9. Maße (weiterführend)

Yichen Han

21. Dezember 2023

Inhalt

Wiederholungsquiz

Abbildung: Quiz 10

Stichwörter

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

Aufgabe 1.1

stündliche Mittelwerte für die Temperatur in $^\circ C$:

7.2	7.2	6.8	7.2	7.4	7.1	7.0	7.2	7.5	7.7	7.9	8.3
7.9	7.6	7.2	7.1	6.9	6.9	6.7	6.2	6.4	6.2	6.2	6.2

(a) Modus, arithmetische Mittel, Median, IQR, Varianz. Charakterisieren.

Aufgabe 1.1

stündliche Mittelwerte für die Temperatur in $^\circ C$:

7.2	7.2	6.8	7.2	7.4	7.1	7.0	7.2	7.5	7.7	7.9	8.3
7.9	7.6	7.2	7.1	6.9	6.9	6.7	6.2	6.4	6.2	6.2	6.2

(a) Modus, arithmetische Mittel, Median, IQR, Varianz. Charakterisieren.
Beobachtungen: $n = 24$
Geordnete Urliste:

6.2	6.2	6.2	6.2	6.4	6.7	6.8	6.9	6.9	7.0	7.1	7.1
7.2	7.2	7.2	7.2	7.2	7.4	7.5	7.6	7.7	7.9	7.9	8.3

stündliche Mittelwerte für die Temperatur in $^\circ C$:

7.2	7.2	6.8	7.2	7.4	7.1	7.0	7.2	7.5	7.7	7.9	8.3
7.9	7.6	7.2	7.1	6.9	6.9	6.7	6.2	6.4	6.2	6.2	6.2

(a) Modus, arithmetische Mittel, Median, IQR, Varianz. Charakterisieren.
Beobachtungen: $n = 24$
Geordnete Urliste:

6.2	6.2	6.2	6.2	6.4	6.7	6.8	6.9	6.9	7.0	7.1	7.1
7.2	7.2	7.2	7.2	7.2	7.4	7.5	7.6	7.7	7.9	7.9	8.3

Modus: $x_{mod} = 7.2\ ^\circ C$

Aufgabe 1.1

stündliche Mittelwerte für die Temperatur in $^\circ C$:

7.2	7.2	6.8	7.2	7.4	7.1	7.0	7.2	7.5	7.7	7.9	8.3
7.9	7.6	7.2	7.1	6.9	6.9	6.7	6.2	6.4	6.2	6.2	6.2

(a) Modus, arithmetische Mittel, Median, IQR, Varianz. Charakterisieren.
Beobachtungen: $n = 24$
Geordnete Urliste:

6.2	6.2	6.2	6.2	6.4	6.7	6.8	6.9	6.9	7.0	7.1	7.1
7.2	7.2	7.2	7.2	7.2	7.4	7.5	7.6	7.7	7.9	7.9	8.3

Modus: $x_{mod} = 7.2\ ^\circ C$
Arithmetisches Mittel: $\bar{x} = \frac{1}{24} \sum_{i=1}^{24} x_i = \frac{1}{24} \cdot 170 \approx 7.08\ ^\circ C$

Aufgabe 1.1

stündliche Mittelwerte für die Temperatur in $^\circ C$:

7.2	7.2	6.8	7.2	7.4	7.1	7.0	7.2	7.5	7.7	7.9	8.3
7.9	7.6	7.2	7.1	6.9	6.9	6.7	6.2	6.4	6.2	6.2	6.2

(a) Modus, arithmetische Mittel, Median, IQR, Varianz. Charakterisieren.
Beobachtungen: $n = 24$
Geordnete Urliste:

6.2	6.2	6.2	6.2	6.4	6.7	6.8	6.9	6.9	7.0	7.1	7.1
7.2	7.2	7.2	7.2	7.2	7.4	7.5	7.6	7.7	7.9	7.9	8.3

Modus: $x_{mod} = 7.2\ ^\circ C$
Arithmetisches Mittel: $\bar{x} = \frac{1}{24} \sum_{i=1}^{24} x_i = \frac{1}{24} \cdot 170 \approx 7.08\ ^\circ C$
Median: $\widetilde{x}_{med} = \frac{1}{2}[x_{(12)} + x_{(13)}] = \frac{1}{2}[7.1 + 7.2] = 7.15\ ^\circ C$
IQR: $d_{Q,X} = \widetilde{x}_{0.75} - \widetilde{x}_{0.25} = \frac{1}{2}\left(x_{(18)} + x_{(19)}\right) - \frac{1}{2}\left(x_{(6)} + x_{(7)}\right) =$
$7.45 - 6.75 = 0.7$

