Einführung II

Kontext und Mehr-Ebenen-Datensätze

Warum?

- Moderne Erklärungen = Mehr-Ebenen-Erklärungen
- Viele Datensätze sind strukturiert (oder können so konstruiert werden)
- Statistische Mehr-Ebenen-Modelle können korrekte Koeffizienten/Standardfehler für Kontextwirkungen schätzen
- (Bilden nicht den komplexen Mechanismus ab)

Was?

1. "random intercept"
2. "contextual effect"
3. "random slope"
4. "cross-level interaction"
5. "frog pond effects"
6. ... ?

Wie?

- Einfache Regressionsmodelle vs
- Echte (statistische) Mehr-Ebenen-Modelle
- Letztere bauen auf den Regressionsmodellen auf, die Sie kennen

"Pooling"

- (Beispiel: Politisches Vertrauen in 27 ESS-Ländern)
- Separate Analysen:
  - 27 Konstanten, 27 Sets von Koeffizienten
  - Keinerlei "pooling" der Daten - ineffizient

"Pooling"

- (Beispiel: Politisches Vertrauen in 27 ESS-Ländern)
- Separate Analysen:
  - 27 Konstanten, 27 Sets von Koeffizienten
  - Keinerlei "pooling" der Daten - ineffizient
- Eine gemeinsame Regressionsanalyse
  - 1 Konstante, 1 Set von Koeffizienten
  - Vollständiges "pooling" der Daten - bias, wenn sich Konstanten unterscheiden (Simpson's Paradox)

"Pooling"

- (Beispiel: Politisches Vertrauen in 27 ESS-Ländern)
- Separate Analysen:
  - 27 Konstanten, 27 Sets von Koeffizienten
  - Keinerlei "pooling" der Daten - ineffizient
- Eine gemeinsame Regressionsanalyse
  - 1 Konstante, 1 Set von Koeffizienten
  - Vollständiges "pooling" der Daten - bias, wenn sich Konstanten unterscheiden (Simpson's Paradox)
- Statistische Mehr-Ebenen-Modelle
  - Optimaler bzw. flexibler Kompromiß bei Kombination
  - "borrowing strength" - vor allem interessant, wenn wenige Personen pro Kontext und/oder große Unterschied in Personenzahl

Simpson's Paradox: künstliche Daten

```
gen c= 1 in 1/4
(52,344 missing values generated)
replace c=2 in 5/8
(4 real changes made)
gen x=.
(52,348 missing values generated)
replace x = _n in 1/8
(8 real changes made)
gen y = 10 +2*x  if c==1
(52,344 missing values generated)
replace y= -8 +2*x if c==2
(4 real changes made)
list c x y in 1/8, sep(4)

     +------------+
     | c   x    y |
     |------------|
  1. | 1   1   12 |
  2. | 1   2   14 |
  3. | 1   3   16 |
  4. | 1   4   18 |
     |------------|
  5. | 2   5    2 |
  6. | 2   6    4 |
  7. | 2   7    6 |
  8. | 2   8    8 |
     +------------+
```

Separate Regressionen

```
reg y x if c==1

      Source |       SS           df       MS      Number of obs   =         4
-------------+----------------------------------   F(1, 2)         =         .
       Model |         20         1          20   Prob > F        =         .
    Residual |          0         2           0   R-squared       =    1.0000
-------------+----------------------------------   Adj R-squared   =    1.0000
       Total |         20         3  6.66666667   Root MSE        =         0

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |          2          .        .       .            .           .
       _cons |         10          .        .       .            .           .
------------------------------------------------------------------------------

reg y x if c==2

      Source |       SS           df       MS      Number of obs   =         4
-------------+----------------------------------   F(1, 2)         =         .
       Model |         20         1          20   Prob > F        =         .
    Residual |          0         2           0   R-squared       =    1.0000
-------------+----------------------------------   Adj R-squared   =    1.0000
       Total |         20         3  6.66666667   Root MSE        =         0

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |          2          .        .       .            .           .
       _cons |         -8          .        .       .            .           .
------------------------------------------------------------------------------
```

