Mixed Models

ILCB Summer School 2024

Royce Anders

Professeur des Universités PhD HDR

Université Paul-Valéry Montpellier 3

ILCB

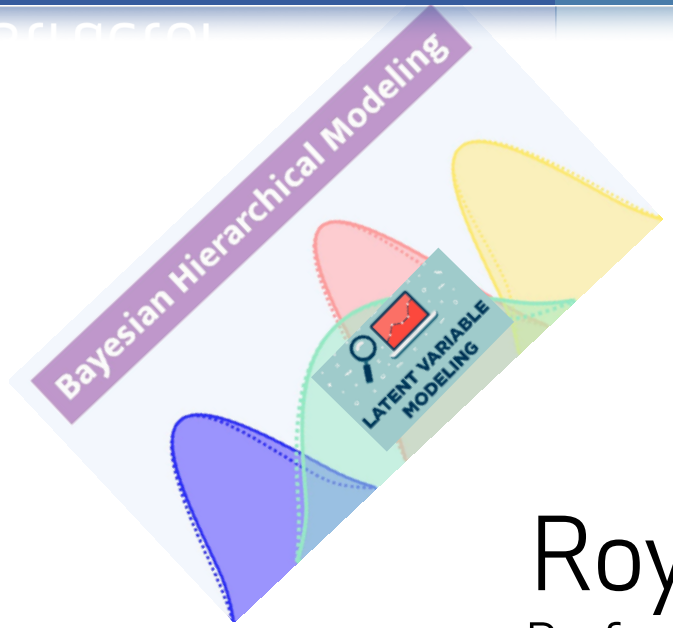
Institute of Language, Communication and the Brain

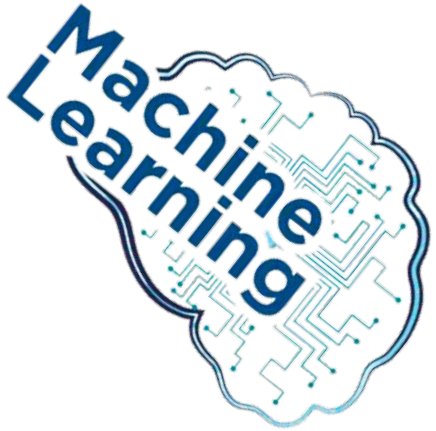
Planning Mixed Models Course

Day	Content	Progression
Monday	**Linear** Mixed Models I	A
Tuesday	**Linear** Mixed Models II	to
Wednesday	**Linear** Mixed Models III	Z
Thursday	**Logistic** Mixed Models I	Concepts and Basic Application

Course Hours: 10:45 – 12:00

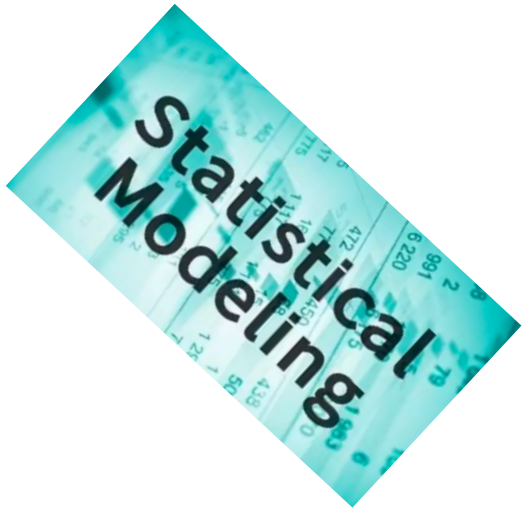
Teaching Objectives	Description
Context	When to and why model?
Data Preparation	Requirements and optimisations for mixed modelling
Implementation	Different ways to model your data
Model Validity	Can I trust this model's results?
Model Performance	How much of my data can the model account for?
Result Interpretation	What does this model have to say about the hypotheses?
Post hoc Considerations	Should I run additional analyses to optimise?
Visualisation	Organise Results in Graphic or Table Form
R	Coding, capacity to implement all in R

