

Zusammenhang zwischen Gini-Koeffizienten und dmft-Mittelwerten

Einbeziehung aller Kinder der Altersklasse 6 (± 1 Jahr)

Bei statistischen Konzentrationsanalysen geht es um die Frage, wie sich die Summe aller Merkmalswerte auf die betrachteten Merkmalsträger verteilt. Wie z.B. verteilt sich der Gesamtumsatz (Merkmalssumme) einer bestimmten Branche auf die beteiligten Unternehmen (Merkmalsträger) ?

In der Kariesstatistik ist die Frage interessant, wie die gesamte dmft-Summe der untersuchten Kinder auf die Kinder verteilt ist. Grafisch läßt sich das durch die Lorenzkurve beurteilen. Die Stärke der Konzentration wird durch den Gini-Koeffizienten repräsentiert.

Fragestellung:

Welcher prinzipielle Zusammenhang besteht zwischen den Gini-Koeffizienten und den dmft-Mittelwerten in der Altersklasse 6 (± 1 Jahr), wenn alle untersuchten Kinder in die Auswertung einbezogen werden ?

Verfügbare Daten:

<u>Quelle</u>	<u>Untersuchungsjahre</u>	<u>Altersklassen</u>	<u>Zahl der Kinder</u>
DAJ	2016, 2009, 2004, 2000, 1997	6 / 7	283.202 Erstklässler
Armfield JM J Public Health Dent (2009)	1977 bis 2002	5 / 6	1.510.044
Kariesdaten aus Südhessen	1960 bis 1982	6	32.418 aus Kindergärten
Becker F und Tempel G GBE Bremen 2016	2013 / 14	6 / 7	3.861 aus 75 Schulen

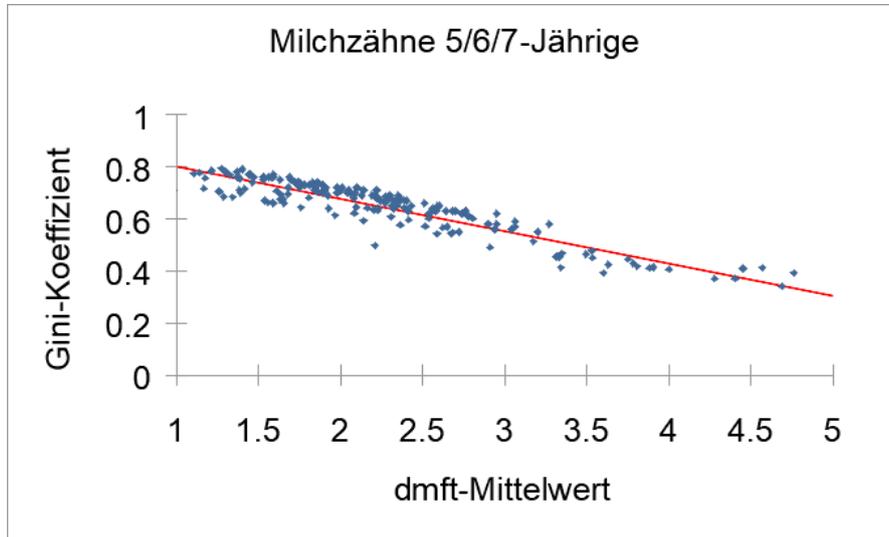
Zahnärztliche Untersuchungen an rund 1,83 Mio. Kindern wurden in Australien (Armfield) und Deutschland im Zeitraum von 1960 bis 2016 durchgeführt bei unterschiedlichen örtlichen Untersuchungsbedingungen, durch verschiedene Untersucher mit ihren diagnostischen Standards, unterschiedlichen Fallzahlen, veränderter Kariesprävalenz mit den Jahren, z.T. unbekannte Randomisierung und biologische Variationen. Daher soll primär die Art des Zusammenhangs dargestellt und zahlenmäßig nur grob geschätzt werden. Soweit nicht in den Quellen bereits angegeben wurden Gini-Koeffizienten und dmft-MW mit Hilfe von Tabellen (wie z.B. Tab. 1 im vorigen Beitrag) berechnet.

Ergebniss:

In der folgenden Grafik sind alle verfügbaren Werte zusammengefasst. Zwischen den Gini-Koeffizienten und den dmft-Mittelwerten besteht offensichtlich ein gegenläufiger linearer Zusammenhang.

Je größer der Gini-Koeffizient und damit die Ungleichheit in der Verteilung der Karies in der Untersuchten Gruppe, desto kleiner wird der dmft-MW dieser Gruppe.

