

Day 1: Introduction to multi-level data problems

Introduction to Multilevel Models
EUI Short Course 4–9 June, 2009
Prof. Kenneth Benoit

June 4, 2009

Course logistics and overview

- ▶ Purpose of the course
 - ▶ introductory
 - ▶ basic mathematical understanding of MLMs
 - ▶ applied, emphasis on Stata
 - ▶ Day 5 covers a few non-linear models

Course logistics and overview

- ▶ Purpose of the course
 - ▶ introductory
 - ▶ basic mathematical understanding of MLMs
 - ▶ applied, emphasis on Stata
 - ▶ Day 5 covers a few non-linear models
- ▶ What we will not do
 - ▶ work with really complicated multi-level structures
 - ▶ deal with estimation issues
 - ▶ use Bayesian methods

Course logistics and overview

- ▶ Purpose of the course
 - ▶ introductory
 - ▶ basic mathematical understanding of MLMs
 - ▶ applied, emphasis on Stata
 - ▶ Day 5 covers a few non-linear models
- ▶ What we will not do
 - ▶ work with really complicated multi-level structures
 - ▶ deal with estimation issues
 - ▶ use Bayesian methods
- ▶ Further caveats

Course logistics and overview

- ▶ Purpose of the course
 - ▶ introductory
 - ▶ basic mathematical understanding of MLMs
 - ▶ applied, emphasis on Stata
 - ▶ Day 5 covers a few non-linear models
- ▶ What we will not do
 - ▶ work with really complicated multi-level structures
 - ▶ deal with estimation issues
 - ▶ use Bayesian methods
- ▶ Further caveats
- ▶ Texts and how to use them
- ▶ Software and datasets
- ▶ Homework format, timing
- ▶ Overview of other course logistics

What is multilevel data?

- ▶ Multilevel data comes from a data structure in the population that is hierarchical, with sample data consisting of a multistage sample from this population

What is multilevel data?

- ▶ Multilevel data comes from a data structure in the population that is hierarchical, with sample data consisting of a multistage sample from this population
- ▶ The classic example is schools and pupils: first we take a sample of schools, then sample pupils within each school

What is multilevel data?

- ▶ Multilevel data comes from a data structure in the population that is hierarchical, with sample data consisting of a multistage sample from this population
- ▶ The classic example is schools and pupils: first we take a sample of schools, then sample pupils within each school
- ▶ We would then say that pupils are *nested* within schools

What is multilevel data?

- ▶ Multilevel data comes from a data structure in the population that is hierarchical, with sample data consisting of a multistage sample from this population
- ▶ The classic example is schools and pupils: first we take a sample of schools, then sample pupils within each school
- ▶ We would then say that pupils are *nested* within schools
- ▶ Other examples:
 - ▶ individuals nested within countries (survey data)
 - ▶ experts nested within countries (expert survey data)
 - ▶ coded documents nested within coders (Comparative Manifesto Project)
 - ▶ political parties within national contexts

What is multilevel data?

- ▶ Multilevel data comes from a data structure in the population that is hierarchical, with sample data consisting of a multistage sample from this population
- ▶ The classic example is schools and pupils: first we take a sample of schools, then sample pupils within each school
- ▶ We would then say that pupils are *nested* within schools
- ▶ Other examples:
 - ▶ individuals nested within countries (survey data)
 - ▶ experts nested within countries (expert survey data)
 - ▶ coded documents nested within coders (Comparative Manifesto Project)
 - ▶ political parties within national contexts
- ▶ Variables may vary at either level

What is multilevel data?

