

# Konzentration und Disparität

## Begleitende Unterlagen zur Übung *Deskriptive Statistik*

Michael Westermann

Universität Essen

### Inhaltsverzeichnis

6	Konzentrations- und Disparitätsmessung . . . . .	2
6.1	Begriff und Einleitung . . . . .	2
6.2	Messung der absoluten Konzentration . . . . .	2
6.3	Messung der relativen Konzentration (Disparität) . . . . .	3
6.4	Aufgaben zu Kapitel 6 . . . . .	5

**Anmerkung:** Bedingt durch den *Streik* der Studierenden konnte meine Übung am 03. Juni 2002 nicht stattfinden bzw. wurde sie nach einer halben Stunde abgebrochen. Hiermit stelle ich den Inhalt, welcher in dieser Veranstaltung behandelt worden wäre, zur selbständigen Nachbereitung zur Verfügung.

## 6 Konzentrations- und Disparitätsmessung

### 6.1 Begriff und Einleitung

Mit Hilfe der Konzentrations- bzw. Disparitätsmessung lassen sich Ballungerscheinungen aufspüren und darstellen. Es geht dabei vornehmlich um die Darstellung, wie ein Merkmal bzw. die Merkmalssumme auf die einzelnen Merkmalsträger verteilt ist und nicht lediglich um die Verteilung des Merkmals selbst. Dabei sind zwei Extremsituationen zu unterscheiden:

1. Von Gleichverteilung spricht man, wenn jeder Merkmalsträger den gleichen Merkmalsbetrag auf sich vereint. ( $\leadsto$  **egalitäre Verteilung**)
2. Von vollkommener Ungleichheit spricht man, wenn ein Merkmalsträger die gesamte Merkmalssumme auf sich vereint und die anderen Merkmalsträger weisen einen Merkmalsbetrag von null auf.

Wir unterscheiden zwei Arten von Konzentration:

*absolute Konzentration:* bezieht die Anteile an der Merkmalssumme auf die **Anzahl** der Merkmalsträger. Starke Konzentration liegt dann vor, wenn eine kleine Anzahl Merkmalsträger einen großen Anteil der Merkmalssumme auf sich vereint.

*relative Konzentration:* ( $\leadsto$  **Disparität**) bezieht die Anteile an der Merkmalssumme hingegen auf den **Anteil** der Merkmalsträger. Eine hohe Disparität liegt dann vor, wenn ein kleiner Anteil der Merkmalsträger einen hohen Anteil an der Merkmalssumme auf sich vereint.

### 6.2 Messung der absoluten Konzentration

Ordnung der  $n$  Beobachtungen des Merkmals  $X$  nach abnehmender Größe

$$x_n \geq x_{n-1} \geq \dots \geq x_2 \geq x_1.$$

Die Merkmalssumme beträgt  $\sum_{i=1}^n x_i = \sum_{j=1}^m x_j n_j = n\bar{x}$ . Der auf den  $i$ -ten Merkmalsträger entfallende Anteil  $q_i$  an der Merkmalssumme beträgt

$$q_i = \frac{x_i}{\sum x_i}, \quad \text{mit } \sum_{i=1}^n q_i = 1.$$

*Herfindahl-Index* Der Herfindahl-Index  $K_H$  ist definiert als die Summe der quadrierten Anteilswerte  $q_i$ :

$$K_H = \sum_{i=1}^n q_i^2, \quad \text{mit } \frac{1}{n} \leq K_H \leq 1.$$

Er weist eine praktische Eigenschaft auf, da er nämlich über die Varianz und das arithmetische Mittel berechnet werden kann. Da diese beiden Maßzahlen für die meisten Datensätze immer berechnet werden, liefern diese beiden Parameter gleichzeitig auch Informationen über die Konzentration der Daten. Es gilt

$$K_H = \sum_{i=1}^n q_i^2 = \sum_{i=1}^n \left( \frac{x_i}{n\bar{x}} \right)^2 = \dots = \frac{\frac{s^2}{\bar{x}^2} + 1}{n} = \frac{V^2 + 1}{n}.$$

