Statistik - 2.3 Konzentrationsanalyse

Dr. Wolfgang Böhme

November 2010

2. Univariate Verteilungsanalyse - Gliederung

- 2.1 Häufigkeitsverteilung
- 2.2 Maßzahlen der Verteilung
- **2.3 Konzentrationsanalyse**

Intention / Gliederung

bisher: Disparitätsmaße: Abweichung der Verteilung von einer Gleichverteilung

jetzt: Untersuchung des Zusammenhanges zwischen der Gesamtsumme aller Merkmalswerte und den Einzelbeobachtungen einer statistischen Gesamtheit

Betrachtet werden **extensive Merkmale**, d.h. die Beobachtungen sind nicht negativ ($x_i \geq 0$) und die Bildung der Gesamtsumme der Merkmalswerte ist sachlogisch plausibel.

Gliederung

- 2.3.1 **absolute statistische Konzentration**
- 2.3.2 relative statistische Konzentration

Konzentrationskoeffizient

Definition: Es seien x_i - (i=1,2,...,n) die der Größe nach absteigend geordneten Beobachtungen eines extensiven Merkmals ($x_i > x_{i+1}$). Die Größen C_r heißen **absolute Konzentrationskoeffizienten**:

$$C_r = \sum_{i=1}^{r} a_i \ mit$$

$$a_i = \frac{x_i}{\sum_{i=1}^{n} x_i}$$

Bemerkung: Man erhält eine Folge von absoluten Konzentrationskoeffizientem C_r (Konzentrationskurve).
Vergleich mit der Gleichverteilungsgerade ($C_r = \frac{r}{n}$).

Interpretationen:

Starke Abweichung von der Gleichverteilungsgerade $\Longrightarrow$ hohe Konzentration.

C_r - kummulativer Anteil der absteigend geordneten Merkmalsausprägungen

Beispiel

Produktionsvolumen in 5 Unternehmen in Mio. EURO

Objekt	x_i
A	20
B	30
C	15
D	70
E	40

Fragestellungen:

Welchen Anteil haben die produktionsstärksten Unternehmen?

Wie kann der Koeffizient C_2 interpretiert werden?

Herfindahl-Index

Definition: Es seien x_i - (i=1,2,...,n) Beobachtungen eines extensiven Merkmals. Die Größen H heißt **Herfindahl-Index**:

$$H = \sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} \left(\frac{x_i}{n \cdot \overline{x}}\right)^2 \; mit$$

$$a_i = \frac{x_i}{\sum_{i=1}^{n} x_i}$$

Bemerkung: $\frac{1}{n} \leq H \leq 1$

Eine minimale Konzentration liegt dabei vor, wenn sich der Absatz auf alle Produzenten gleich verteilt. $H = \frac{1}{n}$

Eine maximale Konzentration liegt vor, wenn der gesamte Absatz auf einen Produzenten entfällt. (H=1)

Beispiel

Produktionsvolumen in 5 Ojekten in Mio. EURO

Objekt	x_i
A	20
B	30
C	15
D	70
E	40

Fragestellung:

Was sagt der Herfindahl-Index áus?

Beispiel Interpretation

$$H = \sum_{i=1}^{n} a_i^2 = 0,262040816$$

Interpretationen: $H > 0,1 \quad \Rightarrow$ starke Konzentration

Fusionsaussage: wie steigt der Konzentrationsgrad, wenn zwei der fünf Unternehmen fusionieren?

$$H = \sum_{i=1}^{n} a_i^2 = 0,444897959$$

Konzentrationsgrad steigt

Gliederung

- 2.3.1 absolute statistische Konzentration
- 2.3.2 **relative statistische Konzentration**

Lorenzkurve

Definition: Es seien x_i - (i=1,2,...,n) die der Größe nach aufsteigend geordneten Beobachtungen eines extensiven Merkmals ($x_i < x_{i+1}$). Es wird der kumulierte Merkmalsträgeranteil A_i den Anteilen der Gleichverteilungsgerade F_i gegenübergestellt:

$$A_i = \frac{\sum_{r=1}^{i} x_r}{\sum_{i=1}^{n} x_i}$$

$$F_i = \frac{i}{n}$$

Die Folge (F_i, A_i) (i=0,1,2,...,n) mit $F_0 = A_0 = 0$ heißt **LORENZ-Kurve**
Bemerkung: Vergleich mit der Gleichverteilungsgerade ($F_i = \frac{i}{n}$).

Starke Abweichung von der Gleichverteilungsgerade $\Longrightarrow$ hohe Konzentration

Beispiel

Produktionsvolumen in 5 Ojekten in Mio. EURO

Objekt	x_i
A	20
B	30
C	15
D	70
E	40

Fragestellung:

Was sagt die Lorenzkurve aus?