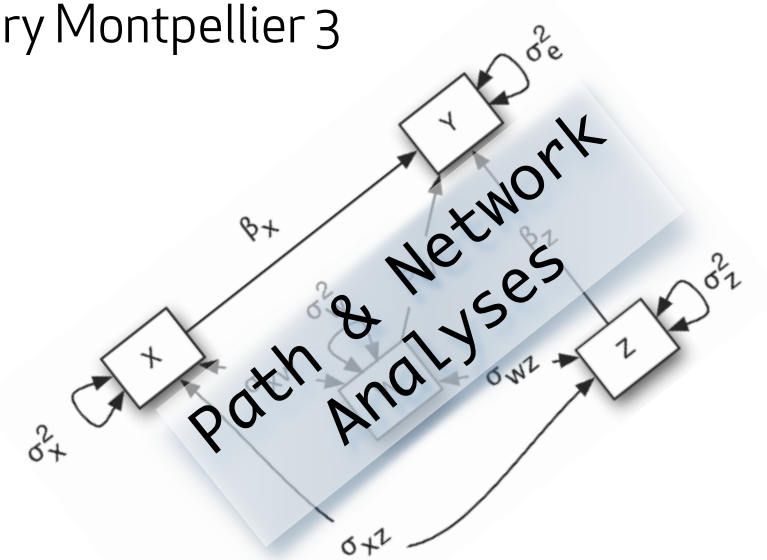




Royce Anders

Professeur des Universités
Université Paul-Valéry Montpellier 3





Some Remarks

- Position in Research activities
- Machine Learning and other analyses

Linear

- Dependent Variable (DV) is continuous

Logistic

- Dependent Variable (DV) is binary ("2 classes")
  - Multinomial for > 2 classes

Installation

- Questions?
- **Step 1 - R Language** https://cran.r-project.org/
- **Step 2 - RStudio** https://www.rstudio.com/products/rstudio/download/
- **Step 3 - Packages in RStudio**

```
install.packages(c("lme4","lmerTest","bestNormalize","MuMIn","sjPlot","sjstats","caret","ROCit"));
install.packages(c("lmtest","nortest","mctest","psych","car","corrplot","BayesFactor"))
```

Link to Course Files

- https://tinyurl.com/ilcbanders

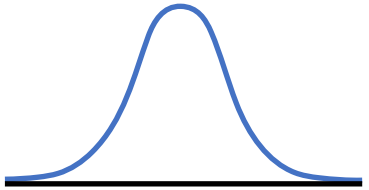
8

Linear Mixed Modelling

Context : When and why to model?

Minimal Requirements

- Dependent Variable (DV) is continuous
- DV is normally-distributed



Objectives

- Predict the DV
  - Know how much each explanatory variable (or IV) influences + or - the DV
- Statistical Testing: does the IV have an effect on the DV?

Mixed vs. Non-mixed Model

- Repeated measures
  - Variation in these measures that you don't have another variable to explain
  - e.g. Variation in the repeated measures of a Participant not explained by Age or IQ

ANOVA vs. Linear Modelling

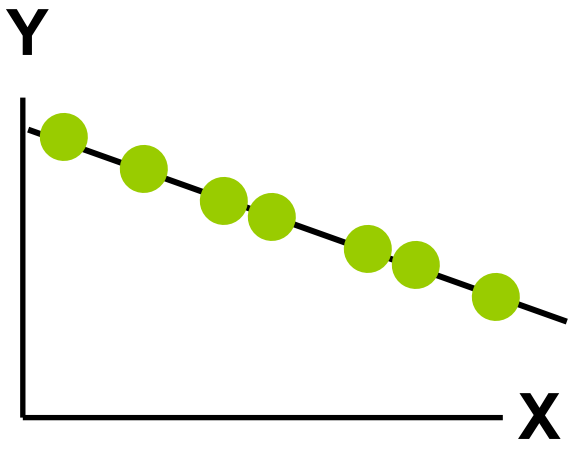
- Typically, the same conclusions in the significance of variables
- Absence of "degree of influence"
- Limited to discrete IV (Young vs. Old instead of actual age submitted to model)
- ANOVA mixed, repeated measures design shares some similarity to Linear Mixed Models