- ▶ Multilevel data comes from a data structure in the population that is hierarchical, with sample data consisting of a multistage sample from this population
- ▶ The classic example is schools and pupils: first we take a sample of schools, then sample pupils within each school
- ▶ We would then say that pupils are *nested* within schools
- ▶ Other examples:
 - ▶ individuals nested within countries (survey data)
 - ▶ experts nested within countries (expert survey data)
 - ▶ coded documents nested within coders (Comparative Manifesto Project)
 - ▶ political parties within national contexts
- ▶ Variables may vary at either level
- ▶ Basic terminology: lowest level is Level 1, higher is Level 2

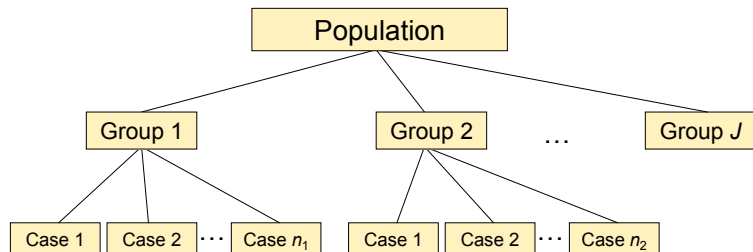
What is multilevel data?

- ▶ Multilevel data comes from a data structure in the population that is hierarchical, with sample data consisting of a multistage sample from this population
- ▶ The classic example is schools and pupils: first we take a sample of schools, then sample pupils within each school
- ▶ We would then say that pupils are *nested* within schools
- ▶ Other examples:
 - ▶ individuals nested within countries (survey data)
 - ▶ experts nested within countries (expert survey data)
 - ▶ coded documents nested within coders (Comparative Manifesto Project)
 - ▶ political parties within national contexts
- ▶ Variables may vary at either level
- ▶ Basic terminology: lowest level is Level 1, higher is Level 2
- ▶ Response variables (Y) always vary at the lowest level

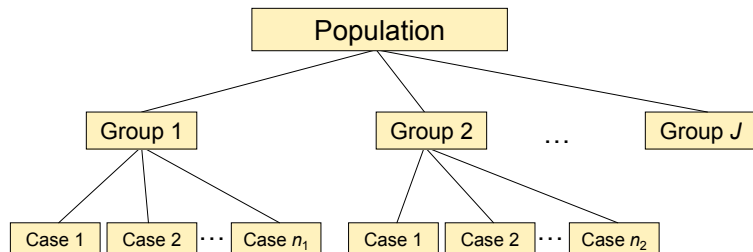
What is multilevel data?

- ▶ Multilevel data comes from a data structure in the population that is hierarchical, with sample data consisting of a multistage sample from this population
- ▶ The classic example is schools and pupils: first we take a sample of schools, then sample pupils within each school
- ▶ We would then say that pupils are *nested* within schools
- ▶ Other examples:
 - ▶ individuals nested within countries (survey data)
 - ▶ experts nested within countries (expert survey data)
 - ▶ coded documents nested within coders (Comparative Manifesto Project)
 - ▶ political parties within national contexts
- ▶ Variables may vary at either level
- ▶ Basic terminology: lowest level is Level 1, higher is Level 2
- ▶ Response variables (Y) always vary at the lowest level

The structure of multilevel data

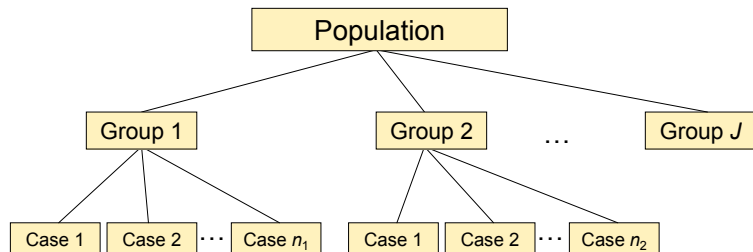


The structure of multilevel data



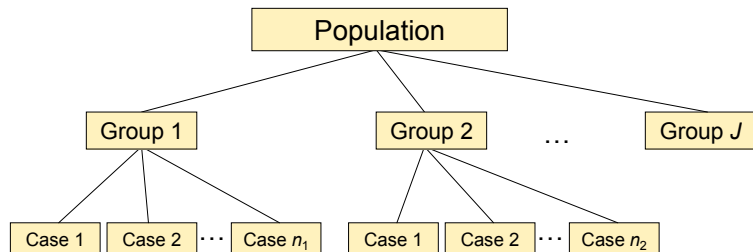
- ▶ a variation on this is *longitudinal* data structure, where the level 1 variable is an observation for a given time, and the level 2 variable is a subject

