Univariate Verfahren

WS 2023-24

DI Emil Marinov | DI David Bechtolf

SPENGERGASSE

ausbildung mit zukunft

Übersicht

SPENGERGASSE
ausbildung mit zukunft

Häufigkeitsverteilungen

Häufigkeitsverteilungen

Absolute Häufigkeitsverteilung

Messwerte:
$x_1, x_2, \dots., x_n$

verschiedene Ausprägungen:
$a_1, a_2, \dots, a_k \ (k \leq n)$

absolute Häufigkeit h(a_j):
wie oft kommt ein Wert a_j vor

Beispiel:

Messwerte: 1, 3, 3, 4, 4, 4, 5, 5, 6, 10

n = 10 Messwerte
$$x_1 = 1, x_2 = 3, x_3 = 3, \dots$$

k = 6 verschiedene Werte
$$a_1 = 1, a_2 = 3, a_3 = 4, \dots$$

absolute Häufigkeiten
$$h(1) = 1, h(3) = 2, h(4) = 3, \dots$$

Häufigkeitsverteilungen

Relative Häufigkeitsverteilung

Messwerte:
$x_1, x_2, \dots., x_n$

verschiedene Ausprägungen:
$a_1, a_2, \dots, a_k \; (k \leq n)$

relative Häufigkeit f(a_j):
prozentueller Anteil $f(a_j) = \frac{h(a_j)}{n}$

Beispiel:

Messwerte: 1, 3, 3, 4, 4, 4, 5, 5, 6, 10

relative Häufigkeiten

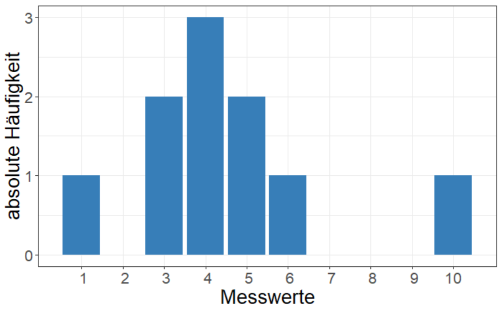
$$f(1) = \frac{1}{10} = 0.1$$
$$f(3) = \frac{2}{10} = 0.2$$

.....

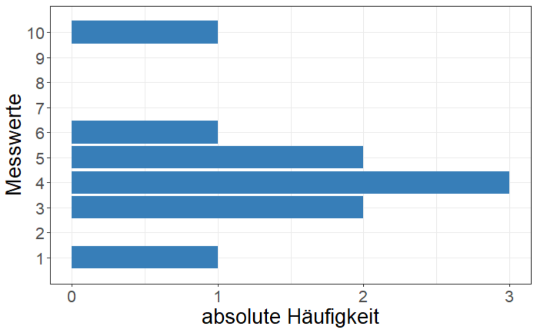
Häufigkeitsverteilungen

Grafische Darstellung

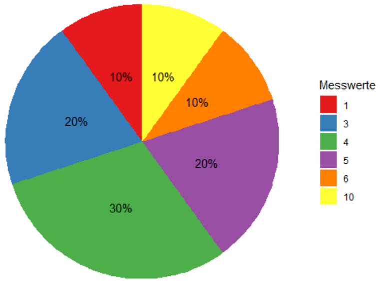
Säulendiagramm



Balkendiagramm



Kreisdiagramm



SPENGERGASSE
ausbildung mit zukunft

Häufigkeitsverteilungen

Kumulierte Häufigkeitsverteilung

- **kumulierte absolute Häufigkeit** H(x): Anzahl der Beobachtungen, die den Wert x nicht überschreiten

Beispiel:

Messwerte: 1, 3, 3, 4, 4, 4, 5, 5, 6, 10

a_i	$h(a_i)$	$H(a_i)$
1	1	1
3	2	3
4	3	6
5	2	8
6	1	9
10	1	10

Häufigkeitsverteilungen

Kumulierte Häufigkeitsverteilung

- **kumulierte relative Häufigkeit** (empirische Verteilungsfunktion) F(x): Anteil der Beobachtungen, die den Wert x nicht überschreiten

Beispiel:

