Karlsruher Institut für Technologie

Institut für Volkswirtschaftslehre (ECON)
Lehrstuhl für Ökonometrie und Statistik

Kapitel VII - Konzentration von Merkmalswerten

Deskriptive Statistik

Prof. Dr. W.-D. Heller
Hartwig Senska
Carlo Siebenschuh

Agenda

1. **Einleitung**
2. Lorenzkurve
3. Gini-Koeffizient
4. Weitere Konzentrationsmaße

Einleitung

Bei vielen Verteilungen sind Lage- und Streuungsparameter für die Analyse einer Häufigkeitsverteilung nicht ausreichend.

Beispiel 7.1
Bei einer Einkommensverteilung ist neben dem arithmetischen Mittel, also dem durchschnittlichen Einkommen, ein Streuungsparameter, z.B. die Standardabweichung, von Bedeutung, die etwas über die Abweichung vom mittleren Einkommen aussagt. Die Standardabweichung ist allerdings nicht genügend aussagekräftig, da die Verteilung in der Regel nicht symmetrisch ist. Es kann z. B. viele mit geringeren Einkommen als dem Durchschnittseinkommen und einige wenige mit wesentlich höheren Einkommen geben. Das Ziel ist es diese Ungleichheit in der Verteilung der Einkommen mit Hilfe eines aussagekräftigen Maßes sichtbar und damit vergleichbar zu machen.

Einleitung

Bei vielen Verteilungen sind Lage- und Streuungsparameter für die Analyse einer Häufigkeitsverteilung nicht ausreichend.

Beispiel 7.1
Bei einer Einkommensverteilung ist neben dem arithmetischen Mittel, also dem durchschnittlichen Einkommen, ein Streuungsparameter, z.B. die Standardabweichung, von Bedeutung, die etwas über die Abweichung vom mittleren Einkommen aussagt. Die Standardabweichung ist allerdings nicht genügend aussagekräftig, da die Verteilung in der Regel nicht symmetrisch ist. Es kann z. B. viele mit geringeren Einkommen als dem Durchschnittseinkommen und einige wenige mit wesentlich höheren Einkommen geben. Das Ziel ist es diese Ungleichheit in der Verteilung der Einkommen mit Hilfe eines aussagekräftigen Maßes sichtbar und damit vergleichbar zu machen.

Einleitung

Ziel: Beschreibung, graphische Darstellung und Messung von Ungleichheiten z.B. bei der Verteilung von "Besitz"

Gegeben: "Besitz"-Verteilung in Form einer nichtnegativen geordneten Urliste

$$0 \leq x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

Gesamtbesitz: Positive Merkmalssumme

$$x = \sum_{i=1}^{n} x_{(i)} > 0$$

Einleitung

Ziel: Beschreibung, graphische Darstellung und Messung von Ungleichheiten z.B. bei der Verteilung von “Besitz”

Gegeben: “Besitz”-Verteilung in Form einer nichtnegativen geordneten Urliste

$$0 \leq x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

Gesamtbesitz: Positive Merkmalssumme

$$x = \sum_{i=1}^{n} x_{(i)} > 0$$

Einleitung

Ziel: Beschreibung, graphische Darstellung und Messung von Ungleichheiten z.B. bei der Verteilung von "Besitz"

Gegeben: "Besitz"-Verteilung in Form einer nichtnegativen geordneten Urliste

$$0 \leq x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

Gesamtbesitz: Positive Merkmalssumme

$$x = \sum_{i=1}^{n} x_{(i)} > 0$$

Einleitung

Frage: Wie ist die Merkmalssumme auf die n Personen verteilt?

Extremfälle:

- Alle besitzen gleich viel.
- Einer besitzt alles.

Einleitung

Frage: Wie ist die Merkmalssumme auf die *n* Personen verteilt?

Extremfälle:

- Alle besitzen gleich viel.
- Einer besitzt alles.

Agenda

Lorenzkurve

Wichtigstes graphisches Hilfsmittel zur Verdeutlichung von Konzentrationsphänomenen. Ausgegangen wird dabei von einer geordneten nichtnegativen statistischen Reihe mit positiver Summe der Beobachtungswerte (Merkmalssumme)

$$0 \leq x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}, \quad x = \sum_{i=1}^{n} x_{(i)} > 0$$

Prinzip: Gegenüberstellung des

- Anteils an der statistischen Masse

und des

- Anteils an der Merkmalssumme der k statistischen Einheiten mit den kleinsten Merkmalswerten

Lorenzkurve

Wichtigstes graphisches Hilfsmittel zur Verdeutlichung von Konzentrationsphänomenen. Ausgegangen wird dabei von einer geordneten nichtnegativen statistischen Reihe mit positiver Summe der Beobachtungswerte (Merkmalssumme)

$$0 \leq x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}, \quad x = \sum_{i=1}^{n} x_{(i)} > 0$$

Prinzip: Gegenüberstellung des

- Anteils an der statistischen Masse

und des

- Anteils an der Merkmalssumme der k statistischen Einheiten mit den kleinsten Merkmalswerten

Lorenzkurve

Anteil an der Merkmalssumme, der auf die k statistischen Einheiten mit den kleinsten Merkmalswerten $(x_{(1)}, \ldots, x_{(k)})$ entfällt:

$$v_k = \frac{\sum\limits_{i=1}^{k} x_{(i)}}{\sum\limits_{i=1}^{n} x_{(i)}} = \frac{x_{(1)} + \ldots + x_{(k)}}{x_{(1)} + \ldots + x_{(n)}}$$

Anteil an der gesamten statistischen Masse:

$$u_k = \frac{k}{n}$$

Lorenzkurve

Anteil an der Merkmalssumme, der auf die k statistischen Einheiten mit den kleinsten Merkmalswerten $(x_{(1)}, \ldots, x_{(k)})$ entfällt:

$$v_k = \frac{\sum\limits_{i=1}^{k} x_{(i)}}{\sum\limits_{i=1}^{n} x_{(i)}} = \frac{x_{(1)} + \ldots + x_{(k)}}{x_{(1)} + \ldots + x_{(n)}}$$