The structure of multilevel data



- ▶ a variation on this is *longitudinal* data structure, where the level 1 variable is an observation for a given time, and the level 2 variable is a subject
- ▶ nesting may be *unintentional*: for instance we could have policy categories from manifestos (level 1) coded by coder (level 2); or survey respondents (level 1) nested within interviewer (level 2)

The structure of multilevel data



- ▶ a variation on this is *longitudinal* data structure, where the level 1 variable is an observation for a given time, and the level 2 variable is a subject
- ▶ nesting may be *unintentional*: for instance we could have policy categories from manifestos (level 1) coded by coder (level 2); or survey respondents (level 1) nested within interviewer (level 2)
- ▶ terminology may vary — here we refer to *multilevel models* generically but terms found in the literature include: variance components models, random-coefficients models and random-effects models, (general) mixed models, and hierarchical linear models

Why would special models be needed for multilevel data?

- ▶ The usual assumptions for causal inference from regression models is that individual observations are independent
- ▶ With nested structures this may not be the case: the correlation between observations within a common unit will be higher than the average correlation of observations between units
- ▶ Consequence is that we will underestimate the uncertainty of causal effects from pooled estimates

Why would special models be needed for multilevel data?

- ▶ The usual assumptions for causal inference from regression models is that individual observations are independent
- ▶ With nested structures this may not be the case: the correlation between observations within a common unit will be higher than the average correlation of observations between units
- ▶ Consequence is that we will underestimate the uncertainty of causal effects from pooled estimates
- ▶ In addition, only multilevel models can help us separate within-unit from between-unit effects, especially the different average effects and the different effects of covariates

The ecological fallacy

- ▶ The ecological fallacy refers to the fallacy of inferring individual behavior from aggregate data – in our context, inferring Level 1 relationships based on Level 2 units

The ecological fallacy

- ▶ The ecological fallacy refers to the fallacy of inferring individual behavior from aggregate data – in our context, inferring Level 1 relationships based on Level 2 units
- ▶ Arises when level 2 variables and level 1 variables reflect different causal processes

The ecological fallacy

- ▶ The ecological fallacy refers to the fallacy of inferring individual behavior from aggregate data – in our context, inferring Level 1 relationships based on Level 2 units
- ▶ Arises when level 2 variables and level 1 variables reflect different causal processes
- ▶ originally from Robinson (1950) who studied the relationship between literacy and race in the US. The correlation between mean literacy rates and mean proportions of the black population was 0.95, but the individual-level correlation ignoring the grouping was just is 0.20

The ecological fallacy

- ▶ The ecological fallacy refers to the fallacy of inferring individual behavior from aggregate data – in our context, inferring Level 1 relationships based on Level 2 units
- ▶ Arises when level 2 variables and level 1 variables reflect different causal processes
- ▶ originally from Robinson (1950) who studied the relationship between literacy and race in the US. The correlation between mean literacy rates and mean proportions of the black population was 0.95, but the individual-level correlation ignoring the grouping was just 0.20
- ▶ A problem in many political research questions, esp. voting behavior inferred from aggregated results

The atomistic fallacy

- ▶ The atomistic fallacy (aka *individualistic fallacy*) may occur when drawing inferences about group-level relationships from individual-level data

The atomistic fallacy

- ▶ The atomistic fallacy (aka *individualistic fallacy*) may occur when drawing inferences about group-level relationships from individual-level data
- ▶ Arises because individual-level associations may differ from those at the group level

The atomistic fallacy

- ▶ The atomistic fallacy (aka *individualistic fallacy*) may occur when drawing inferences about group-level relationships from individual-level data
- ▶ Arises because individual-level associations may differ from those at the group level
- ▶ Example: we might find that individual income is positively associated with decreased mortality from heart disease. From this we should not infer, however, that at the country level, increasing per capita income is associated with decreasing heart disease mortality. In fact, across countries we might actually increase heart disease mortality by increasing income.

