STAT 714 fa 2025 Lec 00

Overview of linear models course

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

One-sample problem

For responses $Y_1, \ldots, Y_n$, assume

$$\boxed{Y_i = \mu + \varepsilon_i, \quad i = 1, \ldots, n,}$$

noise, error terms

where

- μ is the mean
- ε_i are independent Normal$(0, \sigma^2)$

Goal: Make inference on μ.

Exercise: Put equations in matrix form $\boxed{\mathbf{Y} = \mathbf{X}\mathbf{b} + \varepsilon.}$

parameters

$$Y_1 = \mu + \varepsilon_1$$
$$Y_2 = \mu + \varepsilon_2$$
$$\vdots$$
$$Y_n = \mu + \varepsilon_n$$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \mu + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \qquad b = \mu \qquad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Simple linear regression

For responses $Y_1, \ldots, Y_n$, assume

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

where

- β_0 is the intercept
- β_1 is the slope
- ε_i are independent Normal$(0, \sigma^2)$

$$\underset{\sim}{Y} = X \underset{\sim}{b} + \underset{\sim}{\varepsilon}$$

Goal: Make inference on β_0, β_1.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$.

$$Y_1 = \beta_0 + \beta_1 x_{11} + \cdots + \beta_p x_{p1} + \varepsilon_1$$
$$\vdots \qquad\qquad \vdots$$
$$Y_n = \beta_0 + \beta_1 x_{1n} + \cdots + \beta_p x_{pn} + \varepsilon_n$$

Multiple linear regression

For responses $Y_1, \ldots, Y_n$, assume

$$Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i, \quad i = 1, \ldots, n,$$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

where

- β_0 is the intercept
- $\beta_1, \ldots, \beta_p$ are the linear effects of the covariates
- ε_i are independent Normal$(0, \sigma^2)$

Goal: Make inference on β_0, β_1.

Exercise: Put equations in matrix form $\boxed{\mathbf{Y} = \mathbf{Xb} + \varepsilon.}$

Cell-means model (One-way ANOVA)

For responses Y_{ij}, assume

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, n_i$$

where

- μ_i are the treatment means
- ε_{ij} are independent Normal$(0, \sigma^2)$

Goal: Test for differences in treatment means.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{X}\mathbf{b} + \varepsilon$.

$i = 1, 2 \quad (a = 2)$

$n_i = 2$

$$Y_{11} = \mu_1 + \varepsilon_{11}$$
$$Y_{12} = \mu_1 + \varepsilon_{12}$$
$$Y_{21} = \mu_2 + \varepsilon_{21}$$
$$Y_{22} = \mu_2 + \varepsilon_{22}$$

$$\underset{\sim}{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix} \quad \underbrace{X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_{\text{"Design Matrix"}} \quad \underset{\sim}{b} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{bmatrix}$$

Treatment effects model (One-way ANOVA)

For responses Y_{ij}, assume

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, a, \quad j = 1, \dots, n_i$$

$i = 1, 2 \quad (a = 2)$

$n_i = 2$

$$\begin{aligned} Y_{11} &= \mu + \alpha_1 + \varepsilon_{11} \\ Y_{12} &= \mu + \alpha_1 + \varepsilon_{12} \\ Y_{21} &= \mu + \alpha_2 + \varepsilon_{21} \\ Y_{22} &= \mu + \alpha_2 + \varepsilon_{22} \end{aligned}$$

where

- μ is a mean
- α_i are treatment effects
- ε_{ij} are independent Normal$(0, \sigma^2)$

Goal: Test for differences in treatment means.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{X}\mathbf{b} + \varepsilon$.

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \qquad \underset{\sim}{b} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

$$X = \begin{bmatrix} & & & \\ & & & \\ & & & \end{bmatrix} \quad \underset{\sim}{b} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

Treatment effects with continuous covariate (One-way ANCOVA)

For responses Y_{ij}, assume

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, n_i$$

$i = 1, 2$, $a = 2$

$n_i = 2$

$$\begin{aligned}
Y_{11} &= \mu + \alpha_1 + \beta x_{11} + \varepsilon_{11} \\
Y_{12} &= \mu + \alpha_1 + \beta x_{12} + \varepsilon_{12} \\
Y_{21} &= \mu + \alpha_2 + \beta x_{21} + \varepsilon_{21} \\
Y_{22} &= \mu + \alpha_3 + \beta x_{22} + \varepsilon_{22}
\end{aligned}$$

where

- μ is a mean
- α_i are treatment effects
- β is the effect of the covariate
- ε_{ij} are independent Normal$(0, \sigma^2)$

Goal: Test for differences in treatment means.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{X}\mathbf{b} + \varepsilon$.

$4\times4 \quad 4\times1$

$$X = \begin{bmatrix} 1 & 1 & 0 & x_{11} \\ 1 & 1 & 0 & x_{12} \\ 1 & 0 & 1 & x_{21} \\ 1 & 0 & 1 & x_{22} \end{bmatrix} \qquad \underset{\sim}{b} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix}$$

Two-way cell means model

For responses Y_{ijk}, assume

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n_{ij}$$

where

- μ_{ij} are treatment means
- ε_{ijk} are independent Normal$(0, \sigma^2)$

Goal: Test for differences in means.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{X}\mathbf{b} + \varepsilon$.

