Methoden-Workshop
Leading House „Economics of Education“/SKBF
Beispiele zu Cluster Bootstrap

Ben Jann

4. Februar 2016

Inhaltsverzeichnis

1 Daten generieren

10 Klumpen mit je 50 Beobachtungen; Intraklassenkorrelation im Fehlerterm wie auch in `x`.

```
. set seed 4530492
. set obs 10
number of observations (_N) was 0, now 10
. gen id = _n
. generate x = rnormal()
. generate e = rnormal()
. expand 50
(490 observations created)
. replace x = x + rnormal()
(500 real changes made)
. replace e = e + rnormal()
(500 real changes made)
. generate y = 0 + 1 * x + e
```

2 Ergebnisse ohne Berücksichtigung der Klumpenstruktur

```
. regress y x
      Source |       SS           df       MS      Number of obs   =       500
```

```
------------------------------------------------------------------------------
      Model |  1593.33814         1  1593.33814   F(1, 498)       =    813.60
   Residual |  975.273995       498  1.95838152   Prob > F        =    0.0000
------------+----------------------------------   R-squared       =    0.6203
      Total |  2568.61214       499  5.14751931   Adj R-squared   =    0.6195
                                                  Root MSE        =    1.3994

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.236349   .0433447    28.52   0.000     1.151188     1.32151
       _cons |   .1854082   .0626821     2.96   0.003     .0622541    .3085622
------------------------------------------------------------------------------
```

3 Cluster-robuste Standardfehler

```
. regress y x, cluster(id)

Linear regression                               Number of obs     =        500
                                                F(1, 9)           =      48.28
                                                Prob > F          =     0.0001
                                                R-squared         =     0.6203
                                                Root MSE          =     1.3994

                                   (Std. Err. adjusted for 10 clusters in id)
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.236349   .1779247     6.95   0.000     .8338557    1.638843
       _cons |   .1854082   .3157479     0.59   0.572    -.5288633    .8996796
------------------------------------------------------------------------------
```

4 Naiver Custer-Bootstrap

```
. regress y x, vce(boot, reps(500) nodots cluster(id))

Linear regression                               Number of obs     =        500
                                                Replications      =        500
                                                Wald chi2(1)      =      58.01
                                                Prob > chi2       =     0.0000
                                                R-squared         =     0.6203
                                                Adj R-squared     =     0.6195
                                                Root MSE          =     1.3994

                                  (Replications based on 10 clusters in id)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.236349   .1623279     7.62   0.000     .9181925    1.554506
       _cons |   .1854082   .2937799     0.63   0.528    -.3903899    .7612063
------------------------------------------------------------------------------
```

5 Pairs cluster bootstrap-t

Das Programm `clusterbs` kann mit dem Befehl `ssc install clusterbs` installiert werden.

```
. clusterbs regress y x, cluster(id) reps(500)
Cluster variable is id
Number of clusters for model is 10
```

```
Starting 500 Bootstrap Replications
..................................................................................
> ..................................................................................
> ..................................................................................
> ..................................................................................
> ..................................................................................
> ........
Bootstrap iterations completed.  Now storing model results...
The model ran succesfully and stored results in 500 bootstrap iterations.
Model Results
------------------------------------------------------------
              Coefficient Prob>|t|   95%_CI_low  95%_CI_high
------------------------------------------------------------
x               1.2363494      0.0     .6411953    1.8315035
_cons           .18540816      0.6    -.7513213    1.1221376
------------------------------------------------------------
The t-statistics and 95% confidence intervals are generated from the pairs
cluster bootstrap-t procedure and are robust to clustering with a small number
of sampling units.  Please note that the accuracy of the t-statistics and CIs
is conditional on the number of bootstrap replications that were used to
calculate the distribution of t. For p < .05 significance tests, specify reps(500)
 or more. For p < .01 level, specify reps(1000) or more. For p < .001 level, specify
reps(5000) or more.  More iterations will also yield more accurate confidence
intervals.  Post-estimation procedures should not be run on this model.
```

6 Wild cluster bootstrap-t

Installation: `cgmreg.ado` und `cgmreg.hlp` sowie `cgmwildboot.ado` und `cgmwildboot.hlp` von `https://sites.google.com/site/judsoncaskey/data` herunterladen und im Arbeitsverzeichnis speichern. Zudem mit `ssc install unique` das Programm `unique` installieren.

```
. cgmwildboot y x, cluster(id) bootcluster(id) reps(500) null(1)
Bootstrap reps (500)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500
.
Regress with clustered SEs/Wild bootstrap (500 successful resamples)
Number of clustvars=    1                        Number of obs =      500
Num combinations   =    1                        R-squared     =   0.6203
                                                 Adj R-squared =   0.6195
                                                 G(id)         =       10
                                                 (Bootstrapped)
------------------------------------------------------------------------------
           y |      Coef.        Null     p-value    [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.2363494           1         .32   .66877449    1.358281
        cons |   .18540816           .        .588  -.45403808   .83229738
------------------------------------------------------------------------------
```

Option `null(1)` führt dazu, dass für den Effekt von `x` der p-Wert für die Nullhypothese $\beta = 1$ berechnet wird.