Hierarchical Linear Modelling & the General Linear Model

Cyril Pernet, PhD

Centre for Clinical Brain Sciences

The University of Edinburgh

EEGLAB workshop – June 2019

cyril.pernet@ed.ac.uk @CyrilRPernet

Motivations

Motivation for whole channel/IC analyses

- **Data collection** consists in recording electromagnetic events over the whole brain and for a relatively long period of time, with regards to neural spiking. In the majority of cases, **data analysis** consists in looking where we have signal and restrict our analysis to these channels and components.

➢ Are we missing the forest by choosing working on a single, or a few trees?

➢ By analysing where we see an effect, we increase the type 1 FWER because the effect is partly driven by random noise (solved if chosen based on prior results or split the data)

Rousselet & Pernet – It's time to up the Game Front. Psychol., 2011, 2, 107

Motivation for hierarchical models

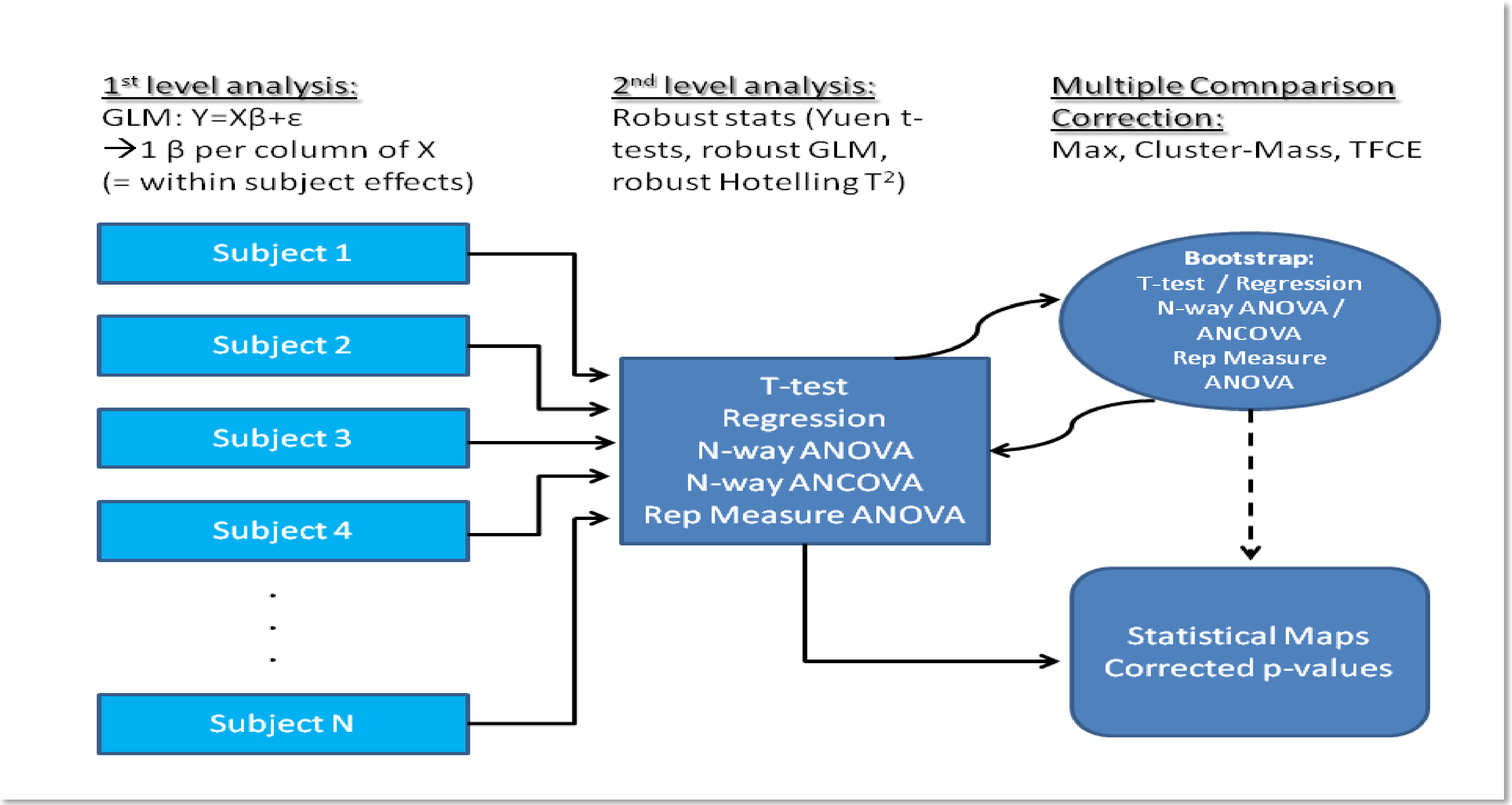
- Most often, we compute averages per condition and do statistics on peak latencies and amplitudes

- ➢ 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

Framework

LIMO Hierarchical Linear Model Framework



www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Electroencephalography -EEG

» Brain imaging

A multi-subject, multi-modal human neuroimaging dataset

Daniel G. Wakeman[1,2] & Richard N. Henson[2]

- *Scientific Data* **2**, Article number: 150001 (2015)
- doi:10.1038/sdata.2015.1

https://www.nature.com/articles/sdata20151

The Data

Famous Unfamiliar Scambled

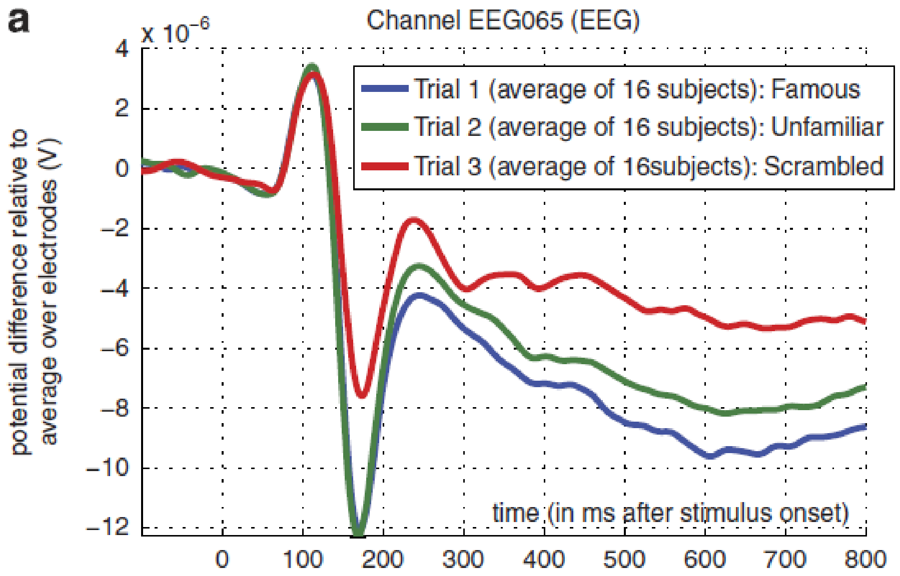
- 3 types of stimuli: Famous faces, Non-famous faces, Scrambled faces
- 3 levels of repetition: 1st time, 2nd time (right after), 3rd time (delayed)

→Priming experiment with a possible interaction with the type of stimuli.

