

Master Class

om

Repeterte Målinger

René Holst

`rene.holst@medisin.uio.no`

Oslo Centre for Biostatistics and Epidemiology

27. Mai 2021

Outline

- ▶ Part I:
 - ▶ Key Concepts
 - ▶ Motivating example (Simulated data)
- ▶ Part II:
 - ▶ Explorative Analysis
 - ▶ Model Specification
 - ▶ Model Estimation
 - ▶ Model Control
 - ▶ Model Report (Not so much)

Part I: Clustered Data - Examples

- ▶ Multiple measurements on pupils within classes within schools within...
- ▶ Multiple raters assessing multiple images
- ▶ Multiple patients attending multiple doctors at multiple hospitals
- ▶ Trials with multiple clinics, some doctors appearing at more hospitals
- ▶ Treatment/Testing on both legs (or eyes or...) on multiple patients on multiple visits
- ▶ Growth curves comparing different traits
- ▶ Multiple measurements on each blood sample from multiple patients
- ▶ Familial data; Twin data

Part I: Statistical Model

A simplified description of the process, by which the data were generated; Rich enough to capture the essential elements of signal and noise.

- ▶ Observation=Signal+Noise=Fixed Part + Random Part
- ▶ $y_i = \mu_i + \epsilon_i; \quad \epsilon_i \sim N(0, \sigma^2)$

Part I: t-test

- ▶ **Independent** Normally distributed data
- ▶ Variance homogeneity
- ▶ No additional data (Besides Group)

$$y_i = \beta_0 + \beta_1 \cdot \text{Group}_i + \epsilon_i$$

$$\text{Group}_i \in \{0, 1\}, \quad \epsilon_{i,j} \sim N(0, \sigma^2)$$

Part I: Ordinary Linear Regression

- ▶ **Independent** Normally distributed data
- ▶ Variance homogeneity
- ▶ Additional data (Covariates: Group, Sex, Age, ...)

$$y_i = \beta_0 + \beta_1 \cdot x_{1i} + \cdots + \beta_p \cdot x_{pi} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

Part I: Mixed Effects Models

- ▶ Clustered Normally distributed data
- ▶ Variance homogeneity (within clusters)
- ▶ Multiple observations within each cluster
- ▶ Potentially more levels (Pupils, Classes, Schools, Regions,...)
- ▶ Additional data (Covariates: : Group, Sex, Age, ...)

$$y_{i,j} = \beta_0 + \beta_1 \cdot x_{1ij} + \cdots + \beta_p \cdot x_{pij} + b_i + \epsilon_{ij}$$

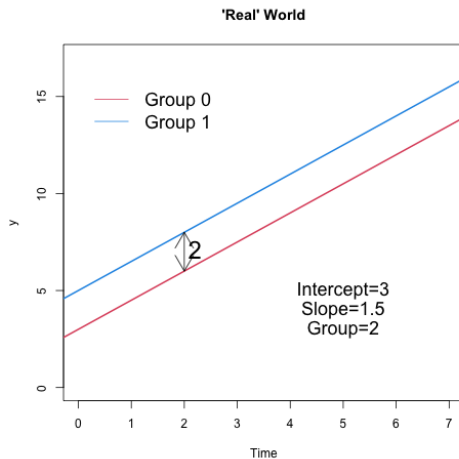
$$b_i \sim N(0, \tau^2), \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

Part I: Example - Simulated Data

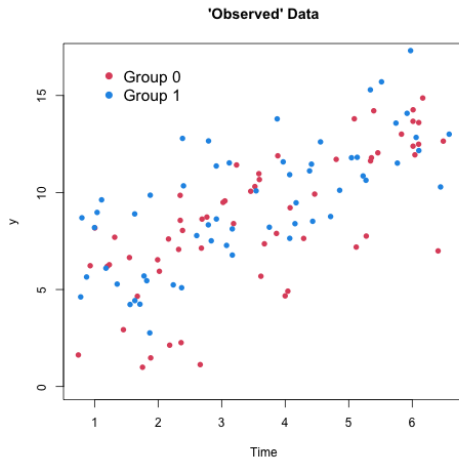
- ▶ 20 individuals
- ▶ 2 Groups
- ▶ Each measured at 6 time points
- ▶ Simulation parameters:
(*Intercept* = 3, *Slope* = 1.5, *Group* = 2)