Gemeinsame Regressionen

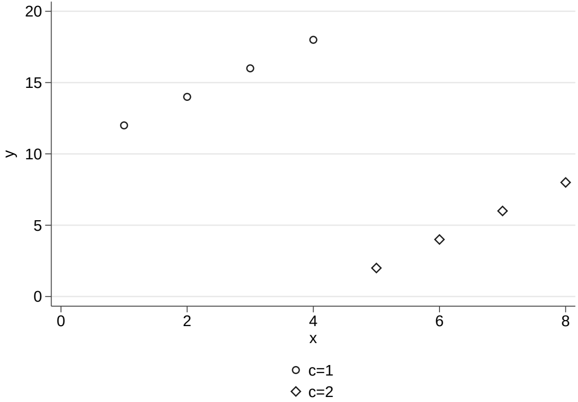
```
reg y x

      Source |       SS           df       MS      Number of obs   =         8
-------------+----------------------------------   F(1, 6)         =      3.33
       Model |  85.7142857         1  85.7142857   Prob > F        =    0.1177
    Residual |  154.285714         6  25.7142857   R-squared       =    0.3571
-------------+----------------------------------   Adj R-squared   =    0.2500
       Total |         240         7  34.2857143   Root MSE        =    5.0709

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  -1.428571   .7824608    -1.83   0.118    -3.343184    .4860412
       _cons |   16.42857   3.951233     4.16   0.006     6.760252    26.09689
------------------------------------------------------------------------------
```

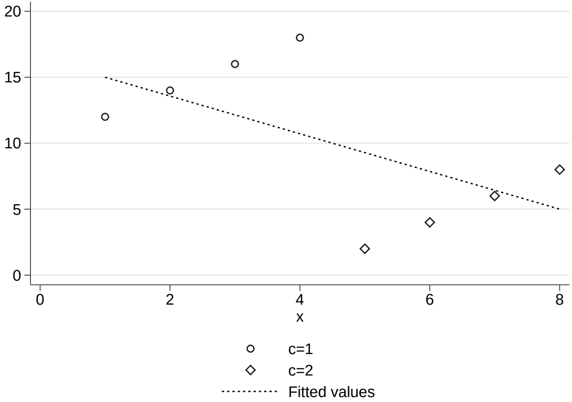
Plot I

```
graph twoway (scatter y x if c==1) (scatter y x if c==2)  , legend(label(1 "c=1") label(2 "c=2"))
graph export plot1.pdf , replace
```



Plot II

```
graph twoway (scatter y x if c==1) (scatter y x if c==2) (lfit y x) , legend(label(1 "c=1") label(2 "c=2"))
graph export plot2.pdf , replace
```



Was geht noch?

- Mehr als zwei Ebenen
- Cross-classification → mehr dazu nächste Woche

Wie schätzt man das?

- Lineare Regressionsmodelle - OLS
- Nicht-lineare Modelle - Maximum Likelihood (Forschungsmethoden VL)
- Für *alle* Mehr-Ebenen-Modelle: Maximum Likelihood oder vergleichbare/verwandte Verfahren
  - Numerisch extrem aufwendig
  - Nicht-lineare Mehr-Ebenen-Modelle möglich (logit, count etc.)

Stata I

- Früher spezielle Software (MLWin, HLM) oder Stata-Addons notwendig
- Heute gängige Modelle mit Stata schätzbar
  - Dokumentation
  - Geschwindigkeit + Stabilität
  - Transaktionskosten + Werkzeuge
- (xt)mixed

Stata II

mixed	lineare Regression
me(qr)logit	binäres Logit-Modell
meologit	ordinales Logit-Modell
me(qr)poisson, menbreg	Count-Modelle
gsem	(u.a.) multinomiales Logit-Modell

Trust

- Interpersonelles Vertrauen (Sozialkapital) $\rightarrow$ *Institutionenvertrauen* (Easton)
- 27 ESS-Länder (Welle 5)

Einfaches Modell, komplettes pooling, keine cluster

```
set more off
use ml-exercise-ess-5 , replace
* komplettes pooling
reg poltrust ppltrst

      Source |       SS           df       MS      Number of obs   =    49,780
-------------+----------------------------------   F(1, 49778)     =   7675.01
       Model |  34520.391         1   34520.391   Prob > F        =    0.0000
    Residual |  223889.722    49,778  4.49776452   R-squared       =    0.1336
-------------+----------------------------------   Adj R-squared   =    0.1336
       Total |  258410.113    49,779  5.19114714   Root MSE        =    2.1208

------------------------------------------------------------------------------
    poltrust |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ppltrst |   .3396662   .0038772    87.61   0.000     .3320669    .3472654
       _cons |   1.719947   .0211187    81.44   0.000     1.678554     1.76134
------------------------------------------------------------------------------
* complete cases markieren, damit empty models vergleichbar sind
gen sample = e(sample)
```