Anteil an der gesamten statistischen Masse:

$$u_k = \frac{k}{n}$$

Lorenzkurve

Damit steht also dem Anteil u_k an der statistischen Masse ein Anteil v_k an der Merkmalssumme gegenüber. Für $k = 1, \ldots, n$ trägt man die Punkte (u_k, v_k) in ein Koordinatenkreuz ein und verbindet sie durch einen Streckenzug, beginnend mit dem Ursprung (0,0):

Lorenzkurve

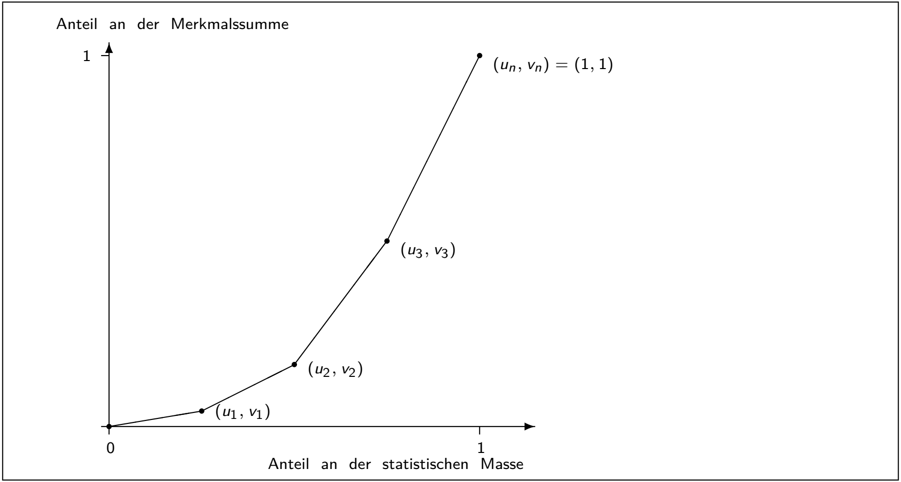


Abbildung - Lorenzkurve.

Lorenzkurve

Beispiel 7.1
Die geordnete statistische Reihe der Monatslöhne in einem mittleren Handwerksbetrieb laute wie folgt in Euro:

$$500, 1900, 2050, 2200, 2250, 2400, 2600, 2950, 4000, 5000.$$

Merkmalssumme ist: 25850. Damit erhält man folgende Tabelle:

u_k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
v_k	0.02	0.09	0.17	0.26	0.34	0.44	0.54	0.65	0.81	1.0

Und die folgende Lorenzkurve ...

Lorenzkurve

Beispiel 7.1
Die geordnete statistische Reihe der Monatslöhne in einem mittleren Handwerksbetrieb laute wie folgt in Euro:

$$500, 1900, 2050, 2200, 2250, 2400, 2600, 2950, 4000, 5000.$$

Merkmalssumme ist: 25850. Damit erhält man folgende Tabelle:

u_k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
v_k	0.02	0.09	0.17	0.26	0.34	0.44	0.54	0.65	0.81	1.0

Und die folgende Lorenzkurve . . .

Lorenzkurve

Beispiel 7.1
Die geordnete statistische Reihe der Monatslöhne in einem mittleren Handwerksbetrieb laute wie folgt in Euro:

$$500, 1900, 2050, 2200, 2250, 2400, 2600, 2950, 4000, 5000.$$

Merkmalssumme ist: 25850. Damit erhält man folgende Tabelle:

u_k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
v_k	0.02	0.09	0.17	0.26	0.34	0.44	0.54	0.65	0.81	1.0

Und die folgende Lorenzkurve ...

Lorenzkurve

Beispiel 7.1
Die geordnete statistische Reihe der Monatslöhne in einem mittleren Handwerksbetrieb laute wie folgt in Euro:

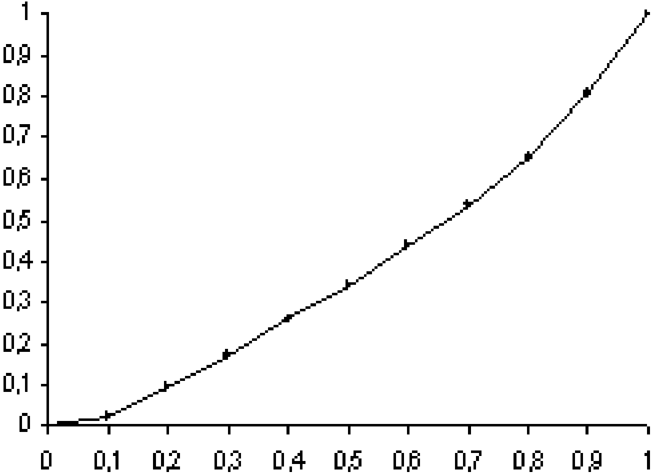
$$500, 1900, 2050, 2200, 2250, 2400, 2600, 2950, 4000, 5000.$$

Merkmalssumme ist: 25850. Damit erhält man folgende Tabelle:

u_k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
v_k	0.02	0.09	0.17	0.26	0.34	0.44	0.54	0.65	0.81	1.0

Und die folgende Lorenzkurve ...

Lorenzkurve



Lorenzkurve

Eigenschaften der Lorenzkurve:

- Die Lorenzkurve beginnt in (0,0) und endet in (1,1).
- Die Lorenzkurve verläuft nirgendwo oberhalb der Diagonalen.
- Die Lorenzkurve steigt monoton.
- Die Lorenzkurve ist konvex.

Lorenzkurve

Eigenschaften der Lorenzkurve:

- Die Lorenzkurve beginnt in (0,0) und endet in (1,1).
- Die Lorenzkurve verläuft nirgendwo oberhalb der Diagonalen.
- Die Lorenzkurve steigt monoton.
- Die Lorenzkurve ist konvex.

