

Panelregression mit R - Teil 1

Benjamin Schlegel

09. Februar 2017

In diesem Artikel wird erklärt, wie man ein lineares Panelmodell in R schätzen kann. Dazu wird der Datensatz "eidgenössische Abstimmungen auf Gemeindeebene" verwendet, welcher unter [Data](#) heruntergeladen werden kann.

Zuerst wird der Datensatz eingelesen und in ein `data.table` Objekt umgewandelt. Anschliessend werden die Anzahl Abstimmungen jedes Abstimmungstermins eruiert [Ein Erklärung wie man in R mit `data.table` aggregieren kann, findet man im Blog [Aggregieren mit data.table in R](#)]. Nach dem aggregieren wird die neu gewonnene Information wieder in den ursprünglichen Datensatz überführt.

```
data = read.csv("gemeinden_abstimmungen.csv", stringsAsFactors = F)
library(data.table)
data = data.table(data)

datum.vorlagen = data[,.(anzahl.vorlagen=round(length(vorlage_nr)/2291)),by=Datum]

data = merge(data, datum.vorlagen, by="Datum")
```

Um pro Abstimmung nur eine Abstimmung zu haben und die Fallzahl zu reduzieren fürs Beispiel, wird für die aktuellsten 60 Abstimmungen im Datensatz jeweils das Thema genommen, wo der Autor davon ausgeht, dass es am meisten mobilisiert hat. [In einer wissenschaftlichen Arbeit müsste die Auswahl begründet werden. Um die Methode zu demonstrieren, soll diese Vorgehensweise genügen.]

```
abstimmungen = data[vorlage_nr %in% c(6010,5970,5940,5910,5880,5860,5840,
5800,5750,5720,5710,5680,5660,5650,5600,5550,5540,5521,5510,
5500,5470,5430,5420,5400,5350,5320,5300,5290,5280,5270,5250,
5220,5200,5190,5170,5160,5130,5080,5060,5010,4930,4910,4891,
4880,4850,4830,4770,4740,4700,4670,4640,4630,4550,4530,4520,
4480,4440,4400,4370,4340), ]
```

Zu guter letzt wird noch die Variable X aus dem Datensatz entfernt und die Variablen Stimmbeteiligung und ja.proz in numerische Variablen konvertiert. Nun sind die Daten bereit.

```
abstimmungen[,c("X", "ja.proz", "Stimmbeteiligung") :=
.(NULL, as.numeric(ja.proz), as.numeric(Stimmbeteiligung))]
```

Als erstes brauchen wir die Bibliothek `plm`. Anschliessend wird bei jedes Gemeinde noch der Kanton angehängt, damit der Name einzigartig ist.

```
library(plm)
abstimmungen[, Gemeinde := paste0(Gemeinde, "-", Kanton)]
```

Nun kann das erste Modell geschätzt werden. Zuerst wird ein pooling-Modell gerechnet, welches identisch ist mit einem normalen linearen OLS-Modell. Es wird jeweils ein die Stimmbeteiligung auf den Typ der Vorlage, das Thema der Vorlage und die Anzahl Vorlagen am Abstimmungstermin regressiert.

```
model.pooling = plm(Stimmbeteiligung ~ Typ + Thema + anzahl.vorlagen,
  data=abstimmungen,index = c("Gemeinde","Datum"), model="pooling")
summary(model.pooling)
```

Pooling Model

Call:

```
plm(formula = Stimmbeteiligung ~ Typ + Thema + anzahl.vorlagen,
  data = abstimmungen, model = "pooling", index = c("Gemeinde",
  "Datum"))
```

Unbalanced Panel: n=2282, T=34-60, N=136882

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-47.900	-7.140	-0.194	6.900	56.200

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	35.434598	0.134330	263.7881	< 2.2e-16 ***
TypObligatorisches Referendum	-4.128468	0.119249	-34.6205	< 2.2e-16 ***
TypVolksinitiative	4.672060	0.068902	67.8076	< 2.2e-16 ***
TypVolksinitiative mit Gegenvorschlag	7.731295	0.176170	43.8855	< 2.2e-16 ***
ThemaGrundlagen der Staatsordnung	1.108754	0.136665	8.1129	4.982e-16 ***
ThemaInfrastruktur und Lebensraum	-2.399483	0.153029	-15.6799	< 2.2e-16 ***
ThemaLandesverteidigung	4.709091	0.162516	28.9762	< 2.2e-16 ***
ThemaÖffentliche Finanzen	0.623276	0.195983	3.1803	0.001472 **
ThemaSchweizerische Aussenpolitik	12.713859	0.140322	90.6048	< 2.2e-16 ***
ThemaSozialpolitik	5.725360	0.122414	46.7705	< 2.2e-16 ***
ThemaWirtschaft	-1.762672	0.165779	-10.6326	< 2.2e-16 ***
anzahl.vorlagen	1.053013	0.024423	43.1161	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 19345000

Residual Sum of Squares: 15470000

R-Squared: 0.20031

Adj. R-Squared: 0.20025

F-statistic: 3116.78 on 11 and 136870 DF, p-value: < 2.22e-16

Als nächstes wird ein fixed Effects Panelmodell geschätzt. Zuerst das gängigere Within-Modell, anschliessend das Between-Modell. Das Modell hat die Annahme, dass es einzigartige Merkmale von Individuen gibt, welche nicht über die Zeit variieren und nicht das Resultat von zufälliger Variation sind. Das Within-Modell schaut sich ein Individuum an, das Between-Modell den Unterschied zwischen Individuen.

```
model.within = plm(Stimmbeteiligung ~ Typ + Thema + anzahl.vorlagen,
  data=abstimmungen,index = c("Gemeinde","Datum"), model="within")
```

```
summary(model.within)
```