Multiple observations within each patient!
Correlated data

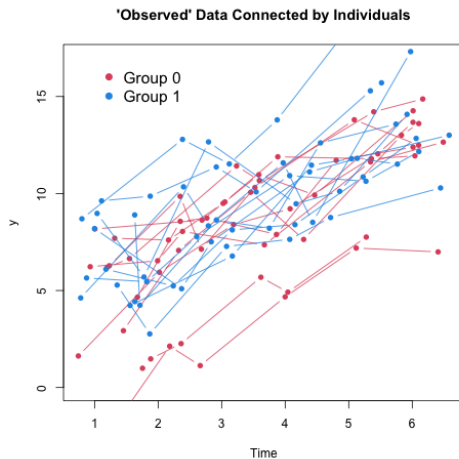
Part I: Example - Simulated Data



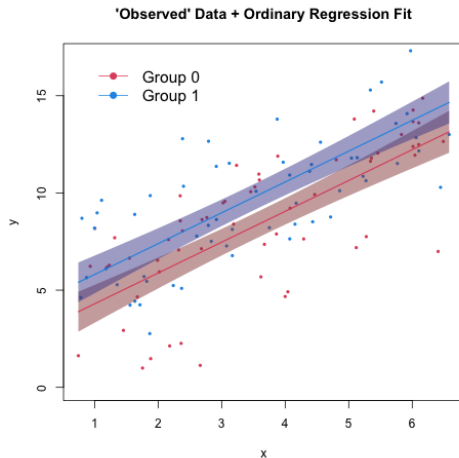
Part I: Example - Simulated Data - Observations



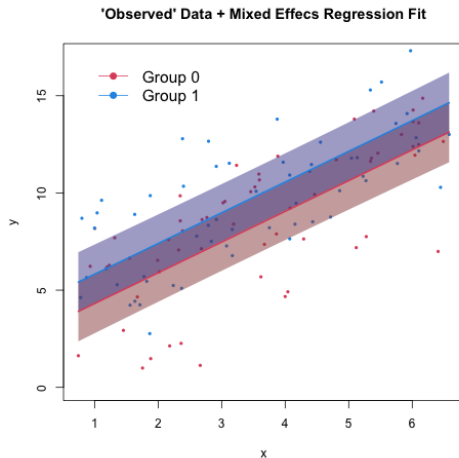
Part I: Example - Simulated Data - Spaghetti



Part I: Motivating Example - Ord. Lin. Regression w Conf Int



Part I: Motivating Example - Mixed Effects Model w Conf Int



Part I: Motivating Example - Tables of estimates

OLR	Estimate	SE	t-value	<i>p</i>
Intercept	2.710	0.601	4.508	0.000
Time	1.586	0.141	11.280	0.000
Group	1.510	0.477	3.164	0.002

Mixed	Estimate	SE	t-value	<i>p</i>
(Intercept)	2.728	0.788	3.462	0.002
Time	1.581	0.080	19.878	0.000
Group	1.510	1.041	1.451	0.164

Part I: Motivating Example - Conclusion

- ▶ Almost identical effect estimates
- ▶ Different SE
- ▶ Affects significance

Neglecting the variance structure may lead to erroneous conclusions.