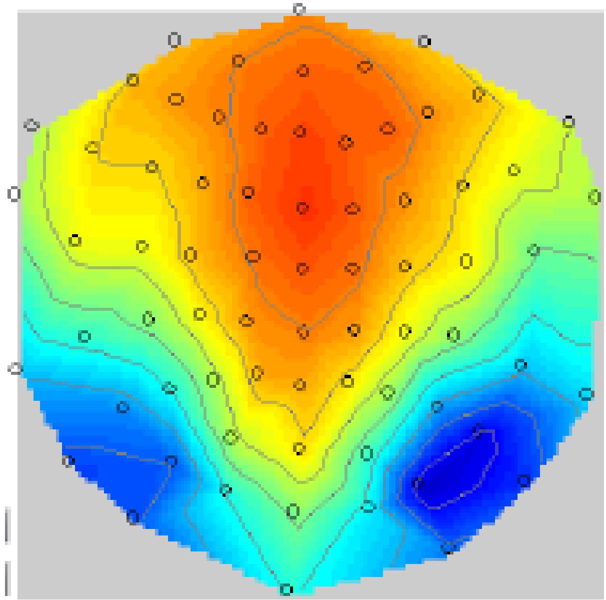
We need the conditions computed per subject (1st level) and then do the repeated measure ANOVA to test main effects and interactions.

What are we going to do?

- 1 – Replicate Henson et al. – faces vs. scrambled



Topography 170 ms



- 2 – learn about HLM and apply multiple comparison corrections

LET'S COMPUTE!

Hierarchical Linear Modelling

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:

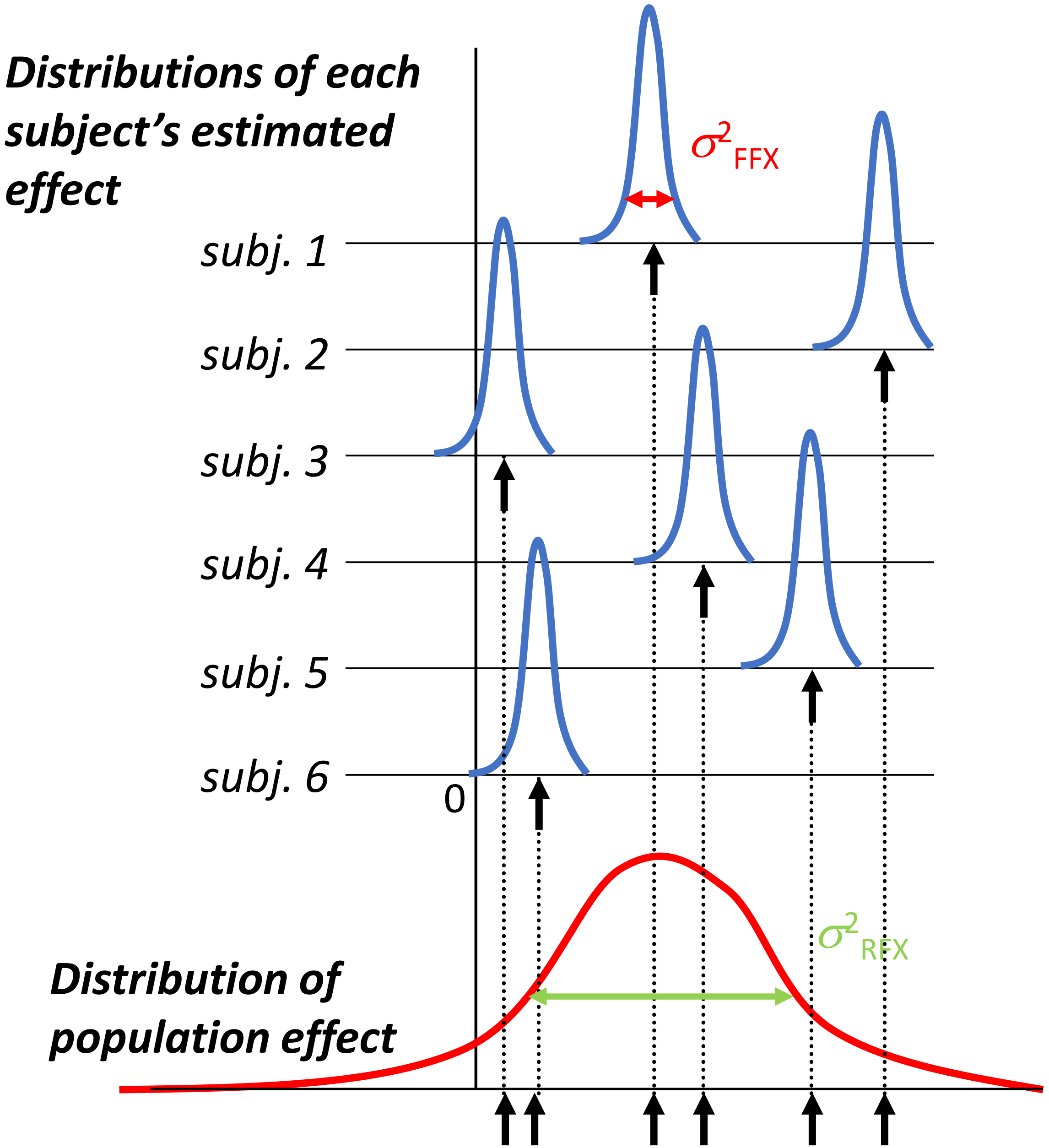
Intra-subjects variation

suggests all these subjects different from zero

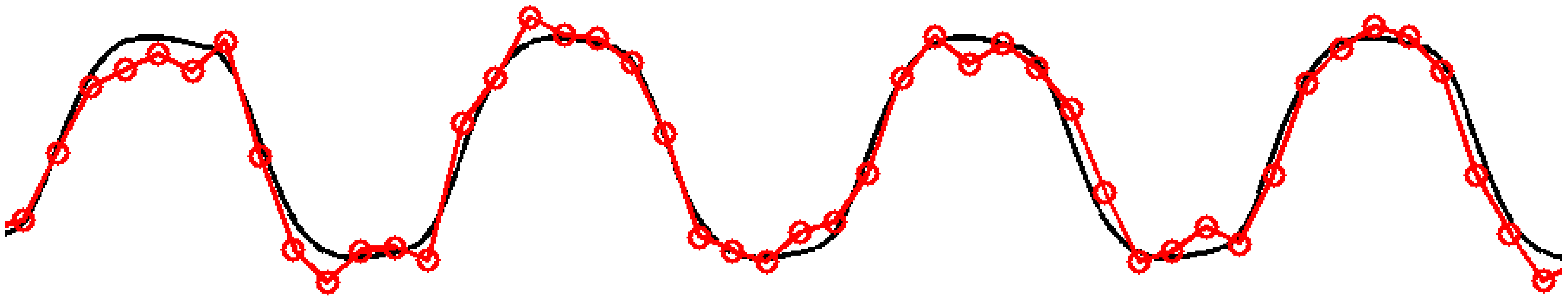
Random effects:

Inter-subjects variation

suggests population not different from zero

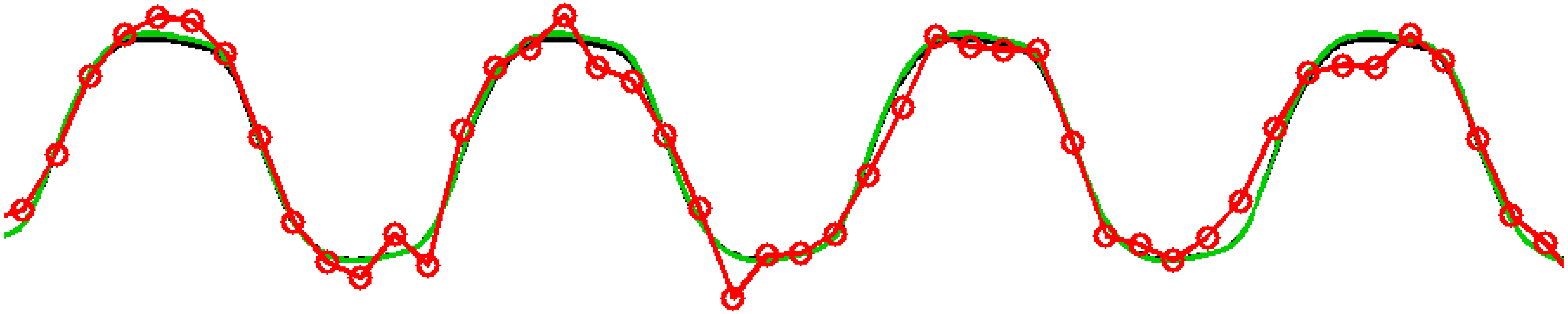


Fixed effects



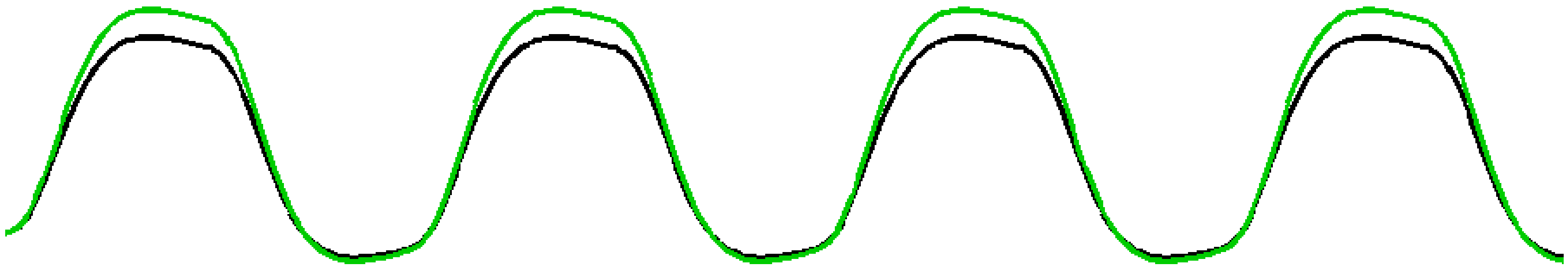
- Only source of variation (over trials) is measurement error
- True response magnitude is *fixed*

Random effects



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

Random effects

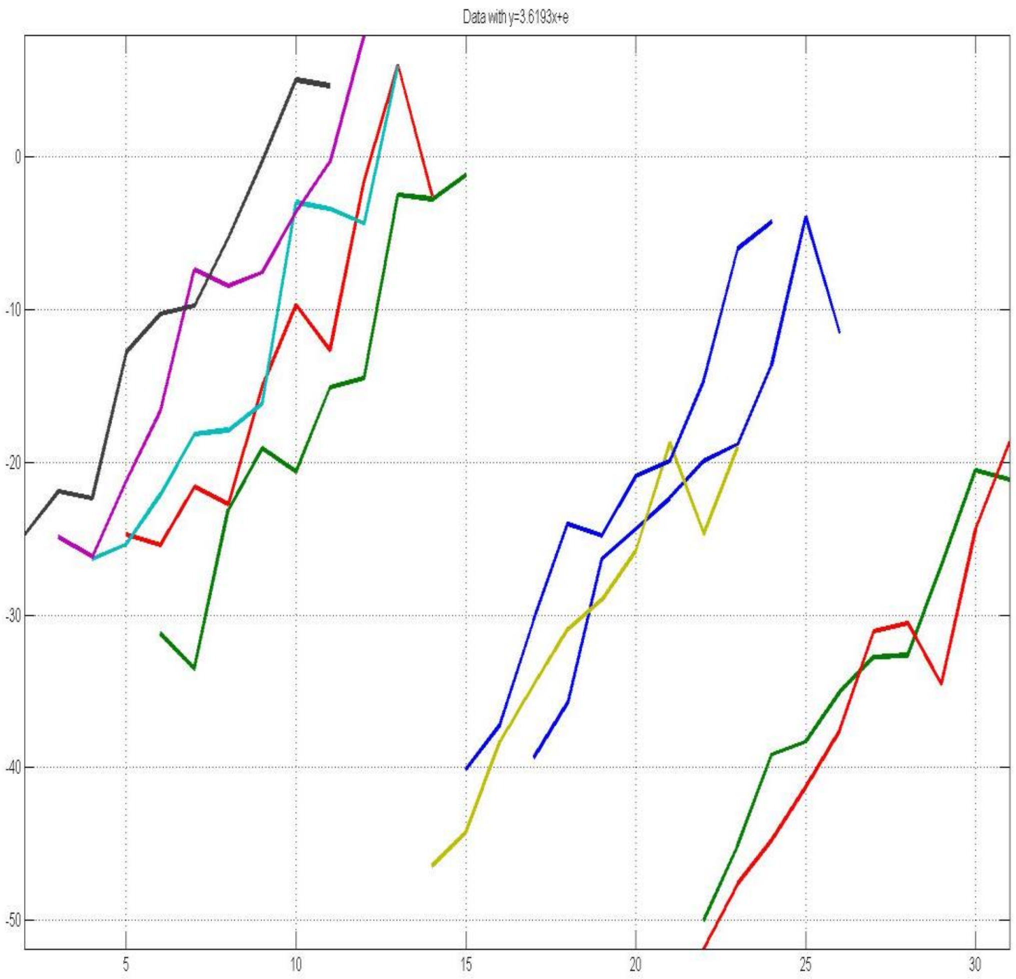


- Two sources of variation
  - measurement errors
  - response magnitude (over subjects)
- Response magnitude is *random*
  - each subject has random magnitude
  - but note, population mean magnitude is *fixed*