Student *t* test vs. Linear Modelling

- Preferably, linear models analyse all variables simultaneously
  - Independent *t* tests may increase Type I or II error rates (false pos or neg)

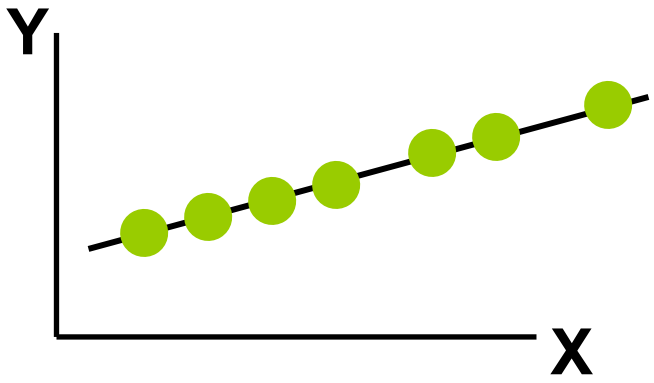
Based on the Pearson linear correlation coefficient *r*

- Generalised to simultaneous correlational modelling… all mapping *f*(X1, X2) etc. to one DV Y

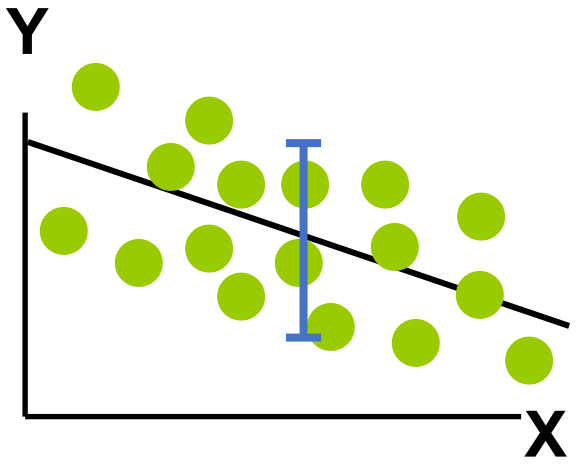


$|r| = 1$

Perfect linear relationship between X and Y:

