

# 1 Lecture 1:Jan 20

## Today

- Introduction
- Course logistics
- Read JF chapter 1, JM Appendix A

## What is this course about?

The term “linear models” describes a wide class of methods for the statistical analysis of multivariate data. The underlying theory is grounded in linear algebra and multivariate statistics, but applications range from biological research to public policy. The objective of this course is to provide a solid introduction to both the theory and practice of linear models, combining mathematical concepts with realistic examples.

## A hierarchy of linear models

- The linear mean model:

$$\underset{n \times 1}{\mathbf{y}} = \underset{n \times p}{\mathbf{X}} \underset{p \times 1}{\boldsymbol{\beta}} + \underset{n \times 1}{\boldsymbol{\epsilon}}$$

where  $\mathbf{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ . Only assumption is that errors have mean 0.

- Gauss-Markov model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where  $\mathbf{E}(\boldsymbol{\epsilon}) = \mathbf{0}$  and  $\mathbf{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$ . Uncorrelated errors with constant variance.

- Aitken model or general linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where  $\mathbf{E}(\boldsymbol{\epsilon}) = \mathbf{0}$  and  $\mathbf{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{V}$ .  $\mathbf{V}$  is fixed and known.

- Variance components models:  $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_1^2 \mathbf{V}_1 + \sigma_2^2 \mathbf{V}_2 + \dots + \sigma_r^2 \mathbf{V}_r)$  with  $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_r$  known.

- General mixed linear Model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where  $\mathbf{E}(\boldsymbol{\epsilon}) = \mathbf{0}$  and  $\mathbf{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ .

- Generalized linear models (GLMs). Logistic regression, probit regression, log-linear model (Poisson regression), ... Note the difference from the general linear model. GLMs are generalization of the *concept* of linear models. They are covered in Math 7360 - Data Analysis class (<https://tulane-math7360.github.io/lectures/>).

## Syllabus

Check course website frequently for updates and announcements.

<https://tulane-math-7260-2021.github.io/>

## HW submission

Through Github with demo on Friday class.