Kapitel V

Konzentrationsanalyse

B. 5. 1.
Im Allgemeinen wird aus statistischer Sicht zwischen

- absoluter und
- relativer

Konzentration unterschieden

Der absolute und relative Aspekt wird noch einmal unterteilt in

- statischer und
- dynamischer

Analyse.

D. 5. 2. (***Absolute und relative Konzentration***)
Von einer *absoluten Konzentration* wird gesprochen, wenn ein Großteil des gesamten Merkmalbetrages auf eine kleine Zahl an Merkmalsträgern entfällt.
Von einer *relativen Konzentration* wird gesprochen, wenn ein Großteil des gesamten Merkmalbetrages auf einen kleinen Anteil der Merkmalsträger entfällt.

BS. 5. 1.
Die 1000 reichsten Bundesbürger erzielen 10% der Einkommen (absolute Konzentration).
Die 1 % reichsten Bundesbürger erzielen 30% der Einkommen (relative Konzentration).

B. 5. 2. (***Absolute Konzentrationsmaße***)
Hierzu gehören:

- Konzentrationsrate
- Herfindahlindex

D. 5. 1. (***Konzentrationsrate***)
Gegeben sei ein metrisch messbares Merkmal X mit den Beobachtungswerten

$$x_1 \geq x_2 \geq \ldots \geq x_n \geq 0 .$$

Als $Konzentrationsrate\, k-ter\ Ordnung$ bezeichnet man:

$$C_k := \frac{\sum_{j=1}^{k} x_j}{\sum_{i=1}^{n} x_i}, \quad k = 1, 2, \ldots, n .$$

(C_k gibt an, wie groß der Anteil der Merkmalssumme ist, der auf diejenigen Merkmalsträger mit den k größten Ausprägungen entfällt.)

BS. 5. 2.
Gegeben seien vier verschiedene Märkte A, B, C und D, in denen jeweils fünf Unternehmen tätig sind. Die Größenverhältnisse der Unternehmen gemessen am Umsatz [Million €] sind in folgender Tabelle dargestellt:

	Verteilung			
Unternehmen	A	B	C	D
1	500	220	250	100
2	0	140	100	100
3	0	100	80	100
4	0	30	50	100
5	0	10	20	100
Insgesamt	500	500	500	500

Zu berechnen sei die Konzentrationsrate 2. Ordnung.

Lösung:

A: $$C_2 = \frac{500+0}{500} = 1.00$$

B: $$C_2 = \frac{220+140}{500} = 0.72$$

C: $$C_2 = \frac{250+100}{500} = 0.70$$

D: $$C_2 = \frac{100+100}{500} = 0.40$$

Damit ist die Konzentrationsrate bei Verteilung A mit 100% am höchsten und bei Verteilung D mit 40% am niedrigsten. Die anderen Verteilungen liegen dazwischen.

D. 5. 2. (***Konzentrationskurve***)
Zeichnet man für $i = 0,1,2,...,n$ die Punkte (i, C_i) in ein Koordinatensystem und verbindet man die Punkte durch einen linearen Streckenzug, so erhält man die *Konzentrationskurve*.

B. 5. 3. (***Eigenschaften der Konzentrationskurve***)

1. Je rascher sich die Konzentrationskurve dem Ordinatenwert 1 nähert, umso größer ist die absolute Konzentration.

2. Es liegt maximale Konzentration vor, falls

$$x_1 = \sum_{i=1}^{n} x_i, \quad x_2 = x_3 = ... = x_n = 0 \text{ und damit bereits } C_1 = 1.$$

3. Es liegt minimale Konzentration vor, falls

$$x_1 = x_2 = \ldots = x_n = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Die Konzentrationskurve ist dann eine Gerade

