Normalverteilung

Für $x \in \mathbb{R}$ heißt

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

Normalverteilungsdichte mit Mittelwert μ und Standardabweichung σ

Für $\mu = 0$ und $\sigma = 1$ erhät man

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}x^2\right)$$

Quantile I

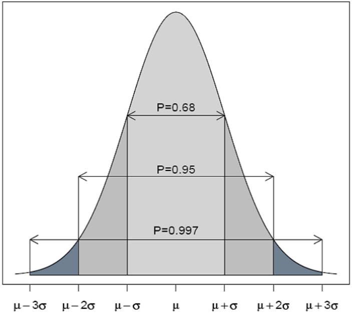
Quantile der Standardnormalverteilung:

p	50%	75%	90%	95%	97.5%	99%
z_p	0.0 (Median)	0.67	1.28	1.64	1.96	2.33

Quantile II

68-95-99.7-Prozent-Regel:

68%	der Beob. liegen im Interv.	$\mu \pm \sigma$
95%	der Beob. liegen im Interv.	$\mu \pm 2\sigma$
99.7%	der Beob. liegen im Interv.	$\mu \pm 3\sigma$

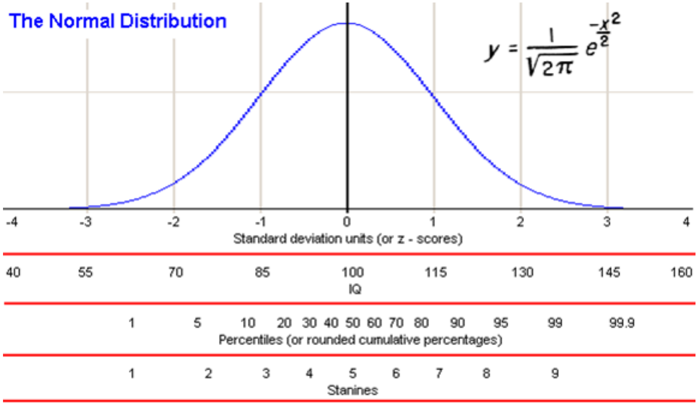


Normalverteilung in der Psychologie

Skalen werden so konstruiert, dass die Verteilung in der Population einer Normalverteilung genügt.

IQ :	Mittelwert = 100, Standardabweichung 15
T- Werte:	Mittelwert 50, Standardabweichung 10
Sta(ndard)nine	Mittelwert 5 Standardabweichung 2

Normalverteilung in der Psychologie II



Strategie zur Skalenbildung

- Ziehe grosse Stichprobe aus der Population
- Ordne Ergebnisse
- Zuordnung von Stanine

Schema

4%	7%	12%	17%	20%	17%	12%	7%	4%
1	2	3	4	5	6	7	8	9

Strategie zur Skalenbildung II

- Ziehe grosse Stichprobe aus der Population
- Standardisierung (Z- Werte)

$$z_i = (x_i - \bar{x})/s_x$$

- Reskalieren durch Multiplikation mit gewünschter Standardabweichung und Addition des gewünschten Mittelwertes

Normalverteilung in der technischen Statistik

- Größen in der Produktion (Längen etc.)
- Messfehler
- Größen nach geeigneter Transformation
- 6 Sigma

Überprüfung der Annahme der Normalverteilung

Fragestellung: Passen die Daten zu einer Normalverteilung?

a) Vergleiche Histogramm, Dichteschätzer mit NV-Dichte ($\mu = \bar{x}, \sigma = S$)

Überprüfung der Annahme der Normalverteilung

Fragestellung: Passen die Daten zu einer Normalverteilung?

a) Vergleiche Histogramm, Dichteschätzer mit NV-Dichte ($\mu = \bar{x}, \sigma = S$)

b) Vergleiche Verteilungsfunktion, d.h. empirische Verteilungsfunktion F mit $\Phi(t) = \int\limits_{-\infty}^{t} f(\mu, \sigma, t)dt$

Überprüfung der Annahme der Normalverteilung

Fragestellung: Passen die Daten zu einer Normalverteilung?

a) Vergleiche Histogramm, Dichteschätzer mit NV-Dichte ($\mu = \bar{x}, \sigma = S$)

b) Vergleiche Verteilungsfunktion, d.h. empirische Verteilungsfunktion F mit $\Phi(t) = \int\limits_{-\infty}^{t} f(\mu, \sigma, t)dt$

c) Schiefe $= 0$?

d) Q-Q-Plot

Normal-Quantil-Plot

Sei $x_{(1)}, \ldots, x_{(n)}$ die geordneten Daten. Für $i = 1, \ldots, n$ werden die $(i - 0.5)/n$-Quantile $z_{(i)}$ der Standardnormalverteilung berechnet.