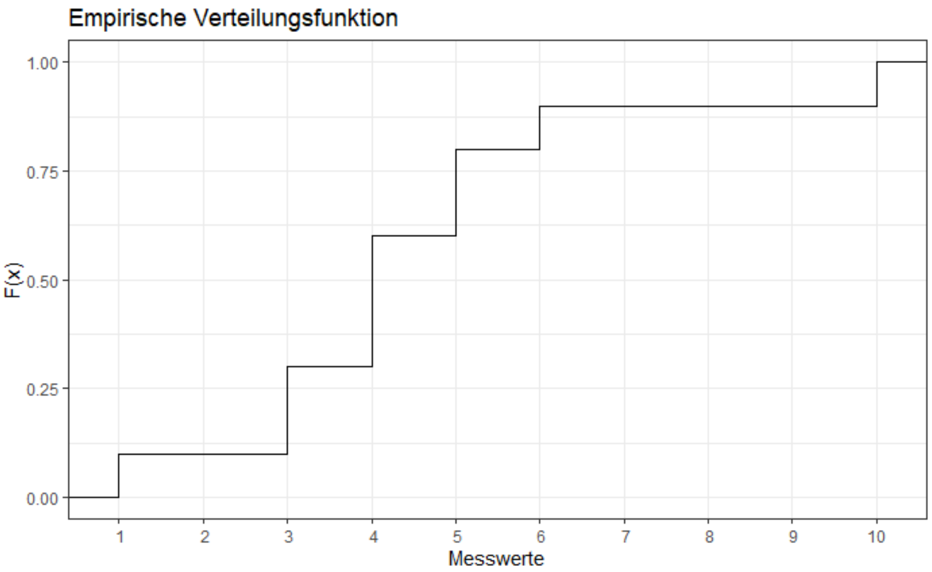
Messwerte: 1, 3, 3, 4, 4, 4, 5, 5, 6, 10

a_i	$f(a_i)$	$F(a_i)$
1	0.1	0.1
3	0.2	0.3
4	0.3	0.6
5	0.2	0.8
6	0.1	0.9
10	0.1	1

Häufigkeitsverteilungen

Grafische Darstellung von kumulierten Häufigkeiten

Stufenfunktion



SPENGERGASSE
ausbildung mit zukunft

Häufigkeitsverteilungen

Klassifizierte Häufigkeitsverteilung

- bei sehr vielen verschiedenen Merkmalswerten werden die Werte in Klassen gruppiert
- bei numerischen Merkmalen(stetig oder diskret mit vielen verschiedenen Werten)

Beispiel:

Messwerte (n = k =26)

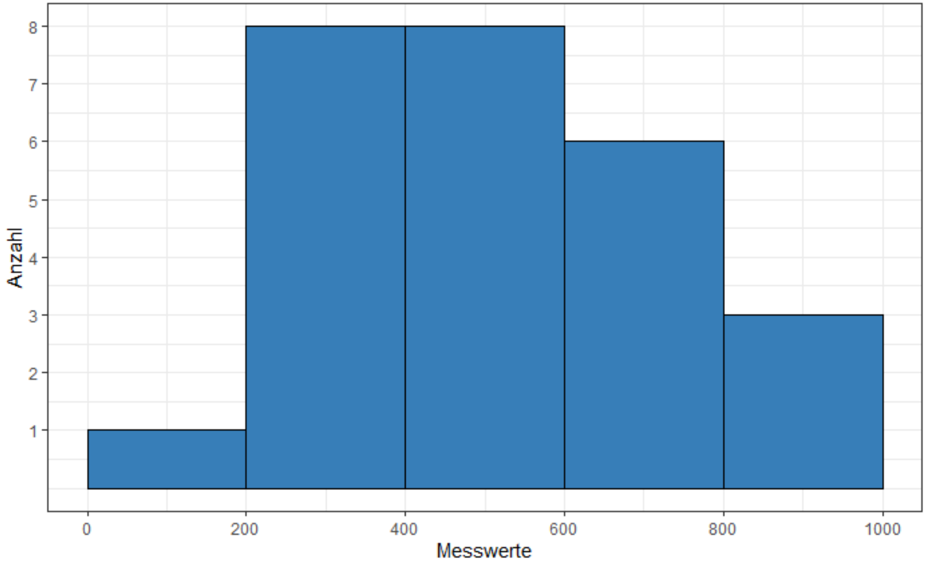
195.35 264.74 271.00 278.15 302.95 312.66
315.69 371.35 399.83 415.31 424.21 461.00
474.82 523.00 556.86 580.00 581.43 638.00
644.05 650.00 657.53 727.03 731.56 816.38
930.00 1000.00