Einfaches Modell, komplettes pooling, aber cluster

```
use ml-exercise-ess-5 , replace
* komplettes pooling
reg poltrust ppltrst, cluster(cntry)

Linear regression                               Number of obs     =     49,780
                                                F(1, 26)          =     121.45
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1336
                                                Root MSE          =     2.1208

                                 (Std. Err. adjusted for 27 clusters in cntry)
------------------------------------------------------------------------------
             |               Robust
    poltrust |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ppltrst |   .3396662   .0308209    11.02   0.000     .2763128    .4030195
       _cons |   1.719947   .1623116    10.60   0.000     1.386311    2.053583
------------------------------------------------------------------------------
est store cluster
```

Einfaches Modell, kein pooling

```
* kein pooling
bysort cntry: reg poltrust ppltrst

-------------------------------------------------------------------------------------------------
-> cntry = BE

      Source |       SS           df       MS      Number of obs   =     1,670
-------------+----------------------------------   F(1, 1668)      =    213.62
       Model |  734.921522         1  734.921522   Prob > F        =    0.0000
    Residual |  5738.47044     1,668     3.44033   R-squared       =    0.1135
-------------+----------------------------------   Adj R-squared   =    0.1130
       Total |  6473.39196     1,669  3.87860513   Root MSE        =    1.8548

------------------------------------------------------------------------------
    poltrust |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ppltrst |    .317986   .0217564    14.62   0.000     .2753132    .3606588
       _cons |   2.459308   .1188338    20.70   0.000     2.226229    2.692387
------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------------
-> cntry = BG

      Source |       SS           df       MS      Number of obs   =     2,291
-------------+----------------------------------   F(1, 2289)      =     98.89
       Model |  407.068824         1  407.068824   Prob > F        =    0.0000
    Residual |  9422.33512     2,289  4.11635436   R-squared       =    0.0414
-------------+----------------------------------   Adj R-squared   =    0.0410
       Total |  9829.40394     2,290  4.29231613   Root MSE        =    2.0289
```

Einfaches Modell, country fixed effects (dummies)

```
use ml-exercise-ess-5 , replace
* country fixed effects
encode cntry , gen(numbercountry)
reg poltrust ppltrst i.numbercountry, cluster(cntry)

Linear regression                               Number of obs     =     49,780
                                                F(0, 26)          =          .
                                                Prob > F          =          .
                                                R-squared         =     0.2866
                                                Root MSE          =      1.925

                                  (Std. Err. adjusted for 27 clusters in cntry)
-------------------------------------------------------------------------------
              |               Robust
     poltrust |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
      ppltrst |   .2308962   .0122556    18.84   0.000     .2057045    .2560878
              |
numbercountry |
          BG  |  -1.575482   .0193904   -81.25   0.000     -1.61534   -1.535625
          CH  |    .998099   .0077862   128.19   0.000     .9820942    1.014104
          CY  |   .0876725   .0135978     6.45   0.000     .0597219    .1156232
          CZ  |  -1.080142   .0069017  -156.50   0.000    -1.094328   -1.065955
          DE  |  -.4013071   .0048993   -81.91   0.000    -.4113778   -.3912364
          DK  |   .8750523   .0223549    39.14   0.000     .8291012    .9210034
          EE  |  -.4639403    .008079   -57.43   0.000    -.4805468   -.4473337
          ES  |  -.8381794   .0011337  -739.33   0.000    -.8405098   -.8358491
          FI  |   .3830721   .0179687    21.32   0.000     .3461369    .4200072
          FR  |  -.4243839   .0088638   -47.88   0.000    -.4426036   -.4061642
          GB  |  -.4842493   .0039486  -122.64   0.000    -.4923657   -.4761329
          GR  |  -2.238518   .0124577  -179.69   0.000    -2.264125   -2.212911
          HR  |  -2.080213    .005917  -351.57   0.000    -2.092376   -2.068051
          HU  |  -.4384424    .006987   -62.75   0.000    -.4528045   -.4240803
          IE  |  -.8429582   .0004863 -1733.49   0.000    -.8439578   -.8419587
          IL  |  -.9373666   .0018617  -503.50   0.000    -.9411934   -.9335398
          LT  |  -1.755159    .003691  -475.53   0.000    -1.762746   -1.747572
          NL  |   .9773006   .0117304    83.31   0.000     .9531884    1.001413
          NO  |   .8609662   .0201895    42.64   0.000     .8194659    .9024664
          PL  |  -1.055359   .0081462  -129.55   0.000    -1.072104   -1.038615
          PT  |  -1.438542   .0162507   -88.52   0.000    -1.471946   -1.405139
          RU  |  -.6123425   .0111056   -55.14   0.000    -.6351703   -.5895146
          SE  |   1.129833    .016004    70.60   0.000     1.096937     1.16273
          SI  |  -1.315145   .0136442   -96.39   0.000    -1.343191   -1.287099
          SK  |  -.7962383   .0129253   -61.60   0.000    -.8228066     -.76967
          UA  |  -1.872426   .0110316  -169.73   0.000    -1.895102   -1.849751
              |
        _cons |   2.898929   .0618649    46.86   0.000     2.771764    3.026095
-------------------------------------------------------------------------------
est store fixedeffects
```