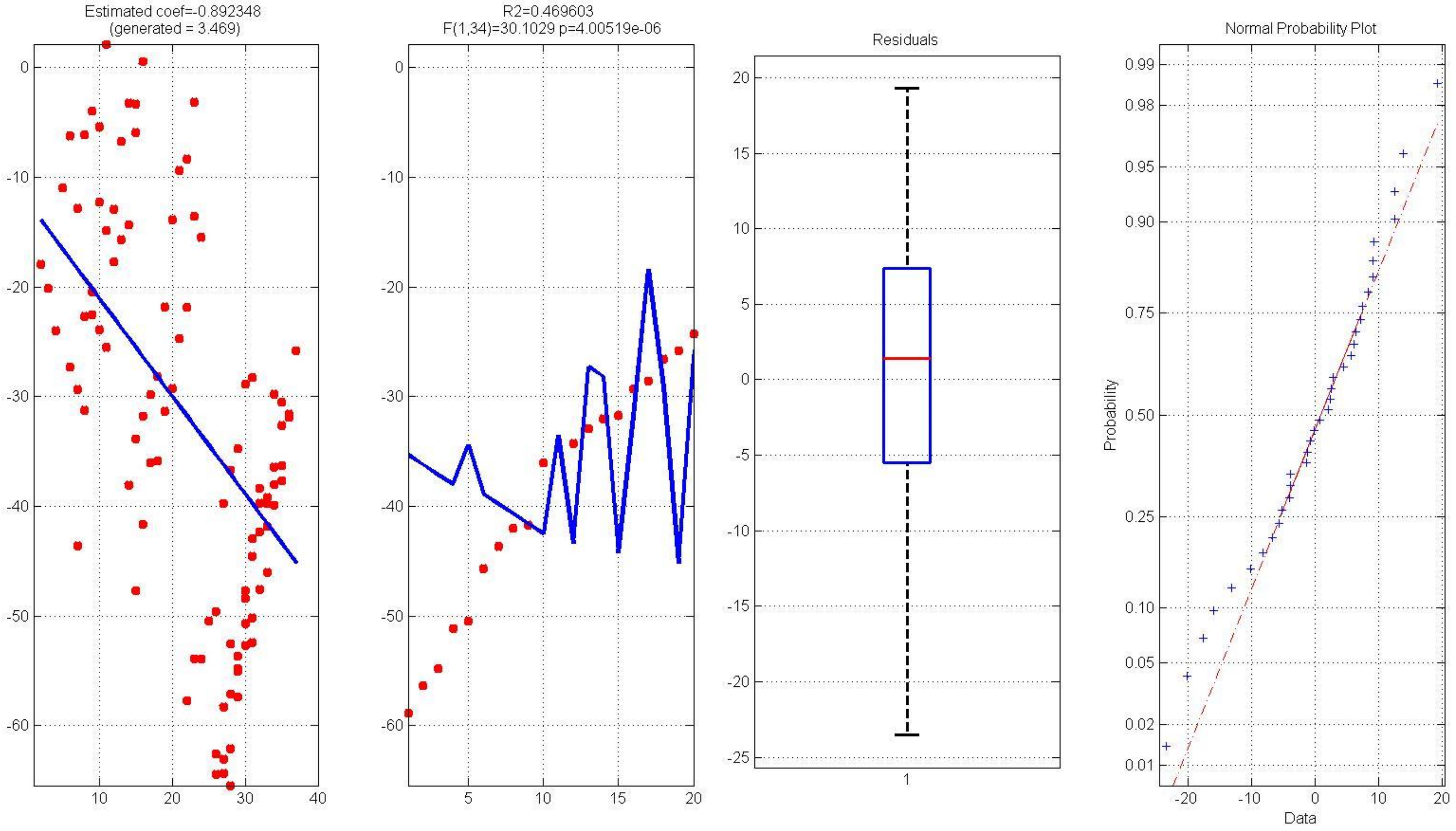
An example

<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

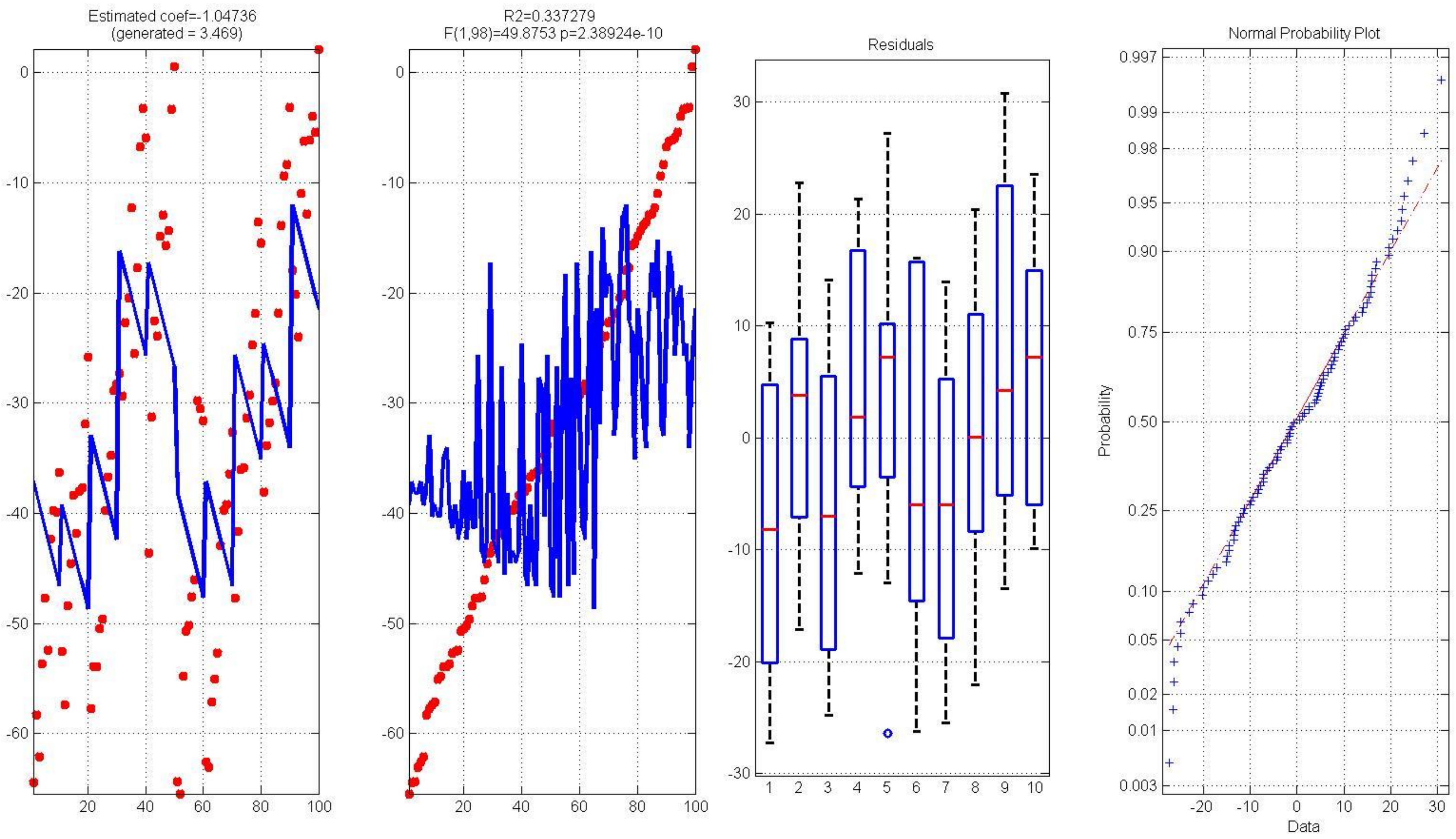