Stata and “robust” clustered standard errors”

- ▶ One method of correcting for the effect of clusters is to specify the `vce(cluster clustvar)` as an option to regression commands

Stata and “robust” clustered standard errors”

- ▶ One method of correcting for the effect of clusters is to specify the `vce(cluster clustvar)` as an option to regression commands
- ▶ This relaxes the requirement that the errors be independent, by allowing them to be correlated within each cluster group

Stata and “robust” clustered standard errors”

- ▶ One method of correcting for the effect of clusters is to specify the `vce(cluster clustvar)` as an option to regression commands
- ▶ This relaxes the requirement that the errors be independent, by allowing them to be correlated within each cluster group
- ▶ The correction only affects the standard errors, not the estimated coefficients, since it operates only on the variance-covariance matrix

Stata and “robust” clustered standard errors”

- ▶ One method of correcting for the effect of clusters is to specify the `vce(cluster clustvar)` as an option to regression commands
- ▶ This relaxes the requirement that the errors be independent, by allowing them to be correlated within each cluster group
- ▶ The correction only affects the standard errors, not the estimated coefficients, since it operates only on the variance-covariance matrix
- ▶ This will not get at the core issues of interest for multilevel models, which have to do with separating between-group effects from within-group effects, and especially not the provision for random intercepts and or slopes

The organization of multilevel data

- ▶ Multilevel data are distinguished by their organization according to multilevel identifying units. Examples:
 - ▶ constituency ID
 - ▶ country ID
 - ▶ school ID

The organization of multilevel data

- ▶ Multilevel data are distinguished by their organization according to multilevel identifying units. Examples:
 - ▶ constituency ID
 - ▶ country ID
 - ▶ school ID
- ▶ There are two basic formats for organizing data that are clustered by identifying units:

The organization of multilevel data

- ▶ Multilevel data are distinguished by their organization according to multilevel identifying units. Examples:
 - ▶ constituency ID
 - ▶ country ID
 - ▶ school ID
- ▶ There are two basic formats for organizing data that are clustered by identifying units:
 - wide format** two columns of data contain the same information, distinguished by different levels

The organization of multilevel data

- ▶ Multilevel data are distinguished by their organization according to multilevel identifying units. Examples:
 - ▶ constituency ID
 - ▶ country ID
 - ▶ school ID
- ▶ There are two basic formats for organizing data that are clustered by identifying units:
 - wide format** two columns of data contain the same information, distinguished by different levels
 - long format** different levels are themselves variables (in their own columns)

Zen and the art of reshaping

- ▶ some things cannot be done in long format. For instance if we want to plot one set of scores against another, e.g. taxes v. spending versus social dimension from the expert surveys

Zen and the art of reshaping

- ▶ some things cannot be done in long format. For instance if we want to plot one set of scores against another, e.g. taxes v. spending versus social dimension from the expert surveys
- ▶ For this we need the **wide** format, where each dimension forms a separate variable and the identifier defines a unique row

Zen and the art of reshaping

- ▶ some things cannot be done in long format. For instance if we want to plot one set of scores against another, e.g. taxes v. spending versus social dimension from the expert surveys
- ▶ For this we need the **wide** format, where each dimension forms a separate variable and the identifier defines a unique row
- ▶ To convert from long to wide (and vice versa), we need the `reshape` command

Zen and the art of reshaping

- ▶ some things cannot be done in long format. For instance if we want to plot one set of scores against another, e.g. taxes v. spending versus social dimension from the expert surveys
- ▶ For this we need the **wide** format, where each dimension forms a separate variable and the identifier defines a unique row
- ▶ To convert from long to wide (and vice versa), we need the `reshape` command
- ▶ The key to using `reshape` is to determine what the logical observation i is and the subobservation j that will be used to organize the data

Zen and the art of reshaping continued

```
(wide form)
i      ..... x_ij .....
id sex  inc80  inc81  inc82
-----
1   0   5000   5500   6000
2   1   2000   2200   3300
3   0   3000   2000   1000
```

```
(long form)
i      j      x_ij
id year  sex  inc
-----
1    80    0  5000
1    81    0  5500
1    82    0  6000
2    80    1  2000
2    81    1  2200
2    82    1  3300
3    80    0  3000
3    81    0  2000
3    82    0  1000
```