Multi-level: Empty Model

```
mixed poltrust  || cntry:  if sample, nolog

Mixed-effects ML regression                     Number of obs     =     49,780
Group variable: cntry                           Number of groups  =         27

Obs per group:
      min =        990
      avg =    1,843.7
      max =      2,933

Wald chi2(0)      =           .
Log likelihood = -105112.39                     Prob > chi2       =          .

------------------------------------------------------------------------------
    poltrust |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   3.465337   .2184091    15.87   0.000     3.037263    3.893411
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cntry: Identity              |
  var(_cons) |   1.285686    .350563      .7534269     2.19396
-----------------------------+------------------------------------------------
       var(Residual) |   3.981859   .0252459      3.932684    4.031649
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 13029.14      Prob >= chibar2 = 0.0000
est store empty
```

ICC

- "Intraclass Correlation Coefficient"
- Varianz der Residuen auf der Kontextebene (random intercepts) / Varianz der Residuen auf der Kontextebene + Varianz der Residuen auf der Individualebene
- Schätzung für die Korrelation zufälliger Einflüsse zwischen zwei Personen aus demselben Kontext - Ähnlichkeit

```
estat icc

Intraclass correlation

------------------------------------------------------------------------------
                       Level |        ICC   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
                       cntry |   .2440769    .0503215      .1590897    .3552833
------------------------------------------------------------------------------
```

Auch das geht: Empty-Model mit drei Ebenen

```
mixed poltrust  || cntry: || ess5_reg:  if sample, nolog

Mixed-effects ML regression                     Number of obs     =     49,780

-------------------------------------------------------------
|     No. of       Observations per Group
 Group Variable |     Groups    Minimum    Average    Maximum
----------------+--------------------------------------------
  cntry |         27         990    1,843.7      2,933
       ess5_reg |        326          5      152.7      2,117
-------------------------------------------------------------

Wald chi2(0)      =          .
Log likelihood = -104861.96                     Prob > chi2       =          .

------------------------------------------------------------------------------
    poltrust |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   3.451006   .2195123    15.72   0.000      3.02077    3.881243
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cntry: Identity              |
  var(_cons) |   1.280021   .3529922      .7455569    2.197625
-----------------------------+------------------------------------------------
ess5_reg: Identity           |
  var(_cons) |   .1088999   .0131177      .0859993    .1378987
-----------------------------+------------------------------------------------
       var(Residual) |   3.909241   .0248672      3.860805    3.958285
------------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 13530.00             Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
est store empty3
```

Modell mit Individualvariable + random intercept

```
mixed poltrust ppltrst, nolog || cntry:

Mixed-effects ML regression                     Number of obs     =     49,780
Group variable: cntry                           Number of groups  =         27

Obs per group:
      min =        990
      avg =    1,843.7
      max =      2,933

Wald chi2(1)      =    3722.74
Log likelihood = -103317.84                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
    poltrust |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ppltrst |   .2311844    .003789    61.01   0.000      .223758    .2386107
       _cons |   2.329236   .1858588    12.53   0.000     1.964959    2.693512
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cntry: Identity              |
  var(_cons) |   .9211896   .2513747      .5396009    1.572626
-----------------------------+------------------------------------------------
       var(Residual) |   3.705399   .0234931      3.659638    3.751732
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 9480.09       Prob >= chibar2 = 0.0000
est store randomi
```