Fixed Effect Model 1: average subjects



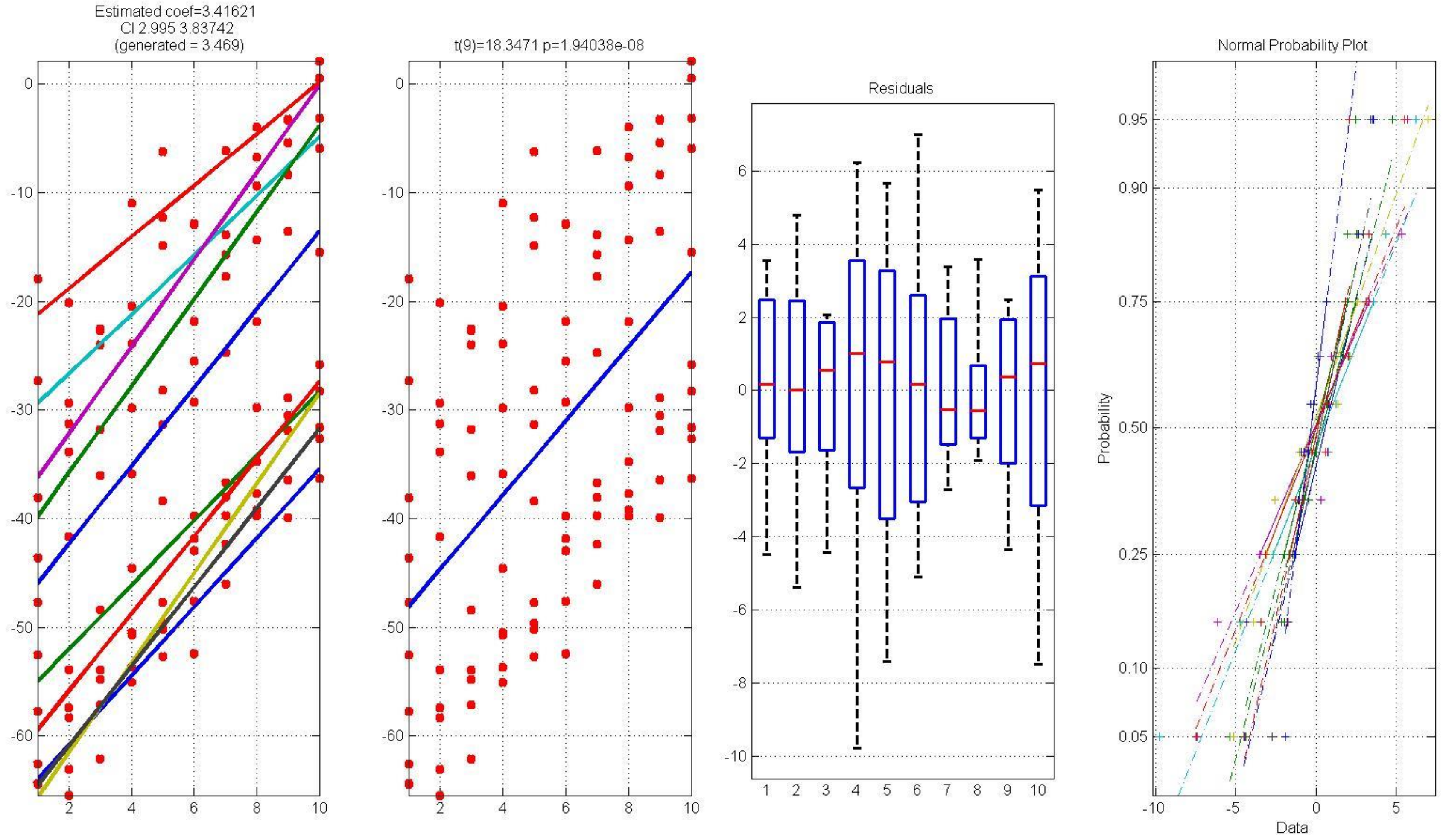
Fixed effect without subject effect → negative effect

