Understanding Hierarchical Linear Models and their application to EEG

Cyril Pernet, PhD

Centre for Clinical Brain Sciences

The University of Edinburgh

EEGLAB Workshop, SanDiego – Nov 2018

cyril.pernet@ed.ac.uk @CyrilRPernet

Motivations

Motivation for hierarchical models

- Most often, we compute averages per condition and do statistics on peak latencies and amplitudes

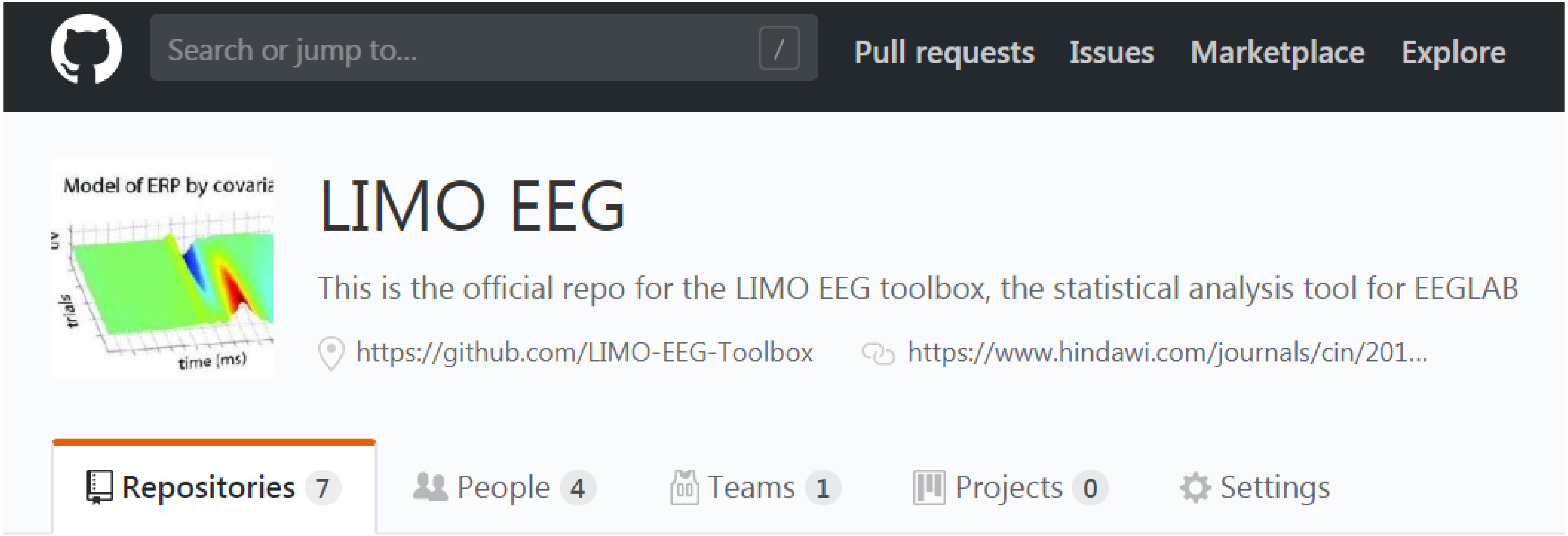
➢ Univariate methods extract information **among trials** in time and/or frequency across space

➢ Multivariate methods extract information across space, time, or both, **in individual trials**

➢ Averages don't account for trial variability, fixed effect can be biased – these methods allow to get around these problems

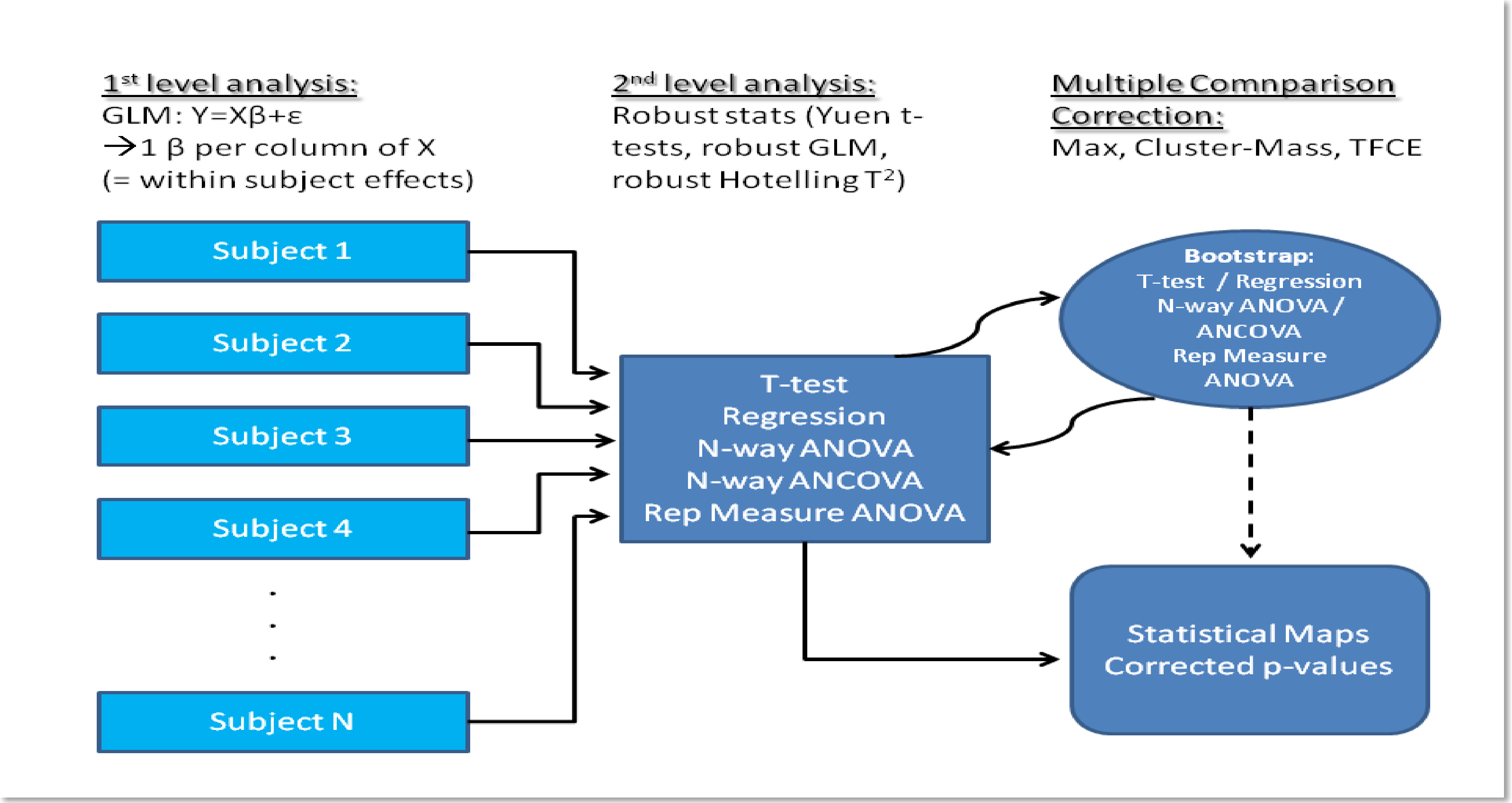
Pernet, Sajda & Rousselet – Single trial analyses, why bother? Front. Psychol., 2011, 2, 322

LIMO EEG Toolbox



Framework

Hierarchical Linear Model Framework



Fixed, Random, Mixed and Hierarchical

Fixed effect: Something the experimenter directly manipulates

y=XB+e data = beta * effects + error

y=XB+u+e data = beta * effects + constant subject effect + error

Random effect: Source of random variation e.g., individuals drawn (at random) from a population. **Mixed effect**: Includes both, the fixed effect (estimating the population level coefficients) and random effects to account for individual differences in response to an effect

Y=XB+Zu+e data = beta * effects + zeta * subject variable effect + error

Hierarchical models are a mean to look at mixed effects.