Lorenzkurve Interpretaion

- Konstruktion einer Grafik (F - Abzisse, A Ordinate)
- Die Gerade durch die Punkte (0,0) und (1,1) ist die Gleichverteilungsgerade
- Lorenzkurve: tatsächliche Verteilung
- Fläche zwischen Lorenzkurve und Gleichverteilungsgerade: Konzentrationsfläche
- je größe die Konzentrationsfläche, desto stärker die relative Konzentration
- gleichmäßige Konzentration: Konzentrationsfläche $\longrightarrow 0$
- starke Konzentration: Konzentrationsfläche $\longrightarrow 0,5$

Lorenzkurve für disjunkte Merkmale

Definition: Es seien ξ_i - (i=1,2,...,n) die der Größe nach aufsteigend geordneten disjunkten Beobachtungen eines extensiven Merkmals ($\xi_i < \xi_{i+1}$) und p_i die zugehörigen relativen Häufigkeiten. Es wird der kumulierte Merkmalsträgeranteil A_i den Anteilen der Gleichverteilungsgerade F_i gegenübergestellt:

$$A_i = \frac{\sum_{r=1}^{i} \xi_r p_r}{\sum_{i=1}^{n} \xi_i p_i}$$

$$F_i = \sum_{r=1}^{i} p_r$$

Die Folge (F_i, A_i) (i=0,1,2,...,n) mit $F_0 = A_0 = 0$ heißt **LORENZ-Kurve**
Bemerkung: Vergleich mit der Gleichverteilungsgerade.

Starke Abweichung von der Gleichverteilungsgerade $\Longrightarrow$ hohe Konzentration

Lorenzkurve für klassierte / gepoolte Merkmale

Definition: Es seien x_i^* - (i=1,2,...,n) bzw. $\overline{x_i}$ - (i=1,2,...,n) die der Größe nach aufsteigend geordneten Klassenmitte bzw. gepoolte Mittelwerte eines extensiven Merkmals und p_i die zugehörigen relativen Häufigkeiten. Es wird der kumulierte Merkmalsträgeranteil A_i den Anteilen der Gleichverteilungsgerade F_i gegenübergestellt:

$$A_i = \frac{\sum_{r=1}^{i} x_r^* p_r}{\sum_{i=1}^{n} x_i^* p_i}$$

$$A_i = \frac{\sum_{r=1}^{i} \overline{x_r} p_r}{\sum_{i=1}^{n} \overline{x_i} p_i}$$

$$F_i = \sum_{r=1}^{i} p_r$$

Die Folge (F_i, A_i) (i=0,1,2,...,n) mit $F_0 = A_0 = 0$ heißt **LORENZ-Kurve** Bemerkung: Vergleich der Fläche zwischen der Gleichverteilungsgerade und der Lorenzkurve.

Starke Abweichung von der Gleichverteilungsgerade $\Longrightarrow$ hohe Konzentration

GINI-Koeffizient

Definition: Es seien x_i - (i=1,2,...,n) die der Größe nach aufsteigend geordneten Beobachtungen eines extensiven Merkmals ($x_i < x_{i+1}$). Aus den kumulierte Merkmalsträgeranteilen A_i wird der **GINI-Koeffizient** berechnet:

$$G = 1 - \frac{1}{n} \sum_{i=1}^{n} (A_i + A_{i-1}) \; mit \; A_0 = 0$$

Bemerkung:

- G nahe 0 $\Longrightarrow$ kaum Konzentration
- G nahe 1 $\Longrightarrow$ hohe Konzentration

GINI-Koeffizient - disjunkte Merkmale

Definition: Es seien ξ_i - (i=1,2,...,n) die der Größe nach aufsteigend geordneten disjunkten Beobachtungen eines extensiven Merkmals ($\xi_i < \xi_{i+1}$) und p_i die zugehörigen relativen Häufigkeiten. Aus den kumulierte Merkmalsträgeranteilen A_i wird der **GINI-Koeffizient** berechnet:

$$G = 1 - \sum_{i=1}^{n} (A_i + A_{i-1}) p_i \ \textit{mit}\ A_0 = 0$$

Bemerkung:

- G nahe 0 $\Longrightarrow$ kaum Konzentration
- G nahe 1 $\Longrightarrow$ hohe Konzentration
- bei gepooltem oder klassiertem Datensatz werden an Stelle der disjunkten Merkmalsausprägungen ξ_i die Mittelwerte der Klassen $\overline{x_i}$ oder die Klassenmitte x_i^* verwendet.

Beispiel -Fortsetzung

Produktionsvolumen in 5 Ojekten in Mio. EURO

Objekt	x_i
A	20
B	30
C	15
D	70
E	40

GINI-Koeffizient- Interpretation

Der GINI-Koeffizient ist aus der Lorenzkurve ableitbar:

$$G = \frac{\textit{Flächeninhalt zwischen Lorenzkurve und Diagonale zwischen } (0,0) \textit{ und } (1,1)}{\textit{Flächeninhalt unterhalb der Diagonale zwischen } (0,0) \textit{ und } (1,1)}$$

$$= \frac{\textit{Flächeninhalt zwischen Lorenzkurve und Diagonale zwischen } (0,0) \textit{ und } (1,1)}{0,5}$$

Beispiel

Objekt	x_i
A	10
B	10
C	30
D	50

Graphisch betrachtet ist der Gini-Koeffizient das Verhältnis der Fläche zwischen Gleichverteilungslinie und Lorenzkurve zur Fläche unterhalb der Gleichverteilungslinie

GINI-Koeffizient- Interpretation