Klasse	*h*	*f*
]0, 200]	1	0.0385
]200, 400]	8	0.308
]400, 600]	8	0.308
]600, 800]	6	0.231
]800, 1000]	3	0.115

Häufigkeitsverteilungen

Grafische Darstellung von klassifizierten Häufigkeiten

Histogramm



SPENGERGASSE
ausbildung mit zukunft

Kontrollfragen

Thema	*Häufigkeitsverteilungen*
Aufgaben	In einer Produktion wurden in den letzten beiden Wochen folgende Anzahl an Ausschussteilen produziert: 5, 3, 2, 1, 5, 4, 10, 2, 1, 2 Erstellen Sie eine Tabelle mit • absoluten Häufigkeiten • relativen Häufigkeiten • kumulierten absoluten Häufigkeiten • kumulierten relativen Häufigkeiten

Lageparameter

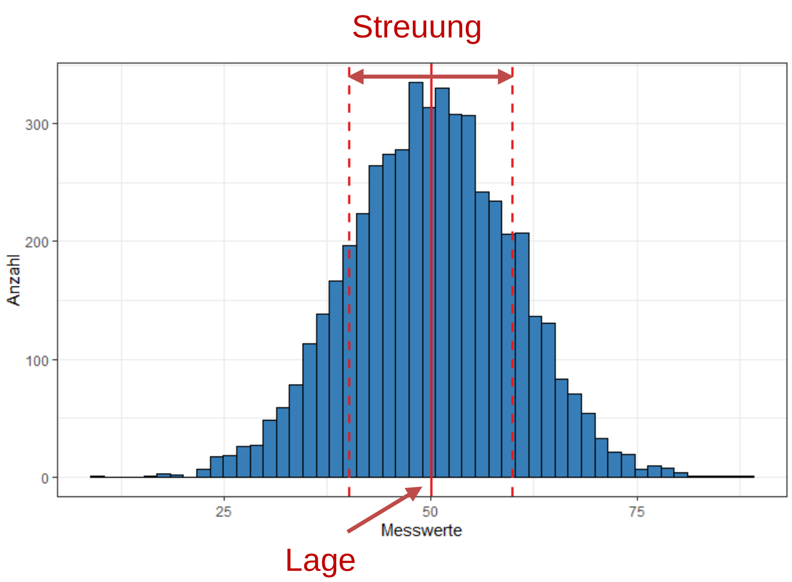
Maßzahlen von Verteilungen

Lageparameter

- beschreiben das Zentrum einer Verteilung durch einen numerischen Wert
- Wahl der Maßzahl ist vom Kontext und vom Skalenniveau des Merkmals abhängig

Streuparameter

- beschreiben die Streuung einer Verteilung durch einen numerischen Wert



SPENGERGASSE
ausbildung mit zukunft

Lageparameter

Arithmetisches Mittel

- Berechnung:
  $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{k} a_i h_i = \sum_{i=1}^{k} a_i f_i$
- Skalenniveau: metrisch
- empfindlich gegen Ausreißer

- Berechnung bei klassifizierten Daten:
  $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} m_i h_i$ (m_i … Klassenmitte)

Beispiel

Messwerte:
1, 3, 3, 4, 4, 4, 5, 5, 6, 10

Arithmetisches Mittel:

$$\bar{x} = \frac{1}{10}(1 + 3 + 3 + 4 + 4 + 4 + \ ... + 10) = 4.5$$

$$\bar{x} = \frac{1}{10}(1 + 2 \cdot 3 + 3 \cdot 4 + 2 \cdot 5 + 6 + 10) = 4.5$$

Lageparameter

Median x_{med}

- liegt in der Mitte der geordneten Stichprobe
- Skalenniveau: ordinal oder metrisch
- nicht empfindlich gegen Ausreißer
- mindestens 50% der Daten sind kleiner oder gleich dem Median
- mindestens 50% der Daten sind größer oder gleich dem Median

