DHBW Mannheim
Informatik

Mannheim, 30.3.2020
Mathematik II/2

Lösungen zu Übungsblatt 1

Aufgabe 1. Vom Merkmal X werden die Merkmalswerte $x_1, \ldots, x_n$ beobachtet. Mit $\overline{x} = \frac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i$ bezeichnen wir das arithmetische Mittel.

a) Zeigen Sie, dass gilt: $\sum\limits_{i=1}^{n} (x_i - \overline{x}) = 0$.

Lösung:

Es gilt

$$\begin{aligned} \sum_{i=1}^{n} (x_i - \overline{x}) &= \left(\sum_{i=1}^{n} x_i \right) - n \cdot \overline{x} \\ &= \left(\sum_{i=1}^{n} x_i \right) - n \cdot \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i \right) \\ &= \left(\sum_{i=1}^{n} x_i \right) - \left(\sum_{i=1}^{n} x_i \right) \\ &= 0 \end{aligned}$$

b) Zeigen Sie, dass für alle $a \in \mathbb{R}$ gilt:

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

und folgern Sie daraus, dass

1. $\sum\limits_{i=1}^{n} (x_i - \overline{x})^2 \leq \sum\limits_{i=1}^{n} (x_i - a)^2$.
2. $\widetilde{s}^2 = \frac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i^2 - \overline{x}^2$.

Lösung:

Es gilt für jedes $a \in \mathbb{R}$:

$$\begin{aligned} \sum_{i=1}^{n} (x_i - a)^2 &= \sum_{i=1}^{n} (x_i - \overline{x} + \overline{x} - a)^2 \\ &= \sum_{i=1}^{n} \left((x_i - \overline{x})^2 + 2(x_i - \overline{x})(\overline{x} - a) + (\overline{x} - a)^2 \right) \\ &= \sum_{i=1}^{n} (x_i - \overline{x})^2 + 2(\overline{x} - a) \cdot \sum_{i=1}^{n} (x_i - \overline{x}) + \sum_{i=1}^{n} (\overline{x} - a)^2 \\ &= \sum_{i=1}^{n} (x_i - \overline{x})^2 + \sum_{i=1}^{n} (\overline{x} - a)^2 \\ &= \sum_{i=1}^{n} (x_i - \overline{x})^2 + n \cdot (\overline{x} - a)^2 \end{aligned}$$

(da $\sum\limits_{i=1}^{n}(x_i - \overline{x}) = \sum\limits_{i=1}^{n} x_i - n \cdot \overline{x} = 0$
Damit gilt

$$\sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 + n \cdot (\overline{x} - a)^2 \geq \sum_{i=1}^{n}(x_i - \overline{x})^2$$

also Aussage 1).

Setzen wir speziell $a = 0$, so folgt

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

also

$$\begin{aligned} \widetilde{s}^2 &= \tfrac{1}{n} \cdot \sum_{i=1}^{n}(x_i - \overline{x})^2 \\ &= \tfrac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2\right) \\ &= \tfrac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - \overline{x}^2 \end{aligned}$$

und das ist die zweite Folgerung.

Aufgabe 2. Die jährliche Haushaltseinkommensverteilung von 10 zufällig ausgewählten Haushalten in zwei Orten (in €1000) ist gegeben durch folgende Tabelle:

Haushalt	1	2	3	4	5	6	7	8	9	10
Ort 1	84	37	60	39	116	88	40	37	93	38
Ort 2	28	42	27	46	32	41	33	26	49	36

a) Bestimmen Sie die Lorenzkurven zu den Einkommensverteilungen in den beiden Orten.

Lösung:

Zunächst sind die Einkommen der Größe nach zu sortieren. Nachdem wir in jedem Ort die ausgewählten Haushalte entsprechend umnummeriert haben, erhalten wir als Tabelle

Haushalt	1	2	3	4	5	6	7	8	9	10
Ort 1	37	37	38	39	40	60	84	88	93	116
Ort 2	26	27	28	32	33	36	41	42	46	49

Damit können nun die Lorenzkurven bestimmt werden. Dazu bestimmen wir zunächst den v–Vektor

