Statistik

Robin Rausch, Florian Maslowski, Ozan Akzebe

10. März 2023

Inhaltsverzeichnis

1 Statistik Grundlagen und Definitionen

1.1 Arithmetischer Mittelwert

Der Arithmetische Mittelwert $\bar{x}$ ist der „normale“ Durchschnitt. Also addiert man alle Werte auf und teilt durch die Anzahl der Werte. Beispiel:

$$\frac{1+2+2+1+4}{5} = 2 \tag{1}$$

1.2 Grundbegriffe

1.2.1 Grundgesamtheit

Beschreibt die Menge gleicherartiger Objekte (z.B. Mietwohnungen).

1.2.2 Stichprobe

Beschreibt die Menge der untersuchten Objekte (z.B. Mietwohnungen in Stuttgart).

1.2.3 Stichprobenumfang

Beschreibt die Anzahl der untersuchten Objekte (z.B. 120 Mietwohnungen in Stuttgart).

1.2.4 Merkmal

Beschreibt ein Kriterium oder eine interessierende Größe (z.B. Anzahl der Zimmer).

1.2.5 Ausprägung

Beschreibt die Werte, die ein Merkmal annehmen kann (z.B. 1 Zimmer, 2 Zimmer, 4 Zimmer, ...).

1.3 Qualitativ und Quantitativ

1.3.1 Qualitativ

Merkmale heißen qualitativ, wenn sie nur ihrer Art nach unterschieden werden können.

1.3.2 Quantitativ

Merkmale heißen quantitativ, wenn sie nur ihrer Größe unterscheidbar sind.

1.4 Diskret und Stetig

1.4.1 Diskret

Merkmale heißen diskret, wenn sie abzählbar sind.

1.4.2 Stetig

Merkmale heißen stetig, wenn sie jeden Wert innerhalb eines Intervalls annehmen können.

1.5 Skalierung von Merkmalen

1.5.1 nominal

Die Skalierung von Merkmalen heißt nominal, wenn sie in verschiedener Reihenfolge aufgelistet werden können (z.B. Augenfarbe).

1.5.2 ordinal

Die Skalierung von Merkmalen heißt ordinal, wenn sie in auf- oder absteigender Reihenfolge aufgezählt werden kann (z.B. Wohnungsgröße).

1.5.3 kardinal

Die Skalierung von Merkmalen heißt kardinal, wenn bei auf- oder absteigender Reihenfolge Abstände definiert werden können (z.B. Geschosshöhe).

1.6 Häufigkeit

Die Häufigkeit h_i ist die Anzahl der jeweils i-ten Ausprägung eines untersuchten Merkmals/Ereignisses. Beispiel:
Wenn es bei einem Experiment 11 Proband*innen gibt und 3 davon die Haarfarbe braun haben, ist die Häufigkeit der Haarfarbe braun 3. Die Häufigkeit aller Haarfarben kombiniert wäre dann 11.

1.7 relative Häufigkeit

Die relative Häufigkeit f_i ist die normale Häufigkeit im Verhältnis zur Gesamtanzahl. Also im o.g. Beispiel:
Die relative Häufigkeit der Haarfarbe braun $f_{braun} = \frac{3}{11}$.

1.8 Häufigkeitsverteilung

Eine Häufigkeit h_i ist die Anzahl von Ausprägungen (oder Ereignissen), also ≥ 1. Eine Häufigkeitsverteilung ist die Darstellung (=Zusammenschau) der Häufigkeiten aller möglichen Ereignissen. Also ein Diagramm oder eine Auflistung aller Häufigkeiten zur Übersicht ihrer Verteilungen zu einander.

1.9 Verteiung der relativen Häufigkeiten

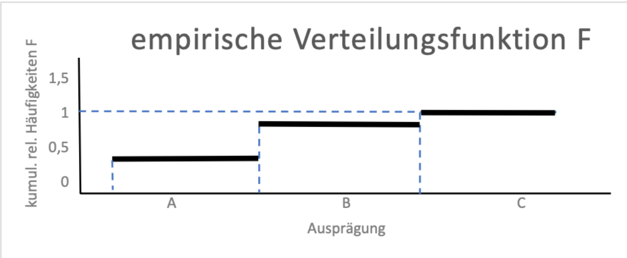
Die Verteilung der relativen Häufigkeit ist die Zusammenschau der relativen Häufigkeiten $f_i = \frac{h_i}{n}$ (n = Stichprobenumfang und h_i = Häufigkeiten) von sämtlichen möglichen Ereignissen (analog zur Häufigkeitsverteilung).

