenz** $(x_i - \bar{x})^2$	**Produkt** $(x_i^2 - \bar{x})^2 \cdot f_i$
1	1	0.0625	0.0625	-5.625	31.6406	1.9775
2	3	0.0625	0.1875	-3.625	13.1406	0.8213
3	4	0.0625	0.25	-2.625	6.8906	0.4307
4	4	0.0625	0.25	-2.625	6.8906	0.4307
5	5	0.0625	0.3125	-1.625	2.6406	0.165
6	5	0.0625	0.3125	-1.625	2.6406	0.165
7	6	0.0625	0.375	-0.625	0.3906	0.0244
8	6	0.0625	0.375	-0.625	0.3906	0.0244
9	7	0.0625	0.4375	0.375	0.1406	0.0088
10	7	0.0625	0.4375	0.375	0.1406	0.0088
11	8	0.0625	0.5	1.375	1.8906	0.1182
12	9	0.0625	0.5625	2.375	5.6406	0.3525
13	9	0.0625	0.5625	2.375	5.6406	0.3525
14	10	0.0625	0.625	3.375	11.3906	0.7119
15	10	0.0625	0.625	3.375	11.3906	0.7119
n = 16	12	0.0625	0.75	5.375	28.8906	1.8057
Summe:	106	1	6.625	0	129.75	8.1094
Mittelwert $\bar{x} = \Sigma x_i/n = \Sigma x_i f_i =$			6.625			
Varianz $s^2 = \Sigma(x_i - \bar{x})^2/n = \Sigma(x_i - \bar{x})^2 f_i =$						8.1094
Standardabweichung $s = (\Sigma(x_i - \bar{x})^2/n)^{1/2} = (\Sigma x_i^2 f_i - (\Sigma x_i f_i)^2)^{1/2} =$						2.8477

i	**Wert** x_i	**Quadrat** x_i^2	**Relative Häufigkeit** $f_i = 1/n$	**Produkt** $x_i \cdot f_i$	**Produkt** $x_i^2 \cdot f_i$
1	1	1	0.0625	0.0625	0.0625
2	3	9	0.0625	0.1875	0.5625
3	4	16	0.0625	0.25	1
4	4	16	0.0625	0.25	1
5	5	25	0.0625	0.3125	1.5625
6	5	25	0.0625	0.3125	1.5625
7	6	36	0.0625	0.375	2.25
8	6	36	0.0625	0.375	2.25
9	7	49	0.0625	0.4375	3.0625
10	7	49	0.0625	0.4375	3.0625
11	8	64	0.0625	0.5	4
12	9	81	0.0625	0.5625	5.0625
13	9	81	0.0625	0.5625	5.0625
14	10	100	0.0625	0.625	6.25
15	10	100	0.0625	0.625	6.25
n = 16	12	144	0.0625	0.75	9
Summe:	106	832	1	6.625	52
Mittelwert $\bar{x} = \Sigma x_i/n = \Sigma x_i f_i =$					6.625
Varianz $s^2 = \Sigma x_i^2/n - (\Sigma x_i)^2/n^2 = \Sigma x_i^2 f_i - (\Sigma x_i f_i)^2 =$					8.1094
Standardabweichung $s = (\Sigma x_i^2/n - (\Sigma x_i)^2/n^2)^{1/2} = (\Sigma x_i^2 f_i - (\Sigma x_i f_i)^2)^{1/2} =$					2.8477

Einfache Stichproben II

Ist eine Stichprobe von n Werten $x_1, x_2, \dots x_n$, $n \varepsilon \mathbf{N}$, mit absoluten Häufigkeiten a_i, i=1, 2, …n, gegeben, errechnen sich die relativen Häufigkeiten als $f_i = \frac{a_i}{\sum_{j=1}^{n} a_j}$ mit $\sum_{i=1}^{n} f_i = 1$. Lage- und Streuungsmaße der Stichprobe stellen sich dann dar als:

Arithmetisches Mittel: $\overline{x} = \sum_{i=1}^{n} x_i f_i$

Varianz: $s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 f_i$

Standardabweichung: $s = \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 f_i}$.

Es gilt hier:

$$s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 f_i = \sum_{i=1}^{n} (x_i^2 - 2x_i \overline{x} + \overline{x}^2) f_i = \sum_{i=1}^{n} x_i^2 f_i - \sum_{i=1}^{n} 2x_i \overline{x} f_i + \sum_{i=1}^{n} \overline{x}^2 f_i =$$

$$\sum_{i=1}^{n} x_i^2 f_i - 2\overline{x} \sum_{i=1}^{n} x_i f_i + \overline{x}^2 \sum_{i=1}^{n} f_i = \sum_{i=1}^{n} x_i^2 f_i - 2\overline{x}^2 + \overline{x}^2 = \sum_{i=1}^{n} x_i^2 f_i - \overline{x}^2 = \sum_{i=1}^{n} x_i^2 f_i - \left(\sum_{i=1}^{n} x_i f_i \right)^2 .$$