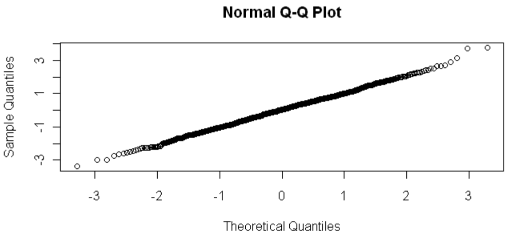
Der Normal-Quantil-Plot (Normal-Q-Q-Plot) besteht aus den Punkten

$$(z_{(1)}, x_{(1)}), \ldots, (z_{(n)}, x_{(n)})$$

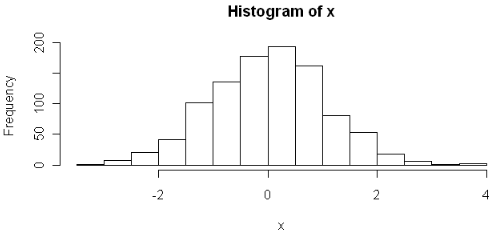
im $z - x-$Koordinatensystem.

Liegen die Punkte auf einer Geraden, so passt die Normalverteilung gut zu den Daten

Beispiele I

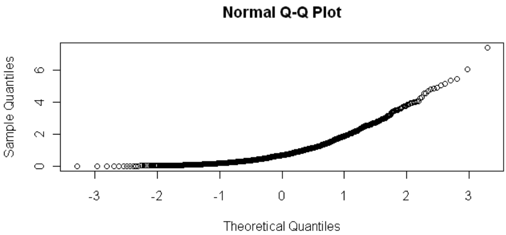


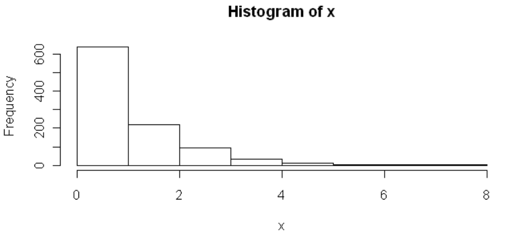
Normal Q-Q Plot
Sample Quantiles
-3
-1
1
3
-3
-2
-1
0
1
2
3
Theoretical Quantiles



Histogram of x
Frequency
0
50
100
200
-2
0
2
4
x

Beispiele II





Konzentrationsmaße

Motivation

Existiert eine Menge, die auf viele Individuen verteilt ist, kann es hilfreich sein zu wissen, wie diese Menge verteilt ist.

Beispiele

- Vermögensverteilung in einem Staat
- Marktanteile von Firmen in einem Segment

Lorenzkurve

Grundidee

Es sollen folgende Aussagen grafisch dargestellt werden:

- Die „Ärmsten“ x% besitzen einen Anteil von y%.
- Die „Reichsten“ x% besitzen einen Anteil von y%.

Lorenzkurve

Definition Lorenzkurve

- Das Merkmal darf nur *positive* Ausprägungen annehmen
- Die Gesamtsumme aller Merkmalswerte ist $\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} x_{(j)}$
- Die Lorenzkurve verbindet Punktepaare bestehend aus den *kumulierten Summen* der nach Größe geordneten Beobachtungswerte $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$ und dem *relativen Anteil* der Individuen, die diese kumulierte Summe besitzen.

Lorenzkurve

Gestaltung

- Es wird festgelegt: $u_{(0)} = 0$ und $v_{(0)} = 0$
- Die x-Achse wird in *gleiche Längen* aufgeteilt, deren Anzahl der der Individuen (Merkmalsausprägungen) entspricht:

$$u_i = \frac{i}{n}, \quad i = 1, \ldots, n$$

- Die y- Werte werden wie folgt berechnet:

$$v_i = \frac{\sum_{j=1}^{i} x_{(j)}}{\sum_{j=1}^{n} x_{(j)}}, \quad i = 1, \ldots, n,$$

  also dem Quotienten aus der kumulierten Summe und der Gesamtsumme.
- Die so errechneten Koordinatenpunkte (u_i, v_i) werden in den Graphen eingetragen und mit Geraden verbunden.

