

# ***Multivariate regression models for Twin data***

***Analysis of Twin Data in Health Science · Session IV***

Prof. Wagner Hugo Bonat

Ômega Data Science | Online School of Data Science



# Motivation

# Heredity and variation

- ▶ Genetic epidemiology is impelled by three basic questions:
  1. Why isn't everyone the same?
  2. Why are children like their parents?
  3. Why aren't children from the same parents all alike?
- ▶ **Main goal: Isolate/Separate sources of variation!**

▶ Variation is everywhere!

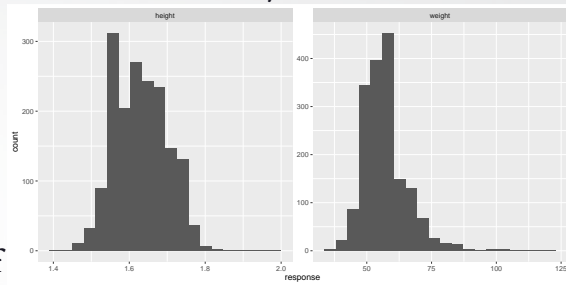


Figure 1. Histogram of height and weight.

## Motivating dataset: Anthropometric measures

- ▶ Anthropometric measurements (weight and height).
- ▶ 861 twin pairs: 327 DZ (dizygotic) and 534 MZ (monozygotic).
- ▶ Bivariate continuous traits.
- ▶ Covariates: age and group.
- ▶ Available as an example in the OpenMx package (Neale, et al., 2016).
- ▶ Easy access from the `mg1m4twin` package.



Figura 2. Photo by Pixabay.

## Motivating dataset: Anthropometric measures

### ► The dataset

```
library(mglim4twin)
data(anthro)
glimpse(anthro)
```

```
## Rows: 1,722
## Columns: 6
## $ weight    <int> 62, 55, 66, 73, 51, 44, 52, 57, 54, 54, 58, 57, ~
## $ height    <dbl> 1.6499, 1.6299, 1.6599, 1.7000, 1.7300, 1.5698, ~
## $ age       <int> 24, 24, 20, 20, 20, 20, 26, 26, 20, 20, 22, 22, ~
## $ Group     <fct> DZ, DZ, DZ, DZ, DZ, DZ, DZ, DZ, DZ, DZ, DZ, DZ, ~
## $ Twin      <int> 535, 535, 536, 536, 537, 537, 538, 538, 539, 539~
## $ Twin_pair <int> 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, ~
```

## Graphing and Quantifying Familial Resemblance

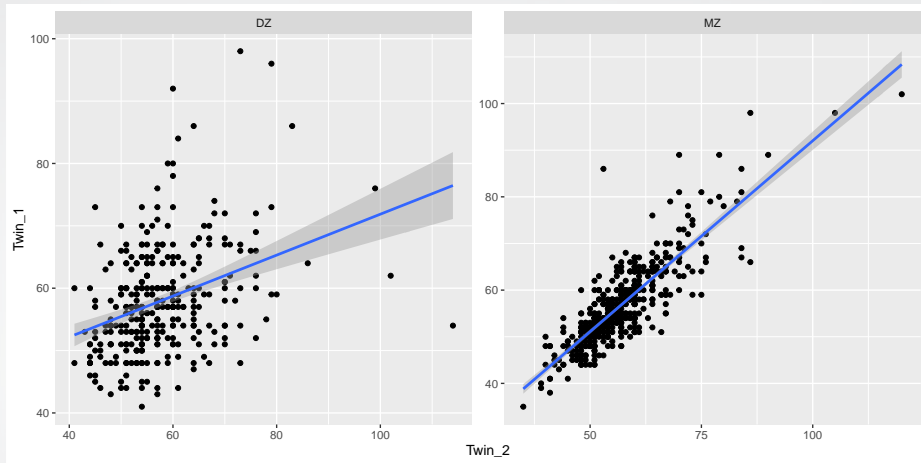


Figure 3. Dispersion diagram by zygosity · Trait weight.

## Multiple traits

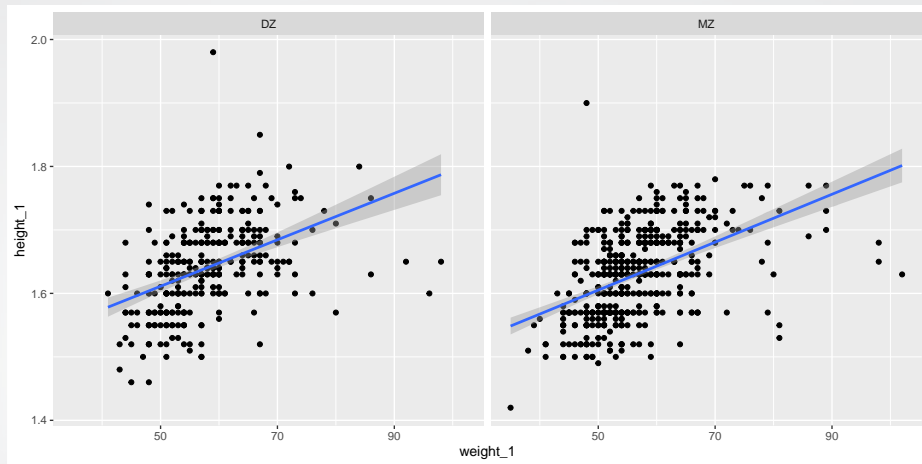


Figure 4. Dispersion diagram by zygosity · Weight vs Height.

## Building and Fitting Models

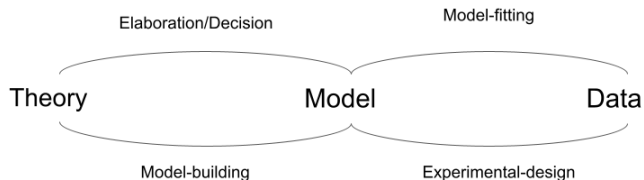


Figura 5. Diagram of the interrelationship between theory, model and empirical observation. Adapted from Neale and Maes (1992).



## Challenges for model-building in Twin data analyses

- ▶ Decompose sources of variation
  1. Genetic Effects.
  2. Environmental Effects.
  3. Genotype-Environment Interaction.
- ▶ Traits types
  1. Binary and binomial data.
  2. Bounded data and continuous proportions.
  3. Under-, equi- and over-dispersed count data.
  4. Semi-continuous data (continuous + mass at zero).
  5. Symmetric and assymetric continuous data.
- ▶ Multiple traits of mixed types.



Figura 6. Photo by Magda Ehlers from Pexels.

## ***Importance and statistical approaches'***

- ▶ Multivariate twin and family studies are important tools to:
  1. Determine traits inheritance;
  2. Determine the influence of genetic and environmental effects on traits.
- ▶ Statistical challenge:
  1. Model the covariance structure to take into account the genetic and environmental structures induced by the twin and family designs.
- ▶ Orthodox approaches:
  1. Structural equation modelling (SEM); 2. Linear mixed models (LMM).
- ▶ Main limitations of SEM and LMM:
  1. Both deal only with Gaussian (symmetric) data;
  2. Standard computational implementations are difficult to adapt for the analysis of twin and family data.