*Rosenbluth-Index* Der Rosenbluth-Index ist definiert als

$$K_R = \left( 2 \sum_{i=1}^n iq_i - 1 \right)^{-1} = \frac{1}{2 \sum_{i=1}^n iq_i - 1}, \quad \text{mit } \frac{1}{n} \leq K_R \leq 1.$$

*Beispiel 1* Für fünf Tageszeitungen werden folgende Marktanteile festgestellt:

Zeitung	A	B	C	D	E
Marktanteil	0,15	0,10	0,25	0,20	0,30

Zur Berechnung der beiden Indizes erstellt man folgende Arbeitstabelle, in der die Marktanteile nach absteigender Größe geordnet worden sind:

$i$	$q_i$	$q_i^2$	$iq_i$
1	0,30	0,0900	0,3
2	0,25	0,0625	0,5
3	0,20	0,0400	0,6
4	0,15	0,0225	0,6
5	0,10	0,0100	0,5
$\Sigma$	1,00	0,2250	2,5

Der Herfindahl-Index beträgt  $K_H = 0,225$  und für den Rosenbluth-Index erhält man  $K_R = \frac{1}{(2 \cdot 2,5) - 1} = 0,25$ . □

### 6.3 Messung der relativen Konzentration (Disparität)

Relative Konzentration setzt Anteile der Merkmalssumme zu Anteilen der Merkmalsträger in Beziehung zueinander. Der Datensatz wird in aufsteigender Größe der Merkmals  $X$  geordnet. Die geordneten Beobachtungen  $x_1 \leq x_2 \leq \dots \leq x_n$  werden in Anteile an der Merkmalssumme

$$q_i = \frac{x_i}{\sum x_i}$$

überführt.

$$Q_i = \sum_{j=1}^i q_j$$

stellt somit den kumulierten Anteil der  $i$  Merkmalsträger mit den kleinsten Merkmalssummenanteilen dar. Der Anteil der Merkmalsträger, welche eine Merkmalsausprägung  $x_i$  repräsentieren, beträgt

$$h_i = \frac{n_i}{n}$$

und der kumulierte Anteil der Merkmalsträger beträgt demnach

$$H_i = \sum_{j=1}^i h_j \quad \text{bzw. bei Einzelbeobachtungen} \quad H_i = \frac{i}{n}.$$

*Die Lorenzkurve* Die graphische Darstellung dieser Verhältnisse erfolgt durch die **Lorenzkurve**. Zu dieser gelangt man, wenn man die Zahlenpaare  $(H_i, Q_i)$  in ein kartesisches Koordinatensystem mit  $H_i$  an der Abszisse und  $Q_i$  an der Ordinate überträgt und die so entstandenen Punkte linear verbindet. Dabei wird im Ursprung begonnen und im Punkt  $(1, 1)$  geendet. Bei Vorliegen einer Häufigkeitsverteilung erhält man die  $q_j$ -Werte bzw.  $h_j$ -Werte durch

$$q_j = \frac{x_j n_j}{\sum_{j=1}^m x_j n_j} \quad \text{bzw.} \quad h_j = \frac{n_j}{n}.$$