Lorenzkurve

Eigenschaften der Lorenzkurve:

- Die Lorenzkurve beginnt in (0,0) und endet in (1,1).
- Die Lorenzkurve verläuft nirgendwo oberhalb der Diagonalen.
- Die Lorenzkurve steigt monoton.
- Die Lorenzkurve ist konvex.

Lorenzkurve

Eigenschaften der Lorenzkurve:

- Die Lorenzkurve beginnt in (0,0) und endet in (1,1).
- Die Lorenzkurve verläuft nirgendwo oberhalb der Diagonalen.
- Die Lorenzkurve steigt monoton.
- Die Lorenzkurve ist konvex.

Lorenzkurve

Eigenschaften der Lorenzkurve:

- Die Lorenzkurve beginnt in (0,0) und endet in (1,1).
- Die Lorenzkurve verläuft nirgendwo oberhalb der Diagonalen.
- Die Lorenzkurve steigt monoton.
- Die Lorenzkurve ist konvex.

Lorenzkurve

Anmerkung: Die Diagonale ist die Bezugskurve zur Lorenzkurve. Sind nämlich alle Beobachtungswerte gleich ...

$$0 < x_1 = x_2 = \ldots = x_n,$$

... so ist für $k = 1, \ldots, n$

$$v_k = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{n} x_i} = \frac{k \cdot x_1}{n \cdot x_1} = \frac{k}{n} = u_k.$$

Die Diagonale gibt also den Zustand wieder, in dem die Merkmalssumme völlig gleichmäßig über die Masse verteilt ist ("Gleichverteilung der Merkmalssumme"). Aus der Sicht der Konzentration der Idealzustand ohne jegliche Konzentration.

Dem Idealzustand der Gleichverteilung entgegengesetzt ist der Extremfall, dass die gesamte Merkmalssumme in einer statistischen Einheit vereint ist:

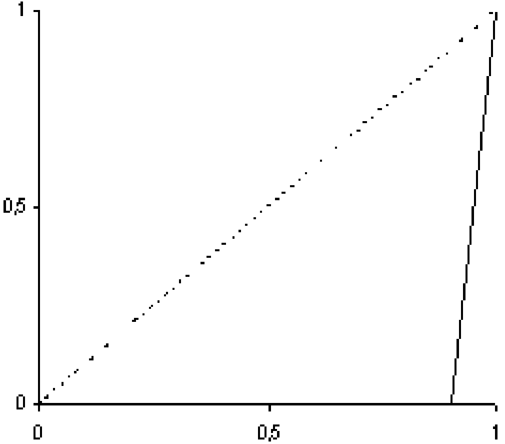
$$0 = x_{(1)} = \ldots = x_{(n-1)} < x_{(n)}.$$

Die Diagonale gibt also den Zustand wieder, in dem die Merkmalssumme völlig gleichmäßig über die Masse verteilt ist ("Gleichverteilung der Merkmalssumme"). Aus der Sicht der Konzentration der Idealzustand ohne jegliche Konzentration.

Dem Idealzustand der Gleichverteilung entgegengesetzt ist der Extremfall, dass die gesamte Merkmalssumme in einer statistischen Einheit vereint ist:

$$0 = x_{(1)} = \ldots = x_{(n-1)} < x_{(n)}.$$

Lorenzkurve



Lorenzkurve

Weitere Anmerkung: Für großes n, also viele statistische Einheiten, erhält man bei vollständiger Konzentration "nahezu" die Katheden des rechtwinkligen Dreiecks mit den Eckpunkten (0,0), (1,0), (1,1).

$\Rightarrow$ Interpretation der Lorenzkurve: Je weiter die Lorenzkurve von der Diagonalen entfernt ist, je mehr die Lorenzkurve also durchhängt, desto größer ist die Konzentration.

Lorenzkurve

Weitere Anmerkung: Für großes n, also viele statistische Einheiten, erhält man bei vollständiger Konzentration "nahezu" die Katheden des rechtwinkligen Dreiecks mit den Eckpunkten (0,0), (1,0), (1,1).

$\Rightarrow$ Interpretation der Lorenzkurve: Je weiter die Lorenzkurve von der Diagonalen entfernt ist, je mehr die Lorenzkurve also durchhängt, desto größer ist die Konzentration.

Lorenzkurve

Beispiel 7.2

Gegeben ist die Häufigkeitsverteilung

a	1	2	3
$h(a)$	2	3	1
$p(a)$	$0.\overline{3}$	0.5	$0.1\overline{6}$

Geordnete Urliste ist dann: 1, 1, 2, 2, 2, 3; Merkmalssumme ist 11.
Damit erhält man die Koordinaten der Lorenzkurve:

u_k	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1
v_k	0	$\frac{1}{11}$	$\frac{2}{11}$	$\frac{4}{11}$	$\frac{6}{11}$	$\frac{8}{11}$	1

Lorenzkurve

Beispiel 7.2

Gegeben ist die Häufigkeitsverteilung

a	1	2	3
$h(a)$	2	3	1
$p(a)$	$0.\overline{3}$	0.5	$0.1\overline{6}$

Geordnete Urliste ist dann: 1, 1, 2, 2, 2, 3; Merkmalssumme ist 11.
Damit erhält man die Koordinaten der Lorenzkurve:

u_k	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1
v_k	0	$\frac{1}{11}$	$\frac{2}{11}$	$\frac{4}{11}$	$\frac{6}{11}$	$\frac{8}{11}$	1

Lorenzkurve

Beispiel 7.2

Gegeben ist die Häufigkeitsverteilung

a	1	2	3
$h(a)$	2	3	1
$p(a)$	$0.\overline{3}$	0.5	$0.1\overline{6}$

Geordnete Urliste ist dann: 1, 1, 2, 2, 2, 3; Merkmalssumme ist 11.
Damit erhält man die Koordinaten der Lorenzkurve:

u_k	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1
v_k	0	$\frac{1}{11}$	$\frac{2}{11}$	$\frac{4}{11}$	$\frac{6}{11}$	$\frac{8}{11}$	1