Beispiel

Messwerte:
1, 3, 3, 4, 4, 4, 5, 5, 6, 10

Median: $x_{med} = 4$

Messwerte:
1, 3, 3, 4, 4, 5, 5, 5, 6, 10

Median: $x_{med} = 4.5$

Lageparameter

Modus x_{mod}

- häufigster Merkmalswert einer Stichprobe
- jedes Skalenniveau möglich
- nicht immer eindeutig
- nicht empfindlich gegen Ausreißer

Beispiel

Messwerte:
1, 3, 3, 4, 4, 4, 5, 5, 6, 10

Modus:

$$x_{mod} = 4$$

Kontrollfragen	
Thema	*Lageparameter*
Aufgaben	In einer Produktion wurden in den letzten beiden Wochen folgende Anzahl an Ausschussteilen produziert: 5, 3, 2, 1, 5, 4, 10, 2, 1, 2 Bestimmen Sie folgende Lageparameter • Modus • Median • arithmetisches Mittel Welchen Lageparameter halten Sie für diese Aufgabe am geeignetsten?

Streuparameter

Streuparameter

Quantile

- **p-Quantil x_p**:
  Merkmalswert, für den mindestens ein Anteil p kleiner oder gleich x_p und mindestens ein Anteil 1-p größer oder gleich x_p ist.
- **spezielle Quantile:**
  - 25%-Quantil $x_{0.25}$: 1. Quartil, unteres Quantil
  - 50%-Quantil $x_{0.5}$: 2. Quartil, Median
  - 75%-Quantil $x_{0.75}$: 3. Quartil, oberes Quartil
- **Interquartilsabstand IQR** (interquartile range):