Hierarchical model = 2-stage LM

Single subject

Each subject's EEG trials are modelled
Single subject parameter estimates

1st level

Single subject parameter estimates or combinations taken to 2nd level

↓

Group/s of subjects

For a given effect, the whole group is modelled
Parameter estimates apply to group effect/s

2nd level

Group level of 2nd level parameter estimates are used to form statistics

Fixed vs Random

Fixed effects:

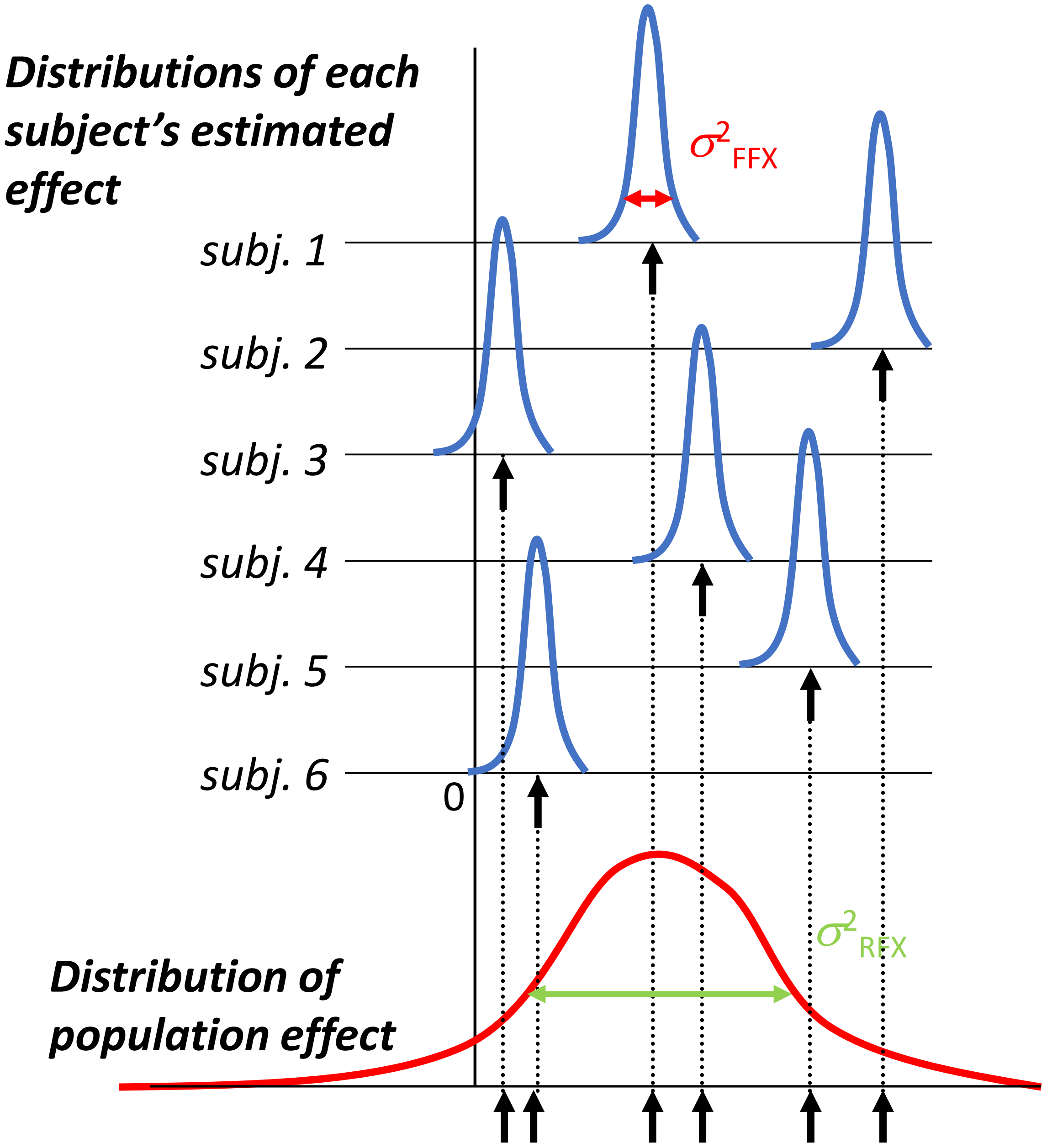
Intra-subjects variation

suggests all these subjects different from zero

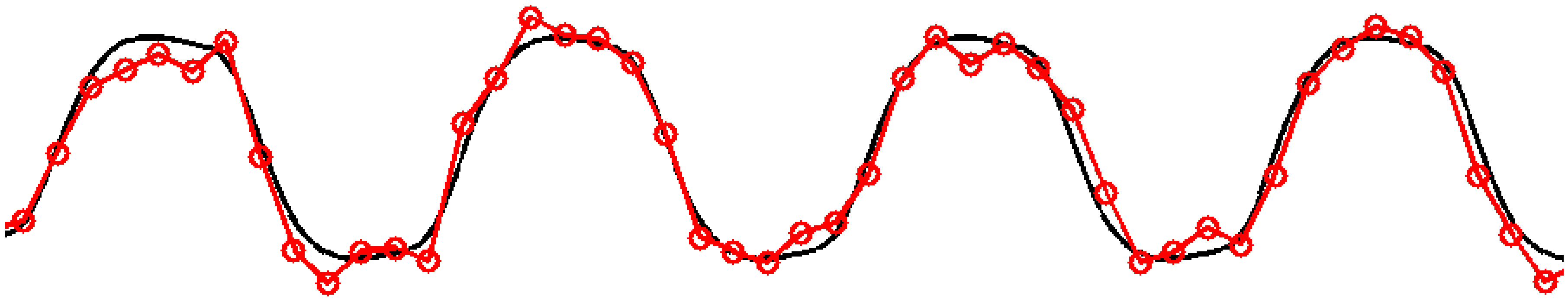
Random effects:

Inter-subjects variation

suggests population not different from zero

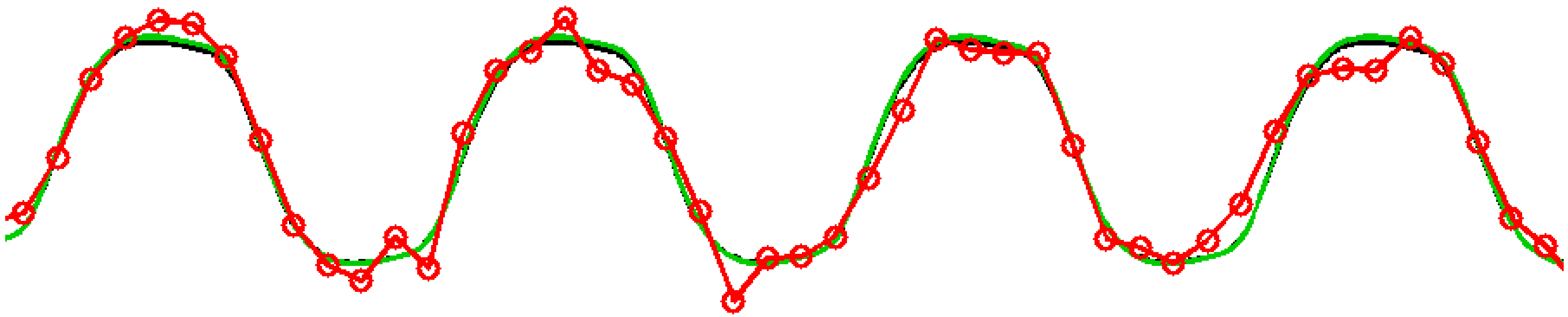


Fixed effects



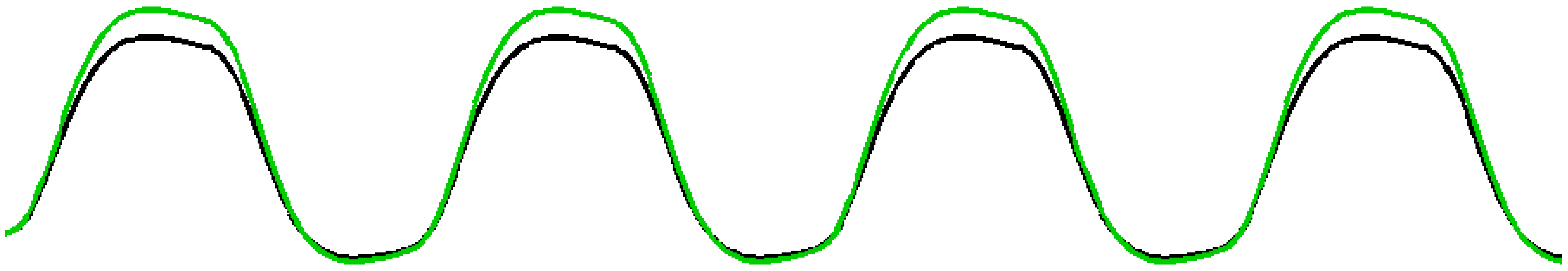
- Only source of variation (over trials) is measurement error
- True response magnitude is *fixed*

Random effects



- Two sources of variation
  - measurement errors
  - response magnitude (over subjects)
- Response magnitude is *random*
  - each subject has random magnitude