Given this data, you could use reshape to convert from one form to the other:

- . reshape long inc, i(id) j(year) (goes from top-form to bottom)
- . reshape wide inc, i(id) j(year) (goes from bottom-form to top)

Example of multilevel data: Benoit and Marsh (2008)

```
. use dail2002spending
```

```
(Irish Dail 2002 from Benoit and Marsh 2008)
```

```
. list constID constituency namelast party votes1st incumb m spent in 6/28, clean
```

	constID	constituency	namelast	party	votes1st	incumb	m	spent
6.	1	Carlow Kilkenny	McGuinness	ff	9343	1	5	19648.3
7.	1	Carlow Kilkenny	Nolan	ff	8711	0	5	24100.27
8.	1	Carlow Kilkenny	Nolan	ind	335	0	5	6544.23
9.	1	Carlow Kilkenny	O'Brien	lab	3732	0	5	8404.43
10.	1	Carlow Kilkenny	Townsend	lab	4272	0	5	10658.21
11.	1	Carlow Kilkenny	White	gp	4961	0	5	12110.11
12.	2	Cavan Monaghan	Boyland	fg	4819	1	5	11217.01
13.	2	Cavan Monaghan	Brennan	ind	1026	0	5	17196.73
14.	2	Cavan Monaghan	Connolly	ind	7722	0	5	17934.79
15.	2	Cavan Monaghan	Crawford	fg	6113	1	5	11124
16.	2	Cavan Monaghan	Cullen	lab	550	0	5	8756.67
17.	2	Cavan Monaghan	Gallagher	ff	3731	0	5	20122.19
18.	2	Cavan Monaghan	Martin	ind	1943	0	5	34542.73
19.	2	Cavan Monaghan	McCabe	gp	1100	0	5	10699.87
20.	2	Cavan Monaghan	McCaughy	pd	1131	0	5	30573.12
21.	2	Cavan Monaghan	O Caolain	sf	10832	1	5	28953.32
22.	2	Cavan Monaghan	O'Hanlon	ff	7204	1	5	21483.37
23.	2	Cavan Monaghan	O'Reilly	fg	4639	0	5	12839.2
24.	2	Cavan Monaghan	Smith	csp	358	0	5	3141.27
25.	2	Cavan Monaghan	Smith	ff	10679	1	5	22383.53
26.	3	Clare	Breen	fg	4541	0	4	11687.46
27.	3	Clare	Breen	ind	9721	0	4	11974.15
28.	3	Clare	Carey	fg	4015	1	4	14195.46

Benoit and Marsh (2008) example continued

```
. desc
```

```
Contains data from dail2002spending.dta
```

```
obs:          463                Irish Dail 2002 from Benoit and Marsh 2008  
vars:         10                18 May 2009 17:45  
size:        26,854 (99.9% of memory free)
```

```
-----  
      storage  display      value  
variable name  type    format    label    variable label  
-----  
constID        byte    %9.0g  
constituency   str20   %20s  
namelast       str15   %15s  
party          byte    %8.0g    party_e  Candidate's party label  
votes1st       int     %9.0g    First preference votes 2002  
incumb         byte    %9.0g    Incumbency status 1/0  
wonseat        byte    %9.0g    Candidate won a seat 1/0  
m              byte    %9.0g    District magnitude  
electorate     float   %9.0g    Registered voters in constituency  
spent          float   %9.0g    Total spending  
-----
```

```
Sorted by:  constID  namelast
```

Constituency-level m, electorate

Candidate-level namelast, votes1st, incumb, wonseat, spent, party

Benoit and Marsh (2008) example continued

```
. desc
```

```
Contains data from dail2002spending.dta
```

```
obs:          463                Irish Dail 2002 from Benoit and Marsh 2008
vars:         10                18 May 2009 17:45
size:        26,854 (99.9% of memory free)
```

variable name	storage type	display format	value label	variable label
constID	byte	%9.0g		Constituency Numeric ID
constituency	str20	%20s		Candidate's constituency
namelast	str15	%15s		Candidate's last name
party	byte	%8.0g	party_e	Candidate's party label
votes1st	int	%9.0g		First preference votes 2002
incumb	byte	%9.0g		Incumbency status 1/0
wonseat	byte	%9.0g		Candidate won a seat 1/0
m	byte	%9.0g		District magnitude
electorate	float	%9.0g		Registered voters in constituency
spent	float	%9.0g		Total spending

```
Sorted by:  constID  namelast
```