Oneway (individual) effect Within Model

Call:

```
plm(formula = Stimmeteiligung ~ Typ + Thema + anzahl.vorlagen,
     data = abstimmungen, model = "within", index = c("Gemeinde",
     "Datum"))
```

Unbalanced Panel: n=2282, T=34-60, N=136882

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-45.8000	-5.3200	0.0818	5.3100	47.8000

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)	
TypObligatorisches Referendum	-4.127813	0.098866	-41.7516	< 2.2e-16	***
TypVolksinitiative	4.671298	0.057124	81.7741	< 2.2e-16	***
TypVolksinitiative mit Gegenvorschlag	7.730434	0.146057	52.9276	< 2.2e-16	***
ThemaGrundlagen der Staatsordnung	1.106975	0.113305	9.7699	< 2.2e-16	***
ThemaInfrastruktur und Lebensraum	-2.400810	0.126872	-18.9231	< 2.2e-16	***
ThemaLandesverteidigung	4.707011	0.134737	34.9348	< 2.2e-16	***
ThemaÖffentliche Finanzen	0.619536	0.162484	3.8129	0.0001374	***
ThemaSchweizerische Aussenpolitik	12.713767	0.116337	109.2842	< 2.2e-16	***
ThemaSozialpolitik	5.724023	0.101490	56.4001	< 2.2e-16	***
ThemaWirtschaft	-1.764134	0.137443	-12.8354	< 2.2e-16	***
anzahl.vorlagen	1.053283	0.020248	52.0187	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 14331000

Residual Sum of Squares: 10456000

R-Squared: 0.2704

Adj. R-Squared: 0.25797

F-statistic: 4534.51 on 11 and 134589 DF, p-value: < 2.22e-16

```
model.between = plm(Stimmeteiligung ~ Typ + Thema + anzahl.vorlagen,
  data=abstimmungen,index = c("Gemeinde","Datum"), model="between")
summary(model.between)
```

Oneway (individual) effect Between Model

Call:

```
plm(formula = Stimmeteiligung ~ Typ + Thema + anzahl.vorlagen,
     data = abstimmungen, model = "between", index = c("Gemeinde",
     "Datum"))
```

Unbalanced Panel: n=2282, T=34-60, N=136882

Observations used in estimation: 2282

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-18.000	-3.970	-0.353	3.470	29.100

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	247.45	1079.43	0.2292	0.8187
TypObligatorisches Referendum	-908.07	3486.69	-0.2604	0.7945
TypVolksinitiative	-399.30	1817.75	-0.2197	0.8261
TypVolksinitiative mit Gegenvorschlag	2787.57	6440.80	0.4328	0.6652

Total Sum of Squares: 83612

Residual Sum of Squares: 83481

R-Squared: 0.0015628

Adj. R-Squared: 0.00024793

F-statistic: 1.18856 on 3 and 2278 DF, p-value: 0.3126

Wie klar erkennbar ist, behält das Between-Modell nur den Typ der Vorlage. Das liegt daran, dass beim Between-Modell die Gemeinden betrachtet werden. Zwischen den einzelnen Gemeinden ist jedoch der Typ, das Thema und die Anzahl Vorlagen immer gleich, wodurch perfekte Kollinearität entsteht. Deshalb wird nur eine dieser Variablen behalten. Auch ist hier nichts signifikant. Hier macht das Between-Modell wenig Sinn.

Zu guter Letzt wird noch das Random-Modell geschätzt. Es hat die Annahme, dass es einzigartige Attribute von Individuen gibt, welche nicht über die Zeit variieren und das Resultat von zufälliger Variation sind.

```
model.random = plm(Stimmbeteiligung ~ Typ + Thema + anzahl.vorlagen,
  data=abstimmungen, index = c("Gemeinde", "Datum"), model="random")
summary(model.random)
```

Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)

Call:

```
plm(formula = Stimmbeteiligung ~ Typ + Thema + anzahl.vorlagen,
  data = abstimmungen, model = "random", index = c("Gemeinde",
  "Datum"))
```

Unbalanced Panel: n=2282, T=34-60, N=136882

Effects:

	var	std.dev	share
idiosyncratic	77.689	8.814	0.687
individual	35.342	5.945	0.313

theta :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.7536	0.8120	0.8120	0.8120	0.8120	0.8120

Residuals :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-44.500	-5.490	0.087	0.000	5.460	47.100

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)	
(Intercept)	35.433874	0.167005	212.1732	< 2.2e-16	***
TypObligatorisches Referendum	-4.127861	0.098866	-41.7519	< 2.2e-16	***
TypVolksinitiative	4.671345	0.057125	81.7746	< 2.2e-16	***
TypVolksinitiative mit Gegenvorschlag	7.730481	0.146057	52.9277	< 2.2e-16	***
ThemaGrundlagen der Staatsordnung	1.107066	0.113305	9.7706	< 2.2e-16	***
ThemaInfrastruktur und Lebensraum	-2.400751	0.126872	-18.9226	< 2.2e-16	***
ThemaLandesverteidigung	4.707117	0.134738	34.9354	< 2.2e-16	***
ThemaÖffentliche Finanzen	0.619743	0.162484	3.8142	0.0001367	***
ThemaSchweizerische Aussenpolitik	12.713762	0.116337	109.2837	< 2.2e-16	***
ThemaSozialpolitik	5.724094	0.101490	56.4006	< 2.2e-16	***
ThemaWirtschaft	-1.764063	0.137443	-12.8349	< 2.2e-16	***
anzahl.vorlagen	1.053267	0.020248	52.0177	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 14508000

Residual Sum of Squares: 10633000

R-Squared: 0.26708

Adj. R-Squared: 0.26703

F-statistic: 4534.29 on 11 and 136870 DF, p-value: < 2.22e-16