Wenn die Daten als klassierte Daten vorliegen, müssen zur Ermittlung von  $q_k$  bzw.  $Q_k$  die Klassenmitten herangezogen werden, also

$$q_k = \frac{m_k n_k}{\sum m_k n_k}.$$

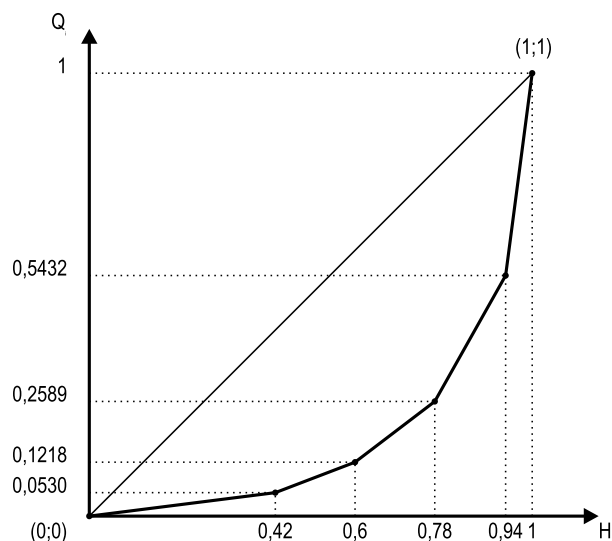
Es existieren folgende Extremfälle der Lorenzkurve:

1. Bei egalitärer Verteilung ist die Lorenzkurve eine Gerade mit der Steigung  $45^\circ$ . Die Lorenzkurve ist dann identisch mit der Gleichverteilungsgeraden.
2. Bei vollkommener Disparität verläuft die Lorenzkurve bis zur Stelle  $\frac{n-1}{n}$  auf der Abszisse und von da an zum Punkt  $P(1, 1)$ .

*Beispiel 2* Man hat in einer Region an 50 landwirtschaftlichen Betrieben die in der Tabelle aufgeführte Größe der landwirtschaftlichen Nutzfläche beobachtet. Um die Lorenzkurvenwerte – und später den Gini-Koeffizienten – zu bestimmen, fertigt man die folgende Arbeitstabelle an:

Fläche (in ha) $[x'_{k-1}; x'_k)$	$n_k$	$h_k$	$H_k$	$\bar{x}_k$	$\bar{x}_k n_k$	$q_k$	$Q_k$	$(H_{k-1} + H_k)q_k$
0 – 5	21	0,42	0,42	2,5	52,5	0,0533	0,0533	0,0224
5 – 10	9	0,18	0,60	7,5	67,5	0,0685	0,1218	0,0699
10 – 20	9	0,18	0,78	15	135,0	0,1371	0,2589	0,1891
20 – 50	8	0,16	0,94	35	280,0	0,2843	0,5432	0,4889
50 – 250	3	0,06	1,00	150	450,0	0,4569	1,00	0,8863
$\sum$	$n = 50$	1			985	1		1,6566

Die Lorenzkurve hat dann folgendes Gestalt:



Zur Interpretation:

Auf die 42% kleinsten Betriebe entfallen 5,3% der Gesamtfläche.

Auf die 60% kleinsten Betriebe entfallen 12,18% der Gesamtfläche.

Auf die 78% kleinsten Betriebe entfallen 25,89% der Gesamtfläche ...

Auf die 22% größten Betriebe entfallen 74,11% der Gesamtfläche.

Der Gini-Koeffizient beträgt

$$D_G = \left[ \sum (H_i + H_{i-1})q_i \right] - 1 = 1,6566 - 1 = 0,6566. \quad \square$$

*Der Gini-Koeffizient* Der Gini-Koeffizient dient zur numerischen Bestimmung der Disparität. Er ist ein Flächenmaß und setzt die Fläche zwischen der Lorenzkurve und der Gleichverteilungsgeraden ( $F$ ) ins Verhältnis zur Dreiecksfläche unter der Gleichverteilungsgeraden. Diese Fläche beträgt  $1/2$ , daher bestimmt sich der Gini-Koeffizient als

$$D_G = \frac{F}{1/2} = 2F,$$

d.h. er ist das zweifache der Fläche  $F$  zwischen Lorenzkurve und Gleichverteilungsgeraden. Konkret wird der Gini-Koeffiziente folgendermaßen errechnet:

– bei Einzelbeobachtungen

$$D_G = \sum_{i=1}^n \frac{2i - n - 1}{n} q_i$$

– bei klassierten Daten

$$D_G = 1 - \left[ \sum (Q_i + Q_{i-1})h_i \right] = \left[ \sum (H_i + H_{i-1})q_i \right] - 1$$

Bei egalitärer Verteilung ist die Fläche gleich Null und bei vollkommener Disparität ist sie am größten, so dass gilt:  $0 \leq D_G \leq 1 - \frac{1}{n}$ .