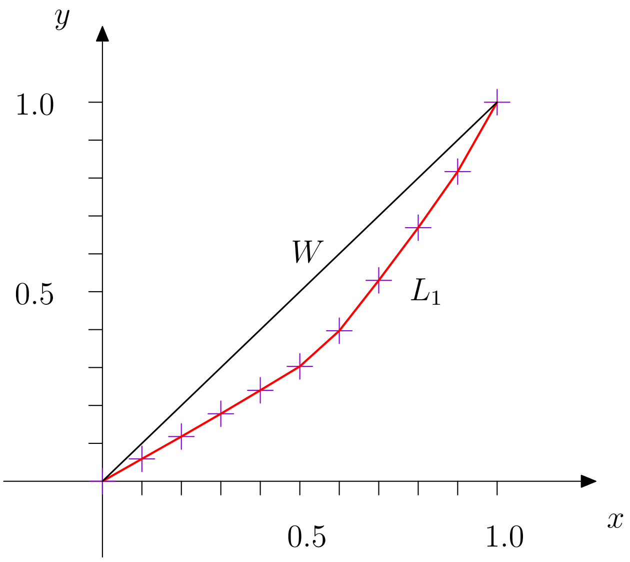
Für Ort 1 erhalten wir

$$v_{\text{Ort 1}} = (0,\ 0.0585,\ 0.1171,\ 01772,\ 0.2389,\ 0.3022,\ 0.3972,\ 0.5301,\ 0.6693,\ 0.8165,\ 1.00)$$

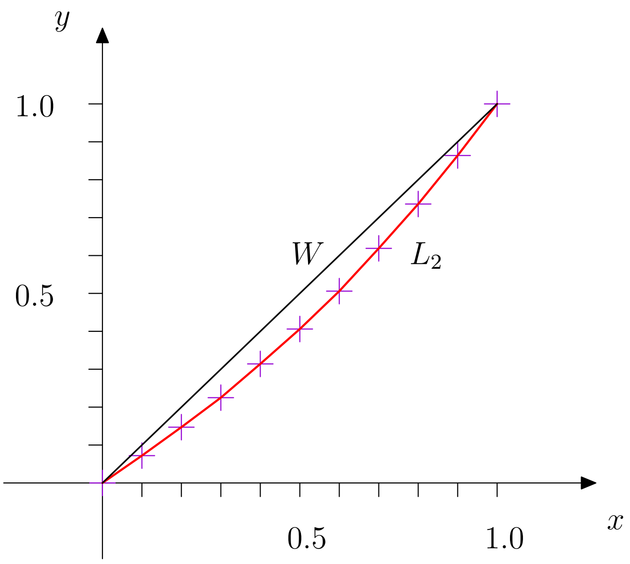
und für Ort 2 ergibt sich

$$v_{\text{Ort 2}} = (0,\ 0.0722,\ 0.1472,\ 0.2250,\ 0.3139,\ 0.4056,\ 0.5056,\ 0.6194,\ 0.7361,\ 0.8639,\ 1.00)$$

Damit erhalten wir für Ort 1 als Lorenzkurve



und als Lorenzkurve für Ort 2



b) Ermitteln Sie, in welchem Ort das Einkommen gleichmäßiger verteilt ist.

Lösung:

Aus den Daten in Teil a) lassen sich die Gini–Koeffizienten leicht ermitteln.

Wir erhalten für Ort 1

$$G_1 = 0.2386, \qquad G_1^* = 0.2651$$

und für Ort 2

$$G_2 = 0.1222, \qquad G_2^* = 0.1358$$

Die Einkommensverteilung ist also in Ort 2 gleichmäßiger als in Ort 1, die Unterschiede sind dabei relativ deutlich.

Aufgabe 3. Neben dem Gini–Koeffizienten gibt es noch weitere Konzentrationsmaße. Eines davon ist der **Herfindahl–Index**, der wie folgt definiert ist: Werden von dem Merkmal X die Merkmalsausprägungen $x_1, \ldots, x_n$ mit $x_i \geq 0$ für alle i beobachtet, so setze

$$M = \sum_{i=1}^{n} x_i, \quad q_i = \frac{x_i}{M}$$

(wobei wir annehmen, dass mindestens ein x_i nicht verschwindet). Dann heißt

$$H := \sum_{i=1}^{n} q_i^2$$

der Herfindahl–Index (dieser Stichprobe). Zeigen Sie:
Es gilt $\frac{1}{n} \leq H \leq 1$, wobei

$$\begin{aligned} H = \tfrac{1}{n} \quad &\Longleftrightarrow \quad x_1 = x_2 = \cdots = x_n \\ H = 1 \quad &\Longleftrightarrow \quad x_i = M \text{ für ein } i \in \{1, \ldots, n\} \end{aligned}$$

Je näher H bei 1 liegt, desto stärker ist also die Konzentration auf dem Markt.