Random effects

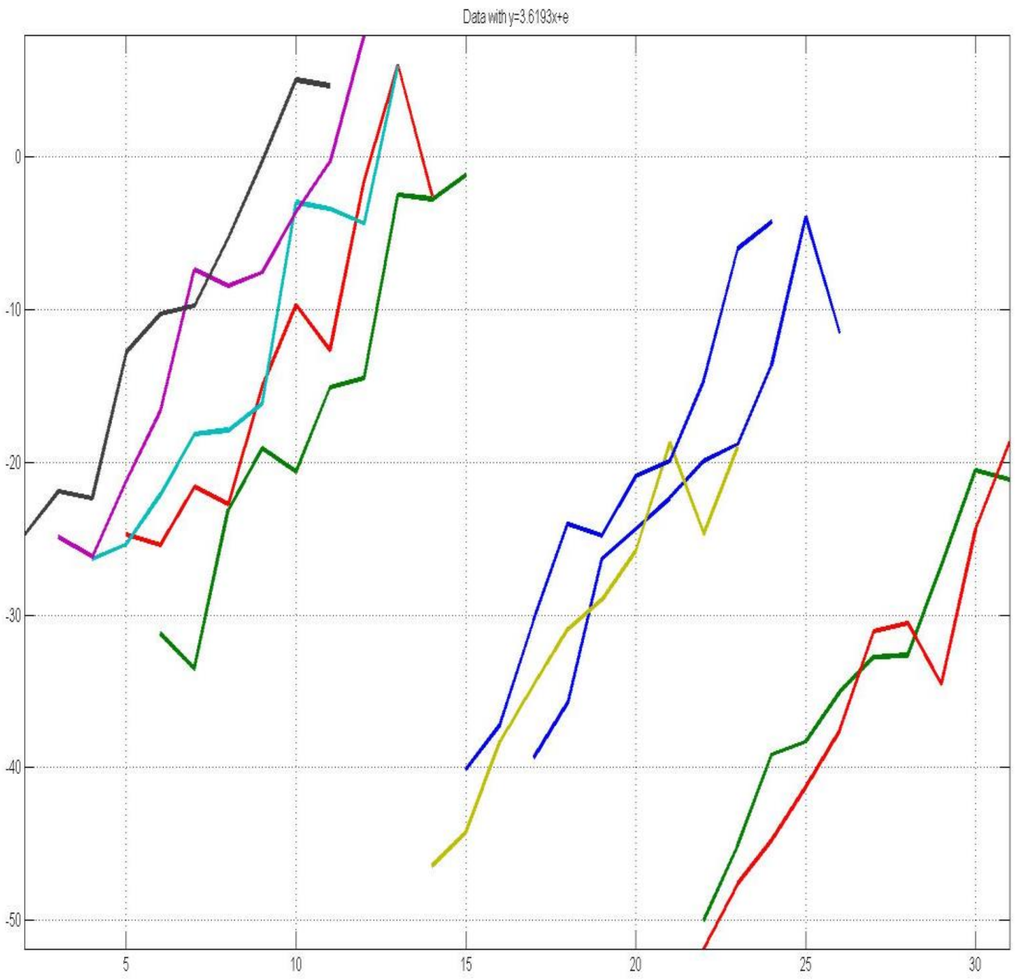


- Two sources of variation
  - measurement errors
  - response magnitude (over subjects)
- Response magnitude is *random*
  - each subject has random magnitude
  - but note, population mean magnitude is *fixed*

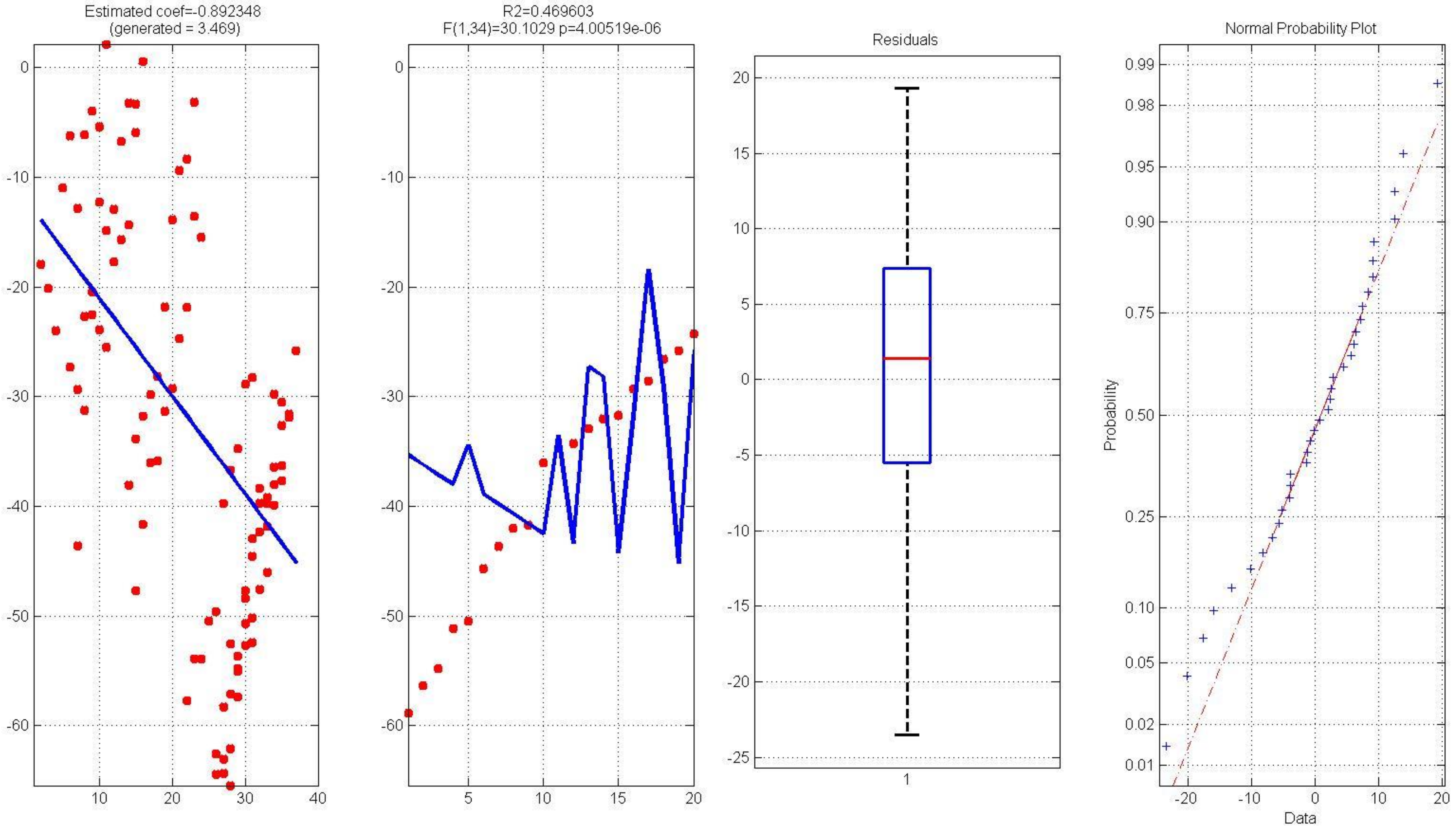
An example

<u>Example</u>: present stimuli from intensity -5 units to +5 units around the subject perceptual threshold and measure RT