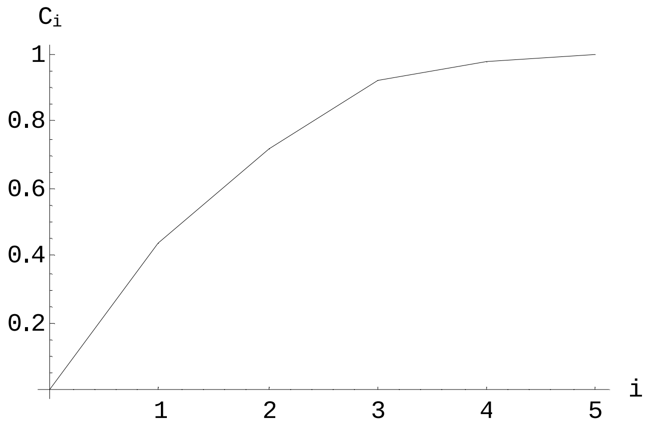
4. Die Konzentrationskurve ist konkav.

<u>BS. 5. 2.</u> (*Fortsetzung*)
Zeichnen Sie die Konzentrationskurve für die Verteilung B.

Lösung:

Arbeitstabelle

i	$\frac{x_i}{\sum_{i=1}^{n} x_i}$	C_i
0	0.00	0.00
1	0.44	0.44
2	0.28	0.72
3	0.20	0.92
4	0.06	0.98
5	0.02	1.00



<u>B. 5. 4.</u> (*Einschätzung der Konzentrationsrate*)

a) Vorteile:

1. Einfache Berechenbarkeit
2. Leichte Datenverfügbarkeit

b) Nachteile:

1. k wird willkürlich gewählt.
2. Informationen der kleineren Merkmalsträger werden nicht berücksichtigt.

(Im obigen Beispiel ist bei $k = 2$ die Konzentrationsrate 72% und bei Verteilung C 70%.Würde man aber $k = 1$ wählen, dann wäre Verteilung B mit 44% weniger stark konzentriert als Verteilung C mit 50%.)

<u>D. 5. 3.</u> (*Herfindahl-Koeffizient*)

$$K_H := \sum_{i=1}^{n} p_i^2 \quad \text{mit} \qquad p_i := \frac{x_i}{\sum_{i=1}^{n} x_i},$$

<u>B. 5. 5.</u> (*Eigenschaften des Herfindahl-Koeffizienten*)

1. $\frac{1}{n} \leq K_H \leq 1$
2. Es gilt $K_H = \frac{1}{n}$ genau dann, wenn minimale Konzentration vorliegt.
3. Es gilt $K_H = 1$ genau dann, wenn maximale Konzentration vorliegt.
4. Es gilt
$$K_H = \frac{1}{n}\left(v^2 + 1\right).$$

<u>BS. 5. 2.</u> (*Fortsetzung*)

Arbeitstabelle

i	p_i	p_i^2
1	0.44	0.1936
2	0.28	0.0784
3	0.20	0.0400
4	0.06	0.0036
5	0.02	0.0004
		0.3160

$K_H := 0.3160$.

<u>B. 5. 6.</u> (*Einschätzung des Herfindahl-Koeffizienten*)
a) Vorteile:

Der Herfindahl-Koeffizient

1. verwendet sämtliche Informationen.
2. berücksichtigt die Anzahl der Merkmalsträger.
3. ist unabhängig gegenüber proportionaler Veränderung aller Werte,

b) Nachteile:

1. Die Quadrierung ist willkürlich
2. Informationen über kleinere Merkmalsträger sind notwendig.

B. 5. 7. (*Relative Konzentrationsmaße*)
Hierzu gehören:

- die Lorenzkurve
- der Gini-Koeffizient.

D. 5. 4. (*Lorenzkurve*)
Gegeben sei ein metrisch messbares Merkmal X mit den Beobachtungswerten

$$0 \le x_1 \le x_2 \le \ldots \le x_n .$$

Als *Lorenzkurve* bezeichnet man ein Streckenzug, der durch die Punkte (u_i, v_i), $i = 0,1,\ldots,n$ läuft. Dabei sind für:

1. Ein nichtgruppiertes Datenmaterial

$$u_i := \begin{cases} 0 & \text{für} \quad i = 0 \\ \dfrac{i}{n} & \text{für} \quad i = 1,2,\ldots,n \end{cases}$$

$$v_i := \begin{cases} 0 & \text{für} \quad i = 0 \\ \dfrac{\sum_{j=1}^{i} x_j}{\sum_{i=1}^{n} x_i} & \text{für} \quad i = 1,2,\ldots,n \end{cases} .$$