Fixed Effect Model 2: constant over subjects



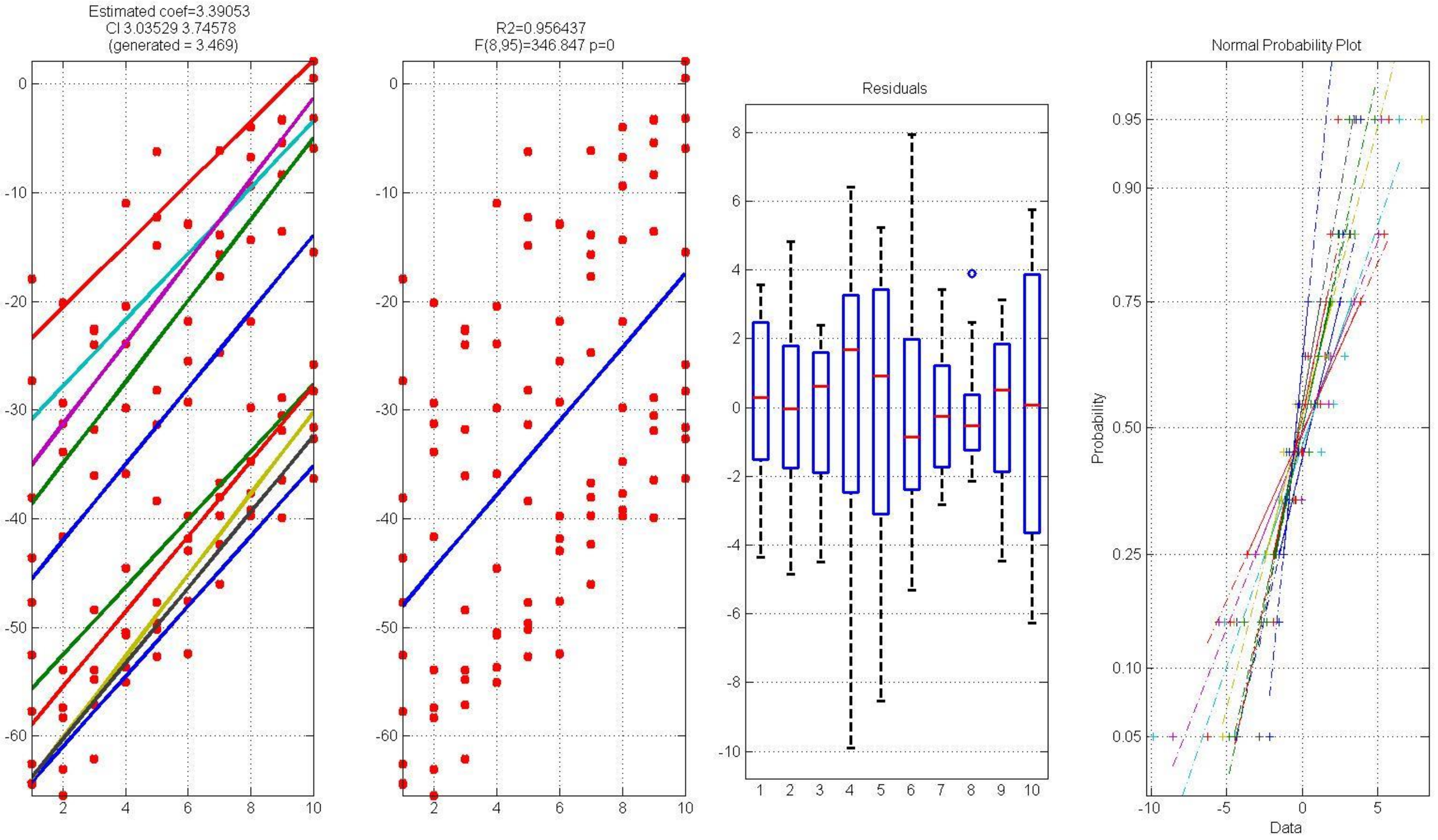
Fixed effect with a constant (fixed) subject effect → positive effect but biased result

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

General Linear Model

Linearity

- Means created by lines
- In maths it refers to equations or functions that satisfy 2 properties: additivity (also called superposition) and homogeneity of degree 1 (also called scaling)
- Additivity → $y = x1 + x2$ (output y is the sum of inputs xs)
- Scaling → $y = \beta\ x1$ (output y is proportional to input x)

http://en.wikipedia.org/wiki/Linear

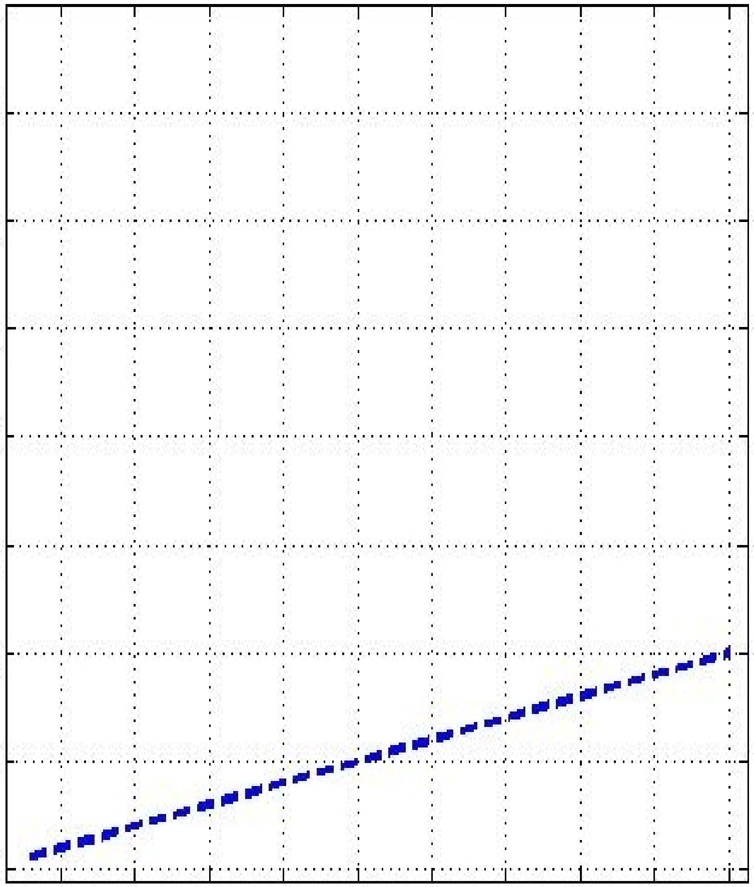
What is a linear model?

- 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

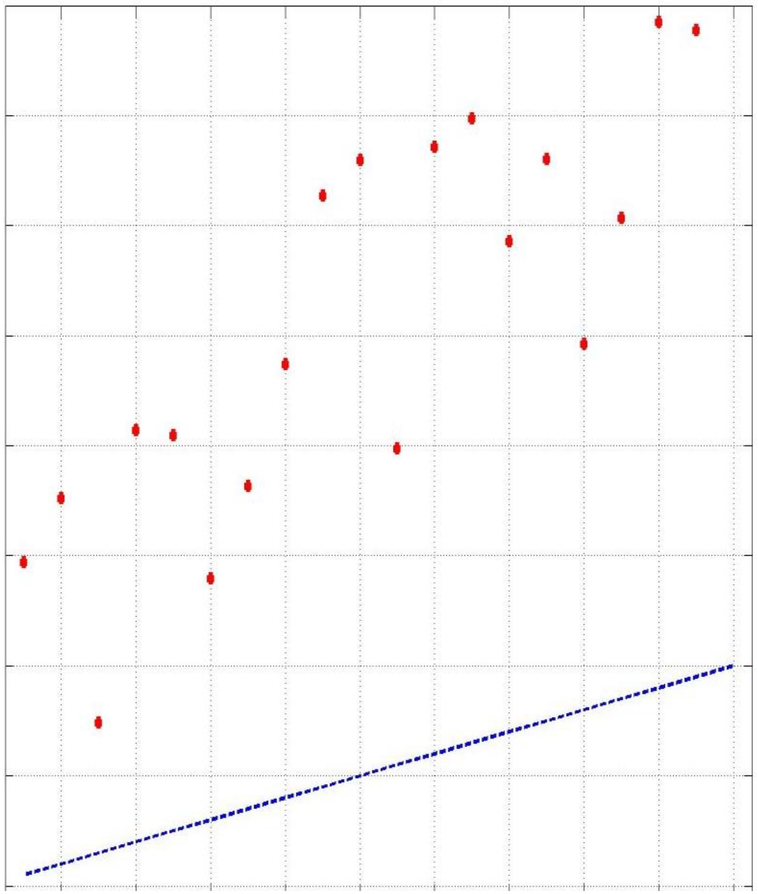
- We have an experimental measure x (e.g. stimulus intensity from 0 to 20)