Die Steigung der Lorenzkurve ist definiert als

$$s_i = \frac{q_i}{h_i} = \frac{x_i}{\bar{x}}.$$

#### 6.4 Aufgaben zu Kapitel 6

#### Übung 17 (↪ Aufgabe 6.7)

In einem Betrieb mit 20 Arbeitnehmern gibt es nur Zank und Krach, weil es drei Lohngruppen  $L_1$ ,  $L_2$  und  $L_3$  mit sehr unterschiedlichen Löhnen gibt:  $L_1 < L_2 < L_3$ . Der Durchschnittslohn beträgt 900 DM. Von der Lorenzkurve seien zwei Punkte bekannt:

Punkt	$H(x)$	$Q(x)$
A	0,4	0,2
B	0,8	0,6

- Man bestimme die drei Löhne  $L_1$ ,  $L_2$  und  $L_3$ .
- Man bestimme den Variationskoeffizient  $V$  der Löhne.
- Man bestimme das Disparitätsmaß  $D_G$  von Gini für die obigen Daten (Punkte A und B).
- Angenommen vom Punkt B sei nur die Koordinate  $H$  bekannt ( $H = 0,8$ ), aber nicht der  $Q$ -Wert. Bestimmen Sie den Punkt B und  $D_G$  so, daß die Voraussetzungen der Lorenzkurve nicht verletzt werden und die Disparität maximal wird.

**Lösung 17**

Bekannt ist, daß gilt  $L_1 < L_2 < L_3$ . Weiterhin ist die Höhe des Durchschnittslohnes bekannt, nämlich

$$\bar{L} = L_1 h_1 + L_2 h_2 + L_3 h_3 = 900.$$

Für die Steigung der Lorenzkurve gilt:

$$s_i = \frac{q_i}{h_i} = \frac{L_i}{\bar{L}} \quad \text{und somit} \quad L_i = s_i \cdot \bar{L}.$$

**Arbeitstabelle:**

Punkt	Gruppe	$H_i$	$Q_i$	$h_i$	$q_i$	$s_i$	$(H_i + H_{i-1})q_i$
<b>A</b>	$L_1$	<b>0,4</b>	<b>0,2</b>	0,4	0,2	0,5	0,08
<b>B</b>	$L_2$	<b>0,8</b>	<b>0,6</b>	0,4	0,4	1,0	0,48
<b>C</b>	$L_3$	1	1	0,2	0,4	2,0	0,72
$\Sigma$				1	1		1,28

a) Die Lohngruppen lauten:

$$L_1 = s_1 \cdot 900 = 0,5 \cdot 900 = 450$$

$$L_2 = s_2 \cdot 900 = 1,0 \cdot 900 = 900$$

$$L_3 = s_3 \cdot 900 = 2,0 \cdot 900 = 1800$$

b) Der Variationskoeffizient ist definiert als

$$V = \frac{s}{\bar{x}}$$

Der Mittelwert beträgt  $\bar{x} = 900$ . Somit errechnet sich die Varianz als

$$s^2 = (450 - 900)^2 0,4 + (900 - 900)^2 0,4 + (1800 - 900)^2 0,2 = 243.000 \text{DM}^2$$