2. Ein gruppiertes Datenmaterial

$$u_i := \begin{cases} 0 & \text{für} \quad i = 0 \\ \sum_{j=1}^{i} h_j & \text{für} \quad i = 1,2,\ldots,n \end{cases}$$

$$v_i := \begin{cases} 0 & \text{für} \quad i = 0 \\ \dfrac{\sum_{j=1}^{i} \bar{x}_j \cdot H_j}{\sum_{i=1}^{p} \bar{x}_i \cdot H_i} & \text{für} \quad i = 1,2,\ldots,n \end{cases} .$$

B. 5. 8.

1. Häufig werden u_i und v_i mit 100 multipliziert um jeweils den prozentualen Anteil zu erhalten.

2. Fehlen die Informationen $\bar{x}_i$, so werden sie durch m_i ersetzt.

B. 5. 9. (*Eigenschaften der Lorenzkurve*)

Die Lorenzkurve

1. gibt Abweichung gegenüber der Gleichverteilung (Diagonale) an (*Disparität*).
2. liegt zwischen den Extremen der *minimalen Disparität* (*absolute Gleichheit*) und der *maximalen Disparität* (*absolute Ungleichheit*).
3. verläuft nirgendwo oberhalb der Diagonalen.
4. ist stückweise linear, streng monoton wachsend und konvex.

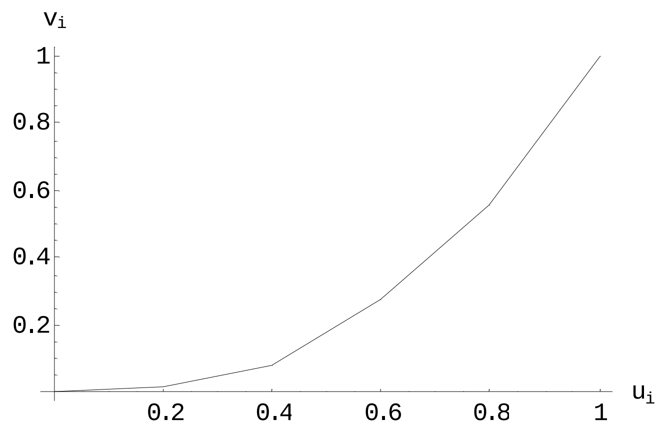
Wenn sich zwei Lorenzkurven nicht schneiden, weist die höhere Lorenzkurve eindeutig weniger Disparität auf als die niedrigere.
Bei praktischen Problemen schneiden sich in vielen Fällen. In solchen Fällen ist kein eindeutiger Disparitätsvergleich möglich. Ein Ausweg bietet sich dann der Disparitätsvergleich durch einen *Index,* z.B. durch den Gini-Index (siehe weiter unten!).

BS. 5. 2. (*Fortsetzung*)
Zeichnen Sie die Lorenzkurve für die Verteilung B.

Lösung:

Arbeitstabelle

i	x_i	u_i	v_i
0	-	0.00	0.00
1	10	0.20	0.02
2	30	0.40	0.08
3	100	0,60	0.28
4	140	0.80	0.56
5	220	1.00	1.00



BS. 5. 3.
Für ein Verbrauchsstudio wurden die Nettojahreseinkommen von 100 Männern wie folgt festgestellt:

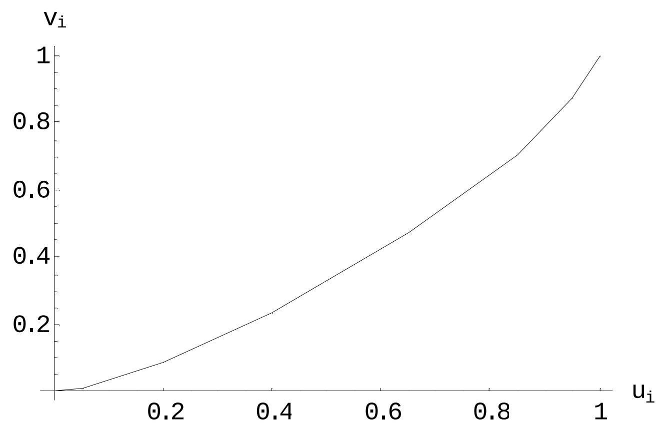
Einkommen [Tsd. €] von… bis unter…	Anzahl
… 10	5
10…20	15
20…25	20
25…30	25
30…40	20
40…60	10
60…85	5

Stellen Sie die entsprechende Lorenzkurve auf.