Part II: Real Data

Data in wide format.

```
. list id age sex ptsd_any3 scl90gsi_1 scl90gsi_2 scl90gsi_3 Barndomstraume IL1ra_1 IL1ra_2 IL1ra_3 in 1/12
```

	id	age	sex	ptsd_any3	scl90g~1	scl90g~2	scl90g~3	Barndo~e	IL1ra_1	IL1ra_2	IL1ra_3
1.	1	36	Women	Har ikke PTS	1.74	.	.	2	39.03	44.55	.
2.	2	24	Women	Har ikke PTS	1.91	1.86	1.29	0	25.48	23.63	27.95
3.	3	19	Women	Har PTSD som	1.77	1.19	.	0	237.1	322.43	279.79
4.	4	58	Men	Har ikke PTS	1.53	1.5	1.48	2	26.72	27.95	23.01
5.	5	57	Women	Har PTSD som	1.97	1.89	1.32	0	32.26	35.96	31.65
6.	6	48	Men	Har ikke PTS	1.22	1.04	.77	0	17.44	16.19	18.68
7.	7	54	Women	Har ikke PTS	1.47	1.46	1.28	1	72.66	79.36	93.95
8.	8	46	Women	Har ikke PTS	1.56	1.48	.24	0	30.42	34.73	37.8
9.	9	35	Men	Har ikke PTS	1.03	.78	.16	0	101.24	77.53	78.14
10.	10	26	Women	Har ikke PTS	1.48	2.06	.47	0	51.9	50.06	55.57
11.	11	44	Men	Har ikke PTS	1	.86	.47	1	230.48	223.86	203.98
12.	12	47	Women	Har ikke PTS	.84	1.19	1.92	1	20.53	14.95	14.95

Part II: Real Data

Reshape Command.

```
. reshape long scl90gsi_ IL1b_ IL1ra_ IL6_ IL10_ IL17A_ IFNy_ MCP1_ TNFa_, i(id) j(visit)
(note: j = 1 2 3)
```

Data	wide	->	long
Number of obs.	147	->	441
Number of variables	50	->	33
j variable (3 values)		->	visit
xij variables:			
scl90gsi_1 scl90gsi_2 scl90gsi_3		->	scl90gsi_
IL1b_1 IL1b_2 IL1b_3		->	IL1b_
IL1ra_1 IL1ra_2 IL1ra_3		->	IL1ra_
IL6_1 IL6_2 IL6_3		->	IL6_
IL10_1 IL10_2 IL10_3		->	IL10_
IL17A_1 IL17A_2 IL17A_3		->	IL17A_
IFNy_1 IFNy_2 IFNy_3		->	IFNy_
MCP1_1 MCP1_2 MCP1_3		->	MCP1_
TNFa_1 TNFa_2 TNFa_3		->	TNFa_

Part II: Real Data

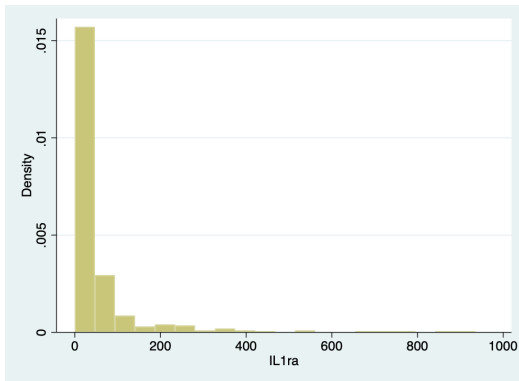
Data in long format.

```
. list id visit age sex ptsd_any3 scl90gsi Barndomstraume IL1ra in 1/12
```

	id	visit	age	sex	ptsd_any3	scl90gsi	Barndome	IL1ra
1.	1	1	36	Women	Har ikke PTS	1.74	2	39.03
2.	1	2	36	Women	Har ikke PTS	.	2	44.55
3.	1	3	36	Women	Har ikke PTS	.	2	.
4.	2	1	24	Women	Har ikke PTS	1.91	0	25.48
5.	2	2	24	Women	Har ikke PTS	1.86	0	23.63
6.	2	3	24	Women	Har ikke PTS	1.29	0	27.95
7.	3	1	19	Women	Har PTSD som	1.77	0	237.1
8.	3	2	19	Women	Har PTSD som	1.19	0	322.43
9.	3	3	19	Women	Har PTSD som	.	0	279.79
10.	4	1	58	Men	Har ikke PTS	1.53	2	26.72
11.	4	2	58	Men	Har ikke PTS	1.5	2	27.95
12.	4	3	58	Men	Har ikke PTS	1.48	2	23.01

Part II: Exploratory Analysis

Distribution of dependent variable, IL1ra



NOT Normal!