Lorenzkurve

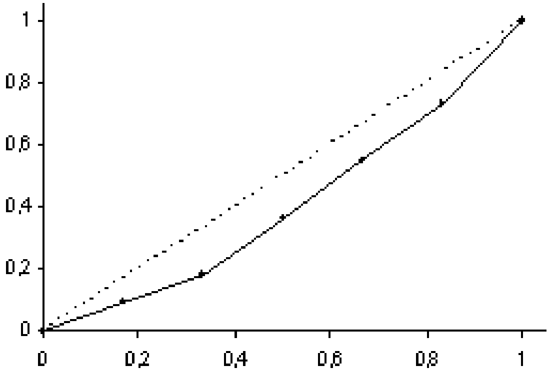


Abbildung: Abbildung 7.4 - Lorenzkurve zu Beispiel 7.2

Anmerkung: Man sieht, dass übereinstimmende Merkmalswerte zu Geradenstücken gleicher Steigung führen. Es genügt also, die Werte für $k = 2$ und $k = 5$ zu berechnen.

Lorenzkurve

Ermittlung der Lorenzkurve aus der absoluten Häufigkeitsverteilung

Merkmalssumme:

$$x = \sum_{a \in M} a \cdot h(a)$$

Zur Berechnung der Koordinaten sind die Merkmalsausprägungen zu ordnen:

$$0 \leq a_1 < a_2 < \ldots < a_m.$$

Lorenzkurve

Ermittlung der Lorenzkurve aus der absoluten Häufigkeitsverteilung

Merkmalssumme:

$$x = \sum_{a \in M} a \cdot h(a)$$

Zur Berechnung der Koordinaten sind die Merkmalsausprägungen zu ordnen:

$$0 \leq a_1 < a_2 < \ldots < a_m.$$

Lorenzkurve

Ermittlung der Lorenzkurve aus der absoluten Häufigkeitsverteilung

Anteil der k niedrigsten Merkmalsausprägungen an der Merkmalssumme:

$$v_k = \frac{\sum_{i=1}^{k} a_i \cdot h(a_i)}{\sum_{a \in M} a \cdot h(a)}$$

Anteil dieser Merkmalsausprägungen an der statistischen Masse:

$$u_k = \frac{\sum_{i=1}^{k} h(a_i)}{\sum_{a \in M} h(a)}$$

Ermittlung der Lorenzkurve aus der absoluten Häufigkeitsverteilung

Anteil der k niedrigsten Merkmalsausprägungen an der Merkmalssumme:

$$v_k = \frac{\sum\limits_{i=1}^{k} a_i \cdot h(a_i)}{\sum\limits_{a \in M} a \cdot h(a)}$$

Anteil dieser Merkmalsausprägungen an der statistischen Masse:

$$u_k = \frac{\sum\limits_{i=1}^{k} h(a_i)}{\sum\limits_{a \in M} h(a)}$$

Lorenzkurve

Beispiel 7.3

Im Beispiel 7.2 erhält man die notwendigen Koordinaten der Lorenzkurve

u_k	0	$\frac{2}{6}$	$\frac{5}{6}$	1
v_k	0	$\frac{2}{11}$	$\frac{8}{11}$	1

Anmerkung: Der Unterschied besteht darin, dass hierbei lediglich die "Knickstellen" der Lorenzkurve berechnet werden.

Lorenzkurve

Beispiel 7.3

Im Beispiel 7.2 erhält man die notwendigen Koordinaten der Lorenzkurve

u_k	0	$\frac{2}{6}$	$\frac{5}{6}$	1
v_k	0	$\frac{2}{11}$	$\frac{8}{11}$	1

Anmerkung: Der Unterschied besteht darin, dass hierbei lediglich die “Knickstellen” der Lorenzkurve berechnet werden.

Lorenzkurve

Ermittlung der Lorenzkurve aus der relativen Häufigkeitsverteilung

Anteil der k niedrigsten Merkmalsausprägungen an der Merkmalssumme:

$$v_k = \frac{\sum_{i=1}^{k} a_i \cdot p(a_i)}{\sum_{a \in M} a \cdot p(a)}$$

Anteil dieser Merkmalsausprägungen an der statistischen Masse:

$$u_k = \sum_{i=1}^{k} p(a_i)$$

Lorenzkurve

Ermittlung der Lorenzkurve aus der relativen Häufigkeitsverteilung

Anteil der k niedrigsten Merkmalsausprägungen an der Merkmalssumme:

$$v_k = \frac{\sum_{i=1}^{k} a_i \cdot p(a_i)}{\sum_{a \in M} a \cdot p(a)}$$

Anteil dieser Merkmalsausprägungen an der statistischen Masse:

$$u_k = \sum_{i=1}^{k} p(a_i)$$

Lorenzkurve

Beispiel 7.4

In Beispiel 7.2 erhält man

$$\sum_{i=1}^{3} a_i p(a_i) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{6} = \frac{11}{6}$$

und damit

$v_1 = \frac{\frac{1}{3}}{\frac{11}{6}} = \frac{2}{11}, v_2 = \frac{(\frac{1}{3}+1)}{\frac{11}{6}} = \frac{8}{11}, v_3 = 1$ wie in Beispiel 7.3.