→ There is a strong positive effect of intensity on responses

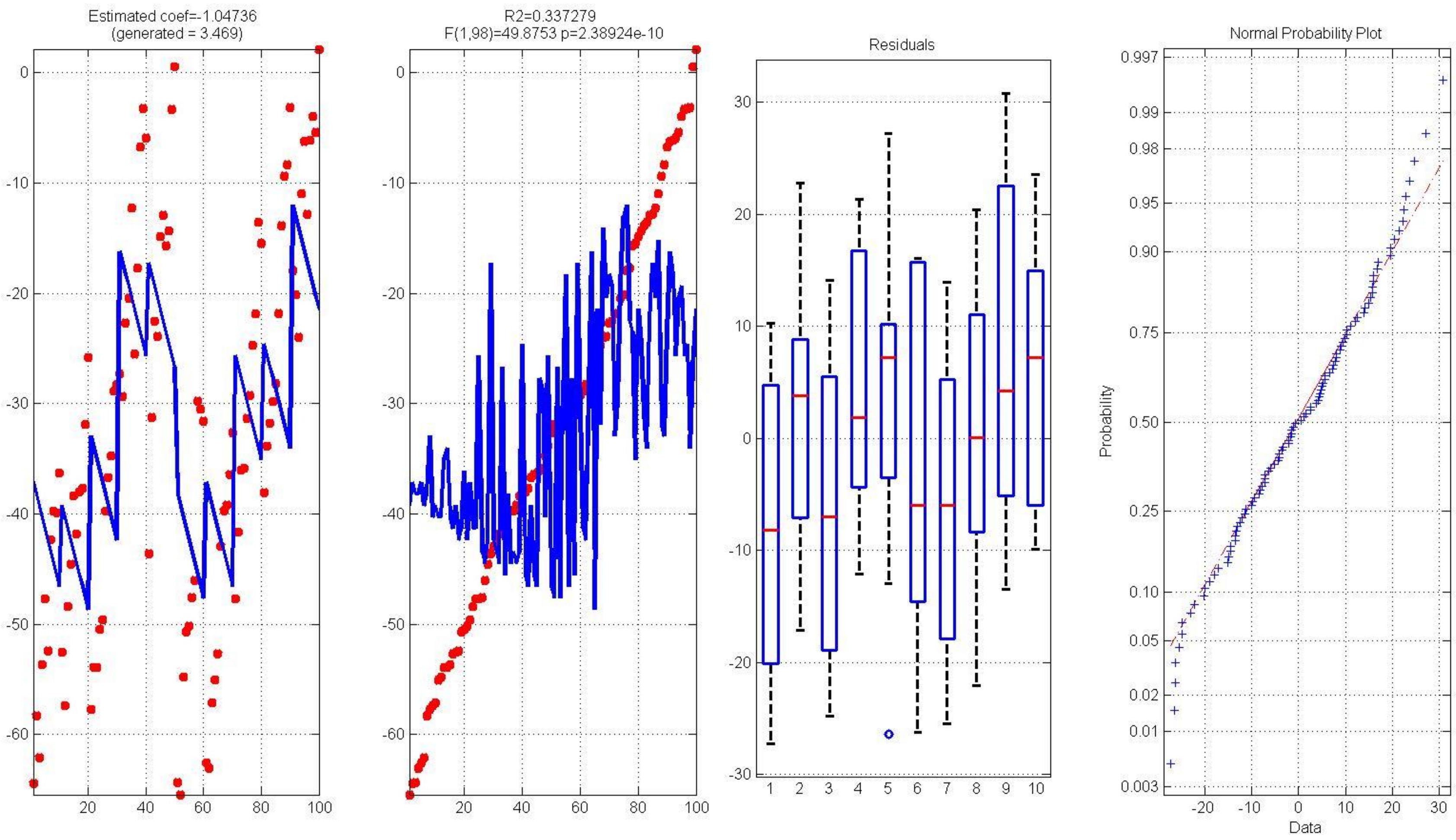


Fixed Effect Model 1: average subjects



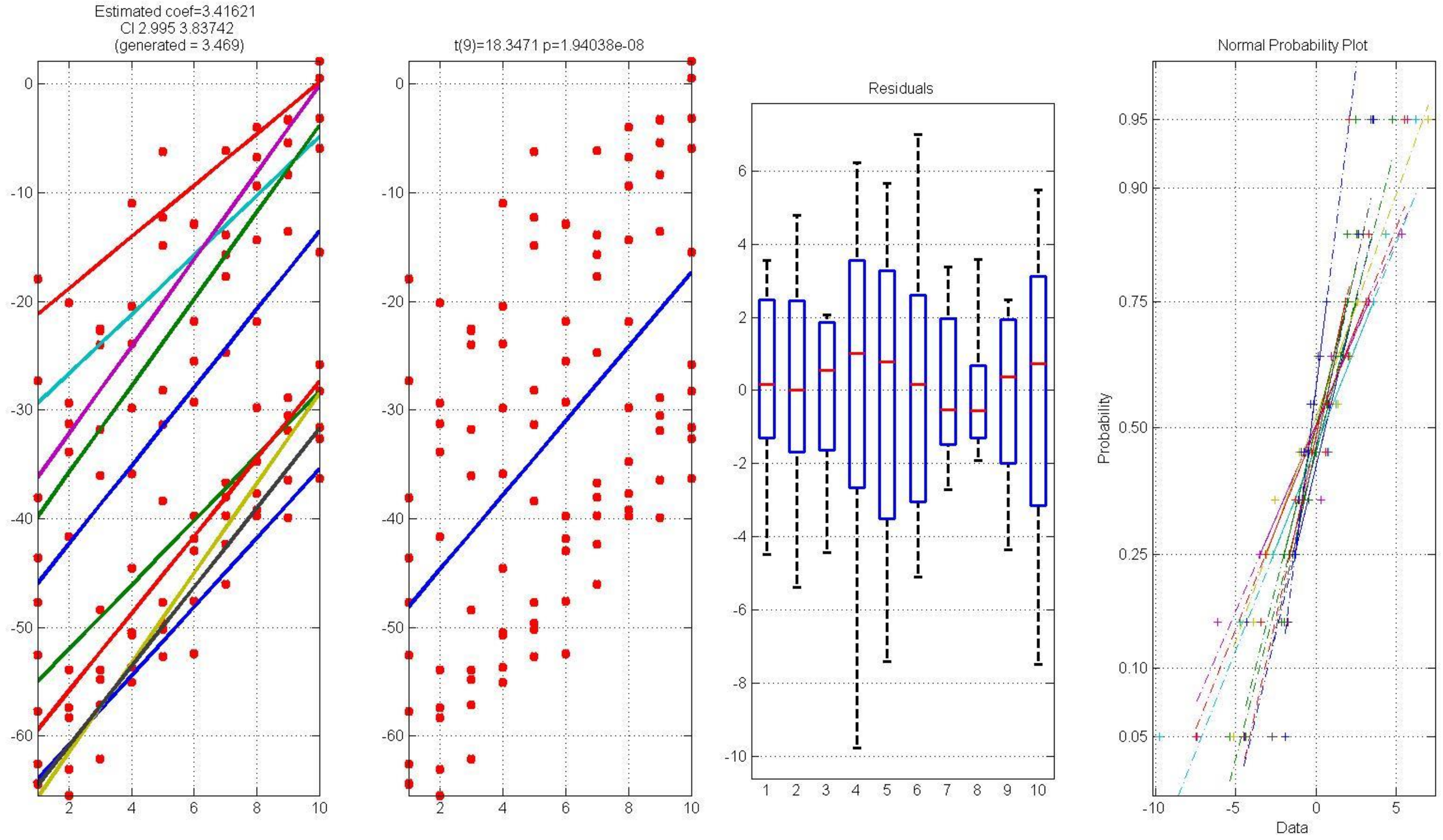
Fixed effect without subject effect → negative effect

Fixed Effect Model 2: constant over subjects



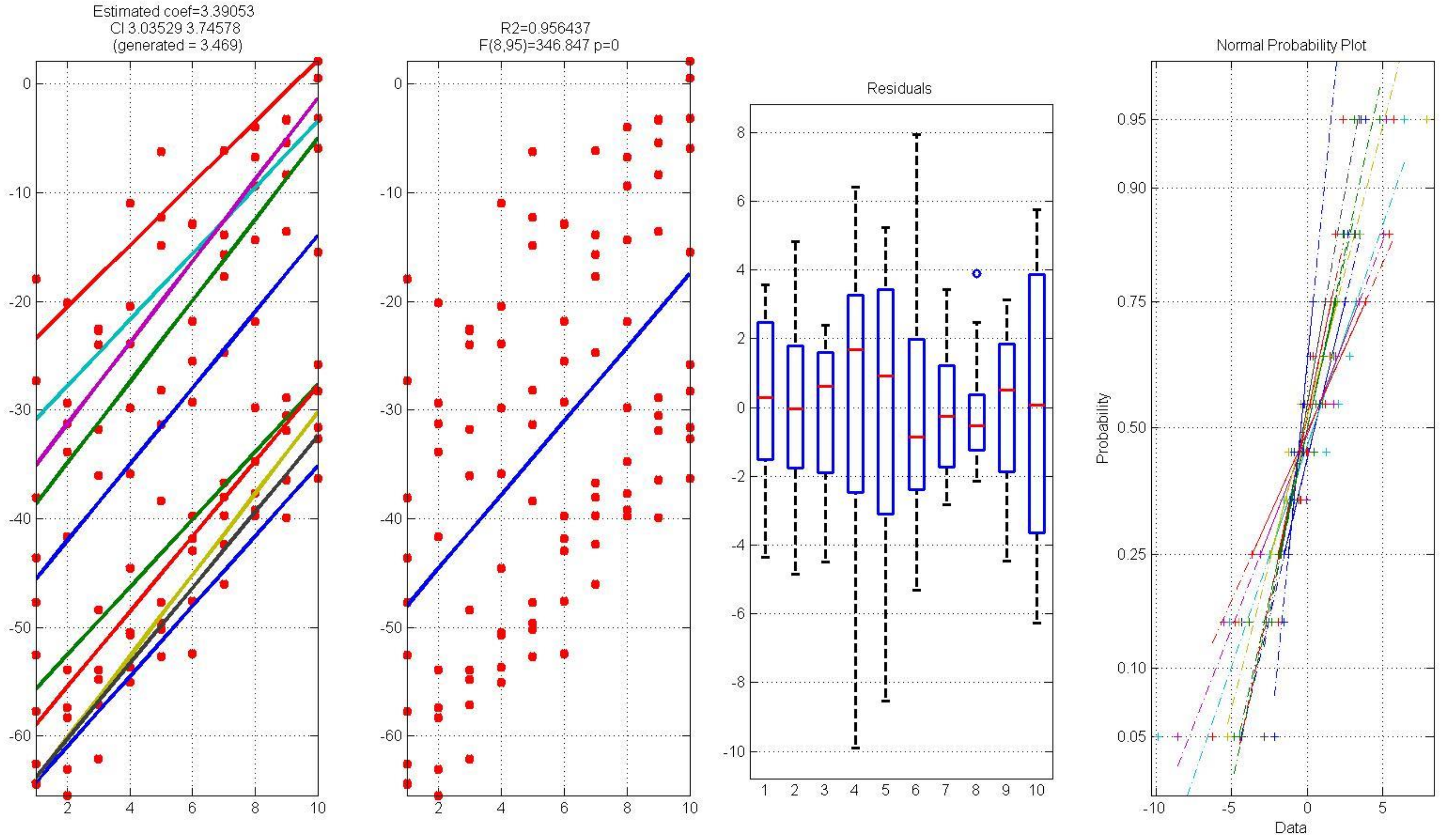
Fixed effect with a constant (fixed) subject effect → positive effect but biased result