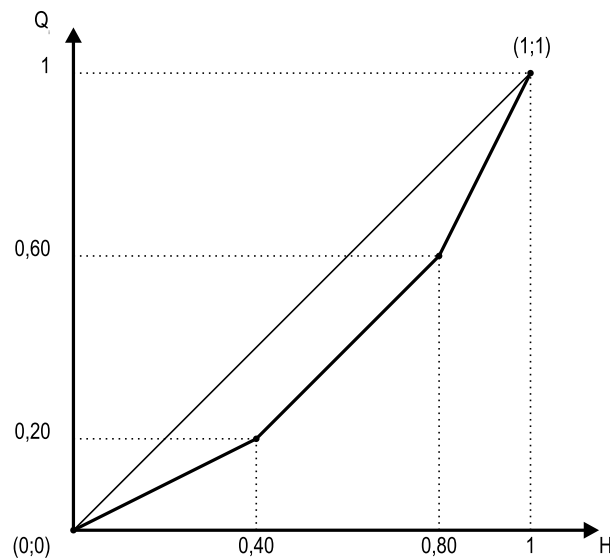
Die Standardabweichung ist dann  $s = 492,95 \text{DM}$  und somit beträgt der Variationskoeffizient

$$V = \frac{s}{\bar{x}} = \frac{492,95}{900} = 0,5477$$

c) Der Gini-Koeffizient beträgt:

$$D_G = \left[ \sum (H_i + H_{i-1})q_i \right] - 1 = 1,28 - 1 = 0,28.$$

**Lorenzkurve:**



d) entfällt!!!

**Übung 18** (↪ **Aufgabe 6.10**)

Gegeben sei die folgende Einkommensverteilung:

Einkommensklasse	Anzahl d. Personen
0 – 1000	100
1000 – 2000	50
2000 – 3000	30
3000 und mehr	20

In jeder Einkommensklasse beträgt das Gesamteinkommen 80.000 DM.

- Berechnen (tabellarische Darstellung) und zeichnen Sie die Lorenzkurve. Berechnen Sie außerdem die Steigung der Lorenzkurve für jede der 4 Klassen.
- Berechnen Sie den Gini-Koeffizienten.
- Ermitteln Sie den Anteil der 25% reichsten Einkommensbezieher am Gesamteinkommen.
- Welche der nachfolgenden Aussagen trifft zu (mehrere richtige Antworten möglich)?

Die Lorenzkurve kann die Gleichverteilungsgerade (GVG)

<input type="checkbox"/>	Schneiden	<input type="checkbox"/>	Nicht schneiden
<input type="checkbox"/>	Tangieren	<input type="checkbox"/>	Nicht tangieren

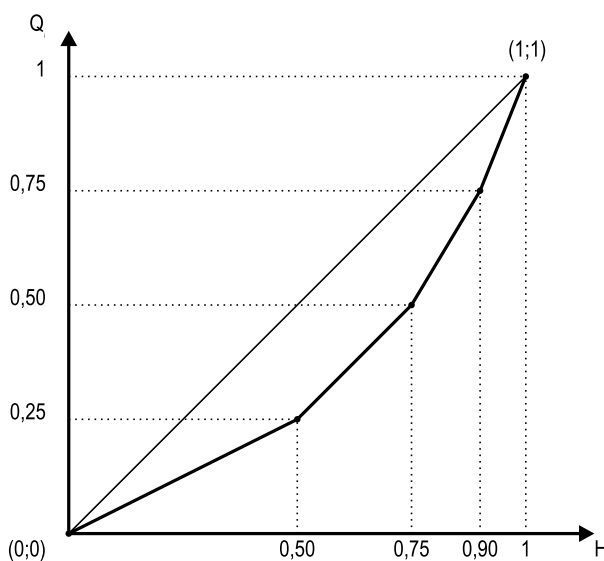
**Lösung 18**

Insgesamt wurden  $n = 200$  Personen untersucht. Das Gesamteinkommen (die Merkmalssumme) beträgt  $4 \cdot 80.000 = 320.000$  DM, da in jeder der vier Einkommensklassen ein Gesamteinkommen von 80.000DM erzielt wird.