Lösung:

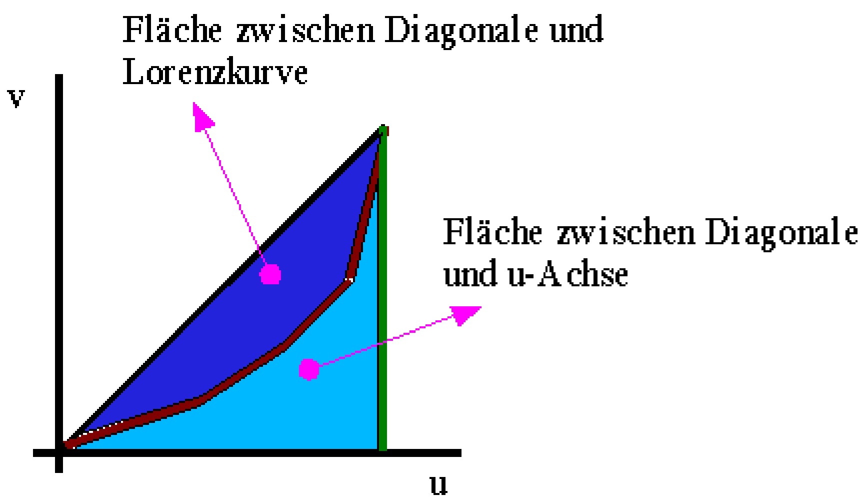
Arbeitstabelle

i	g_i	G_i	H_i	h_i	$\sum_{j=1}^{i} h_j$	m_i	$m_i \cdot H_i$	$\sum_{j=1}^{i} m_j \cdot H_j$	$\sum_{j=1}^{i} m_j \cdot H_j \Big/ \sum_{i=1}^{p} m_i \cdot H_i$
1	0	10	5	0.05	0.05	5.0	25.0	25.0	0.008
2	10	20	15	0.15	0.20	15.0	225.0	250.0	0.085
3	20	25	20	0.20	0.40	22.5	450.0	700.0	0.237
4	25	30	25	0.25	0.65	27.5	687.5	1387.5	0.470
5	30	40	20	0.20	0.85	35.0	700.0	2087.5	0.708
6	40	60	10	0.10	0.95	50.0	500.0	2587.5	0.877
7	60	85	5	0.05	1.00	72.5	362.5	2950.0	1.000
			100	1.00			2950.0		



D. 5. 5. (***Gini-Koeffizient*** **bzw.** ***Gini-Index.***)

$$G := \frac{\text{Inhalt der Fläche zwischen der Diagonalen und Lorenzkurve}}{\text{Inhalt der Fläche unterhalb der Diagonalen}}$$



$$G = \frac{\text{Fläche zwischen Diagonale und Lorenzkurve}}{\text{Fläche zwischen Diagonale und u-Achse}}$$

<u>B. 5. 10.</u> (*Gini-Koeffizient: (Berechnungsformeln)*)
Folgende Formeln lassen sich zur Berechnung des Gini-Koeffizienten herleiten:

1. <u>Nichtgruppiertes Datenmaterial</u>

$$\begin{aligned} G &= \frac{2}{n} \cdot \frac{\sum_{i=1}^{n} i \cdot x_i}{\sum_{i=1}^{n} x_i} - \frac{n+1}{n} \\ &= 1 - \frac{2}{n}\left(\sum_{i=1}^{n} v_i - \frac{1}{2}\right), \qquad 0 \le G \le 1 - \frac{1}{n}. \end{aligned}$$

