

Zusammenhang zwischen Gini-Koeffizienten und dmft-Mittelwerten

Einbeziehung aller Kinder der Altersgruppe 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. die gesamte dmft-Summe der untersuchten Kinder einer Gruppe (z.B. Altersgruppe, regionale Gruppe) auf die Kinder verteilt ist. Grafisch lässt sich das durch die Lorenzkurve beurteilen. Die Stärke der Konzentration wird durch den Gini-Koeffizienten repräsentiert (siehe obiger Beitrag).

Fragestellung für 6-Jährige:

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

Verfügbare Daten:

Quelle	Untersuchungsjahre	Altersgruppe	Zahl der Kinder
DAJ	2016, 2009, 2004, 2000, 1997	6 / 7	76 Gruppen mit gesamt 283.202 Erstklässlern
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

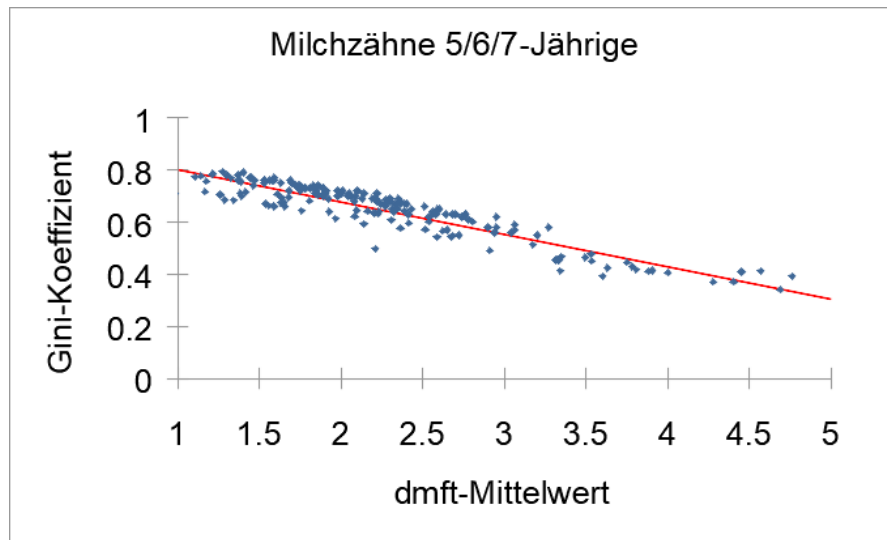
Zahnärztliche Untersuchungen an rund 1,83 Mio. Kindern wurden in Australien (Armfield) und Deutschland im Zeitraum von 1960 bis 2016 durchgeführt bei unterschiedlichen örtlichen Untersuchungsbedingungen, durch verschiedene Untersucher mit ihren diagnostischen Standards, unterschiedlichen Fallzahlen, 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 (siehe Anhang).

```
. 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: $\text{gini} = 0,93 - 0,12 \cdot \text{dmft-MW}$ $R^2 = 0,93$ (Bestimmtheitsmaß)

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:

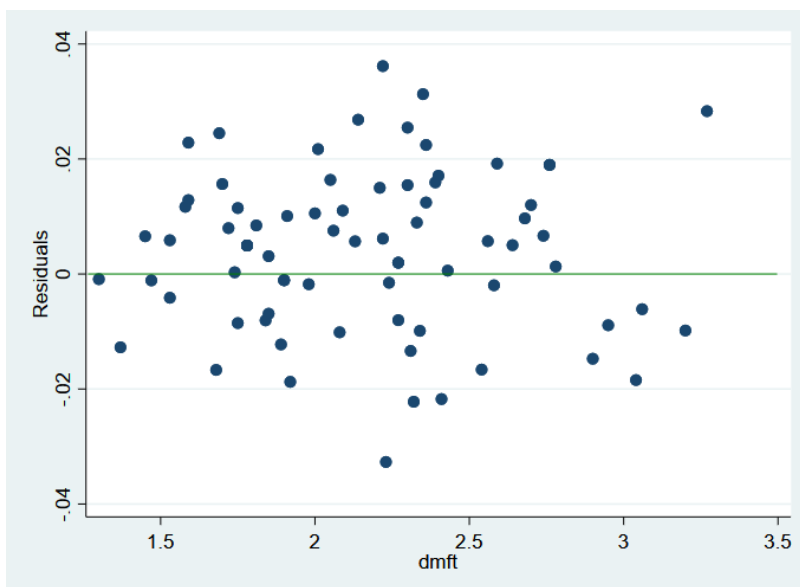
```
predict x, residuals
. swilk x
```

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/>

Anhang: Auszug aus dem DAJ Datensatz [1]

Bundesland	Quelle	Gini	dmft-MW	n