Varianz:

$$\tilde{S}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1211.86}{24} - 7.08^2 \approx 0.32\ ^\circ C^2$$

Aufgabe 1.2

Varianz:

$$\tilde{S}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1211.86}{24} - 7.08^2 \approx 0.32\ {}^\circ C^2$$

Charakterisierung der Verteilung:

Beobachtung: Modus $\approx$ Median $\approx$ Mittelwert

Varianz:

$$\tilde{S}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1211.86}{24} - 7.08^2 \approx 0.32\ {}^\circ C^2$$

Charakterisierung der Verteilung:

Beobachtung: Modus $\approx$ Median $\approx$ Mittelwert

Da arithmetisches Mittel, Median und Modus sehr ähnlich sind, kann von einer **symmetrischen** Verteilung ausgegangen werden.

- Y: Stündliches Temperaturmittel in München am 27.11.2012 (in °F)
- Modus:

$$y_{mod} = a + bx_{mod} = 32 + \frac{9}{5} \cdot 7.2 = 44.96 \quad [°F]$$

- Arithmetisches Mittel:

$$\bar{y} = a + b\bar{x} = 32 + \frac{9}{5} \cdot 7.08 \approx 44.74 \quad [°F]$$

- Median:

$$\widetilde{y}_{med} = a + b\widetilde{x}_{med} = 32 + \frac{9}{5} \cdot 7.15 = 44.87 \quad [°F]$$

- Unteres Quartil:

$$\widetilde{y}_{0.25} = a + b\widetilde{x}_{0.25} = 32 + \frac{9}{5} \cdot 6.75 = 44.15 \quad [°F]$$

- Oberes Quartil:

$$\widetilde{y}_{0.75} = a + b\widetilde{x}_{0.75} = 32 + \frac{9}{5} \cdot 7.45 = 45.41 \quad [°F]$$

- Interquartilsabstand:

$$d_{Q,Y} = \widetilde{y}_{0.75} - \widetilde{y}_{0.25} = a + b\widetilde{x}_{0.75} - (a + b\widetilde{x}_{0.25}) = b(\widetilde{x}_{0.75} - \widetilde{x}_{0.25}) = bd_{QX} = \frac{9}{5} \cdot 0.7 = 1.26 \quad [°F]$$

- Varianz

$$\tilde{S}_Y^2 = b^2 \tilde{S}_X^2 = \left(\frac{9}{5}\right)^2 \cdot 0.32 \approx 1.04 \quad [°F^2]$$

(c) Geben Sie den MAD und den MedAD für die Temperatur in $^\circ$C an.

(c) Geben Sie den MAD und den MedAD für die Temperatur in °C an.

- MAD:

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| = \frac{1}{24}\sum_{i=1}^{24}|x_i - 7.08| \approx 0.44$$

(c) Geben Sie den MAD und den MedAD für die Temperatur in °C an.

- MAD:

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| = \frac{1}{24}\sum_{i=1}^{24}|x_i - 7.08| \approx 0.44$$

- MedAD:

$$MedAD = median(|x_i - x_{med}|) = median(|x_i - 7.15|)$$

(c) Geben Sie den MAD und den MedAD für die Temperatur in °C an.

- MAD:

$$MAD = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}| = \frac{1}{24} \sum_{i=1}^{24} |x_i - 7.08| \approx 0.44$$

- MedAD:

$$MedAD = median(|x_i - x_{med}|) = median(|x_i - 7.15|)$$

Geordnete Urliste der $|x_i - 7.15|$:

0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.15	0.25	0.25	0.25	0.35
0.35	0.45	0.45	0.55	0.75	0.75	0.75	0.95	0.95	0.95	0.95	1.15

(c) Geben Sie den MAD und den MedAD für die Temperatur in °C an.

- MAD:

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| = \frac{1}{24}\sum_{i=1}^{24}|x_i - 7.08| \approx 0.44$$

- MedAD:

$$MedAD = median(|x_i - x_{med}|) = median(|x_i - 7.15|)$$

Geordnete Urliste der $|x_i - 7.15|$:

0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.15	0.25	0.25	0.25	0.35
0.35	0.45	0.45	0.55	0.75	0.75	0.75	0.95	0.95	0.95	0.95	1.15

$\Rightarrow$

$$MedAD = \frac{1}{2}(0.35 + 0.35) = 0.35$$

(d) Der Variationskoeffizient gilt als skalierungsunabhängiges Streuungsmaß. Kann daraus abgeleitet werden, dass die Variationskoeffizienten für die Temperatur in $^\circ$F und in $^\circ$C gleich sind? Begründen Sie.