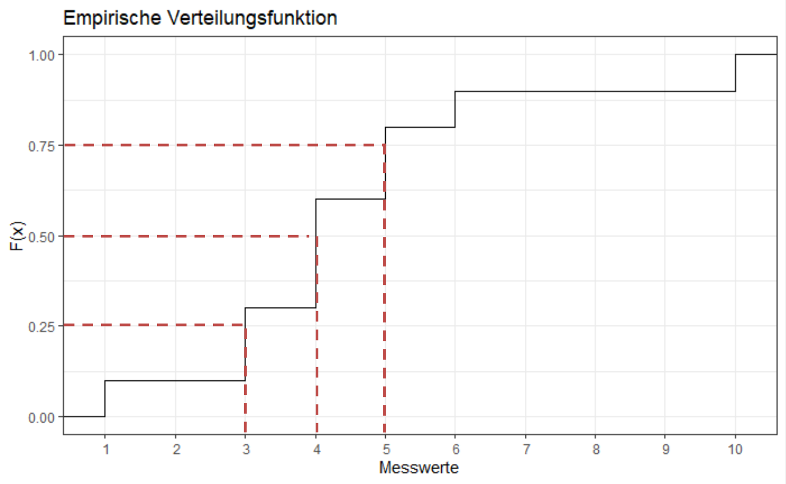
$$IQR = x_{0.75} - x_{0.25}$$

Streuparameter

Beispiel

Messwerte:
1, 3, 3, 4, 4, 4, 5, 5, 6, 10

Quantile:
$x_{0.25} = 3$
$x_{0.5} = 4$
$x_{0.75} = 5$



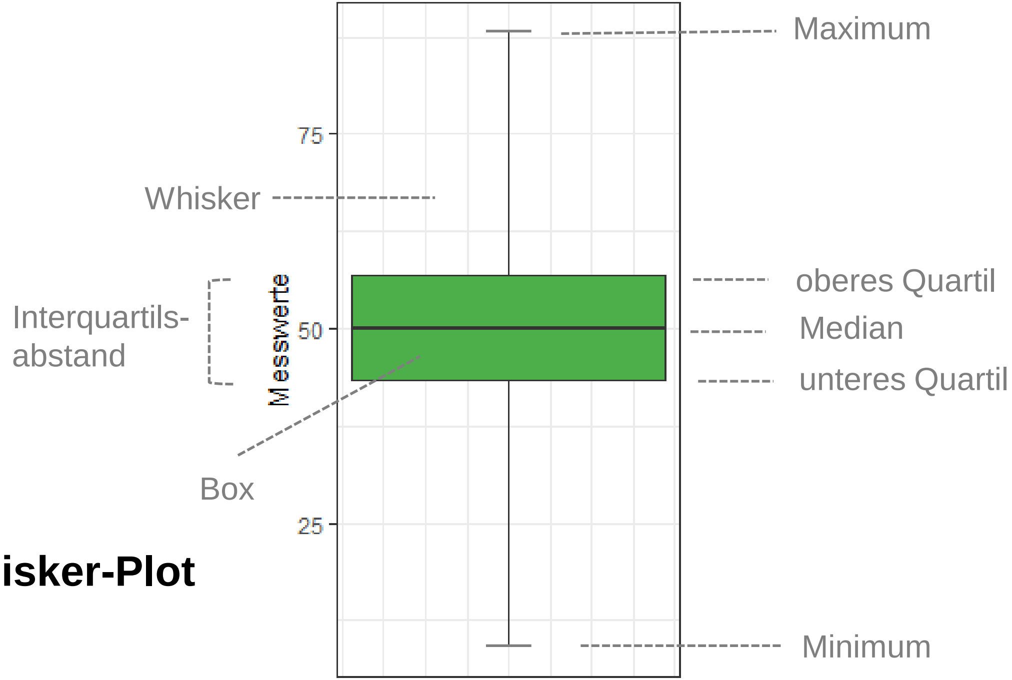
SPENGERGASSE
ausbildung mit zukunft

Streuparameter

Fünf-Punkte-Zusammenfassung

- Minimum x_{min}
- unteres Quartil $x_{0.25}$
- Median $x_{0.5}$
- oberes Quartil $x_{0.75}$
- Maximum x_{max}

Grafische Darstellung als **Box-Whisker-Plot**



SPENGERGASSE
ausbildung mit zukunft

Streuparameter

Spannweite

- Differenz zwischen Maximum und Minimum

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

- Skalenniveau: metrisch

Varianz

- mittlere quadratische Abweichung vom Mittelwert

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

- Skalenniveau: metrisch

Streuparameter

Standardabweichung

- Wurzel aus der Varianz

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

- Skalenniveau: metrisch
- empirische Standardabweichung:

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

Variationskoeffizient

- Maß für relative Streuung

$$v_x = \frac{\sigma_x}{\bar{x}}$$

- dimensionslose Größe
- Skalenniveau: metrisch
- geeignet für den Vergleich von Streuungen bei unterschiedlichen Merkmalen

Streuparameter

Beispiel

Messwerte:
1, 3, 3, 4, 4, 4, 5, 5, 6, 10

- Mittelwert: $\bar{x} = 4.5$
- Spannweite: $R = 10 - 1 = 9$
- Varianz: $\sigma_x^2 = \frac{1}{10}(12.25 + 2 \cdot 2.25 + 3 \cdot 0.25 + 2 \cdot 0.25 + 2.25 + 30.25) = 5.05$
- Standardabweichung: $\sigma_x = \sqrt{5.05} = 2.25$
- Variationskoeffizient: $v_x = \frac{2.25}{4.5} = 0.5$

a_i	$h(a_i)$	$(a_i - \bar{x})^2$
1	1	12.25
3	2	2.25
4	3	0.25
5	2	0.25
6	1	2.25
10	1	30.25

Kontrollfragen	
Thema	*Streuparameter*
Aufgaben	In einer Produktion wurden in den letzten beiden Wochen folgende Anzahl an Ausschussteilen produziert: 5, 3, 2, 1, 5, 4, 10, 2, 1, 2 Bestimmen Sie folgende Streuparameter • Varianz • Standardabweichung • Variationskoeffizient

SPENGERGASSE
ausbildung mit zukunft