Die Berechnung der statistischen Maßzahlen x^- und s kann damit tabellarisch über die Summen der Produkte $x_i f_i$ und $x_i^2 f_i$ erfolgen, was einfacher ist als über die quadratischen Differenzen $(x_i - x^-)^2 f_i$, i=1, 2, …n. Es ergibt sich damit die Vorgehensweise gemäß folgender Tabelle:

i	**Wert x_i**	**Quadrat x_i^2**	**Absolute Häufigkeit a_i**	**Relative Häufigkeit f_i**	**Produkt $x_i \cdot f_i$**	**Produkt $x_i^2 \cdot f_i$**
1	x_1	x_1^2	a_1	$f_1 = \frac{a_1}{\sum_{j=1}^{n} a_j}$	$x_1 f_1$	$x_1^2 f_1$
2	x_2	x_2^2	a_2	$f_2 = \frac{a_2}{\sum_{j=1}^{n} a_j}$	$x_2 f_2$	$x_2^2 f_2$
…	…	…	…	…	…	…
n	x_n	x_n^2	a_n	$f_n = \frac{a_n}{\sum_{j=1}^{n} a_j}$	$x_n f_n$	$x_n^2 f_n$
Summe:	-	-	$\sum_{i=1}^{n} a_i$	1	$A = \sum_{i=1}^{n} x_i f_i$	$B = \sum_{i=1}^{n} x_i^2 f_i$
Mittelwert $\overline{x} = \sum_{i=1}^{n} x_i f_i$					A	

Varianz $s^2 = \sum_{i=1}^{n} x_i^2 f_i - \bar{x}^2 = \sum_{i=1}^{n} x_i^2 f_i - \left(\sum_{i=1}^{n} x_i f_i \right)^2$	C = B – A^2
Standardabweichung $s = \sqrt{\sum_{i=1}^{n} x_i^2 f_i - \bar{x}^2} = \sqrt{\sum_{i=1}^{n} x_i^2 f_i - \left(\sum_{i=1}^{n} x_i f_i \right)^2}$	D = $\sqrt{C}$

Beispiel (einfache Stichprobe)

Ein Würfel wird 100 Mal geworfen. Die Augenzahlen 1, 2, 3, 4, 5, 6 treten mit absoluten Häufigkeiten 18, 12, 21, 10, 15, 24 auf.

Stichprobe: Stichprobe (geordnet) (mit Häufigkeiten): 1 (18); 2 (12); 3 (21); 4 (10); 5 (15); 6 (24)

i	**Wert x_i**	**Absolute Häufigkeit a_i**	**Relative Häufigkeit f_i**	**Produkt $x_i \cdot f_i$**	**Differenz x_i-x^-**	**Quadrat der Differenz $(x_i-x^-)^2$**	**Produkt $(x_i^2-x^-)^2 \cdot f_i$**
1	1	18	0.18	0.18	-2.64	6.9696	1.2545
2	2	12	0.12	0.24	-1.64	2.6896	0.3228
3	3	21	0.21	0.63	-0.64	0.4096	0.086
4	4	10	0.1	0.4	0.36	0.1296	0.013
5	5	15	0.15	0.75	1.36	1.8496	0.2774
6	6	24	0.24	1.44	2.36	5.5696	1.3367
Summe:	-	100	1	3.64	-	-	3.2904
Mittelwert $x^- = \Sigma x_i f_i$ =				3.64			
Varianz $s^2 = \Sigma(x_i-x^-)^2 f_i$ =							3.2904
Standardabweichung $s = (\Sigma(x_i-x^-)^2 f_i)^{1/2}$ =							1.8139

i	**Wert x_i**	**Quadrat x_i^2**	**Absolute Häufigkeit a_i**	**Relative Häufigkeit f_i**	**Produkt $x_i \cdot f_i$**	**Produkt $x_i^2 \cdot f_i$**
1	1	1	18	0.18	0.18	0.18
2	2	4	12	0.12	0.24	0.48
3	3	9	21	0.21	0.63	1.89
4	4	16	10	0.1	0.4	1.6
5	5	25	15	0.15	0.75	3.75
6	6	36	24	0.24	1.44	8.64
Summe:	-	-	100	1	3.64	16.54
Mittelwert $x^- = \Sigma x_i f_i$ =					3.64	
Varianz $s^2 = \Sigma x_i^2 f_i - (\Sigma x_i f_i)^2$ =						3.2904
Standardabweichung $s = (\Sigma x_i^2 f_i - (\Sigma x_i f_i)^2)^{1/2}$ =						1.8139

Literaturhinweise: REINHARDT, F., dtv-Atlas zur Schulmathematik. Definitionen, Beweise, Sätze, München [3]2003, S.226f (Statistische Grundlagen); SACHS, L., Angewandte Statistik. Anwendungen statistischer Methoden, Berlin-Heidelberg-N.Y. [9]1999, S.11-15, 130-133 (Methoden der Statistik, einfache Stichproben).

Michael Buhlmann, www.michael-buhlmann.de 04.2023