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20		
2	20		
3	20		
4	20		
5	20		

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	$\frac{1}{5}$	$\frac{20}{100}$
2	20		
3	20		
4	20		
5	20		

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	0,2	0,2
2	20		
3	20		
4	20		
5	20		

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	0,2	0,2
2	20	$\frac{2}{5}$	$\frac{40}{100}$
3	20		
4	20		
5	20		

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	0,2	0,2
2	20	0,4	0,4
3	20		
4	20		
5	20		

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	0,2	0,2
2	20	0,4	0,4
3	20	$\frac{3}{5}$	$\frac{60}{100}$
4	20		
5	20		

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	0,2	0,2
2	20	0,4	0,4
3	20	0,6	0,6
4	20		
5	20		

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	0,2	0,2
2	20	0,4	0,4
3	20	0,6	0,6
4	20	$\frac{4}{5}$	$\frac{80}{100}$
5	20		

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	0,2	0,2
2	20	0,4	0,4
3	20	0,6	0,6
4	20	0,8	0,8
5	20		

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	0,2	0,2
2	20	0,4	0,4
3	20	0,6	0,6
4	20	0,8	0,8
5	20	$\frac{5}{5}$	$\frac{100}{100}$

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

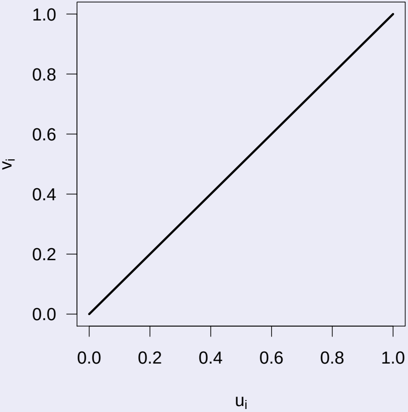
i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	0,2	0,2
2	20	0,4	0,4
3	20	0,6	0,6
4	20	0,8	0,8
5	20	1	1

Lorenzkurve

Beispiel

5 Bauern teilen sich eine Ackerfläche von 100 ha zu je 20 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	20	0,2	0,2
2	20	0,4	0,4
3	20	0,6	0,6
4	20	0,8	0,8
5	20	1	1



Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0		
2	0		
3	0		
4	0		
5	100		

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	$\frac{1}{5}$	$\frac{0}{100}$
2	0		
3	0		
4	0		
5	100		

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	0,2	0
2	0		
3	0		
4	0		
5	100		

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	0,2	0
2	0	$\frac{2}{5}$	$\frac{0}{100}$
3	0		
4	0		
5	100		

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	0,2	0
2	0	0,4	0
3	0		
4	0		
5	100		

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	0,2	0
2	0	0,4	0
3	0	$\frac{3}{5}$	$\frac{0}{100}$
4	0		
5	100		

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	0,2	0
2	0	0,4	0
3	0	0,6	0
4	0		
5	100		

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	0,2	0
2	0	0,4	0
3	0	0,6	0
4	0	$\frac{4}{5}$	$\frac{0}{100}$
5	100		

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	0,2	0
2	0	0,4	0
3	0	0,6	0
4	0	0,8	0
5	100		

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	0,2	0
2	0	0,4	0
3	0	0,6	0
4	0	0,8	0
5	100	$\frac{5}{5}$	$\frac{100}{100}$

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

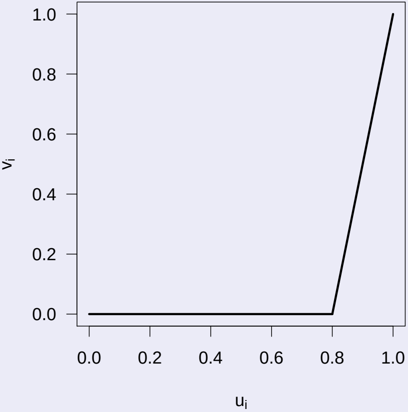
i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	0,2	0
2	0	0,4	0
3	0	0,6	0
4	0	0,8	0
5	100	1	1