BW_6_7	DAJ_2016	0.72	1.85	2310
Bayern	DAJ_2016	0.76	1.37	2874
Berlin	DAJ_2016	0.69	2.13	21934
BRB	DAJ_2016	0.71	1.85	17227
HB	DAJ_2016	0.69	1.92	1888
HH	DAJ_2016	0.75	1.7	3458
Hessen	DAJ_2016	0.73	1.81	2324
MeckPom	DAJ_2016	0.64	2.23	2643
Nieders	DAJ_2016	0.73	1.78	2281
Nordrhein	DAJ_2016	0.77	1.59	5268
RP	DAJ_2016	0.76	1.53	2610
Saarl	DAJ_2016	0.75	1.53	1941
Sachsen	DAJ_2016	0.72	1.75	28989
Sachs-Anh	DAJ_2016	0.65	2.31	13965
Schl-Holst	DAJ_2016	0.76	1.47	24083
Thüring	DAJ_2016	0.68	2.08	14566
Westfalen	DAJ_2016	0.73	1.78	3194
Bayern	DAJ_2009	0.68	2.36	1914
Berlin	DAJ_2009	0.67	2.4	2249
BRB	DAJ_2009	0.69	2.21	1197
HB	DAJ_2009	0.67	2.39	566
HH	DAJ_2009	0.72	1.68	487
Hessen	DAJ_2009	0.73	1.74	2323
MeckPom	DAJ_2009	0.66	2.27	1799
Nieders	DAJ_2009	0.73	1.78	1473
Nordrhein	DAJ_2009	0.74	1.75	4526
RP	DAJ_2009	0.73	1.78	2096
Saarl	DAJ_2009	0.78	1.3	481
Sachsen	DAJ_2009	0.7	1.89	2705
Sachs-Anh	DAJ_2009	0.64	2.32	1029
Schl-Holst	DAJ_2009	0.77	1.45	1002
Thüring	DAJ_2009	0.64	2.56	762
Westfalen	DAJ_2009	0.71	1.9	4544
BW_6_7	DAJ_2004	0.76	1.58	1535
Bayern	DAJ_2004	0.69	2.35	1930
Berlin	DAJ_2004	0.62	2.74	1466
BRB	DAJ_2004	0.63	2.76	931
HB	DAJ_2004	0.63	2.76	970
HH	DAJ_2004	0.71	1.84	510
Hessen	DAJ_2004	0.7	2.06	1938
MeckPom	DAJ_2004	0.63	2.58	1323
Nieders	DAJ_2004	0.7	2.09	855
Nordrhein	DAJ_2004	0.71	2.05	4348
RP	DAJ_2004	0.72	2.01	2437
Saarl	DAJ_2004	0.71	2	497
Sachsen	DAJ_2004	0.65	2.34	24656
Sachs-Anh	DAJ_2004	0.58	2.9	1890
Schl-Holst	DAJ_2004	0.76	1.69	1145
Thüring	DAJ_2004	0.61	2.78	1330
Westfalen	DAJ_2004	0.67	2.27	4268
.....				