Constituency-level m, electorate

Candidate-level namelast, votes1st, incumb, wonseat, spent, party
(and we could view party as having a special status)

Long v. wide data format: *PPMD* example

```
. use PPMD_detail, clear  
(Party Policy in Modern Democracies, Kenneth Benoit and Michael Laver)
```

```
. sample 20, count  
(206945 observations deleted)
```

```
. list Country Party Dimension Scale Survey_Label_ID Score Vote_Share Election_Date, clean
```

	Country	Party	Dimension	Scale	Survey_Label_ID	Score	Vote_Share	Election_Date
1.	SE	MP	Taxes v. Spending	Position	408	11	4.6	2002
2.	ES	CiU	Taxes v. Spending	Position	191	12	3.2	2004
3.	SE	M	EU: Peacekeeping	Position	892	5	15.2	2002
4.	DE	CDU/CSU	EU: Peacekeeping	Importance	1689	3	38.51	2002
5.	MD	PDAM	Environment	Importance	12	6	1.9	2001
6.	SI	SNS	Urban-Rural	Importance	44	16	4.4	2000
7.	NO	KrF	NATO/Peacekeeping	Position	74	8	12.5	2001
8.	FR	UDF	Taxes v. Spending	Importance	16	14	4.8	2002
9.	SR	DSS	Left-Right	Position	1	13	18	2003
10.	CA	LPC	Sympathy	Position	820	17	40.8	2000
11.	JP	JCP	Defense policy	Importance	3	5	7.7	2003
12.	CA	GPC	Sympathy	Position	437	7	.8	2000
13.	RO	PD	Social	Position	574	6	7.03	2000
14.	HU	MUNKS	Media Freedom	Importance	823	19	2.8	2002
15.	IL	Merz	Palestinian State	Importance	543	20	5.2	2003
16.	DE	GRU	EU: Peacekeeping	Importance	1262	13	8.6	2002
17.	IT	SDI	Deregulation	Importance	120	14	1.1	2001
18.	BE	PS	Environment	Position	547	10	13	2003
19.	IT	UDC	EU: Accountability	Importance	189	12	3.2	2001
20.	CZ	SZ	Social	Position	31	12	2.36	2002

Long v. wide data format: *PPMD* example

- ▶ The PPMD dataset is organized as long data, where the basic unit of variation is the Score variable

Long v. wide data format: *PPMD* example

- ▶ The PPMD dataset is organized as long data, where the basic unit of variation is the Score variable
- ▶ Score represents the placement on a 1–20 point scale of either the left-right location or the low–high importance

Long v. wide data format: *PPMD* example

- ▶ The PPMD dataset is organized as long data, where the basic unit of variation is the Score variable
- ▶ Score represents the placement on a 1–20 point scale of either the left-right location or the low–high importance
- ▶ The different variables are:
 - Country a code designating the country
 - Party a country-specific alphanumeric identifier for party
 - Dimension one of 40-odd policy dimensions
 - Scale either Position or Importance
 - Survey_Label_ID country-specific respondent ID

Long v. wide data format: *PPMD* example

- ▶ The PPMD dataset is organized as long data, where the basic unit of variation is the Score variable
- ▶ Score represents the placement on a 1–20 point scale of either the left-right location or the low–high importance
- ▶ The different variables are:
 - Country a code designating the country
 - Party a country-specific alphanumeric identifier for party
 - Dimension one of 40-odd policy dimensions
 - Scale either Position or Importance
 - Survey_Label_ID country-specific respondent ID
- ▶ This is a useful way to *store* the data, but may not be useful for analyzing it, although this depends