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

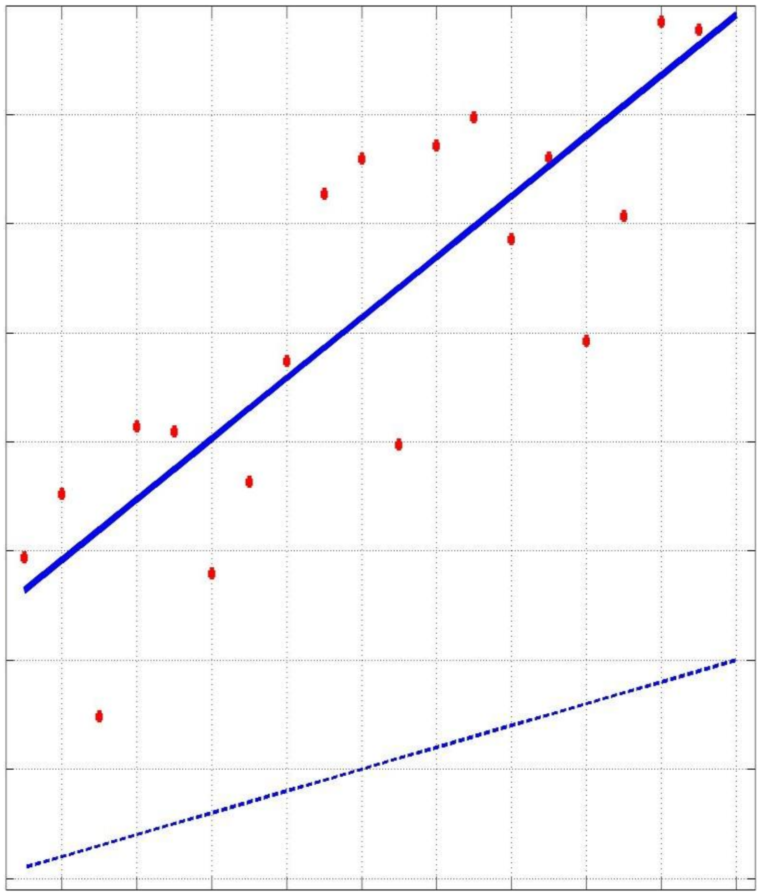
- We have an experimental measure x (e.g. stimulus intensity from 0 to 20)

- 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 $\beta 2$

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

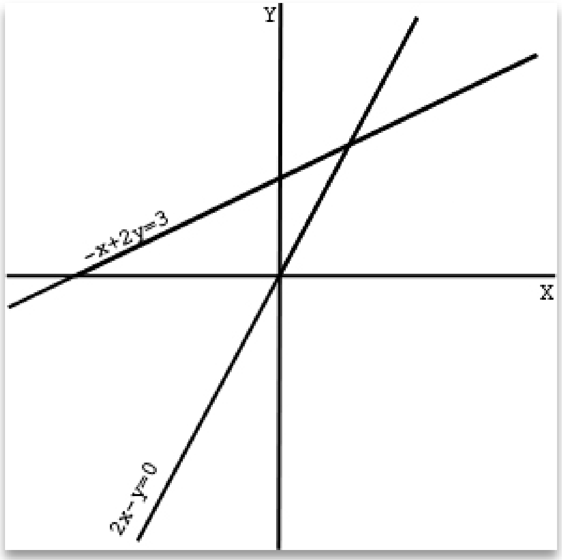


Linear algebra for regression

- 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 $y = x_1 \beta_1 + x_2 \beta 2$

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

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



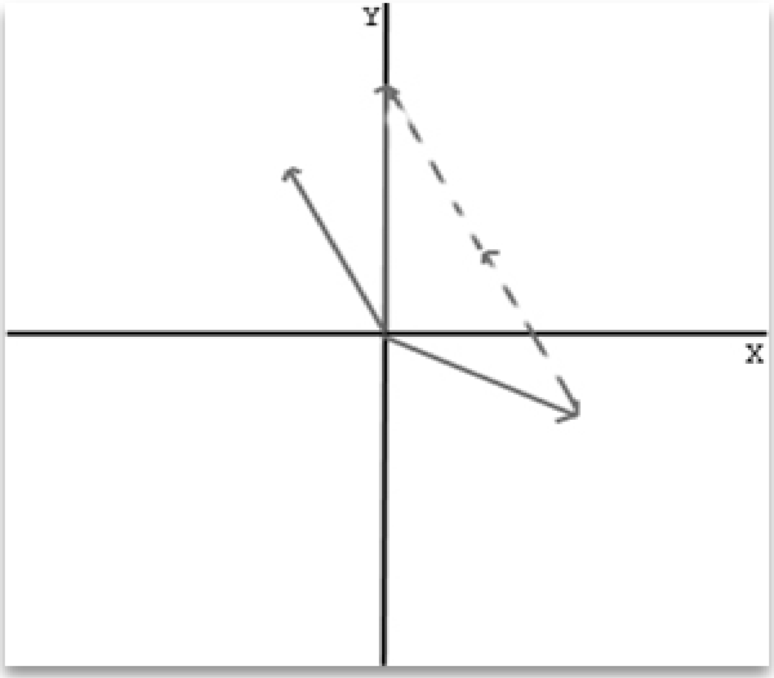
Linear algebra for regression

- 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

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

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

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



Linear algebra for ANOVA

- In text books we have $y = u + xi + \varepsilon$, 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

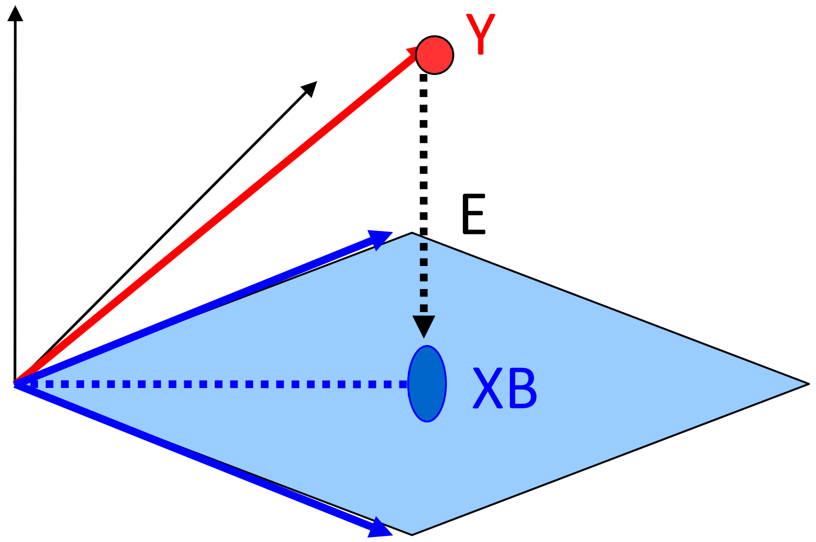
$$y(1..3)1 = 1x1+0x2+0x3+0x4+c+e11$$
$$y(1..3)2 = 0x1+1x2+0x3+0x4+c+e12$$
$$y(1..3)3 = 0x1+0x2+1x3+0x4+c+e13$$
$$y(1..3)4 = 0x1+0x2+0x3+1x4+c+e13$$