HLM: random subject effect



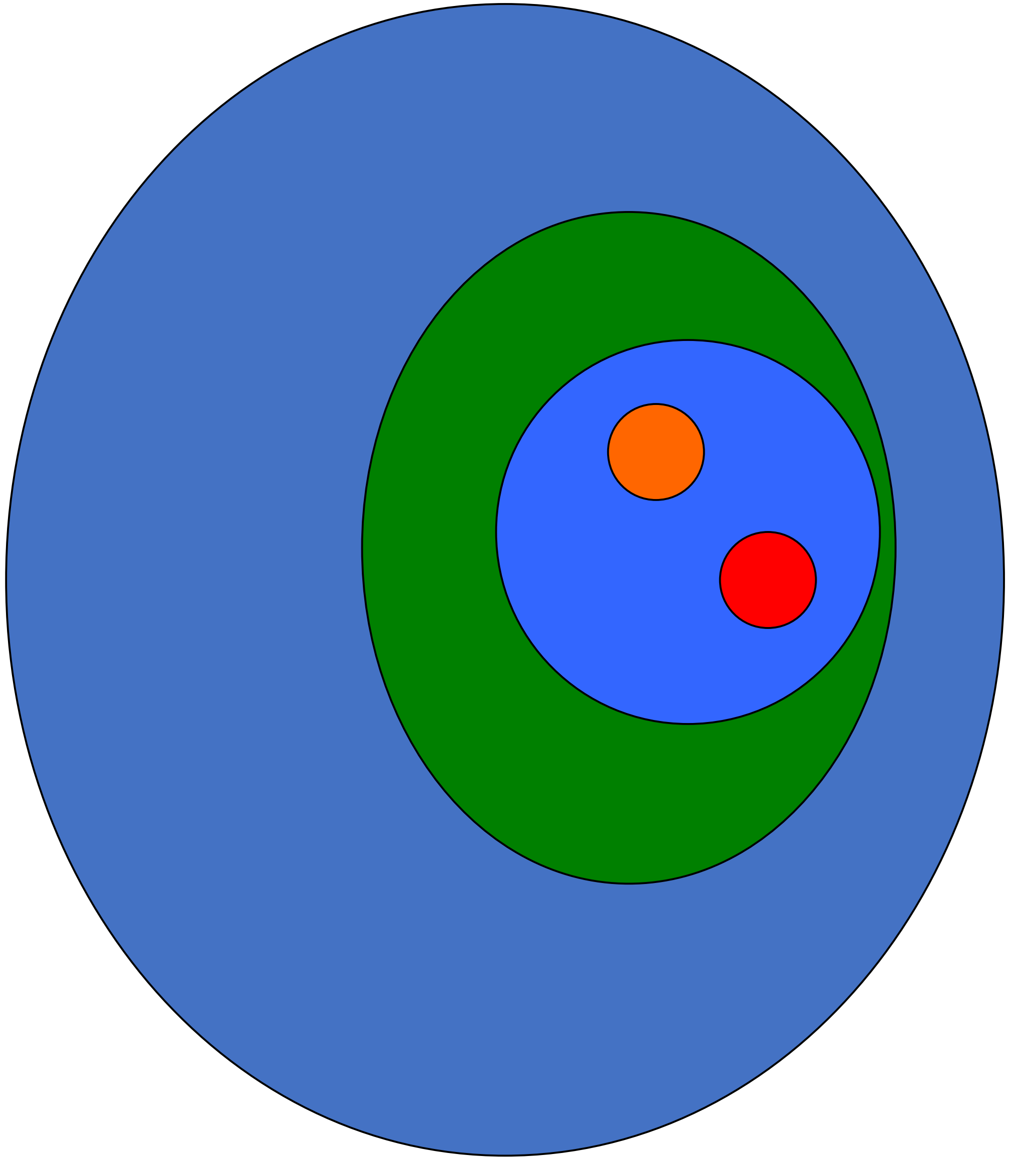
Mixed effect with a random subject effect → positive effect with good estimate of the truth

MLE: random subject effect



Mixed effect with a random subject effect → positive effect with good estimate of the truth

The GLM Family



T-tests

Simple regression

ANOVA

Multiple regression

General linear model

- Mixed effects/hierarchical
- Timeseries models (e.g., autoregressive)
- Robust regression
- Penalized regression (LASSO, Ridge)

Generalized linear models

- Non-normal errors
- Binary/categorical outcomes (logistic regression)

One-step solution

Iterative solutions (e.g., IWLS)

Tor Wager’s slide

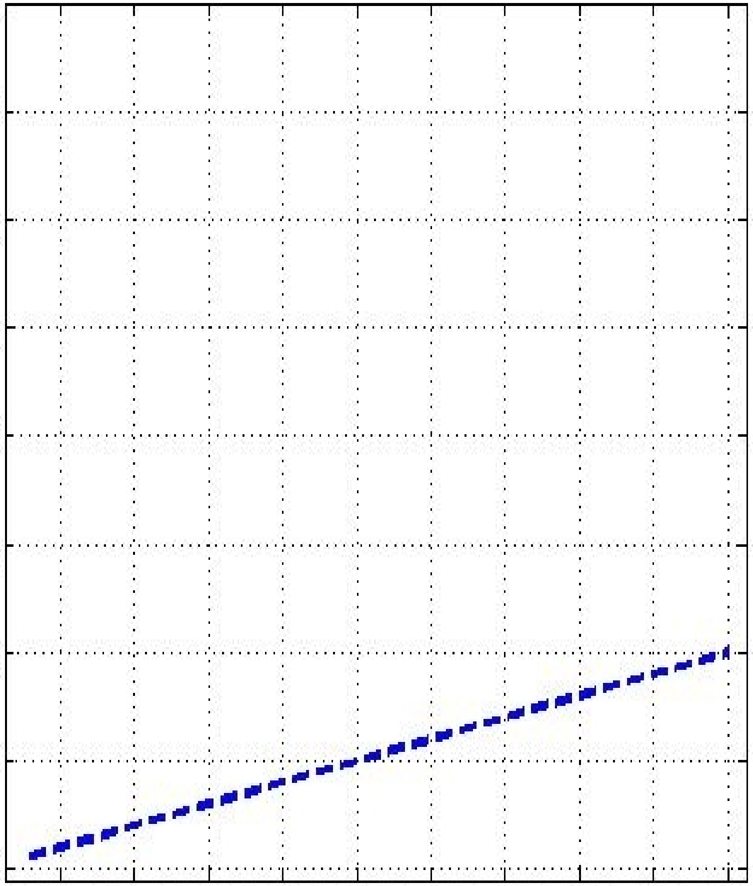
What is a linear model?

- An equation or a set of equations that models data and which corresponds geometrically to straight lines, planes, hyper-planes and satisfy the properties of additivity and scaling.

- Simple regression: $y = \beta 1x+\beta 2+\varepsilon$
- Multiple regression: $y = \beta 1x1+\beta 2x2+\beta 3+\varepsilon$
- One way ANOVA: $y = u+\alpha i+\varepsilon$
- Repeated measure ANOVA: $y=u+\alpha i+\varepsilon$
- ...

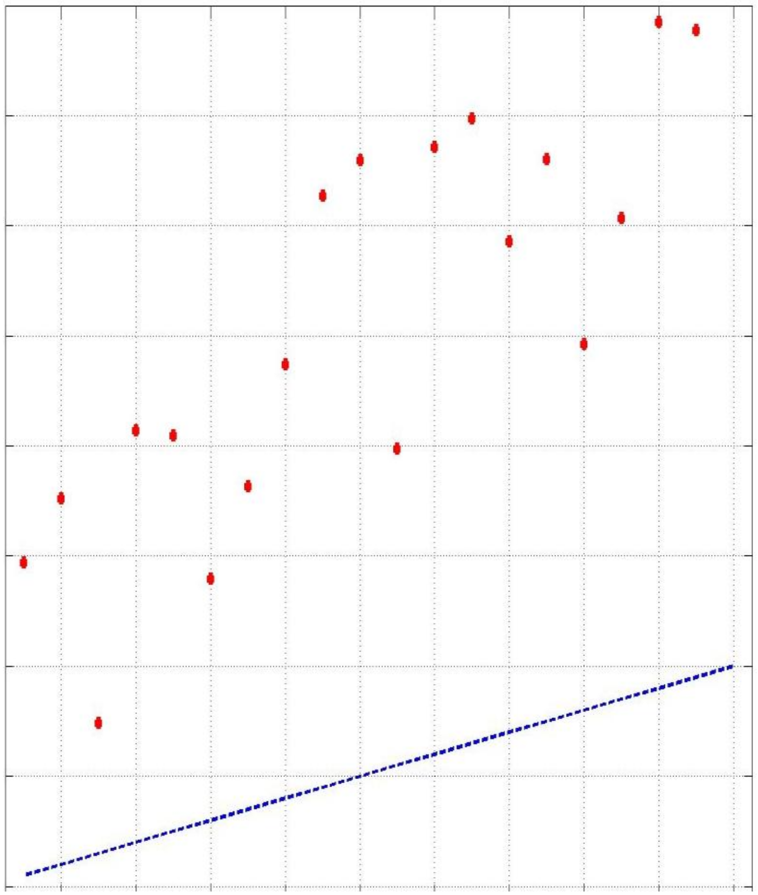
A regression is a linear model

- We have an experimental measure x (e.g. stimulus intensity from 0 to 20)



A regression is a linear model

- We have an experimental measure x (e.g. stimulus intensity from 0 to 20)
- We then do the expe and collect data y (e.g. RTs)