**Arbeitstabelle:**

Einkommen	$n_k$	$h_k$	$H_k$	$x_k n_k$	$q_k$	$Q_k$	$s_k$	$(H_k + H_{k-1})q_k$
0 - 1000	100	0,5	0,5	80000	0,25	0,25	0,5	0,125
1000 - 2000	50	0,25	0,75	80000	0,25	0,5	1,0	0,3125
2000 - 3000	30	0,15	0,9	80000	0,25	0,75	1,6667	0,4125
$\geq 3000$	20	0,1	1	80000	0,25	1	2,5	0,475
$\Sigma$	200	1		320000	1			1,325

a) **Lorenzkurve**



Die Steigung der Lorenzkurve ist definiert als

$$s_k = \frac{q_k}{h_k}.$$

b) Der Gini-Koeffizient für die obige Einkommensverteilung beträgt

$$D_G = \left[ \sum (H_k + H_{k-1})q_k \right] - 1 = 1,325 - 1 = 0,325.$$

c) Der Anteil der 25% reichsten Einkommensbezieher am Gesamteinkommen beträgt 50% .

d) Die Lorenzkurve kann die Gleichverteilungsgerade (GG) **nicht** schneiden und **nicht** tangieren.

### Übung 19 (↔ Aufgabe 6.14)

Gegeben seien drei Einkommensklassen  $E_1 < E_2 < E_3$  und ein Durchschnittseinkommen in Höhe von 2000 DM. Es gelte  $h_1 = 0,5$ ,  $h_2 = 0,4$  und  $h_3 = 0,1$  und für die Steigungen  $s$  der Lorenzkurve  $s_1 = 0,4$ ,  $s_2 = 0,75$  und  $s_3 = 5$ . Man bestimme die Einkommen  $E_1, E_2$  und  $E_3$ !

### Lösung 19

Arbeitstabelle:

Einkommen	$h_i$	$H_i$	$q_i$	$Q_i$	$s_i$	$(H_i + H_{i-1})q_i$
$E_1$	0,5	0,5	0,2	0,2	0,4	0,10
$E_2$	0,4	0,9	0,3	0,5	0,75	0,42
$E_3$	0,1	1	0,5	1	5	0,95
$\Sigma$	1		1			1,47

Die Lösung ist ähnlich der in Aufgabe 17. Es ist bekannt, daß gilt  $E_1 < E_2 < E_3$ . Das Durchschnittseinkommens beträgt

$$\bar{E} = E_1h_1 + E_2h_2 + E_3h_3 = 2000.$$

Mit Hilfe der bekannten Steigungen kann man zur Bestimmung der Einkommen sowie der Anteile am Gesamteinkommen folgende Beziehung ausnutzen:

$$s_i = \frac{q_i}{h_i} = \frac{E_i}{\bar{E}} \Rightarrow E_i = s_i \bar{E} \Rightarrow q_i = s_i h_i$$

Daraus ergeben folgende Lohnhöhen und Anteile am Gesamteinkommen:

$$\begin{aligned} E_1 &= s_1 \cdot \bar{E} = 0,40 \cdot 2000 = 800 & q_1 &= s_1 \cdot h_1 = 0,40 \cdot 0,5 = 0,20 \\ E_2 &= s_2 \cdot \bar{E} = 0,75 \cdot 2000 = 1500 & q_2 &= s_2 \cdot h_2 = 0,75 \cdot 0,4 = 0,30 \\ E_3 &= s_3 \cdot \bar{E} = 5 \cdot 2000 = 10000 & q_3 &= s_3 \cdot h_3 = 5 \cdot 0,1 = 0,50 \end{aligned}$$

Man erhält einen Gini-Koeffizienten in Höhe von

$$D_G = \left[ \sum (H_i + H_{i-1})q_i \right] - 1 = 1,47 - 1 = 0,47.$$