Modell mit Individualvariable, random intercept und Kontextvariable (corruption protection)

```
mixed poltrust ppltrst c_ticpi_2009, nolog || cntry:

Mixed-effects ML regression                     Number of obs     =     49,780
Group variable: cntry                           Number of groups  =         27

Obs per group:
      min =        990
      avg =    1,843.7
      max =      2,933

Wald chi2(2)      =    3793.57
Log likelihood = -103303.55                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
    poltrust |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ppltrst |    .231046   .0037889    60.98   0.000     .2236199    .2384721
c_ticpi_2009 |   .3776843    .053013     7.12   0.000     .2737806    .4815879
       _cons |  -.0467665   .3509429    -0.13   0.894     -.734602     .641069
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cntry: Identity              |
  var(_cons) |   .3184155   .0872196       .186138     .544695
-----------------------------+------------------------------------------------
       var(Residual) |   3.705398   .0234931      3.659637     3.75173
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 4087.44       Prob >= chibar2 = 0.0000
est store randomikon
```

Modell mit Individualvariablen, random intercept, Kontextvariable (corruption protection) + random slope ppltrst

```
mixed poltrust ppltrst c_ticpi_2009, nolog || cntry: ppltrst

Mixed-effects ML regression                     Number of obs     =     49,780
Group variable: cntry                           Number of groups  =         27

Obs per group:
      min =        990
      avg =    1,843.7
      max =      2,933

Wald chi2(2)      =     398.93
Log likelihood = -103214.01                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
    poltrust |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ppltrst |   .2366623   .0122872    19.26   0.000     .2125798    .2607447
c_ticpi_2009 |   .2791271   .0566023     4.93   0.000     .1681887    .3900655
       _cons |   .5105324   .3730705     1.37   0.171    -.2206724    1.241737
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cntry: Independent           |
var(ppltrst) |   .0036517   .0011126      .0020097     .006635
  var(_cons) |   .3521383   .0990664      .2028833    .6111956
-----------------------------+------------------------------------------------
       var(Residual) |   3.687217   .0233846      3.641668    3.733336
------------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 4266.51              Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
est store randomikonrandoms
```

Modelle im Überblick

```
est tab cluster empty empty3 randomi randomikon randomikonrandoms

-----------------------------------------------------------------------------------------------
    Variable |  cluster        empty        empty3       randomi      randomikon   randomik~s
-------------+---------------------------------------------------------------------------------
_            |
     ppltrst |  .33966617
       _cons |  1.7199469
-------------+---------------------------------------------------------------------------------
poltrust     |
     ppltrst |                                        .23118438    .23104601    .23666226
c_ticpi_2009 |                                                     .37768426     .2791271
       _cons |               3.4653374    3.4510065   2.3292358   -.0467665    .51053239
-------------+---------------------------------------------------------------------------------
lns1_1_1     |
       _cons |               .12564621     .1234383  -.04104472  -.57219903    -2.806288
-------------+---------------------------------------------------------------------------------
lnsig_e      |
       _cons |               .69087439    .68167165   .65489542   .65489528    .65243604
-------------+---------------------------------------------------------------------------------
lns2_1_1     |
       _cons |                           -1.1086631
-------------+---------------------------------------------------------------------------------
lns1_1_2     |
       _cons |                                                                -.52186567
-----------------------------------------------------------------------------------------------
```

Sparsamkeit und Fit

```
est stats cluster empty empty3 randomi randomikon randomikonrandoms

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
     cluster |     49,780   -111627  -108057.9       2    216119.8   216137.4
       empty |     49,780         .  -105112.4       3    210230.8   210257.2
      empty3 |     49,780         .    -104862       4    209731.9   209767.2
     randomi |     49,780         .  -103317.8       4    206643.7   206678.9
  randomikon |     49,780         .  -103303.5       5    206617.1   206661.2
randomikon~s |     49,780         .    -103214       6      206440   206492.9
-----------------------------------------------------------------------------
       Note: N=Obs used in calculating BIC; see [R] BIC note.
```

Fazit

- Mit Stata lassen sich Mehr-Ebenen-Modelle schätzen ...
- ... die komplizierter als die meisten unserer Theorien sein können
- Meist keine spezielle Software nötig

Ausblick

- Nächste Woche: Was tun mit Multiple-Country Repeated Cross-Sectional Designs?
- Dann: viele *Anwendungen*
- Fragen?