A regression is a linear model

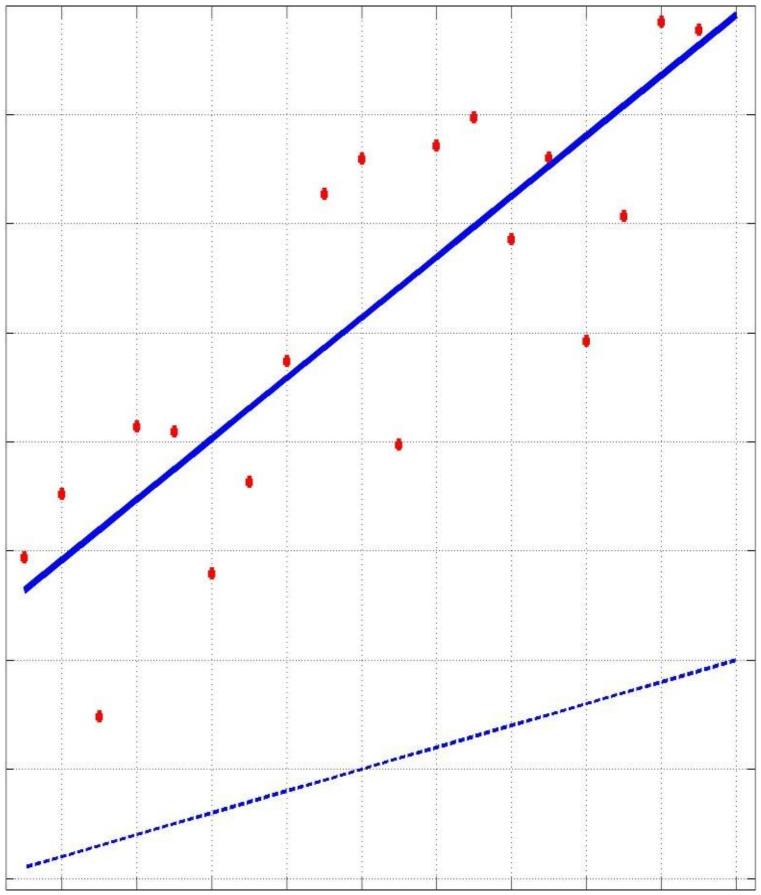
- We have an experimental measure x (e.g. stimulus intensity from 0 to 20)

- We then do the expe and collect data y (e.g. RTs)

- Model: $y = \beta 1 x + \beta 2$

- Do some maths / run a software to find $\beta 1$ and $\beta 2$

- $\hat{y} = 2.7x + 23.6$

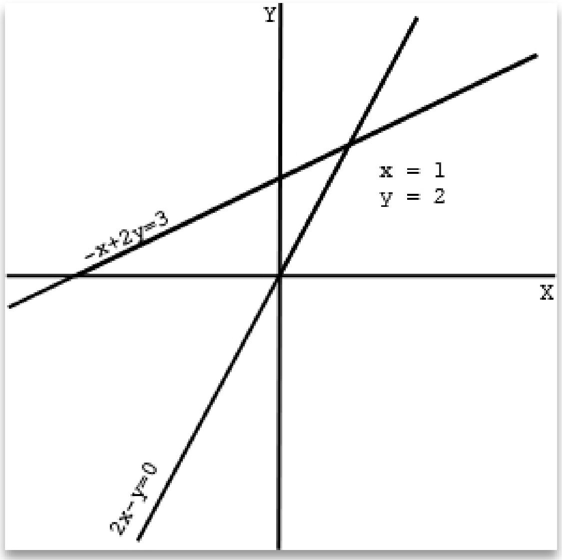


Linear algebra for regression

- Linear algebra has to do with solving linear systems, i.e. a set of linear equations
- For instance we have observations (y) for a stimulus characterized by its properties x_1 and x_2 such as $y = x_1 \beta_1 + x_2 \beta 2$

$$2\beta 1 - \beta 2 = 0$$
$$-\beta 1 + 2\beta 2 = 3$$

$$\beta 1 = 1 \; ; \; \beta 2 = 2$$



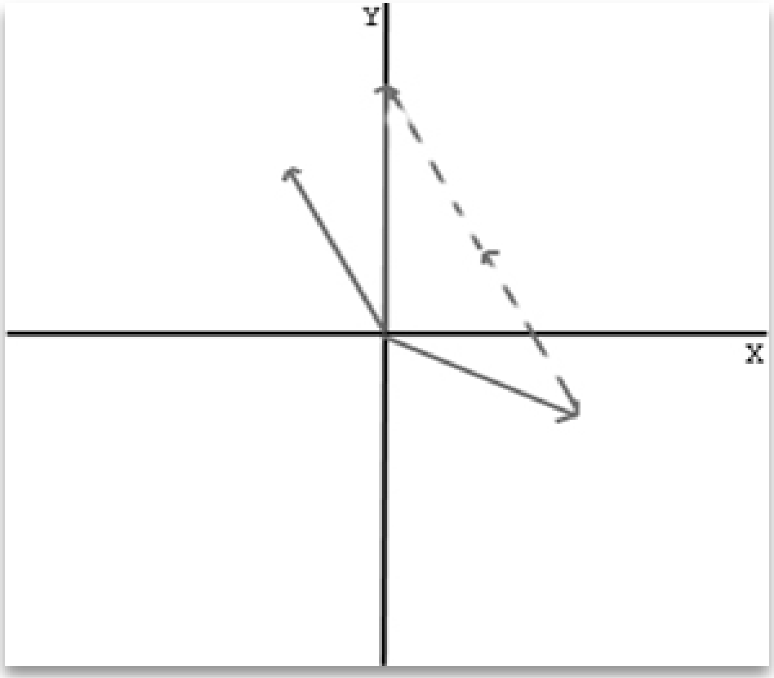
Linear algebra for regression

- With matrices, we change the perspective and try to combine columns instead of rows, i.e. we look for the coefficients with allow the linear combination of vectors

$$2\beta1 - \beta2 = 0$$
$$-\beta1 + 2\beta2 = 3$$

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} [\beta1 \beta2] = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$\beta1 = 1 \; ; \; \beta2 = 2$$



Linear algebra for ANOVA

- In text books we have y = u + xi + ε, that is to say the data (e.g. RT) = a constant term (grand mean u) + the effect of a treatment (xi) and the error term (ε)
- In a regression xi takes several values like e.g. [1:20]
- In an ANOVA xi is designed to represent groups using 1 and 0

Linear algebra for ANOVA

Y	Gp
8	1
9	1
7	1
5	2
7	2
3	2
3	3
4	3
1	3
6	4
4	4
9	4

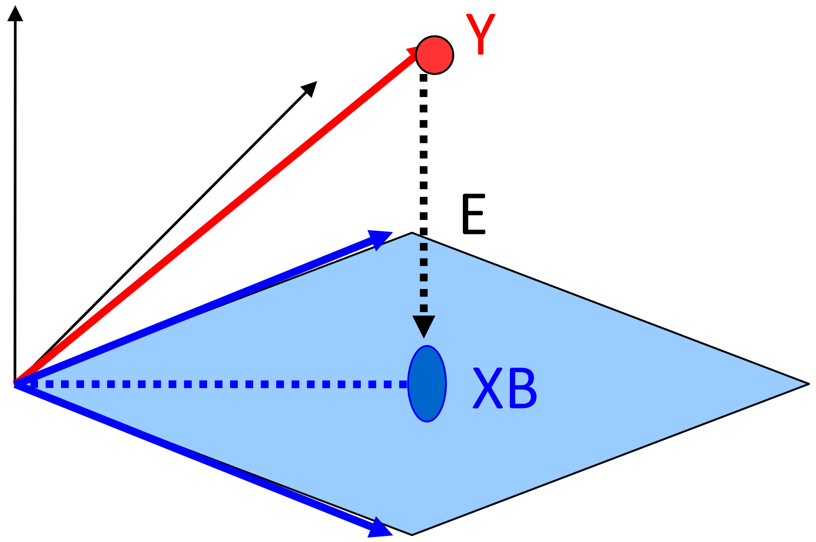
y(1..3)1= 1x1+0x2+0x3+0x4+c+e11
y(1..3)2= 0x1+1x2+0x3+0x4+c+e12
y(1..3)3= 0x1+0x2+1x3+0x4+c+e13
y(1..3)4= 0x1+0x2+0x3+1x4+c+e13

$$\begin{Bmatrix} 8 \\ 9 \\ 7 \\ 5 \\ 7 \\ 3 \\ 3 \\ 4 \\ 1 \\ 6 \\ 4 \\ 9 \end{Bmatrix} = \begin{Bmatrix} 1\,0\,0\,0\,1 \\ 1\,0\,0\,0\,1 \\ 1\,0\,0\,0\,1 \\ 0\,1\,0\,0\,1 \\ 0\,1\,0\,0\,1 \\ 0\,1\,0\,0\,1 \\ 0\,0\,1\,0\,1 \\ 0\,0\,1\,0\,1 \\ 0\,0\,1\,0\,1 \\ 0\,0\,0\,1\,1 \\ 0\,0\,0\,1\,1 \\ 0\,0\,0\,1\,1 \end{Bmatrix} * \begin{Bmatrix} \beta 1 \\ \beta 2 \\ \beta 3 \\ \beta 4 \\ c \end{Bmatrix} + \begin{Bmatrix} e1 \\ \vdots \\ e13 \end{Bmatrix}$$