Lorenzkurve

Beispiel 7.4

In Beispiel 7.2 erhält man

$$\sum_{i=1}^{3} a_i p(a_i) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{6} = \frac{11}{6}$$

und damit

$$v_1 = \frac{\frac{1}{3}}{\frac{11}{6}} = \frac{2}{11}, v_2 = \frac{(\frac{1}{3} + 1)}{\frac{11}{6}} = \frac{8}{11}, v_3 = 1 \text{ wie in Beispiel 7.3.}$$

Lorenzkurve

Ermittlung der Lorenzkurve bei klassierten Merkmalen

Weder u_k noch v_k können gebildet werden. Seien I_j die Klassen und $h(I_j)$ bzw. $p(I_j)$ die absoluten bzw. relativen Häufigkeiten. Die Klasse I hat damit den relativen Anteil $p(I) = \frac{h(I)}{n}$ an der statistischen Masse. Seien also die Klassen $I_1, \ldots, I_m$ nach ihren Klassengrenzen geordnet, dann kann man statt $u_1, \ldots, u_n$ die folgenden Werte verwenden:

$$p(I_1), p(I_1) + p(I_2), \ldots, \sum_{j=1}^{m} p(I_j)$$

Lorenzkurve

Ermittlung der Lorenzkurve bei klassierten Merkmalen

Der Anteil an der Merkmalssumme einer Klasse I lässt sich nur anhand der Urliste oder der Häufigkeitsverteilung der unklassierten Daten feststellen. Geht man davon aus, dass die Klassenmitte z_I das arithmetische Mittel der Merkmalswerte der Klasse ist, so ist

$$z_I h(I): \quad \text{Merkmalssumme der Klasse } I$$

und

$$\sum_{j=1}^{m} z_j h(I_j): \quad \text{Merkmalssumme der Gesamtmasse.}$$

Lorenzkurve

Ermittlung der Lorenzkurve bei klassierten Merkmalen

Der Anteil an der Merkmalssumme einer Klasse I lässt sich nur anhand der Urliste oder der Häufigkeitsverteilung der unklassierten Daten feststellen. Geht man davon aus, dass die Klassenmitte z_I das arithmetische Mittel der Merkmalswerte der Klasse ist, so ist

$$z_I h(I): \quad \text{Merkmalssumme der Klasse } I$$

und

$$\sum_{j=1}^{m} z_j h(I_j): \quad \text{Merkmalssumme der Gesamtmasse.}$$

Lorenzkurve

Ermittlung der Lorenzkurve bei klassierten Merkmalen

Punkte der Lorenzkurve für absolute Häufigkeiten:

$$(u_k, v_k) = \left(\sum_{j=1}^{k} p(I_j), \frac{\sum_{j=1}^{k} z_j h(I_j)}{\sum_{j=1}^{m} z_j h(I_j)} \right) \qquad \text{für } k = 0, \ldots, m$$

Punkte der Lorenzkurve für relative Häufigkeiten:

$$(u_k, v_k) = \left(\sum_{j=1}^{k} p(I_j), \frac{\sum_{j=1}^{k} z_j p(I_j)}{\sum_{j=1}^{m} z_j p(I_j)} \right) \qquad \text{für } k = 0, \ldots, m$$

Lorenzkurve

Ermittlung der Lorenzkurve bei klassierten Merkmalen

Punkte der Lorenzkurve für absolute Häufigkeiten:

$$(u_k, v_k) = \left(\sum_{j=1}^{k} p(I_j), \frac{\sum\limits_{j=1}^{k} z_j h(I_j)}{\sum\limits_{j=1}^{m} z_j h(I_j)} \right) \qquad \text{für } k = 0, \ldots, m$$

Punkte der Lorenzkurve für relative Häufigkeiten:

$$(u_k, v_k) = \left(\sum_{j=1}^{k} p(I_j), \frac{\sum\limits_{j=1}^{k} z_j p(I_j)}{\sum\limits_{j=1}^{m} z_j p(I_j)} \right) \qquad \text{für } k = 0, \ldots, m$$

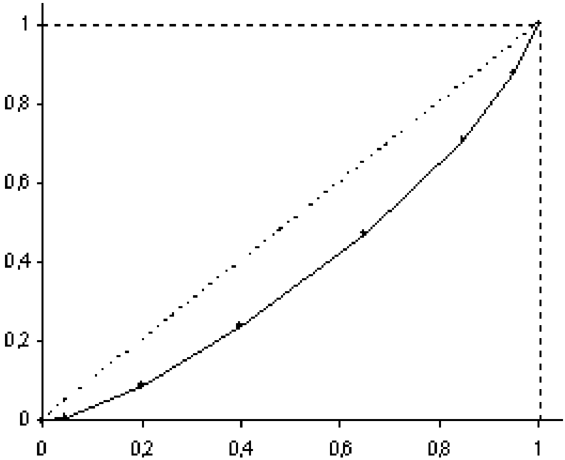
Lorenzkurve

Beispiel 7.5

Für eine Verbrauchsstudie wurden die Nettojahreseinkommen von 100 Männern festgestellt:

	Männer					
Einkommen in TEuro	abs. H.	rel. H.	u_k	$z_j h(I_j)$	$\sum z_j h(I_j)$	v_k
0 bis unter 10	5	0.05	0.05	25	25	0.008
10 bis unter 20	15	0.15	0.20	225	250	0.085
20 bis unter 25	20	0.20	0.40	450	700	0.237
25 bis unter 30	25	0.25	0.65	687.5	1387.5	0.470
30 bis unter 40	20	0.20	0.85	700	2087.5	0.708
40 bis unter 60	10	0.10	0.95	500	2587.5	0.877
60 bis unter 85	5	0.05	1	362.5	2950	1
$\sum$	100	1				

Lorenzkurve



Agenda

1. Einleitung
2. Lorenzkurve
3. **Gini-Koeffizient**
4. Weitere Konzentrationsmaße

Gini-Koeffizient

Anteil der Fläche zwischen der Diagonalen und der Lorenzkurve an Gesamtfläche unterhalb der Diagonalen. Er ist ein Maß für die Konzentration, die eben gerade der Abweichung der Lorenzkurve von der Diagonalen entspricht.