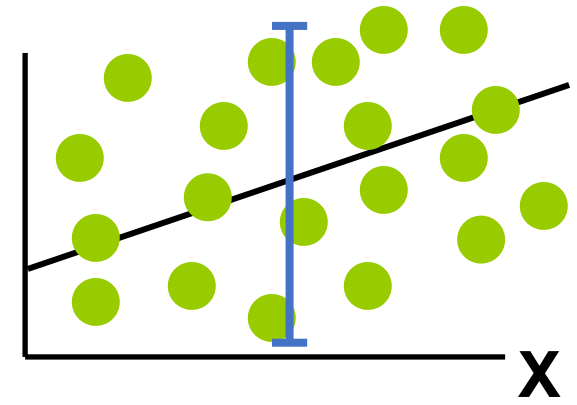


100% of the variation in Y is explained by variation in X



$0 < |r| < 1$

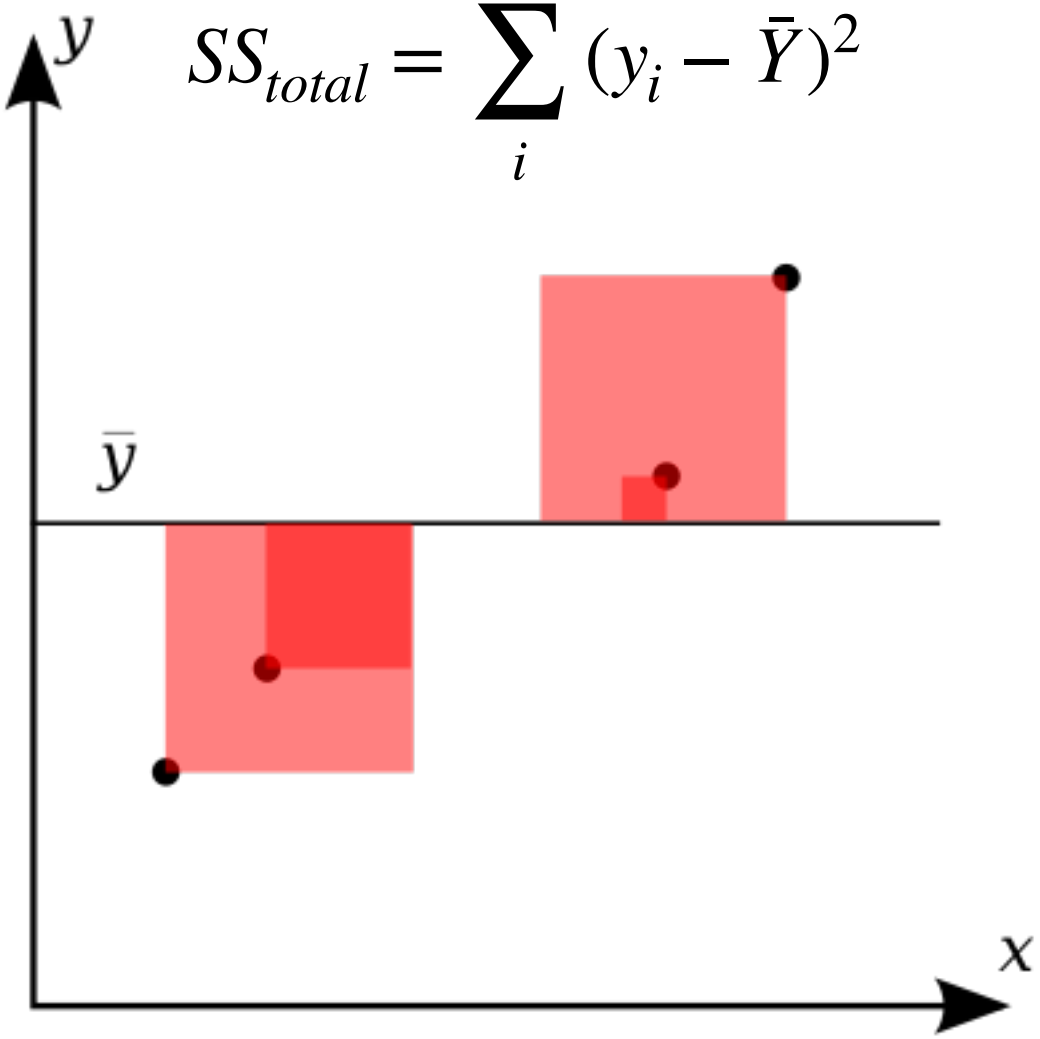
Weaker linear relationships between X and Y:



Some but not all of the variation in Y is explained by variation in X

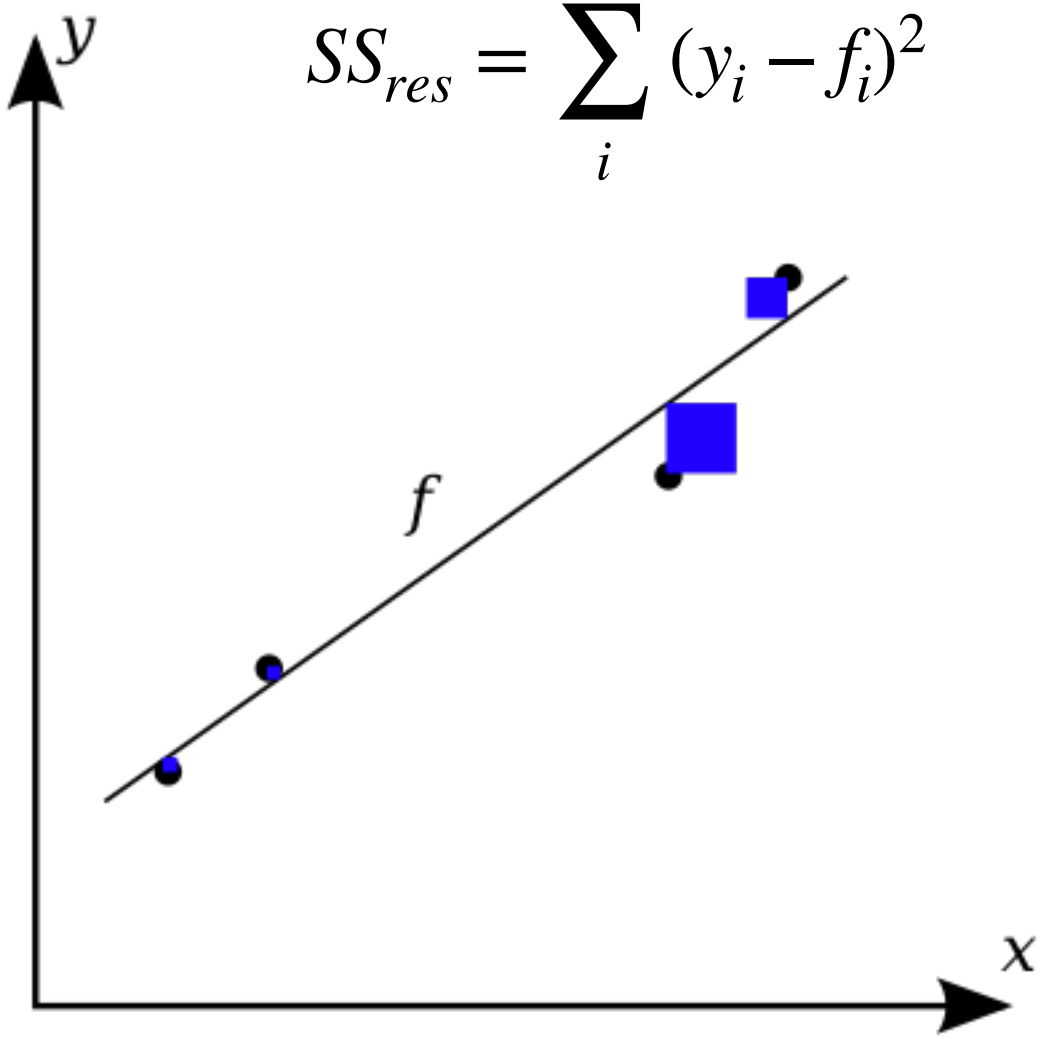
How is a linear model fit?

