8. Lagemaße und Streuungsmaße

Yichen Han

14. Dezember 2023

Inhalt

Wiederholungsquiz

Abbildung: Quiz 9

Stichwörter

1. Modus, Median, Quantil, Mittelwerte, Erwartungswert
2. Spannweite, IQR, Standardabweichung, Varianz, MAD, MedAD, Schiefe, Kurtosis
3. Konzentrationsmaße: Lorenz-Kurve, Gini-Koeffizient, Herfindahl-Index
4. Verschiebungssatz, Tschebyscheffsche Ungleichung, Boxplot

Beispiele

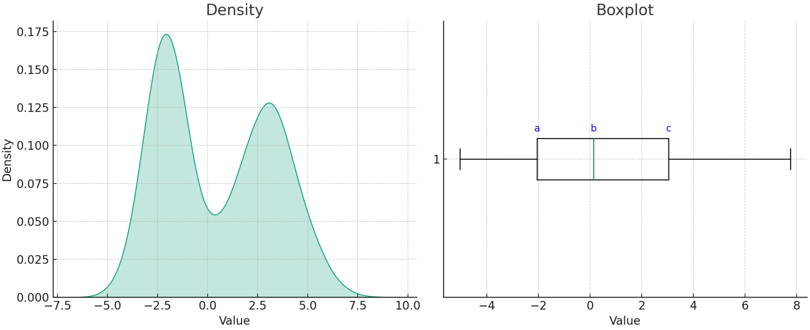

Abbildung: Beispiel 1: eine bimodale Verteilung

Beispiel 2: siehe R-Simulation.

Inhalt

Aufgabe 1.1

Jahr	2012	2013	2014	2015	2016
Umsatz (in Euro)	510000	530700	590200	640800	?

Tabelle: Umsatzentwicklung über die Jahre

(a) Angenommen, der Umsatz im Jahr 2016 konnte um 20 Prozent im Vergleich zum Vorjahr gesteigert werden. Ermitteln Sie den durchschnittlichen Umsatz für die Jahre 2012 bis 2016.

Aufgabe 1.1

Jahr	2012	2013	2014	2015	2016
Umsatz (in Euro)	510000	530700	590200	640800	?

Tabelle: Umsatzentwicklung über die Jahre

(a) Angenommen, der Umsatz im Jahr 2016 konnte um 20 Prozent im Vergleich zum Vorjahr gesteigert werden. Ermitteln Sie den durchschnittlichen Umsatz für die Jahre 2012 bis 2016.

$U_{2016} = (1 + 0.2) \times U_{2015} = 1.2 \times 640800 = 768960,$

$\overline{U} = \frac{1}{5} \sum_{i=1}^{5} U_i = 608132.$

Aufgabe 1.2

Jahr	2012	2013	2014	2015	2016
Umsatz (in Euro)	510000	530700	590200	640800	768960

Tabelle: Umsatzentwicklung über die Jahre

(b) Berechnen Sie das mittlere jährliche Umsatzwachstum in Prozent.

Aufgabe 1.2

Jahr	2012	2013	2014	2015	2016
Umsatz (in Euro)	510000	530700	590200	640800	768960

Tabelle: Umsatzentwicklung über die Jahre

(b) Berechnen Sie das mittlere jährliche Umsatzwachstum in Prozent.

Der t-te Wachstumsfaktor ist definiert als: $i_t = \frac{x_t}{x_{t-1}} \quad$ für $t = 1, \ldots, T$.

Aufgabe 1.2

Jahr	2012	2013	2014	2015	2016
Umsatz (in Euro)	510000	530700	590200	640800	768960

Tabelle: Umsatzentwicklung über die Jahre

(b) Berechnen Sie das mittlere jährliche Umsatzwachstum in Prozent.

Der t-te Wachstumsfaktor ist definiert als: $i_t = \frac{x_t}{x_{t-1}}$ für $t = 1, \ldots, T$.

Das passende Mittel: Geometrisches Mittel! (Wachstumsfaktor ist ein multiplikativer Faktor).

Jahr	2012	2013	2014	2015	2016
Umsatz (in Euro)	510000	530700	590200	640800	768960

Tabelle: Umsatzentwicklung über die Jahre

(b) Berechnen Sie das mittlere jährliche Umsatzwachstum in Prozent.

Der t-te Wachstumsfaktor ist definiert als: $i_t = \frac{x_t}{x_{t-1}}$ für $t = 1, \ldots, T$.