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

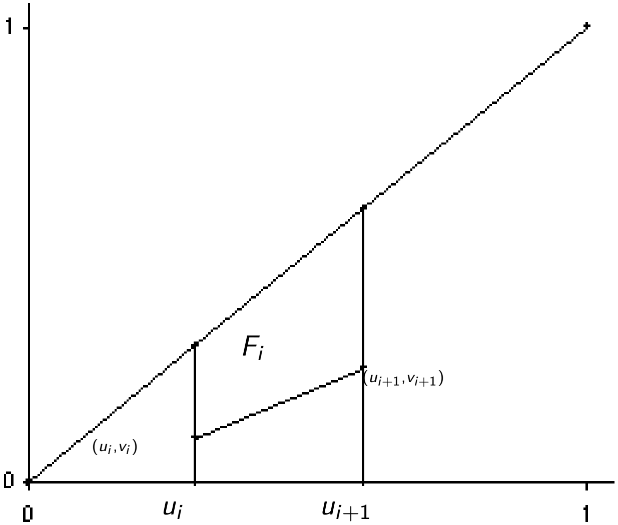


Abbildung 7.6 - Zur Berechnung des Gini-Koeffizienten

Gini-Koeffizient

Berechnung des Gini-Koeffizienten

Für F_i gilt

$$F_i = (u_{i+1} - u_i) \cdot (\frac{u_i - v_i}{2} + \frac{u_{i+1} - v_{i+1}}{2}),$$

da F_i (um 90^o gedreht) ein Trapez ist, mit der Höhe $u_{i+1} - u_i$ und der Mittellinie

$$0.5((u_i - v_i) + (u_{i+1} - v_{i+1})).$$

Gini-Koeffizient

Berechnung des Gini-Koeffizienten

Für F_i gilt

$$F_i = (u_{i+1} - u_i) \cdot (\frac{u_i - v_i}{2} + \frac{u_{i+1} - v_{i+1}}{2}),$$

da F_i (um 90^o gedreht) ein Trapez ist, mit der Höhe $u_{i+1} - u_i$ und der Mittellinie

$$0.5((u_i - v_i) + (u_{i+1} - v_{i+1})).$$

Gini-Koeffizient

Damit ist

$$
\begin{aligned}
G &= \frac{\frac{1}{2}\sum\limits_{i=0}^{n-1}(u_{i+1}-u_i)(u_i-v_i+u_{i+1}-v_{i+1})}{\frac{1}{2}} \\
&= \sum_{i=0}^{n-1}(u_{i+1}-u_i)(u_i-v_i+u_{i+1}-v_{i+1})
\end{aligned}
$$

Setzt man die Daten aus der geordneten Urliste ein, so erhält man nach einigem Rechenaufwand

$$G = \frac{2\cdot\sum\limits_{i=1}^{n} i\cdot x_{(i)} - (n+1)\cdot\sum\limits_{i=1}^{n} x_{(i)}}{n\cdot\sum\limits_{i=1}^{n} x_{(i)}}.$$

Gini-Koeffizient

Damit ist

$$
\begin{aligned}
G &= \frac{\frac{1}{2}\sum\limits_{i=0}^{n-1}(u_{i+1}-u_i)(u_i-v_i+u_{i+1}-v_{i+1})}{\frac{1}{2}} \\
&= \sum_{i=0}^{n-1}(u_{i+1}-u_i)(u_i-v_i+u_{i+1}-v_{i+1})
\end{aligned}
$$

Setzt man die Daten aus der geordneten Urliste ein, so erhält man nach einigem Rechenaufwand

$$
G = \frac{2\cdot\sum\limits_{i=1}^{n} i\cdot x_{(i)} - (n+1)\cdot\sum\limits_{i=1}^{n} x_{(i)}}{n\cdot\sum\limits_{i=1}^{n} x_{(i)}}.
$$

Gini-Koeffizient

Beispiel 7.6

(1) In Beispiel 7.1 erhält man:

$$\begin{aligned} G &= \frac{1}{10} \cdot (2 \cdot 0.08 + 2 \cdot 0.11 + 2 \cdot 0.13 + 2 \cdot 0.14 + 2 \cdot 0.16 + 2 \cdot 0.16 + \\ &\quad +2 \cdot 0.16 + 2 \cdot 0.15 + 2 \cdot 0.09) = \frac{2}{10} \cdot 1.18 = 0.236. \end{aligned}$$

Nach der zweiten Formel erhält man:

$$G = \frac{2 \cdot 172700 - 11 \cdot 25850}{10 \cdot 25850} = 0.236.$$

Gini-Koeffizient

Beispiel 7.6

(1) In Beispiel 7.1 erhält man:

$$\begin{aligned} G &= \frac{1}{10} \cdot (2 \cdot 0.08 + 2 \cdot 0.11 + 2 \cdot 0.13 + 2 \cdot 0.14 + 2 \cdot 0.16 + 2 \cdot 0.16 + \\ &\quad +2 \cdot 0.16 + 2 \cdot 0.15 + 2 \cdot 0.09) = \frac{2}{10} \cdot 1.18 = 0.236. \end{aligned}$$

Nach der zweiten Formel erhält man:

$$G = \frac{2 \cdot 172700 - 11 \cdot 25850}{10 \cdot 25850} = 0.236.$$

Gini-Koeffizient

Beispiel 7.6

(1) In Beispiel 7.1 erhält man:

$$
\begin{aligned}
G &= \frac{1}{10} \cdot (2 \cdot 0.08 + 2 \cdot 0.11 + 2 \cdot 0.13 + 2 \cdot 0.14 + 2 \cdot 0.16 + 2 \cdot 0.16 + \\
&\quad +2 \cdot 0.16 + 2 \cdot 0.15 + 2 \cdot 0.09) = \frac{2}{10} \cdot 1.18 = 0.236.
\end{aligned}
$$