(d) Der Variationskoeffizient gilt als skalierungsunabhängiges Streuungsmaß. Kann daraus abgeleitet werden, dass die Variationskoeffizienten für die Temperatur in °F und in °C gleich sind? Begründen Sie.

$$y[^\circ F] = \frac{9}{5}x[^\circ C] + 32$$

(d) Der Variationskoeffizient gilt als skalierungsunabhängiges Streuungsmaß. Kann daraus abgeleitet werden, dass die Variationskoeffizienten für die Temperatur in °F und in °C gleich sind? Begründen Sie.

$$y[^\circ F] = \frac{9}{5}x[^\circ C] + 32$$

$\Rightarrow$ Nein! Die Streuung der beiden Verteilungen kann ohne vorheriges Umrechnen nicht mit Hilfe des Variationskoeffizienten verglichen werden. $v_X = v_Y$ gilt nur, wenn $a = 0$ gilt.

Aufgabe 2.1

Folien 385, 388-390

Schüler	1	2	3	4	5
Taschengeld	50	80	20	65	40

(a) Berechnen und interpretieren Sie ein auf den Bereich $[0, 1]$ normiertes Maß für die Konzentration des Taschengeldes. Stellen Sie die Situation graphisch dar.

Aufgabe 2.1

Folien 385, 388-390

Schüler	1	2	3	4	5
Taschengeld	50	80	20	65	40

(a) Berechnen und interpretieren Sie ein auf den Bereich $[0, 1]$ normiertes Maß für die Konzentration des Taschengeldes. Stellen Sie die Situation graphisch dar.

normierte Gini-Koeffizient $G^{+} = \frac{n}{n-1} G \in [0, 1]$.

Aufgabe 2.2

i	$x_{(i)}$	$u_i = \frac{i}{n}$	$v_i = \frac{\sum_{j=1}^{i} x_{(j)}}{\sum_{j=1}^{n} x_{(j)}}$
1	20	$\frac{1}{5} = 0.2$	$\frac{20}{255} \approx 0.078$
2	40	$\frac{2}{5} = 0.4$	$\frac{60}{255} \approx 0.235$
3	50	$\frac{3}{5} = 0.6$	$\frac{110}{255} \approx 0.431$
4	65	$\frac{4}{5} = 0.8$	$\frac{175}{255} \approx 0.686$
5	80	$\frac{5}{5} = 1$	$\frac{255}{255} = 1$

Aufgabe 2.2

i	$x_{(i)}$	$u_i = \frac{i}{n}$	$v_i = \frac{\sum_{j=1}^{i} x_{(j)}}{\sum_{j=1}^{n} x_{(j)}}$
1	20	$\frac{1}{5} = 0.2$	$\frac{20}{255} \approx 0.078$
2	40	$\frac{2}{5} = 0.4$	$\frac{60}{255} \approx 0.235$
3	50	$\frac{3}{5} = 0.6$	$\frac{110}{255} \approx 0.431$
4	65	$\frac{4}{5} = 0.8$	$\frac{175}{255} \approx 0.686$
5	80	$\frac{5}{5} = 1$	$\frac{255}{255} = 1$

$$G = 1 - \frac{1}{n}\sum_{j=1}^{n}(v_{j-1} + v_j) = 1 - \frac{1}{5}((0 + 0.078) + \ldots) \approx 0.227$$

Aufgabe 2.2

i	$x_{(i)}$	$u_i = \frac{i}{n}$	$v_i = \frac{\sum_{j=1}^{i} x_{(j)}}{\sum_{j=1}^{n} x_{(j)}}$
1	20	$\frac{1}{5} = 0.2$	$\frac{20}{255} \approx 0.078$
2	40	$\frac{2}{5} = 0.4$	$\frac{60}{255} \approx 0.235$
3	50	$\frac{3}{5} = 0.6$	$\frac{110}{255} \approx 0.431$
4	65	$\frac{4}{5} = 0.8$	$\frac{175}{255} \approx 0.686$
5	80	$\frac{5}{5} = 1$	$\frac{255}{255} = 1$

$$G = 1 - \frac{1}{n}\sum_{j=1}^{n}(v_{j-1} + v_j) = 1 - \frac{1}{5}((0 + 0.078) + \ldots) \approx 0.227$$