Das passende Mittel: Geometrisches Mittel! (Wachstumsfaktor ist ein multiplikativer Faktor).

$$\bar{x}_G = \sqrt[T]{i_1 \cdot \ldots \cdot i_T} = \sqrt[T]{\frac{x_1}{x_0} \cdot \frac{x_2}{x_1} \cdot \ldots \cdot \frac{x_T}{x_{T-1}}} = \sqrt[T]{\frac{x_T}{x_0}}$$

$$= \sqrt[4]{\frac{x_4}{x_0}} = \sqrt[4]{\frac{768960}{510000}} \approx 1.1081.$$

Aufgabe 2.1

Feinstaub: 20, 27, 22, 20, 22, 20, 19, 20, 17, 27, 23, 27, 22, 27, 25.

(a) Geben Sie das arithmetische Mittel sowie den Median und die beiden Quartile an.

Feinstaub: 20, 27, 22, 20, 22, 20, 19, 20, 17, 27, 23, 27, 22, 27, 25.

(a) Geben Sie das arithmetische Mittel sowie den Median und die beiden Quartile an.

Stichprobenumfang $n = 15$,
Geordnete Werte: $17, 19, 20, 20, 20, 20, 22, 22, 22, 23, 25, 27, 27, 27, 27$.

$$\bar{x} = \frac{1}{15}\sum_{i=1}^{15} x_i = \frac{1}{15} \cdot 338 \approx 22.53$$

Feinstaub: 20, 27, 22, 20, 22, 20, 19, 20, 17, 27, 23, 27, 22, 27, 25.

(a) Geben Sie das arithmetische Mittel sowie den Median und die beiden Quartile an.

Stichprobenumfang $n = 15$,
Geordnete Werte: $17, 19, 20, 20, 20, 20, 22, 22, 22, 23, 25, 27, 27, 27, 27.$

$$\bar{x} = \frac{1}{15} \sum_{i=1}^{15} x_i = \frac{1}{15} \cdot 338 \approx 22.53$$

p-Quantile:

$$\widetilde{x}_p = \begin{cases} x_{(k)} & np \notin \mathbb{Z}, \; k \text{ kleinste ganze Zahl} > np \\ \in [x_{(k)}; x_{(k+1)}] & k = np \in \mathbb{Z} \end{cases}$$

$$\widetilde{x}_{0.25} = x_{(4)} = 20, \quad \widetilde{x}_{med} = x_{(8)} = 22, \quad \widetilde{x}_{0.75} = x_{(12)} = 27$$

Aufgabe 3.1

Sei Z eine Zufallsvariable, die das Minimum von drei Würfen eines sechsseitigen Würfels beschreibt.

$$F(z) = \begin{cases} 0 & \text{für } z < 1 \\ \frac{91}{216} & \text{für } 1 \leq z < 2 \\ \frac{152}{216} & \text{für } 2 \leq z < 3 \\ \frac{189}{216} & \text{für } 3 \leq z < 4 \\ \frac{208}{216} & \text{für } 4 \leq z < 5 \\ \frac{215}{216} & \text{für } 5 \leq z < 6 \\ 1 & \text{für } z \geq 6. \end{cases}$$

Berechnen Sie $E(Z)$.

Aufgabe 3.2

$$f(z) = \begin{cases} 0 & \text{für } z \notin \{1, 2, \ldots, 6\} \\ \frac{91}{216} & \text{für } z = 1 \\ \frac{152-91}{216} = \frac{61}{216} & \text{für } z = 2 \\ \frac{189-152}{216} = \frac{37}{216} & \text{für } z = 3 \leq z < 4 \\ \frac{208-189}{216} = \frac{19}{216} & \text{für } z = 4 \leq z < 5 \\ \frac{215-208}{216} = \frac{7}{216} & \text{für } z = 5 \leq z < 6 \\ \frac{216-215}{216} = \frac{1}{216} & \text{für } z = 6. \end{cases}$$

$$f(z) = \begin{cases} 0 & \text{für } z \notin \{1, 2, \ldots, 6\} \\ \frac{91}{216} & \text{für } z = 1 \\ \frac{152-91}{216} = \frac{61}{216} & \text{für } z = 2 \\ \frac{189-152}{216} = \frac{37}{216} & \text{für } z = 3 \leq z < 4 \\ \frac{208-189}{216} = \frac{19}{216} & \text{für } z = 4 \leq z < 5 \\ \frac{215-208}{216} = \frac{7}{216} & \text{für } z = 5 \leq z < 6 \\ \frac{216-215}{216} = \frac{1}{216} & \text{für } z = 6. \end{cases}$$

$$E(Z) = \sum_{z \in T_Z} z f(z) = 1 \cdot \frac{91}{216} + 2 \cdot \frac{61}{216} + \ldots + 6 \cdot \frac{1}{216} \approx 2.042$$

Aufgabe 4.1

Stunden	[0,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)	[9,10)	Σ
1960	5	3	10	9	13	18	21	27	12	3	$n_1 = 121$
1980	6	7	5	20	29	27	13	5	3	2	$n_2 = 117$
2000	35	24	13	8	9	4	2	1	0	1	$n_3 = 97$