$$\begin{Bmatrix} 8\\9\\7\\5\\7\\3\\3\\4\\1\\6\\4\\9 \end{Bmatrix} = \begin{Bmatrix} 1\,0\,0\,0\,1\\1\,0\,0\,0\,1\\1\,0\,0\,0\,1\\0\,1\,0\,0\,1\\0\,1\,0\,0\,1\\0\,1\,0\,0\,1\\0\,0\,1\,0\,1\\0\,0\,1\,0\,1\\0\,0\,1\,0\,1\\0\,0\,0\,1\,1\\0\,0\,0\,1\,1\\0\,0\,0\,1\,1 \end{Bmatrix} * \begin{Bmatrix} \beta1\\\beta2\\\beta3\\\beta4\\c \end{Bmatrix} + \begin{Bmatrix} e1\\ \vdots \\ e13 \end{Bmatrix}$$

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

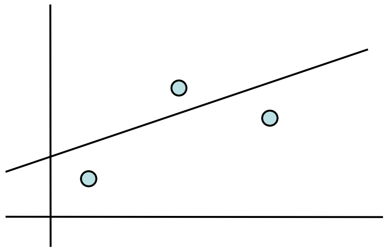
Linear Algebra, geometry and Statistics

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



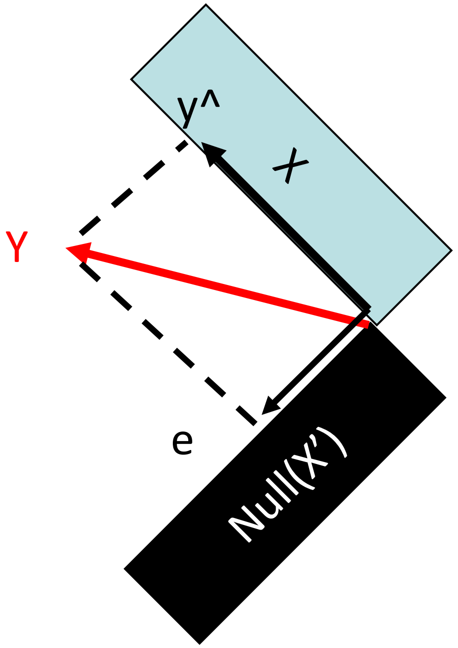
SS total = variance in Y
SS effect = variance in XB
SS error = variance in E
R2 = SS effect / SS total
F = SS effect/df / SS error/dfe

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));

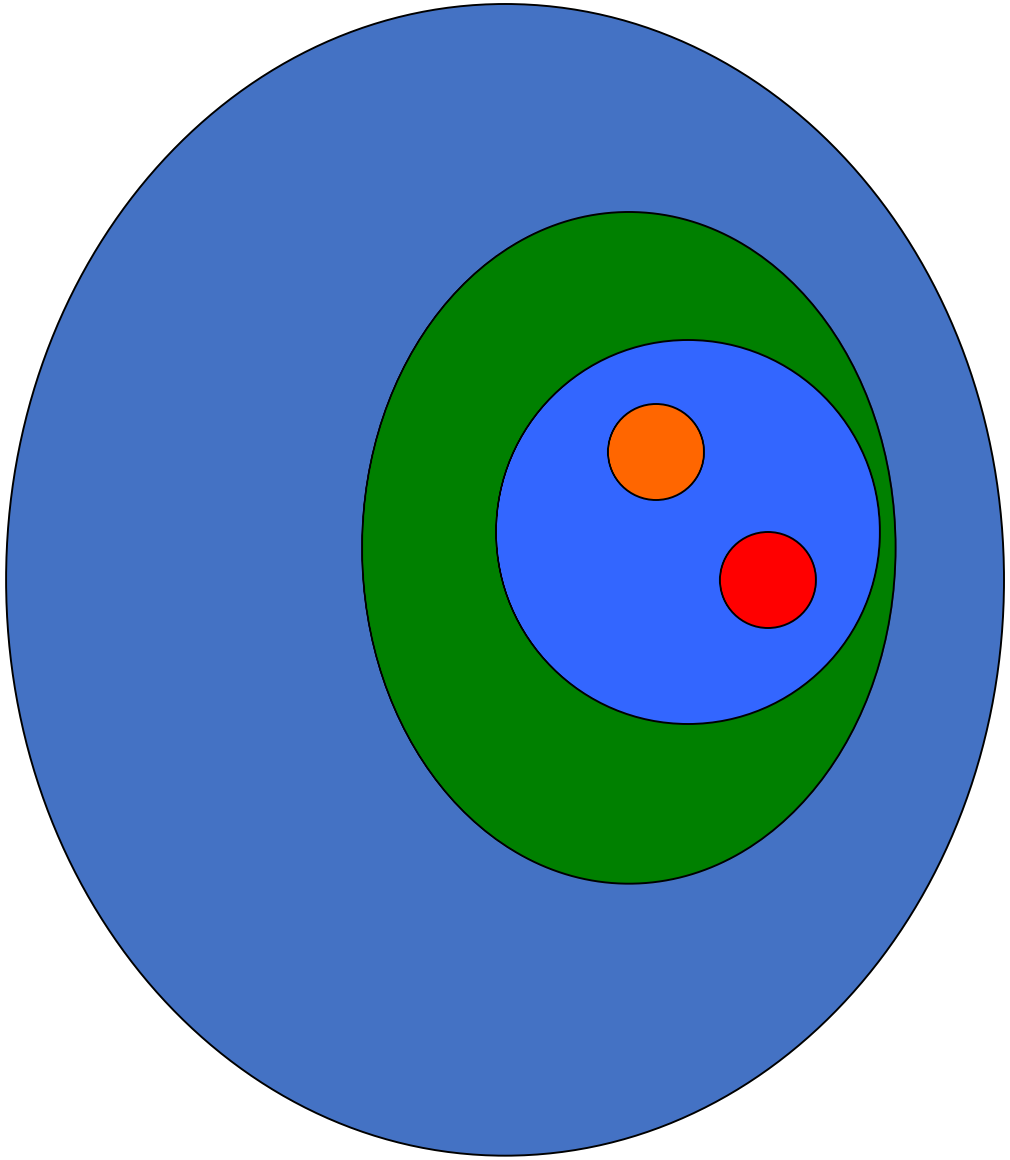
P = R0 - R;

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)

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


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