1.10 Empirische Verteilungsfunktion

Die Empirische Verteilungsfunktion ist die Funktion der kumulierten relativen Häufigkeiten.
Beispiel:

$$F(x) = \sum_i f_i \tag{2}$$

WÜRFEL mit 2 As, 3 Bs und 1 C (vgl. S. 18)



h(A) = h_1 = 2 **f_1 = 2/6**
h(B) = h_2 = 3 **f_2 = 3/6** **F = $\sum f_i$= 6/6 =1**
h(B) = h_3 = 1 **f_3 = 1/6**

F = F(x) ist Sprungfunktion:

$$F(x) = \begin{cases} 0 & \text{für } [\,0\,;\,A\,) \\ 2/6 & \text{für } [\,A\,;\,B\,) \\ 5/6 & \text{für } [\,B\,;\,C\,) \\ 6/6 & \text{für } [\,C\,;\,\infty\,) \end{cases}$$

1.11 Modalwert

Der Modalwert ist der Messwert eines Datensatzes, welcher am häufigsten auftritt.

1.12 Median

Der Medianm $\tilde{x}$ ist der Wert, der genau in der Mitte eines **geordneten** Datensatzes liegt.
Beispiel:
Datensatz = { 1, 2, 2, 4, 70}
Median = 2

1.13 Kennwerte

1.13.1 Quantil und Quartil

Der Median ist ebenfalls ein Quantil:
$x_{0,5} = \tilde{x}$
Das Quantil kann aber auch ein Quartil oder anderes sein:
unteres Quartil: $x_{0,25}$
oberes Quartil: $x_{0,75}$

$$x_p = x_{0,75} = 0,5(x_8 + x_9) = \frac{1}{2}(3+3) = 3 = x_{0,75} = q_{obere} \tag{3}$$

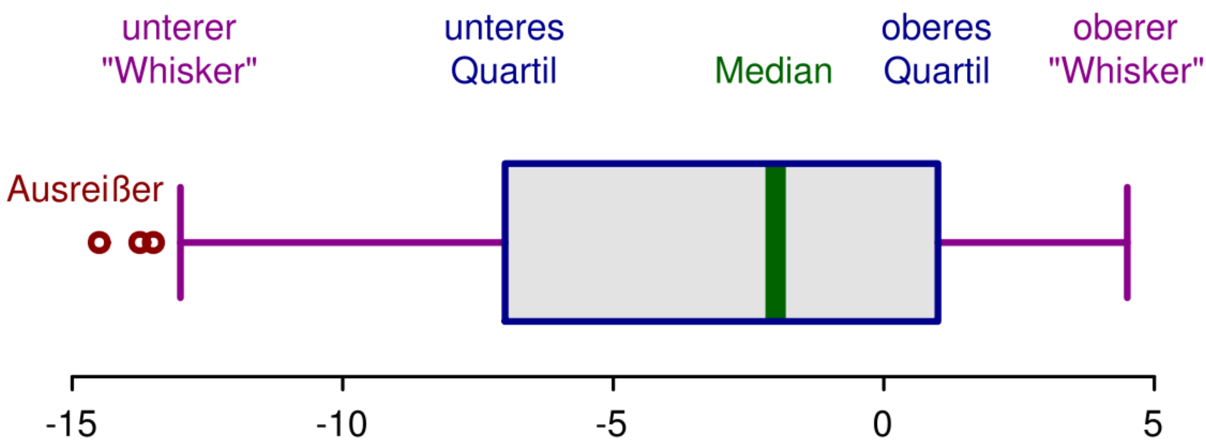
1.13.2 Decil

Als Decil werden alle Quantile mit $p = k * 0,1; k = 1; 2; ...; 10$
Also $\{x_{0,1}; x_{0,2}; x_{0,3};$ ...; $x_{0,8}; x_{0,9}; x_{1,0};\}$

1.13.3 Percentil

Als Percentil werden alle Quantile mit $p = k * 0,01; k = 1; 2; ...; 100$
Also $\{x_{0,01}; x_{0,02}; x_{0,03};$ ...; $x_{0,98}; x_{0,99}; x_{1,0};\}$

1.14 Boxplot



Whisker sind minimale und maximale Werte.

1.15 Empirische Varianz

Diese gibt an, wie stark die einzelnen Werte um den Mittelwert abweichen.

1.16 Empirische Standardabweichung

Die empirische Standardabweichung SD(x) lässt sich aus der empirischen Varianz berechnen:

$$SD(x) = s = \sqrt{VAR} = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \tilde{x})^2} \tag{4}$$

Immer positiv!