Gruppierte Lagemaße

- **Modus:** Bestimme Modalklasse (Klasse mit der größten Beobachtungszahl) und verwende Klassenmitte (m_i) als Modus.
- **Median:** Bestimme Einfallsklasse $[c_{i-1}, c_i)$ des Medians und daraus

$$\tilde{x}_{\text{med,grupp}} = c_{i-1} + \frac{d_i \cdot (0.5 - F(c_{i-1}))}{f_i}.$$

- **Arithmetisches Mittel:**

$$\bar{x}_{\text{grupp}} = \sum_{i=1}^{k} f_i m_i.$$

Quelle: Fahrmeir et al., Kapitel 2.2, Seite 55

Stunden	**[0,1)**	**[1,2)**	**[2,3)**	**[3,4)**	**[4,5)**	**[5,6)**	**[6,7)**	**[7,8)**	**[8,9)**	**[9,10)**	Σ
1960	5	3	10	9	13	18	21	27	12	3	$n_1 = 121$
1980	6	7	5	20	29	27	13	5	3	2	$n_2 = 117$
2000	35	24	13	8	9	4	2	1	0	1	$n_3 = 97$

(a) Bestimmen Sie aus den klassierten Daten jeweils Modus und Median für die drei Jahre.

Aufgabe 4.2

Stunden	**[0,1)**	**[1,2)**	**[2,3)**	**[3,4)**	**[4,5)**	**[5,6)**	**[6,7)**	**[7,8)**	**[8,9)**	**[9,10)**	Σ
1960	5	3	10	9	13	18	21	**27**	12	3	$n_1 = 121$
1980	6	7	5	20	**29**	27	13	5	3	2	$n_2 = 117$
2000	**35**	24	13	8	9	4	2	1	0	1	$n_3 = 97$

(a) Bestimmen Sie aus den klassierten Daten jeweils Modus und Median für die drei Jahre.

Stunden	[0,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)	[9,10)	Σ
1960	5	3	10	9	13	18	21	**27**	12	3	$n_1 = 121$
1980	6	7	5	20	**29**	27	13	5	3	2	$n_2 = 117$
2000	**35**	24	13	8	9	4	2	1	0	1	$n_3 = 97$

(a) Bestimmen Sie aus den klassierten Daten jeweils Modus und Median für die drei Jahre.

Jahr	Modalklasse	Modus (Klassenmitte)
1960	$[7, 8)$	7.5
1980	$[4, 5)$	4.5
2000	$[0, 1)$	0.5

Stunden	**[0,1)**	**[1,2)**	**[2,3)**	**[3,4)**	**[4,5)**	**[5,6)**	**[6,7)**	**[7,8)**	**[8,9)**	**[9,10)**	Σ
1960	5	3	10	9	13	18	21	**27**	12	3	$n_1 = 121$
1980	6	7	5	20	**29**	27	13	5	3	2	$n_2 = 117$
2000	**35**	24	13	8	9	4	2	1	0	1	$n_3 = 97$

(a) Bestimmen Sie aus den klassierten Daten jeweils Modus und Median für die drei Jahre.

Jahr	Modalklasse	Modus (Klassenmitte)
1960	$[7, 8)$	7.5
1980	$[4, 5)$	4.5
2000	$[0, 1)$	0.5

Median:
Berechnung der relativen und kumulierten relativen Häufigkeiten für die einzelnen Klassen: (nächste Seite)

Aufgabe 4.3

Klasse j Stunden	1 [0,1)	2 [1,2)	3 [2,3)	4 [3,4)	5 [4,5)	6 [5,6)	7 [6,7)	8 [7,8)	9 [8,9)	10 [9,10)
f_m^{1960}	0.04	0.02	0.08	0.07	0.11	0.15	0.17	0.22	0.10	0.02
F_m^{1960}	0.04	0.07	0.15	0.22	0.33	0.48	**0.65**	0.88	0.98	1.00
f_m^{1980}	0.05	0.06	0.04	0.17	0.25	0.23	0.11	0.04	0.03	0.02
F_m^{1980}	0.05	0.11	0.15	0.32	**0.57**	0.80	0.91	0.96	0.98	1.00
f_m^{2000}	0.36	0.25	0.13	0.08	0.09	0.04	0.02	0.01	0.00	0.01
F_m^{2000}	0.36	**0.61**	0.74	0.82	0.92	0.96	0.98	0.99	0.99	1.00