Long v. wide data format: *PPMD* example continued

For data analysis based on tables, the long format is required.
Example:

```
. use PPMD_detail, clear
(Party Policy in Modern Democracies, Kenneth Benoit and Michael Laver)

. table Party Dimension if Country=="IT":cntryLab & Scale==1 & Dimension<15, c(mean Score) format(%9.1f)
```

Party abbreviat ion	Policy dimension				
	Taxes v. Spending	Social	Environment	Decentralization	Left-Right
AN	10.1	18.3	13.5	14.9	16.9
DS	6.7	5.0	7.3	7.4	6.0
FI	17.5	12.9	17.2	8.9	15.6
Green	4.9	3.4	1.7	9.5	4.0
It.Val.	8.6	9.9	8.3	9.1	10.1
LN	15.1	17.1	15.3	2.4	16.9
MSFT	6.7	18.5	10.7	16.2	19.0
Marg	8.5	11.9	8.3	8.1	8.0
PDCI	3.9	4.2	6.4	12.5	3.3
Pann	15.2	2.0	9.3	6.8	12.0
RC	2.9	3.7	5.6	13.4	2.1
SDI	9.3	7.1	9.6	8.9	8.6
UDC	10.6	16.0	11.7	10.5	12.4

Long v. wide data format: *PPMD* example continued

```
. table Country Dimension if Dimension<5 & Country>60
```

Country name	Policy dimension		
	Taxes v. Spending	Social	EU joining
GR	127	128	
IS	144	144	
IE	626	632	
IL	668	682	
IT	1,189	1,196	
LU	48	47	
MT	35	38	38
NL	387	387	
NZ	301		
NI	152	154	
NO	330	333	336
PT	246	243	
ES	753	751	
SE	937	936	
CH	934	956	897
TR	412	411	423
US	662	662	
EU	273	287	
JP	696	696	

Long v. wide data format: *PPMD* example continued

For data analysis based on tables, the long format is required.
Example:

```
. ttest Score if Dimension>13, by(Scale)
```

```
Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Position	39587	10.65799	.0292697	5.823644	10.60062	10.71536
Importan	30031	12.98175	.0274553	4.75786	12.92794	13.03557
combined	69618	11.66039	.0208877	5.511278	11.61945	11.70133
diff		-2.323758	.0412452		-2.404599	-2.242918

```
diff = mean(Position) - mean(Importan)          t = -56.3401
Ho: diff = 0                                   degrees of freedom = 69616
```

```
Ha: diff < 0
Pr(T < t) = 0.0000
```

```
Ha: diff != 0
Pr(|T| > |t|) = 0.0000
```

```
Ha: diff > 0
Pr(T > t) = 1.0000
```

Reshape example with expert survey data

```
. reshape wide Score, i(Country Survey_Label_ID Party Vote_Share Scale) j(Dimension)
(note: j = 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40)
```

```
Data                long  ->  wide
-----
Number of obs.      206970 ->  21029
Number of variables    12  ->   50
j variable (40 values) Dimension -> (dropped)
xij variables:
                        Score  ->  Score1 Score2 ... Score99
-----
```

```
. list Country Party Scale Survey_Label_ID Vote_Share Score1-Score4 in 1/10, clean
```

	Country	Party	Scale	Survey_Label_ID	Vote_Share	Score1	Score2	Score3	Score4
1.	AL	PBDNJ	Position	1	2.6	9	9	9	14
2.	AL	PBDNJ	Importance	1	2.6	4	1	3	17
3.	AL	PD	Position	1	19.36	12	9	15	17
4.	AL	PD	Importance	1	19.36	16	4	15	17
5.	AL	PDr	Position	1	5.1	13	9	15	17
6.	AL	PDr	Importance	1	5.1	16	6	15	17
7.	AL	PLL	Position	1	4.03	14	11	15	17
8.	AL	PLL	Importance	1	4.03	14	4	15	17
9.	AL	PR	Position	1	4.83	13	11	15	17
10.	AL	PR	Importance	1	4.83	16	4	16	17

Reshape example with expert survey data

```
. use PPMd_summary_day1, clear
(Party Policy in Modern Democracies, K. Benoit and M. Laver, Summary Data)