→ This is like the multiple regression except that we have ones and zeros instead of 'real' values so we can solve the same way

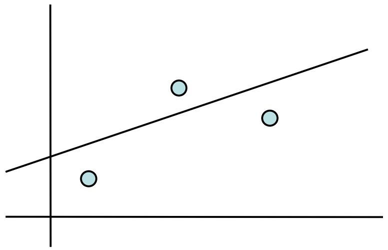
Linear Algebra, geometry and Statistics

- Y = 3 observations X = 2 regressors
- Y = XB+E → B = inv(X'X)X'Y → Y^=XB



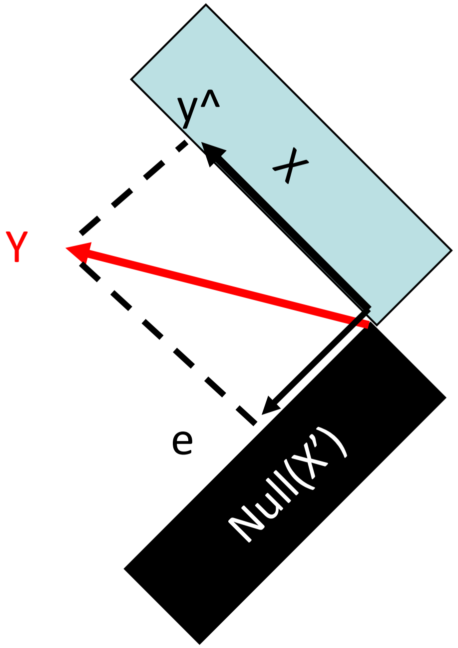
SS total = variance in Y
SS effect = variance in XB
SS error = variance in E
R2 = SS effect / SS total
F = SS effect/df / SS error/dfe

Linear Algebra, geometry and Statistics



$y = \beta x + c$

Projecting the points on the line at perpendicular angles minimizes the distance^2



Y = y^+e

P = X inv(X'X) X'

y^ = PY

e = (I-P)Y

An 'effect' is defined by which part of X to test (i.e. project on a subspace)

R0 = I - (X0*pinv(X0));

P = R0 - R;

Effect = (B'*X'*P*X*B);

Linear Algebra, geometry and Statistics

- Projections are great because we can now constrain Y^ to move along any combinations of the columns of X
- Say you now want to contrast gp1 vs gp2 in a ANOVA with 3 gp, do C = [1 -1 0 0]
- Compute B so we have XB based on the full model X then using P(C(X)) we project Y^ onto the constrained model (think doing a multiple regression gives different coef than multiple simple regression → project on different spaces)


UH-OH...LOOKS LIKE MATH ANXIETY...
GET THE INTEGRAL SIGN.

Application for EEG

Design considerations

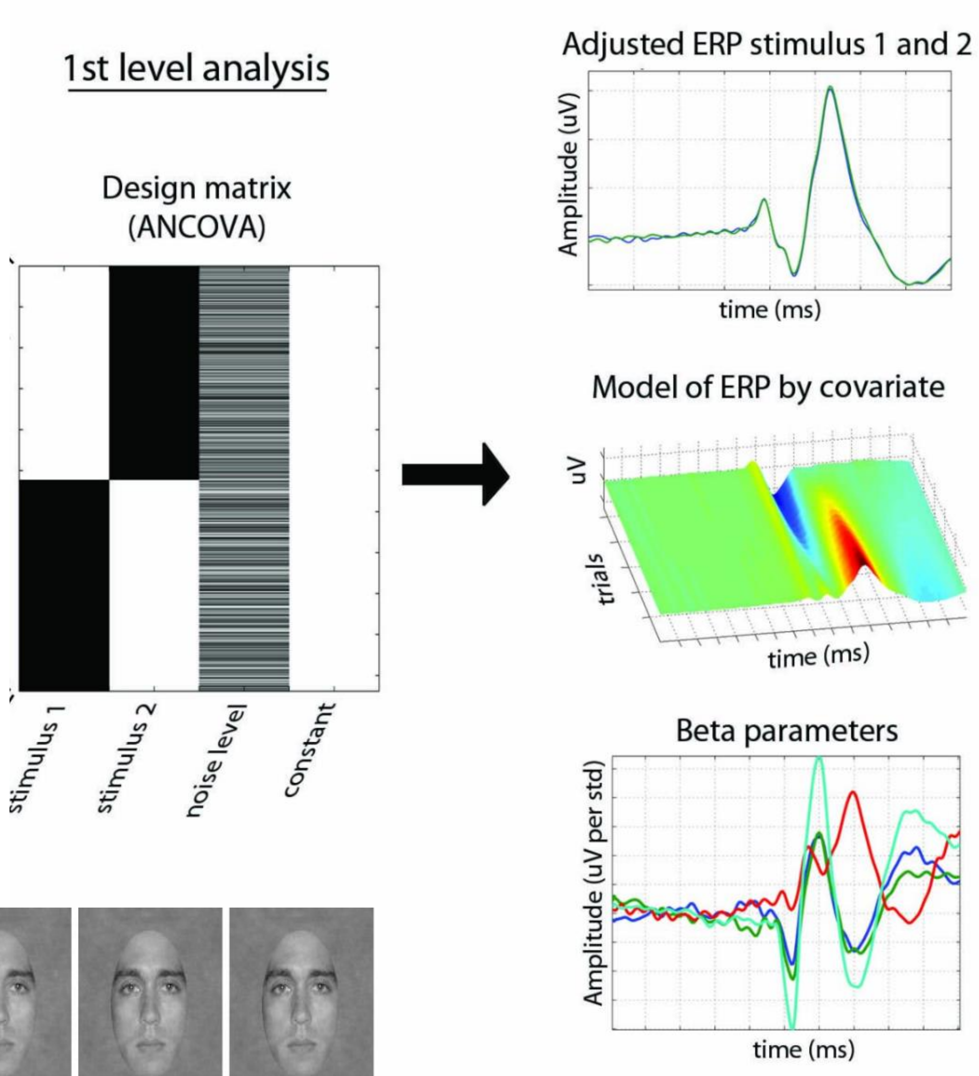
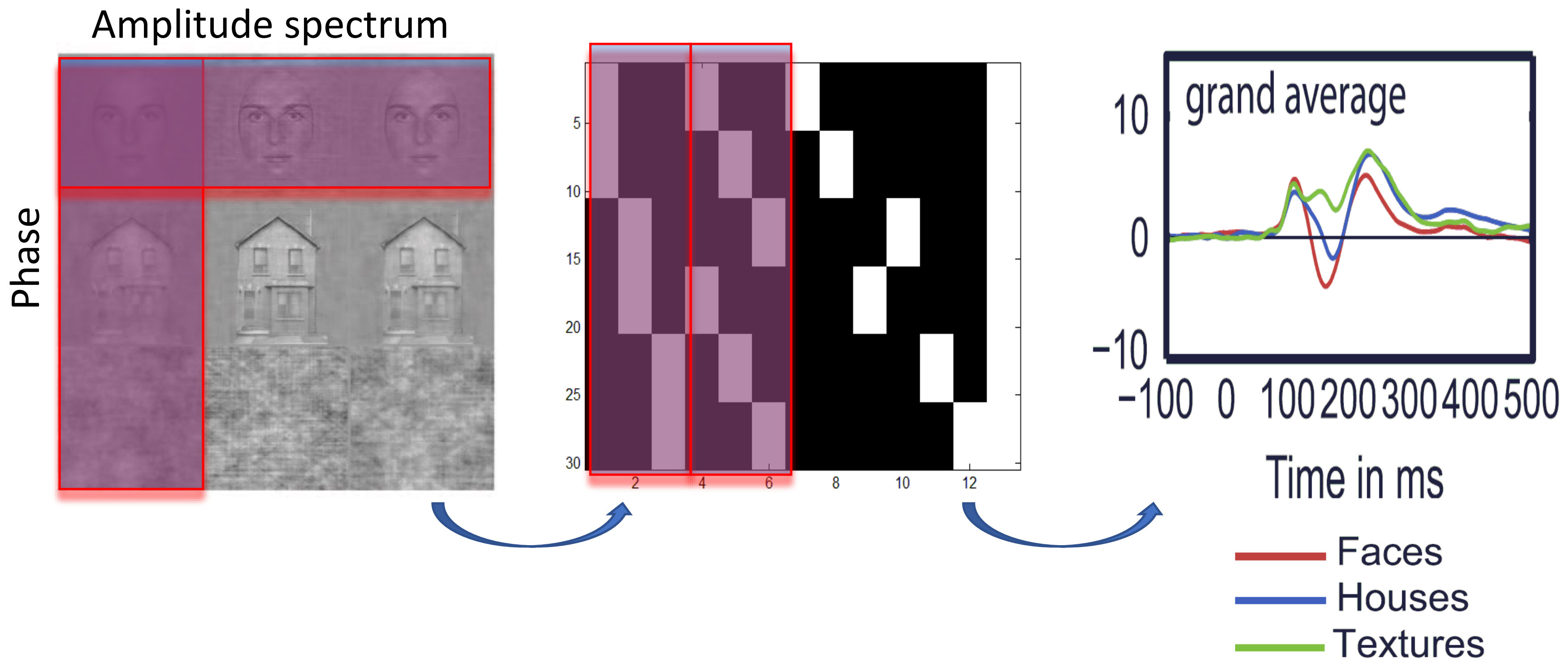