Two-way treatment effects model (Two-way ANOVA)

For responses Y_{ijk}, assume

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n_{ij}$$

where

- μ is a mean
- α_i are treatment effects for factor 1
- β_j are treatment effects for factor 2
- $(\alpha\beta)_{ij}$ are interaction effects
- ε_{ijk} are independent Normal$(0, \sigma^2)$

Goal: Make inference on μ, $\alpha_1, \dots, \alpha_a$.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{X}\mathbf{b} + \varepsilon$.

$a = 2$, $b = 2$, $n_i = 2$

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

$$
\begin{aligned}
Y_{111} &= \mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11} + \varepsilon_{11} \\
Y_{112} &= \mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11} + \\
Y_{121} &= \mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12} \\
Y_{122} &= \mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12} \\
Y_{211} &= \mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21} \\
Y_{212} &= \mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21} \\
Y_{221} &= \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22} \\
Y_{222} &= \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}
\end{aligned}
$$

$$
X = \begin{bmatrix}
1 & 1 & 0 & 1 & 0 & 1 & & & \\
1 & 1 & 0 & 1 & 0 & 1 & & & \\
1 & 1 & 0 & 0 & 1 & & 1 & & \\
1 & 1 & 0 & 0 & 1 & & 1 & & \\
1 & 0 & 1 & 1 & 0 & & & 1 & \\
1 & 0 & 1 & 1 & 0 & & & 1 & \\
1 & 0 & 1 & 0 & 1 & & & & 1 \\
1 & 0 & 1 & 0 & 1 & & & & 1
\end{bmatrix}
\qquad
\underset{\sim}{b} = \begin{bmatrix}
\mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22}
\end{bmatrix}
$$

"fixed"

"random effects"

One-way random effects model

For responses Y_{ij}, assume

$$Y_{ij} = \mu + A_i + \varepsilon_{ij}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, n_i$$

$i = 1, 2 \quad (a = 2)$

$n_i = 2$

$$\begin{aligned} Y_{11} &= \mu + A_1 + \varepsilon_{11} \\ Y_{12} &= \mu + A_1 + \varepsilon_{12} \\ Y_{21} &= \mu + A_2 + \varepsilon_{21} \\ Y_{22} &= \mu + A_2 + \varepsilon_{22} \end{aligned}$$

where

- ε_{ij} are independent Normal$(0, \sigma_\varepsilon^2)$
- A_i are independent Normal$(0, \sigma_A^2)$
- A_i and ε_{ij} are independent

Goal: Test if treatment effect variance σ_A^2 is zero.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{Xb} + \mathbf{Zu} + \varepsilon$.

$$\underset{\sim}{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix} \quad X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \underset{\sim}{b} = [\mu] \quad Z = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \underset{\sim}{u} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$$

Two-way random effects model

For responses Y_{ijk}, assume

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, n_{ij}$$

where

- ε_{ijk} are independent Normal$(0, \sigma^2_\varepsilon)$
- A_i are independent Normal$(0, \sigma^2_A)$
- B_i are independent Normal$(0, \sigma^2_B)$
- $(AB)_{ij}$ are independent Normal$(0, \sigma^2_{AB})$
- A_i,B_i, $(AB)_{ij}$, and ε_{ij} are independent

Goal: Test whether variance components σ^2_A, σ^2_B, σ^2_{AB} are equal to zero.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{Xb} + \mathbf{Zu} + \varepsilon$.

Two-way mixed effects model (Randomized complete block design)

For responses Y_{ijk}, assume

$$Y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n_{ij}$$

where

- μ is a mean
- α_i are treatment effects
- ε_{ijk} are independent Normal$(0, \sigma^2_\varepsilon)$
- B_i are independent Normal$(0, \sigma^2_B)$
- $(\alpha B)_{ij}$ are independent Normal$(0, \sigma^2_{AB})$
- B_i, $(AB)_{ij}$, and ε_{ij} are independent

Goal: Test if variance components are equal to zero; test for treatment effects.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{Xb} + \mathbf{Zu} + \varepsilon$.

One-way treatment effects model with subsampling

For responses Y_{ijk}, assume

$$Y_{ijk} = \underbrace{\mu + \alpha_i}_{} + \varepsilon_{ij} + d_{ijk}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, n_i, \quad k = 1, \ldots, m_{ij}$$

$\mathbb{E}\, Y_{ijk} = \mu + \alpha_i$

where

- μ is a mean
- α_i are treatment effects
- ε_{ij} are independent Normal$(0, \sigma^2_\varepsilon)$
- d_{ijk} are independent Normal$(0, \sigma^2_d)$
- ε_{ij} and d_{ijk} are independent

Goal: Test for differences in the treatments.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{Xb} + \mathbf{Zu} + \varepsilon$.

A *linear model* gives the expected value of the response as a linear function of some parameters.

E.g. the model $Y_i = \beta_0 e^{\beta_1 X_i} + \varepsilon_i$, $i = 1, \dots, n$ is a non-linear model.

$\mathbb{E} Y_i$

The goal is that by the end of the course you should:

1. Know how to conduct inference in various linear models.
2. Understand theoretical justifications of these inference methods
   (Like why you should use an F-test with such-and-such degrees of freedom).
3. Have a solid understanding of linear algebra and its role in statistics.