. reshape wide Mean, i(Country Party Scale) j(Dimension)
(note: j = 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 36 37)

Data
-----
                long  ->  wide
-----
Number of obs.          8106  ->   739
Number of variables      8    ->   44
j variable (38 values)  Dimension -> (dropped)
xij variables:
                Mean  ->  Mean1 Mean2 ... Mean99
-----

. graph twoway (qfitci Mean24 Mean13) (scatter Mean24 Mean13, msize(small) m(oh)) if Scale==1,
> xtitle(Left-Right) ytitle(EU Integration) legend(off)
```

Reshape example with expert survey data

```
. use PPMd_summary_day1, clear  
(Party Policy in Modern Democracies, K. Benoit and M. Laver, Summary Data)
```

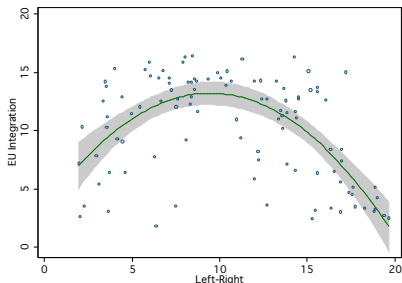
```
. reshape wide Mean, i(Country Party Scale) j(Dimension)
```

```
(note: j = 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 36 37)
```

```
Data
```

	long	->	wide
Number of obs.	8106	->	739
Number of variables	8	->	44
j variable (38 values)	Dimension	->	(dropped)
xij variables:			
	Mean	->	Mean1 Mean2 ... Mean99

```
. graph twoway (qfitci Mean24 Mean13) (scatter Mean24 Mean13, msize(small) m(oh)) if Scale==1,  
> xtitle(Left-Right) ytitle(EU Integration) legend(off)
```



Introducing variance decomposition models

- ▶ Standard model without covariates:

$$y_{ij} = \beta + \xi_{ij}$$

Introducing variance decomposition models

- ▶ Standard model without covariates:

$$y_{ij} = \beta + \xi_{ij}$$

- ▶ We can model the dependence within subjects j by splitting ξ_{ij} into two components ζ_j and ϵ_{ij} :

$$y_{ij} = \beta + \zeta_j + \epsilon_{ij}$$

Introducing variance decomposition models

- ▶ Standard model without covariates:

$$y_{ij} = \beta + \xi_{ij}$$

- ▶ We can model the dependence within subjects j by splitting ξ_{ij} into two components ζ_j and ϵ_{ij} :

$$y_{ij} = \beta + \zeta_j + \epsilon_{ij}$$

- ▶ ζ_j represent level-2 effects, also known as “random intercepts”, with variance ψ :

$$\zeta_j \sim N(0, \psi)$$

Introducing variance decomposition models

- ▶ Standard model without covariates:

$$y_{ij} = \beta + \xi_{ij}$$

- ▶ We can model the dependence within subjects j by splitting ξ_{ij} into two components ζ_j and ϵ_{ij} :

$$y_{ij} = \beta + \zeta_j + \epsilon_{ij}$$

- ▶ ζ_j represent level-2 effects, also known as “random intercepts”, with variance ψ :

$$\zeta_j \sim N(0, \psi)$$

- ▶ ϵ_{ij} are level-1 errors, with variance θ

$$\epsilon_{ij} \sim N(0, \theta)$$

Introducing variance decomposition models

- ▶ Standard model without covariates:

$$y_{ij} = \beta + \xi_{ij}$$

- ▶ We can model the dependence within subjects j by splitting ξ_{ij} into two components ζ_j and ϵ_{ij} :

$$y_{ij} = \beta + \zeta_j + \epsilon_{ij}$$

- ▶ ζ_j represent level-2 effects, also known as “random intercepts”, with variance ψ :

$$\zeta_j \sim N(0, \psi)$$

- ▶ ϵ_{ij} are level-1 errors, with variance θ

$$\epsilon_{ij} \sim N(0, \theta)$$

- ▶ More complicated models will be explored later, such as random coefficients (involving β differences at level-2)