Aufgabe 4.3

Klasse j Stunden	1 [0,1)	2 [1,2)	3 [2,3)	4 [3,4)	5 [4,5)	6 [5,6)	7 [6,7)	8 [7,8)	9 [8,9)	10 [9,10)
f_m^{1960}	0.04	0.02	0.08	0.07	0.11	0.15	0.17	0.22	0.10	0.02
F_m^{1960}	0.04	0.07	0.15	0.22	0.33	0.48	**0.65**	0.88	0.98	1.00
f_m^{1980}	0.05	0.06	0.04	0.17	0.25	0.23	0.11	0.04	0.03	0.02
F_m^{1980}	0.05	0.11	0.15	0.32	**0.57**	0.80	0.91	0.96	0.98	1.00
f_m^{2000}	0.36	0.25	0.13	0.08	0.09	0.04	0.02	0.01	0.00	0.01
F_m^{2000}	0.36	**0.61**	0.74	0.82	0.92	0.96	0.98	0.99	0.99	1.00

$$1960\text{: } \widetilde{x}_{med,1} = 6 + \frac{0.5 - 0.48}{0.17} \approx 6.12$$

$$1980\text{: } \widetilde{x}_{med,2} = 4 + \frac{0.5 - 0.32}{0.25} \approx 4.72$$

$$2000\text{: } \widetilde{x}_{med,3} = 1 + \frac{0.5 - 0.36}{0.25} \approx 1.56$$

(b) Wie drücken sich die im Laufe der Zeit veränderten Hörgewohnheiten durch die drei unter (a) berechneten Lagemaße aus?

(b) Wie drücken sich die im Laufe der Zeit veränderten Hörgewohnheiten durch die drei unter (a) berechneten Lagemaße aus?

Jahr	Modus (Klassenmitte)	Median
1960	7.5	6.12
1980	4.5	4.72
2000	0.5	1.56

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

(b) Wie drücken sich die im Laufe der Zeit veränderten Hörgewohnheiten durch die drei unter (a) berechneten Lagemaße aus?

Jahr	Modus (Klassenmitte)	Median
1960	7.5	6.12
1980	4.5	4.72
2000	0.5	1.56

Modus und Median wurden über die Zeit kleiner, dies drückt einen rückgängigen Radiokonsum im Zeitverlauf aus. Dieser Effekt fällt beim Modus stärker aus.

(c) Wie können Sie aus diesen klassierten Daten arithmetische Mittelwerte für jedes Jahr berechnen? Welche zusätzlichen Annahmen müssen Sie dafür treffen?

Arithmetisches Mittel (Verwendung der Klassenmittelpunkte):

$$\text{1960: } \bar{x}_1 = \frac{1}{121}(0.5 \cdot 5 + \ldots + 9.5 \cdot 3) \approx 5.71$$

$$\text{1980: } \bar{x}_2 = \frac{1}{117}(0.5 \cdot 6 + \ldots + 9.5 \cdot 2) \approx 4.63$$

$$\text{2000: } \bar{x}_3 = \frac{1}{97}(0.5 \cdot 35 + \ldots + 9.5 \cdot 1) \approx 2.13$$

Die Formeln aus Fahrmeir sind im Prinzip *lineare Interpolation* der Daten.

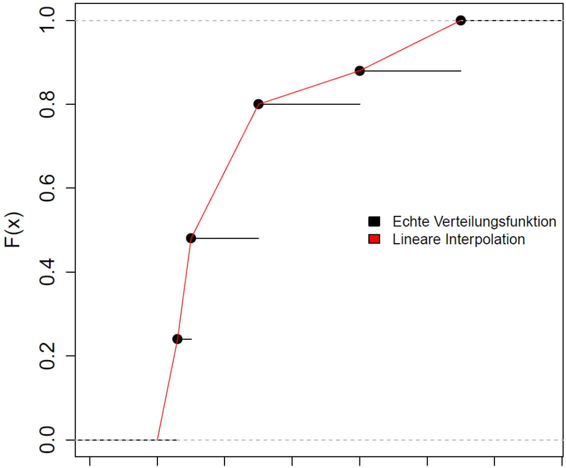

Abbildung: Beispiel: Lineare Interpolation einer empirischen Verteilung

Aufgabe 4.7

$$\widetilde{x}_{\text{med,grupp}} = \overbrace{c_{i-1}}^{\text{Basis}} + \underbrace{\frac{d_i \cdot (0.5 - F(c_{i-1}))}{f_i}}_{\text{Modifikation unter Gleichverteilung}}$$

$$\widetilde{x}_{\text{med,grupp}} = \overbrace{c_{i-1}}^{\text{Basis}} + \underbrace{\frac{d_i \cdot (0.5 - F(c_{i-1}))}{f_i}}_{\text{Modifikation unter Gleichverteilung}}$$

Annahmen der gruppierten Lagemaße

- gleiche Klassenbreite (optional)
- Die Merkmalswerte sind innerhalb aller Klassen gleichverteilt. (zentral)