Illustration with a set of studies looking at the effect of stimulus phase information

[Rousselet, Pernet, Bennet, Sekuler (2008). Face phase processing. *BMC Neuroscience 9:98*]

Factorial Designs: N*N*N*…

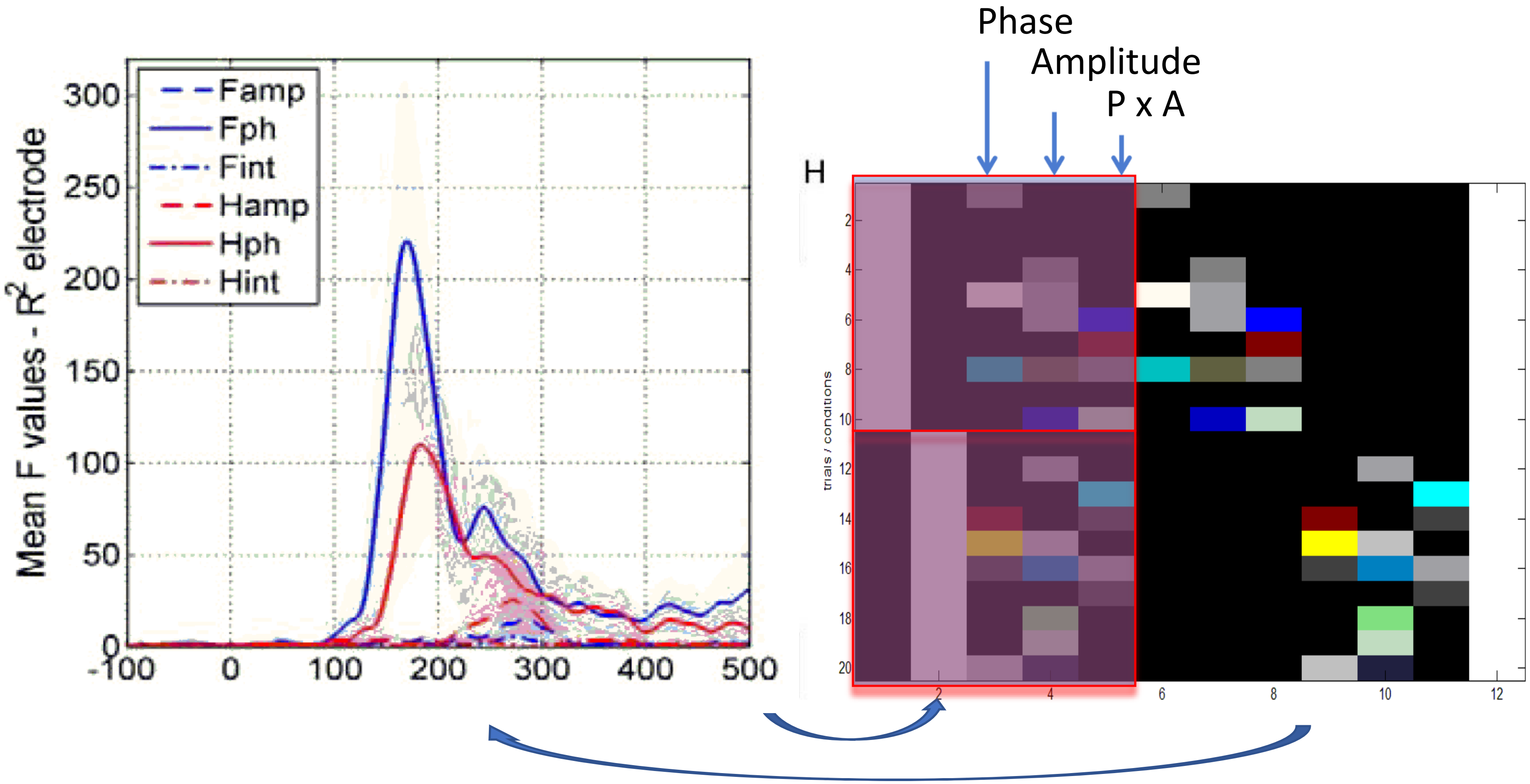
Categorical designs: Group level analyses of course but also Individual analyses with bootstrap



[Bienek, Pernet, Rousselet (2012). Phase vs Amplitude Spectrum. Journal of Vision 12(13), 1–24]

Regression based designs

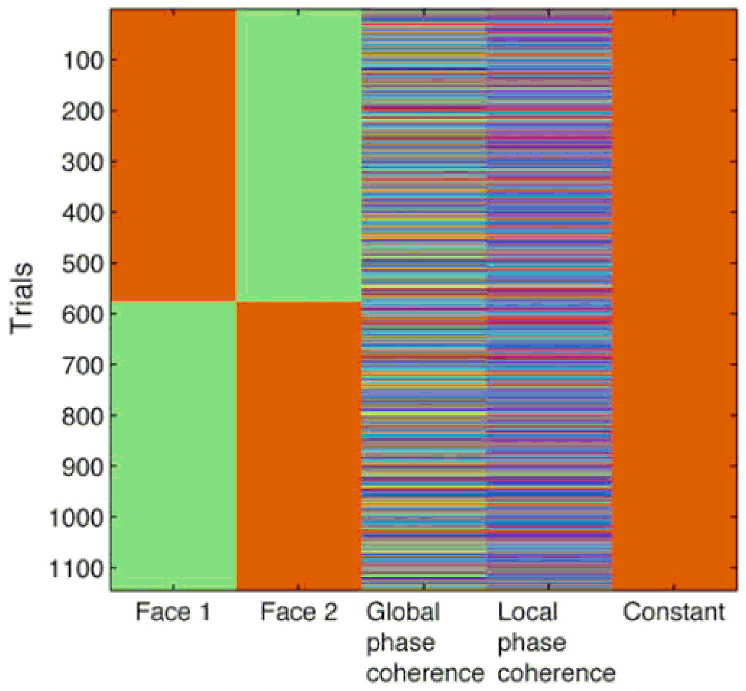
Mixed design: Control of low level physical properties

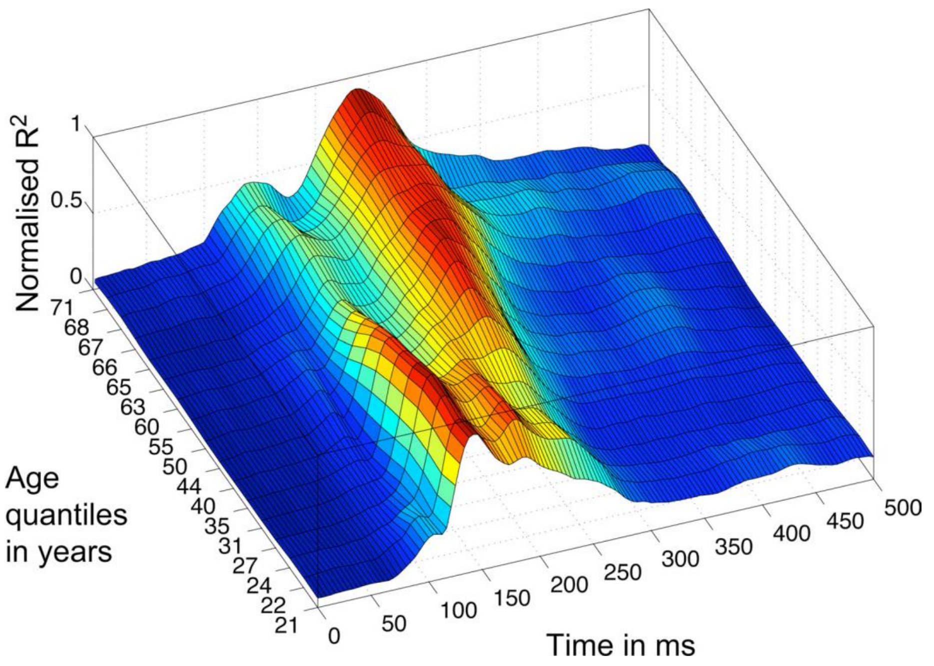


Bienek, Pernet, Rousselet (2012). Phase vs Amplitude Spectrum. Journal of Vision 12(13), 1–24

Regression based designs (2 levels)

Parametric designs:
study the effect of stimulus properties within subjects
effect of aging between subjects





Rousselet, Gaspar, Pernet, Husk, Bennett, Sekuler (2010). Aging and face perception. Front Psy

Conclusion

- HLM allows you to model any designs
- Not just designs, also confounds (e.g. stimulus properties)
- 1st level is like getting averages for each condition but better because (i) it removes subjects effect (ii) accounts for trial variability
- GLM is just your usual statistics but using generic approach, i.e. it's better because more flexible