Lösung:
Es ist $0 \leq q_i \leq 1$ für alle i mit $\sum_{i=1}^{n} q_i = 1$. Daher gilt $q_i^2 \leq q_i$ für alle i und damit

$$0 \leq \sum_{i=1}^{n} q_i^2 \leq \sum_{i=1}^{n} q_i = 1$$

(also $0 \leq H \leq 1$), mit Gleichheit genau dann, wenn $q_i^2 = q_i$ für alle i, also genau dann, wenn $q_i = 0$ oder $q_i = 1$ für alle i. Wegen $\sum_{i=1}^{n} q_i = 1$ bedeutet das aber, dass für genau ein i gelten muss $q_i = 1$ (und für alle $j \neq i$ dann $q_j = 0$), und das wiederum bedeutet, dass $x_i = M$ für ein i (und damit $x_j = 0$ für $j \neq i$).

Für die Abschätzung nach unten setzen wir $a_i = q_i - \frac{1}{n}$, sodass also

$$q_i = \frac{1}{n} + a_i$$

Dann gilt

$$1 = \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} \left(\frac{1}{n} + a_i \right) = 1 + \sum_{i=1}^{n} a_i$$

so dass also notwendig $\sum\limits_{i=1}^{n} a_i = 0$. Damit erhalten wir

$$\begin{aligned}
\sum_{i=1}^{n} q_i^2 &= \sum_{i=1}^{n} \left(\tfrac{1}{n} + a_i\right)^2 \\
&= \sum_{i=1}^{n} \left(\tfrac{1}{n^2} + \tfrac{2}{n} \cdot a_i + a_i^2\right) \\
&= \sum_{i=1}^{n} \tfrac{1}{n^2} + \tfrac{2}{n} \cdot \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} a_i^2 \\
&= \sum_{i=1}^{n} \tfrac{1}{n^2} + \sum_{i=1}^{n} a_i^2 \\
&= \tfrac{1}{n} + \sum_{i=1}^{n} a_i^2
\end{aligned}$$

und damit gilt

$$H \geq \frac{1}{n} \text{ mit Gleichheit genau dann, wenn } a_i = 0 \iff q_i = \frac{1}{n} (\forall i)$$

Da $q_i = \frac{1}{n}$ genau dann, wenn $x_i = \frac{M}{n}$, bedeutet das, dass

$$H = \frac{1}{n} \iff x_1 = x_2 = \cdots = x_n$$

Aufgabe 4. Wir betrachten zwei Märkte, auf denen jeweils 10 Unternehmen tätig sind. Auf dem ersten Markt haben 5 Unternehmen jeweils einen Marktanteil von 0.5 % und 5 Unternehmen einen Marktanteil von 19.5 %, auf dem zweiten Markt haben 9 Unternehmen einen Marktanteil von jeweils 5.5 % und ein Unternehmen einen Marktanteil von 50.5 %.
Berechnen Sie für beide Märkte den Gini–Koeffizienten und den Herfindahl–Index (vergleiche Aufgabe 3) und interpretieren Sie die Ergebnisse.

Lösung:
Zunächst ist zu beachten, dass der Gini–Koeffizient für Absolutwerte definiert ist. Allerdings folgt aus der Berechnungsformel sofort, dass er auch über die Prozentangaben ermittelt werden kann.
Damit erhalten wir für Markt 1:

$$G_1 = 0.4756, \qquad G_1^* = 0.5278$$

und für Markt 2

$$G_2 = 0.4050, \qquad G_2^* = 0.4500$$

Aus Sicht des Ginikoeffizienten ist die Verteilung in Markt 1 also ungleichmäßiger als in Markt 2.
Für den Herfindahlkoeffizienten gilt im Markt 1:

$$H_1 = 0.1903$$

und im Markt 2:

$$H_2 = 0.2823$$

d.h. für den Herfindahlkoeffizienten ist die Verteilung in Markt 2 also ungleichmäßiger als in Markt 1.
Der Grund hierfür liegt darin, dass der Herfindahlkoeffizient (aufgrund der quadrierten Prozentanteile) sehr stark auf Monopolbildungen untersucht. Hat z.B. ein Markteilnehmen 100 %, so ist $H = 1$, haben zwei Teilnehmer je 50 %, so sinkt H bereits auf $\frac{1}{4}$, unabhängig davon, wie viele weitere Teilnehmer (mit 0 %) noch auf dem Markt sind. Der Gini-Koeffizient dagegen berücksichtigt alle Marktteilnehmer und ihre Abweichung von der Gleichverteilung gleichmäßig und berücksichtigt, wie viel den $5, 10, 20, \ldots$ kleinsten Teilnehmern insgesamt zu einem einer Gleichverteilung angemessenen Anteil fehlt.