$$SS_{total} = \sum_{i}^{N} (y_i - \bar{Y})^2$$



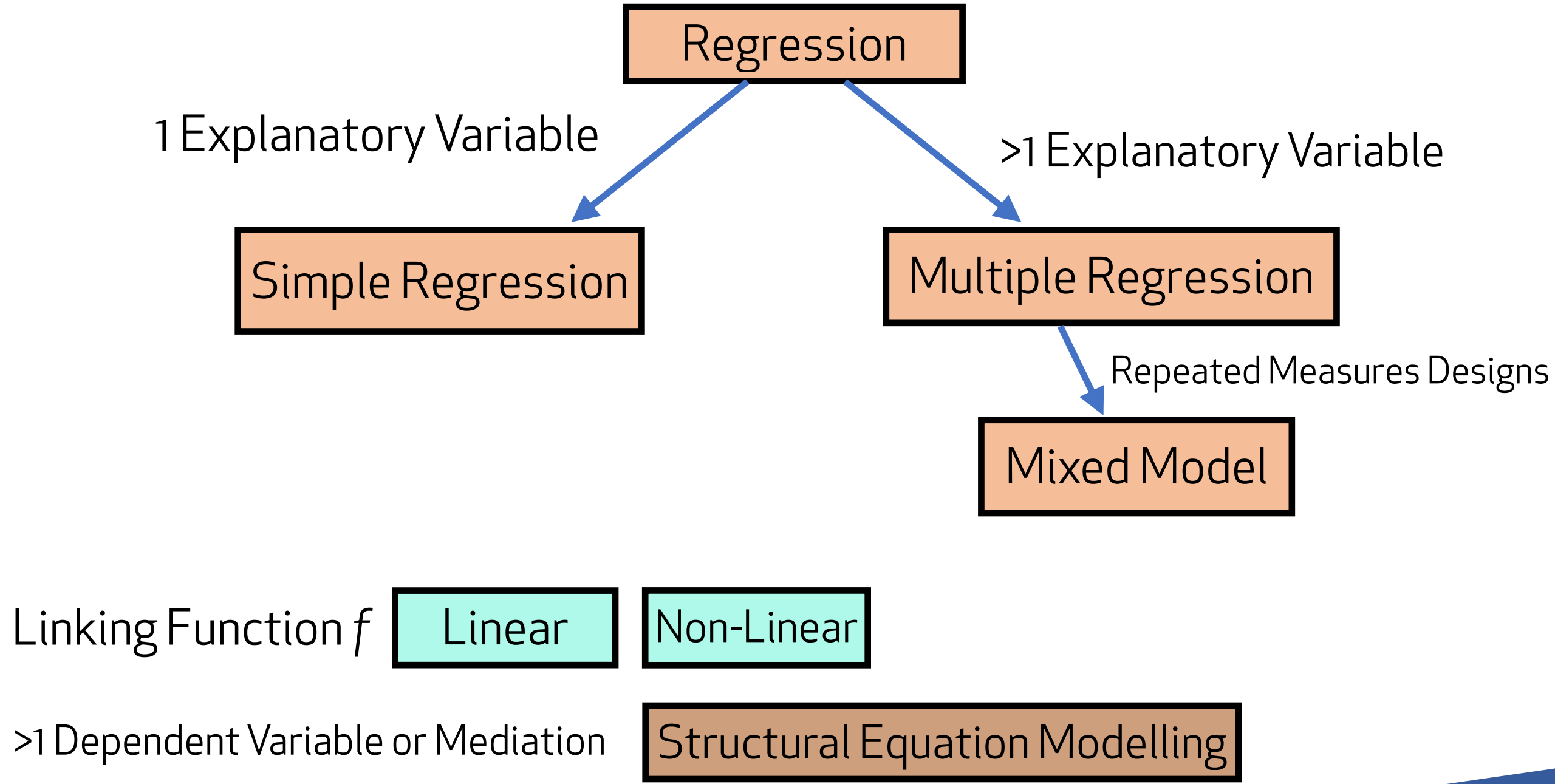
Average is the model

$$SS_{res} = \sum_{i}^{N} (y_i - f_i)^2$$



Predictors as the model

$$y = f(B, X) = B_0 + B_1 x_1 + B_2 x_2 + \ldots + B_k x_k + \epsilon$$


Regression
1 Explanatory Variable
>1 Explanatory Variable
Simple Regression
Multiple Regression
Repeated Measures Designs
Mixed Model
Linking Function *f*
Linear
Non-Linear
>1 Dependent Variable or Mediation
Structural Equation Modelling

- Simple

$$y = B_0 + B_1 x_1 + \epsilon$$

Intercept

Slope

Random Effects ?

Fixed Effects

(Do not vary as a function of the item or individual modelled)

16

Data Preparation

Principal Steps

- Load data into R as a data.frame (verify variable coding and format)
- Filter subjects based on Inclusion/Exclusion Criteria (also missing data decisions)
- Outlier Detection and Elimination in the DV (winsorizing, IQR, MAD, later: Cook's Distance)
- Determine which variables are factors (discrete predictors) and indicate them
  - Factors are categorical where category # should not have any sense
- Normalise all data except for factors
- Standardise all data …