2. <u>Gruppiertes Datenmaterial</u>

$$\begin{aligned} G &= 1 - \frac{1}{n}\sum_{i=1}^{p} H_i\left(v_{i-1} + v_i\right) \\ &= 1 - \sum_{i=1}^{p} h_i\left(v_{i-1} + v_i\right), \qquad 0 \le G \le 1 - \frac{H_p}{n}. \end{aligned}$$

<u>B. 5. 11.</u> (*Eigenschaften des Gini-Koeffizienten*)

1. Es gilt $G = 0$ genau dann, wenn *minimale Disparität* (*absolute Gleichheit*) vorliegt.
2. Es gilt $G = 1 - \frac{1}{n}$ genau dann, wenn *maximale Disparität* (*absolute Ungleichheit*) vorliegt.

<u>D. 5. 6.</u> (*Normierter (bzw. Standardisierter) Gini-Koeffizient*)

$$G_s := G \cdot \frac{n}{n-1}, \qquad 0 \le G_s \le 1.$$

(Die Berechnung des normierten Gini-Koeffizienten ist nur bei kleinen Gesamtheiten sinnvoll.)

<u>BS. 5. 2.</u> (*Fortsetzung***)**
Berechnen Sie die Gini-Koeffizienten für die Verteilungen: A - D.

Lösung:

A: $$G = \frac{2}{5} \cdot \frac{1 \cdot 0 + 2 \cdot 0 + 3 \cdot 0 + 4 \cdot 0 + 5 \cdot 500}{500} - \frac{5+1}{5} = 0.800,$$

B: $$G = \frac{2}{5} \cdot \frac{1 \cdot 10 + 2 \cdot 30 + 3 \cdot 100 + 4 \cdot 140 + 5 \cdot 220}{500} - \frac{5+1}{5} = 0.424,$$

C: $$G = \frac{2}{5} \cdot \frac{1 \cdot 20 + 2 \cdot 50 + 3 \cdot 80 + 4 \cdot 100 + 5 \cdot 250}{500} - \frac{5+1}{5} = 0.408,$$

D: $$G = \frac{2}{5} \cdot \frac{1 \cdot 100 + 2 \cdot 100 + 3 \cdot 100 + 4 \cdot 100 + 5 \cdot 100}{500} - \frac{5+1}{5} = 0.000.$$

<u>BS. 5. 3.</u> (*Fortsetzung***)**
Berechnen Sie den Gini-Koeffizienten und den normierten Gini-Koeffizienten.

Lösung:

$$G = 1 - \left(0.05 \cdot 0.008 + 0.15 \cdot 0.093 + 0.20 \cdot 0.322 + 0.25 \cdot 0.707 + 0.20 \cdot 1.178 + 0.10 \cdot 1.585 + 0.05 \cdot 1.877\right)$$

$$= 0.25655.$$

$$G_s = 0.2655 \cdot \frac{100}{99} = 0.259141414.$$

<u>B. 5. 12.</u> (*Einschätzung des Gini-Koeffizienten*)

a) Vorteil

Der Gini-Koeffizient bietet eine Vergleichsmöglichkeit im Falle, dass sich die Lorenzkurven schneiden

b) Nachteil

1. Unterschiedliche Lorenzkurven können zu demselben Gini-Koeffizienten führen.
2. Es besteht eine starke Abhängigkeit des Gini-Koeffizienten von der Zahl der einbezogenen statistischen Einheiten. Weglassen von kleinen Merkmalswerten verringert den Gini-Koeffizienten.

B. 5. 13. (*Statische und dynamische Konzentration*)
Meistens bezieht sich der Konzentrationsbegriff auf eine gegebene Verteilung, d. h. es wird die Konzentration eines Bestandes untersucht (*statische Konzentration*).
Unter *dynamischer Konzentration* dagegen versteht man die zunehmende Konzentration des Bestandes im Zeitablauf, d. h. also einen Konzentrationsprozess.
(Die Problematik wird hier nicht behandelt.)

(Letzte Aktualisierung: 15.09.2011)