Nach der zweiten Formel erhält man:

$$G = \frac{2 \cdot 172700 - 11 \cdot 25850}{10 \cdot 25850} = 0.236.$$

Gini-Koeffizient

Beispiel 7.6

(2) Aus den Daten von Beispiel 7.2 bzw. 7.3 erhält man:

$$\begin{aligned} G &= \frac{2}{6}\cdot\left(\frac{2}{6}-\frac{2}{11}\right)+\frac{3}{6}\cdot\left(\frac{2}{6}+\frac{5}{6}-\frac{2}{11}-\frac{8}{11}\right)+\frac{1}{6}\cdot\left(\frac{5}{6}+1-\frac{8}{11}-1\right) \\ &= \frac{2}{6}\cdot\frac{10}{66}+\frac{3}{6}\cdot\frac{17}{66}+\frac{1}{6}\cdot\frac{7}{66}=\frac{78}{6\cdot 66}=\frac{13}{66}=0.20. \end{aligned}$$

(3) Für die Einkommensverteilung der Männer lautet der Gini-Koeffizient:

$$\begin{aligned} G &= 0.05\cdot 0.042+0.15\cdot 0.157+0.2\cdot 0.278+0.25\cdot 0.343+0.2\cdot 0.322 \\ &\quad +0.1\cdot 0.215+0.05\cdot 0.073 \\ &= 0.257. \end{aligned}$$

Gini-Koeffizient

Beispiel 7.6

(2) Aus den Daten von Beispiel 7.2 bzw. 7.3 erhält man:

$$\begin{aligned} G &= \frac{2}{6}\cdot\Big(\frac{2}{6}-\frac{2}{11}\Big)+\frac{3}{6}\cdot\Big(\frac{2}{6}+\frac{5}{6}-\frac{2}{11}-\frac{8}{11}\Big)+\frac{1}{6}\cdot\Big(\frac{5}{6}+1-\frac{8}{11}-1\Big) \\ &= \frac{2}{6}\cdot\frac{10}{66}+\frac{3}{6}\cdot\frac{17}{66}+\frac{1}{6}\cdot\frac{7}{66}=\frac{78}{6\cdot 66}=\frac{13}{66}=0.20. \end{aligned}$$

(3) Für die Einkommensverteilung der Männer lautet der Gini-Koeffizient:

$$\begin{aligned} G &= 0.05\cdot 0.042+0.15\cdot 0.157+0.2\cdot 0.278+0.25\cdot 0.343+0.2\cdot 0.322 \\ &\quad +0.1\cdot 0.215+0.05\cdot 0.073 \\ &= 0.257. \end{aligned}$$

Gini-Koeffizient

Betrachte: Der Maximalwert des Gini-Koeffizienten ist für

$$0 = x_{(1)} = \ldots = x_{(n-1)}, x_{(n)} = \sum_{i=1}^{n} x_{(i)}$$

nach der zweiten Formel

$$G_{\max} = \frac{2 \cdot n \cdot x_{(n)} - (n+1) \cdot x_{(n)}}{n \cdot x_{(n)}} = \frac{n-1}{n}$$

Bei einer Maßzahl geht man üblicherweise von einem Maximalwert 1 aus. Aus diesem Grund normiert man den Ginikoeffizienten.

Gini-Koeffizient

Betrachte: Der Maximalwert des Gini-Koeffizienten ist für

$$0 = x_{(1)} = \ldots = x_{(n-1)}, x_{(n)} = \sum_{i=1}^{n} x_{(i)}$$

nach der zweiten Formel

$$G_{\max} = \frac{2 \cdot n \cdot x_{(n)} - (n+1) \cdot x_{(n)}}{n \cdot x_{(n)}} = \frac{n-1}{n}$$

Bei einer Maßzahl geht man üblicherweise von einem Maximalwert 1 aus. Aus diesem Grund normiert man den Ginikoeffizienten.

Gini-Koeffizient

Normierter Gini-Koeffizient

$$G_{\text{norm}} = \frac{n}{n-1} \cdot G.$$

Es gilt

$$0 \leq G_{\text{norm}} \leq 1$$

und

$$\begin{aligned} G_{\text{norm}} &= 1 \quad \text{bei vollständiger Konzentration,} \\ G_{\text{norm}} &= 0 \quad \text{bei gleichmäßiger Verteilung der Merkmalssumme.} \end{aligned}$$

Gini-Koeffizient

Normierter Gini-Koeffizient

$$G_{\text{norm}} = \frac{n}{n-1} \cdot G.$$

Es gilt

$$0 \leq G_{\text{norm}} \leq 1$$

und

$$\begin{aligned} G_{\text{norm}} &= 1 \quad \text{bei vollständiger Konzentration,} \\ G_{\text{norm}} &= 0 \quad \text{bei gleichmäßiger Verteilung der Merkmalssumme.} \end{aligned}$$

Gini-Koeffizient

Normierter Gini-Koeffizient

$$G_{\text{norm}} = \frac{n}{n-1} \cdot G.$$

Es gilt

$$0 \leq G_{\text{norm}} \leq 1$$

und

$$\begin{aligned} G_{\text{norm}} &= 1 \quad \text{bei vollständiger Konzentration,} \\ G_{\text{norm}} &= 0 \quad \text{bei gleichmäßiger Verteilung der Merkmalssumme.} \end{aligned}$$

Gini-Koeffizient

Kritikpunkte

- Unterschiedliche Lorenzkurven können zu dem selben Gini-Koeffizienten führen.
- Es besteht eine starke Abhängigkeit des Gini-Koeffizienten von der Zahl der einbezogenen statistischen Einheiten. Weglassen von kleinen Merkmalswerten verringert G.
- Der Gini-Koeffizient ist nur ein Maß für die **relative Konzentration**, nicht für die absolute Konzentration.

Gini-Koeffizient

Kritikpunkte

- Unterschiedliche Lorenzkurven können zu dem selben Gini-Koeffizienten führen.
- Es besteht eine starke Abhängigkeit des Gini-Koeffizienten von der Zahl der einbezogenen statistischen Einheiten. Weglassen von kleinen Merkmalswerten verringert G.
- Der Gini-Koeffizient ist nur ein Maß für die **relative Konzentration**, nicht für die absolute Konzentration.