Lorenzkurve

Beispiel eines Monopols

Von 5 Bauern besitzt einer die gesamten 100 ha.

i	$x_{(i)}$	u_i	v_i
0	-	0	0
1	0	0,2	0
2	0	0,4	0
3	0	0,6	0
4	0	0,8	0
5	100	1	1



Lorenzkurve

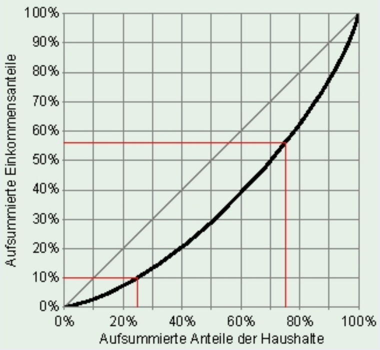
Erscheinungsbild von Lorenzkurven

- Die Kurve bildet auf einen quadratischen Graphen mit Kantenlänge 1 ab.
- Die Koordinate $(u_0; v_0)$ ist *immer* $(0; 0)$.
- Die Koordinate $(u_n; v_n)$ ist *immer* $(1; 1)$.
- Der konstruierte Polygonzug verläuft *immer unterhalb* (im Grenzfall auf) der Winkelhalbierenden.
- Der konstruierte Polygonzug ist *(streng) monoton steigend*.
- Die Steigung des nächsten Polygonsegments ist entweder *gleich groß* oder *größer* als die Steigung des letzten Polygonsegments.

Lorenzkurve

Beispiel

Bruttohaushaltseinkommen 2003 in der Schweiz



Es zeigt sich, dass das ärmste Viertel der Schweizer Bevölkerung nur 10%, das reichste Viertel jedoch über 40% des gesamten Bruttoeinkommens verdient.

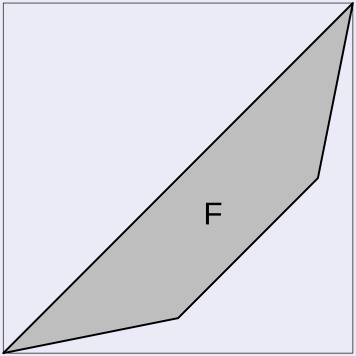
Gini-Koeffizient

Definition Gini-Koeffizient

Der Gini-Koeffizient bzw. das Lorenzsche Konzentrationsmaß ist eine Maßzahl, die das *Ausmaß* der Konzentration beschreibt. Er ist definiert als

$$G = 2 \cdot F,$$

wobei F die Fläche zwischen der Diagonalen und der Lorenzkurve ist.



Gini-Koeffizient

Berechnung des Gini-Koeffizienten

Für die praktische Berechnung von G aus den Wertepaaren $(u_i; v_i)$ stehen folgende alternative Formeln zur Verfügung:

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

oder alternativ

$$G = 1 - \frac{1}{n}\sum_{i=1}^{n}(v_{i-1} + v_i)$$

Wertebereich des Gini-Koeffizienten

$$0 \leq G \leq \frac{n-1}{n}$$

Gini-Koeffizient

Normierter Gini-Koeffizient G^+

Der Gini-Koeffizient wird auf folgende Weise normiert:

$$G^+ = \frac{n}{n-1} G$$

Er hat somit den Wertebereich

$$0 \leq G \leq 1,$$

wobei 0 für *keine Konzentration* (Gleichverteilung) und 1 für *vollständige Konzentration* (Monopol) steht.

Herfindahl-Index

Herfindahl-Index

$x_1, \ldots x_n$ seien die Daten mit $x_i \geq 0$.
Die Anteile der Einheiten i sind wie folgt definiert:

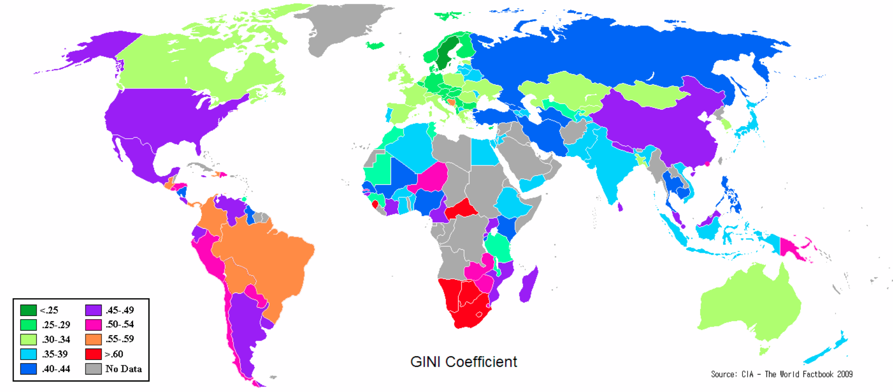
$$p_i := \frac{x_i}{\sum_{j=1}^{n} x_j}$$

Der Herfindahhl-Index ist

$$H_i := \sum_{j=1}^{n} p_i^2$$

Der Wertebereich ist von $1/n$ (Identische x) bis 1 (Monopol)

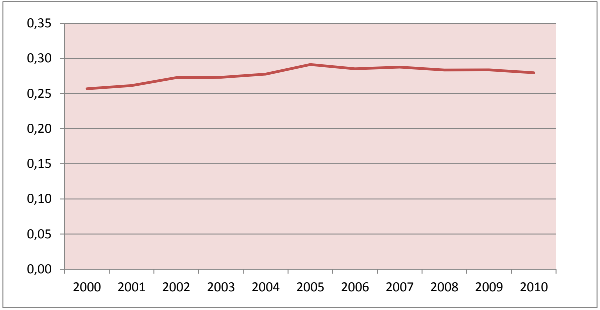
GINI Einkommen nach CIA report 2009



Bericht der Bundesregierung

\- X -

Einkommensverteilung (Gini-Koeffizient)



Quelle: Berechnungen des DIW Berlin auf Basis SOEP 2011.

Ein weiteres Verteilungsmaß ist der Gini-Koeffizient. Er beschreibt auf einer Skala von null bis eins die Ungleichheit der Verteilung. Je höher der Wert, umso ungleicher ist die Verteilung. Dieses Maß zeigt eine nach 2007 rückläufige Ungleichheit der Nettoäquivalenzeinkommen auf Haushaltsebene an. Dies umfasst alle Einkommensarten (insbesondere Einkommen aus Erwerb, Renten und Pensionen, aus Vermögen und Sozialtransfers). Der Trend einer Zunahme zwischen 2000 und 2005 hat sich also in der Zeit danach umgekehrt. Die Ungleichheit der Einkommen nimmt derzeit ab.

Bericht der Bundesregierung

- 325 -

personenhaushalt berücksichtigt. Die Verteilung der so ermittelten Nettoäquivalenzeinkommen hat sich, gemessen am Gini-Koeffizienten und den Anteilen der Dezile, nach den Daten der EVS zwischen 2003 und 2008 leicht weiter gespreizt.

Tabelle C I.1.2:
Verteilung der Nettoäquivalenzeinkommen 2003 und 2008

Jahr	Dezil										Gini-Koeffizient
	1	2	3	4	5	6	7	8	9	10	
	Anteile (%) am Volumen des Nettoäquivalenzeinkommens										
2003	3,9	5,5	6,5	7,5	8,4	9,4	10,5	12,0	14,3	22,0	0,267
2008	3,6	5,1	6,3	7,3	8,3	9,3	10,5	12,2	14,7	22,7	0,284

Quelle: EVS; Statistisches Bundesamt.

Während die unteren sechs Dezile gegenüber 2003 einen geringeren Anteil aufweisen, haben die obersten drei Dezile Zuwächse erfahren. Der Gini-Koeffizient stieg von 0,267 auf 0,284 und damit um rund sechs Prozent (Tabelle C I.1.2). Nach den Daten des SOEP zeigt dieses Maß eine nach 2007 rückläufige Ungleichheit der Nettoäquivalenzeinkommen auf Haushaltsebene an. Der Trend einer Zunahme zwischen 2000 und 2005 hat sich also in der Zeit danach umgekehrt. Die Ungleichheit der Einkommen nimmt derzeit ab (Schaubild C I.1.1).[338]

Schaubild C I.1.1:
Ungleichheit der Einkommensverteilung in Deutschland, 2000-2011 (Gini-Koeffizient)