Part II: Exploratory Analysis

Suitable transformation? Use Box-Cox analysis:

```
.boxcox IL1ra age sex visit1 visit2 visit3.
```

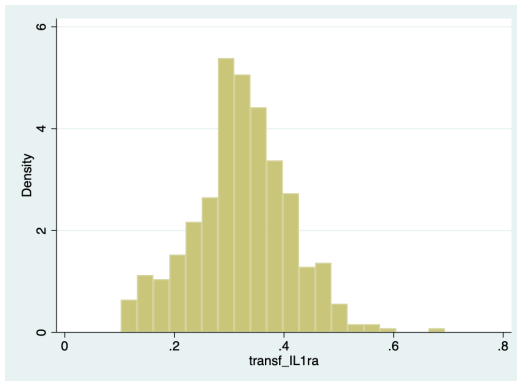
IL1ra	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
/theta	-.3175804	.0420247	-7.56	0.000	-.3999473	-.2352135

Generate New variable:

```
.gen transIL1ra = IL1ra(-1/3)
```

Part II: Exploratory Analysis

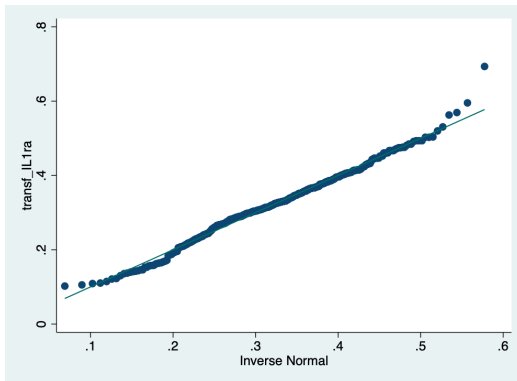
Distribution of transformed dependent variable?



Not perfect, but improved

Part II: Exploratory Analysis

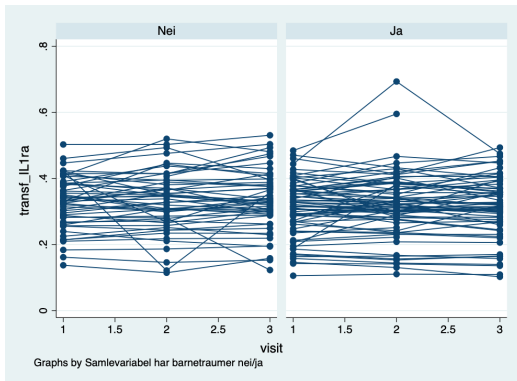
QQ-plot



Acceptable

Part II: Exploratory Plot

Does (transformed)IL1ra vary with "Barndoms-traumer" (0/1)?