Gini-Koeffizient

Kritikpunkte

- Unterschiedliche Lorenzkurven können zu dem selben Gini-Koeffizienten führen.
- Es besteht eine starke Abhängigkeit des Gini-Koeffizienten von der Zahl der einbezogenen statistischen Einheiten. Weglassen von kleinen Merkmalswerten verringert G.
- Der Gini-Koeffizient ist nur ein Maß für die **relative Konzentration**, nicht für die absolute Konzentration.

Gini-Koeffizient

Beispiel 7.7

Die Messung der Wettbewerbskonzentration mit Hilfe des Gini-Koeffizienten auf zwei verschiedenen Märkten ergibt den selben Wert, obwohl die Märkte nicht identisch sind:

- $G = 0$ für zwei Firmen mit je 50% Marktanteil
- $G = 0$ für 20 Firmen mit je 5% Marktanteil

Gini-Koeffizient

Beispiel 7.7

Die Messung der Wettbewerbskonzentration mit Hilfe des Gini-Koeffizienten auf zwei verschiedenen Märkten ergibt den selben Wert, obwohl die Märkte nicht identisch sind:

- $G = 0$ für zwei Firmen mit je 50% Marktanteil
- $G = 0$ für 20 Firmen mit je 5% Marktanteil

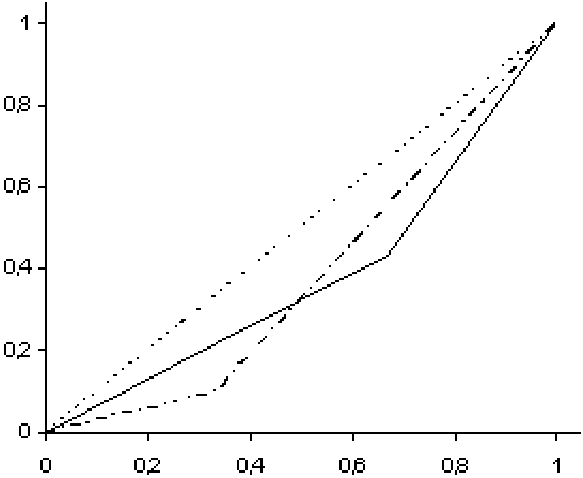
Gini-Koeffizient

Beispiel 7.7

Die Messung der Wettbewerbskonzentration mit Hilfe des Gini-Koeffizienten auf zwei verschiedenen Märkten ergibt den selben Wert, obwohl die Märkte nicht identisch sind:

- $G = 0$ für zwei Firmen mit je 50% Marktanteil
- $G = 0$ für 20 Firmen mit je 5% Marktanteil

Gini-Koeffizient



Agenda

1. Einleitung
2. Lorenzkurve
3. Gini-Koeffizient
4. **Weitere Konzentrationsmaße**

Weitere Konzentrationsmaße

Konzentrationskoeffizient

$$CR_g = \frac{\sum\limits_{i=n-g+1}^{n} x_{(i)}}{\sum\limits_{i=1}^{n} x_{(i)}} \quad \text{für } g = (1,2,)3,\ldots,n$$

CR_g gibt an, welchen Anteil der Merkmalssumme die g letzten Merkmalswerte der geordneten statistischen Reihe in sich vereinen. Die Vorgehensweise entspricht der Konstruktion der Lorenzkurve, wobei die geordnete Urliste von rechts nach links, also in umgekehrter Reihenfolge abgearbeitet wird.

Die zugehörige Kurve wird üblicherweise als **Paretokurve** bezeichnet.

Weitere Konzentrationsmaße

Konzentrationskoeffizient

$$CR_g = \frac{\sum\limits_{i=n-g+1}^{n} x_{(i)}}{\sum\limits_{i=1}^{n} x_{(i)}} \quad \text{für } g = (1, 2,)3, \ldots, n$$

CR_g gibt an, welchen Anteil der Merkmalssumme die g letzten Merkmalswerte der geordneten statistischen Reihe in sich vereinen. Die Vorgehensweise entspricht der Konstruktion der Lorenzkurve, wobei die geordnete Urliste von rechts nach links, also in umgekehrter Reihenfolge abgearbeitet wird.

Die zugehörige Kurve wird üblicherweise als **Paretokurve** bezeichnet.

Weitere Konzentrationsmaße

Konzentrationskoeffizient

$$CR_g = \frac{\sum\limits_{i=n-g+1}^{n} x_{(i)}}{\sum\limits_{i=1}^{n} x_{(i)}} \quad \text{für } g = (1, 2,)3, \ldots, n$$

CR_g gibt an, welchen Anteil der Merkmalssumme die g letzten Merkmalswerte der geordneten statistischen Reihe in sich vereinen. Die Vorgehensweise entspricht der Konstruktion der Lorenzkurve, wobei die geordnete Urliste von rechts nach links, also in umgekehrter Reihenfolge abgearbeitet wird.

Die zugehörige Kurve wird üblicherweise als **Paretokurve** bezeichnet.

Weitere Konzentrationsmaße

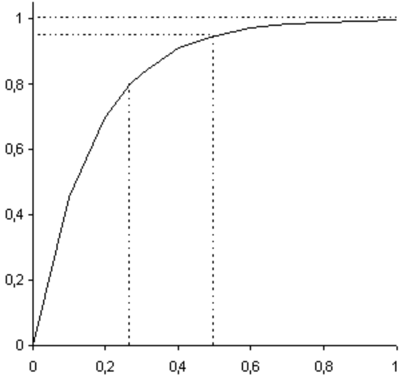


Abbildung: Paretokurve zu Beispiel 7.8

Weitere Konzentrationsmaße

Herfindahl-Index

$$H := \sum_{i=1}^{n} \left(\frac{x_i}{\sum\limits_{i=1}^{n} x_i} \right)^2$$

H ist die Summe der quadrierten individuellen Anteile an der Merkmalssumme. Je größer H ist, desto größer ist die Konzentration.