Normierter Gini-Koeffizient: $G^+ = \frac{n}{n-1} G = \frac{5}{4} \cdot 0.227 \approx 0.284$

Aufgabe 2.2

i	$x_{(i)}$	$u_i = \frac{i}{n}$	$v_i = \frac{\sum_{j=1}^{i} x_{(j)}}{\sum_{j=1}^{n} x_{(j)}}$
1	20	$\frac{1}{5} = 0.2$	$\frac{20}{255} \approx 0.078$
2	40	$\frac{2}{5} = 0.4$	$\frac{60}{255} \approx 0.235$
3	50	$\frac{3}{5} = 0.6$	$\frac{110}{255} \approx 0.431$
4	65	$\frac{4}{5} = 0.8$	$\frac{175}{255} \approx 0.686$
5	80	$\frac{5}{5} = 1$	$\frac{255}{255} = 1$

$$G = 1 - \frac{1}{n}\sum_{j=1}^{n}(v_{j-1} + v_j) = 1 - \frac{1}{5}((0 + 0.078) + \ldots) \approx 0.227$$

Normierter Gini-Koeffizient: $G^{+} = \frac{n}{n-1} G = \frac{5}{4} \cdot 0.227 \approx 0.284$
Relativ geringe Konzentration des monatlichen Taschengeldes.

(b) Wie ändert sich das in (a) berechnete Konzentrationsmaß, wenn jeder Schüler

(i) 10 Euro mehr

(ii) das doppelte von seinem ursprünglichen Taschengeld

bekommt?

(b) Wie ändert sich das in (a) berechnete Konzentrationsmaß, wenn jeder Schüler
   (i) 10 Euro mehr
   (ii) das doppelte von seinem ursprünglichen Taschengeld

   bekommt?

(i) Der normierte Gini-Koeffizient wird geringer, wenn jeder Schüler 10 Euro mehr Taschengeld bekommt. Die Konzentration nimmt ab, da der Anteil, den die Schüler mit dem geringsten Taschengeld erhalten, größer wird.

(b) Wie ändert sich das in (a) berechnete Konzentrationsmaß, wenn jeder Schüler
   (i) 10 Euro mehr
   (ii) das doppelte von seinem ursprünglichen Taschengeld

   bekommt?

(i) Der normierte Gini-Koeffizient wird geringer, wenn jeder Schüler 10 Euro mehr Taschengeld bekommt. Die Konzentration nimmt ab, da der Anteil, den die Schüler mit dem geringsten Taschengeld erhalten, größer wird.

(ii) Der normierte Gini-Koeffizient bleibt gleich, da die Konzentration unverändert bleibt. Ändern sich die Werte aller Schüler um einen konstanten Faktor, so bleiben die Anteile am gesamten Taschengeld für alle identisch.

(c) Statt fünf Schülern werden nun 485 Schüler betrachtet. Ändert sich das Konzentrationsmaß, wenn jeweils 97 Schüler ein monatliches Taschengeld von 20 Euro, von 40 Euro, von 50 Euro, von 65 Euro und von 80 Euro bekommen?

(c) Statt fünf Schülern werden nun 485 Schüler betrachtet. Ändert sich das Konzentrationsmaß, wenn jeweils 97 Schüler ein monatliches Taschengeld von 20 Euro, von 40 Euro, von 50 Euro, von 65 Euro und von 80 Euro bekommen?

- Der Gini-Koeffizient ändert sich nicht, da die prozentuale Aufteilung der Taschengeldbeträge auf fünf Gruppen aus jeweils 97 Schülern identisch zur Aufteilung auf fünf Schüler ist.

Aufgabe 2.4

(c) Statt fünf Schülern werden nun 485 Schüler betrachtet. Ändert sich das Konzentrationsmaß, wenn jeweils 97 Schüler ein monatliches Taschengeld von 20 Euro, von 40 Euro, von 50 Euro, von 65 Euro und von 80 Euro bekommen?

- Der Gini-Koeffizient ändert sich nicht, da die prozentuale Aufteilung der Taschengeldbeträge auf fünf Gruppen aus jeweils 97 Schülern identisch zur Aufteilung auf fünf Schüler ist.
- Der normierte Gini-Koeffizient ändert sich jedoch, da dieser von der Stichprobengröße abhängt.

Aufgabe 3,4

Siehe Tafel

Frohe Weihnachten