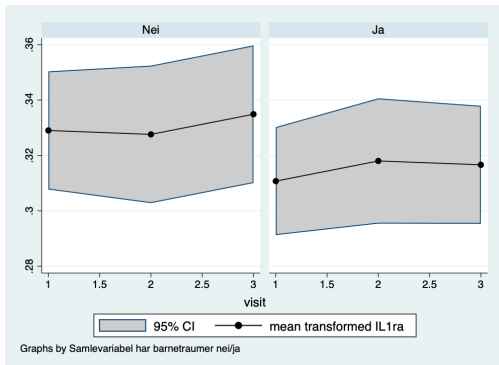
Difficult to read

Part II: Exploratory Plot

Collapsed data: Mean and 95% Conf Int for each trauma group and each visit

```
. collapse (mean) middel=transf_IL1ra (sd) stddev =transf_IL1ra (count) n=transf_IL1ra ,/*  
> */ by (visit Barnetraumer)  
  
. generate High=middel+2*stddev/sqrt(n)  
  
. generate Low=middel-2*stddev/sqrt(n)  
  
. twoway (rarea Low High visit, bfcOLOR(gs12) sort)/*  
> */ (connected middel visit, mcolor(black) /*  
> */ clcolor(black) sort) , by(Barnetraumer)/*  
> */ legend(order(1 "95% CI" 2 "mean transformed IL1ra"))
```

Part II: Exploratory Plot



Easy to read

Part II: Analysis

First Model

```
.xtmixed transfILlra age sex i.visit##i.Barnetraumer  
c.scl90gsi##i.Barnetraumer ||id:
```

transf_ILlra	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.0000487	.0006784	-0.07	0.943	-.0013784	.001281
sex	.0210534	.0173653	1.21	0.225	-.012982	.0550888
visit						
2	-.0004321	.0074142	-0.06	0.954	-.0149636	.0140995
3	.005738	.0077441	0.74	0.459	-.0094401	.0209161
Barnetraumer						
Ja	-.0072582	.0235255	-0.31	0.758	-.0533673	.0388509
visit#Barnetraumer						
2#Ja	.0069384	.0094682	0.73	0.464	-.0116189	.0254958
3#Ja	-.001213	.0098799	-0.12	0.902	-.0205773	.0181512
scl90gsi	.001515	.0081944	0.18	0.853	-.0145456	.0175757
Barnetraumer#c.scl90gsi						
Ja	-.0106267	.0114091	-0.93	0.352	-.0329881	.0117347
_cons	.295985	.0481968	6.14	0.000	.201521	.390449

Some model reduction....
Similar to ordinary regressions

Part II: Analysis

Last Model

```
.xtmixed transfl1ra age sex ||id:
```

transf_IL1ra	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.000087	.0006801	-0.13	0.898	-.00142	.0012461
sex	.0171074	.0172231	0.99	0.321	-.0166493	.0508642
_cons	.2967146	.0474756	6.25	0.000	.2036642	.3897651

Random-effects Parameters		Estimate	Std. Err.	[95% Conf. Interval]	
id: Identity					
	var(_cons)	.0065853	.0008746	.0050761	.0085432
	var(Residual)	.0012236	.0001097	.0010264	.0014586

LR test vs. linear model: `chibar2(01) = 332.97` Prob >= `chibar2 = 0.0000`

Part II: Analysis

Intra Class Correlation

```
.estat icc
```

Residual intraclass correlation

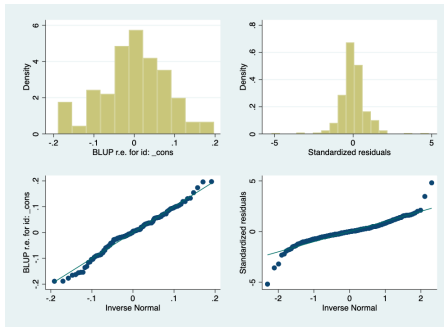
Level	ICC	Std. Err.	[95% Conf. Interval]	
id	.8433111	.0216114	.7961684	.8811776

So 84% of the total random variation can be attributed to heterogeneity between patients

Part II: Model Validation

Check for normality

- ▶ residuals (within individuals) : `.predict res, rstandard`
- ▶ random effects (between individuals): `.predict ranef, reffects`



Part II: Other Issues

- ▶ Prediction from model - Original scale
- ▶ More complex variance structures
 - ▶ Multilevel
 - ▶ Nested
 - ▶ Crossed
 - ▶ Combination of these
- ▶ Various correlation structures
- ▶ Other types of data: Poisson, Binomial, Gamma, Survival,....
- ▶ Missing data and imputations

Thank you