Ein Hauptziel zahnmedizinischer Gruppenprophylaxe, die Reduktion der dmft-Mittelwerte, ist daher immer mit einer Zunahme der Konzentration (Polarisation) der Karies verbunden.



Unter bestimmten Voraussetzungen ist es mit Hilfe der linearen Regression möglich, aus den dmft-Mittelwerten entsprechende Gini-Koeffizienten zu schätzen. Im folgenden werden hierfür lediglich die DAJ-Daten [1] verwendet.

```
. regress gini dmft [aweight = n]
(sum of wgt is 283,202)
```

Source	SS	df	MS	Number of obs	=	76
Model	.145030628	1	.145030628	F(1, 74)	=	1006.89
Residual	.010658805	74	.000144038	Prob > F	=	0.0000
Total	.155689433	75	.002075859	R-squared	=	0.9315
				Adj R-squared	=	0.9306
				Root MSE	=	.012

	gini	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	dmft	-.1163525	.0036668	-31.73	0.000	-.1236588 - .1090463
	_cons	.9321508	.0076016	122.63	0.000	.9170043 .9472974

Konfidenzintervalle der Regressionskoeffizienten stehen im STATA-Ausdruck.

Regressionsgleichung: $gini = 0,93 - 0,12 \cdot dmft\text{-MW}$ $R^2 = 0,93$

Für einen dmft-MW von 3,0 z.B. schätzt man einen Gini-K. von 0,57. Eine Reduktion des dmft-MW von 3,0 auf 2,0 bewirkt eine Erhöhung der Konzentration von 0,57 auf 0,69 also um 0,12 Einheiten.

Prüfung der Modellvoraussetzungen für die Regression:

1. Mit $R^2 = 0,93$ kann man von einem starken linearen Zusammenhang ausgehen.

2. Die Normalverteilung der abhängigen Variablen *gini* wird mit dem Shapiro-Wilk Test geprüft:

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
gini	76	0.97199	1.844	1.337	0.09069

3. Die Residuen erhält man in STATA aus:

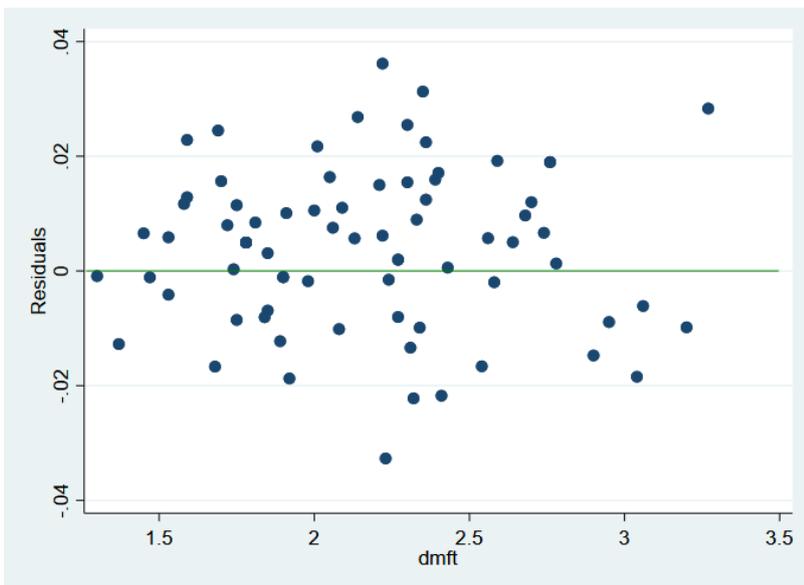
```
predict x, residuals
. swilk x
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
x	76	0.99447	0.364	-2.207	0.98634

und mit `swilk x` die Bestätigung ihrer Normalverteilung.

4. Die Varianzhomogenität (Homoskedastizität) lässt sich z.B. im Residuenplot erkennen.



Es ist kein Muster zu erkennen, das auf Abweichung von der Homoskedastizität hindeutet.

Damit sind die Modellvoraussetzungen erfüllt.

Bei Regressionsanalysen von zusammengeführten Daten aus verschiedenen Quellen (siehe Grafik auf Seite 2) kann es trotz eines starken linearen Zusammenhangs zu Verletzungen der Modellvoraussetzungen kommen.

[1] Epidemiologische Begleituntersuchungen zur Gruppenprophylaxe, DAJ, Bonn
<https://daj.de/gruppenprophylaxe/epidemiologische-studien/>