Applied in R

- Let's practice

Visualisations

- May often help for making decisions
- Histograms
- QQplot

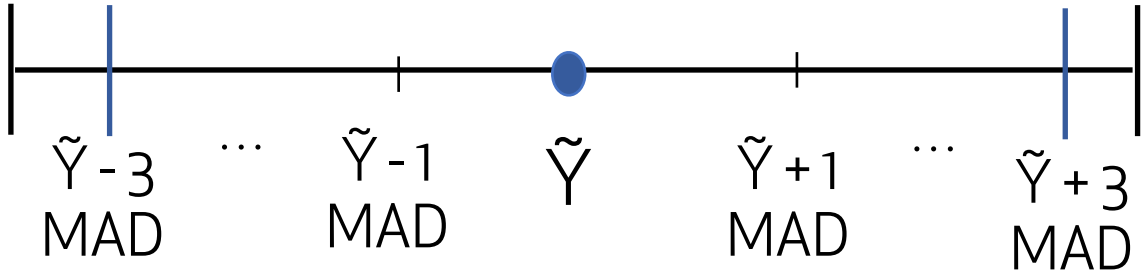
Other Approaches

- Sort the variable and look at the first or last 20 values
- Tabulate the variable
- See the unique values of the variable (and potentially sort)

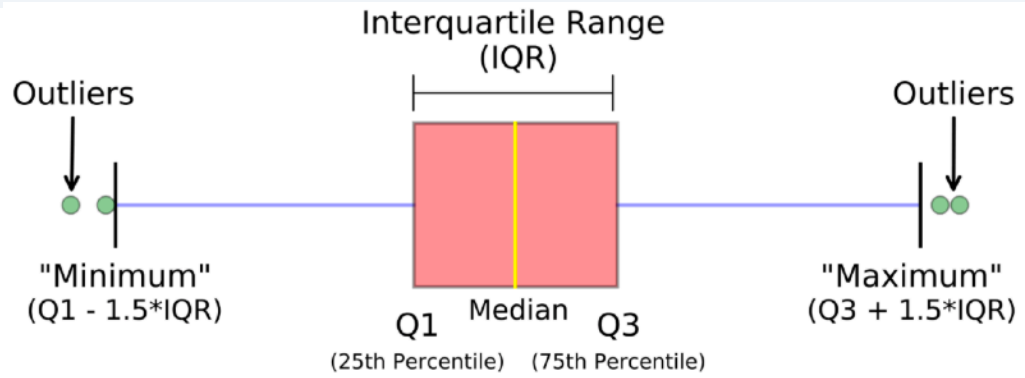
Applied in R

- Let's practice

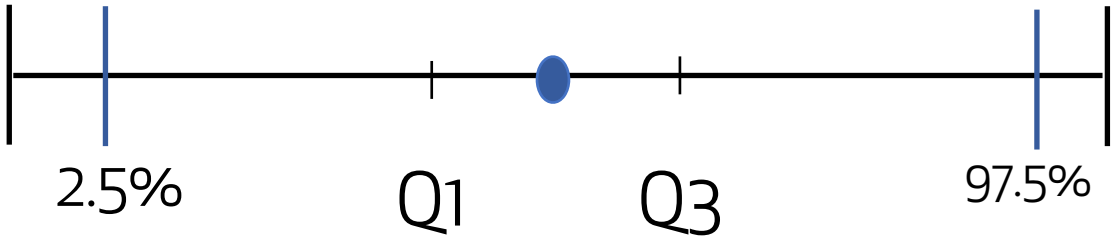
Median ± n × MAD



Quartiles ± 1.5 × IQR



Winsorizing Quantiles < 2.5% > 97.5%



Data Import

- read.csv() or read_excel() #Load data from library(readxl)

Inclusion/Exclusion Criteria

- which(condition) # Indices that match a condition being met
- subset() # Filter several conditions simultaneously
- is.na() # Are there missing values? combine with sum() or colSums() or na.omit()

Outlier Approaches

- sort() # order the values
- hist() #histogram density() and plot(density())
- median() and mad() or iqr() quantile( )

Recoding Variables

- df$yourvar = as.factor(df$yourvar) # for categorical variables and relevel() # to set the reference level
- cut() # to set as ordinal variable

Normalize your data

- yeojohnson() # from library(bestNormalize) combine with apply()

Standardize your data

- scale()