THE CLASSICAL LINEAR MODEL

- Most commonly used statistical models
- Flexible models
- Well-developed and understood properties
- Ease of interpretation
- Building block for more general models

  1. General Linear Model
  2. Generalized Linear Model
  3. Generalized Estimating Equations
  4. Generalized Linear Mixed Model, etc.
  5. Heirarchical Generalized Linear Mixed Model, etc.

EXAMPLES:

1. EXAMPLE 1. *Simple linear regression.*

   OBJECTIVE: Relate weight to blood pressure.

   Consider a random sample of n individuals. The i-th patient has weight x_i and blood pressure Y_i $(i = 1, 2, \ldots, n)$.

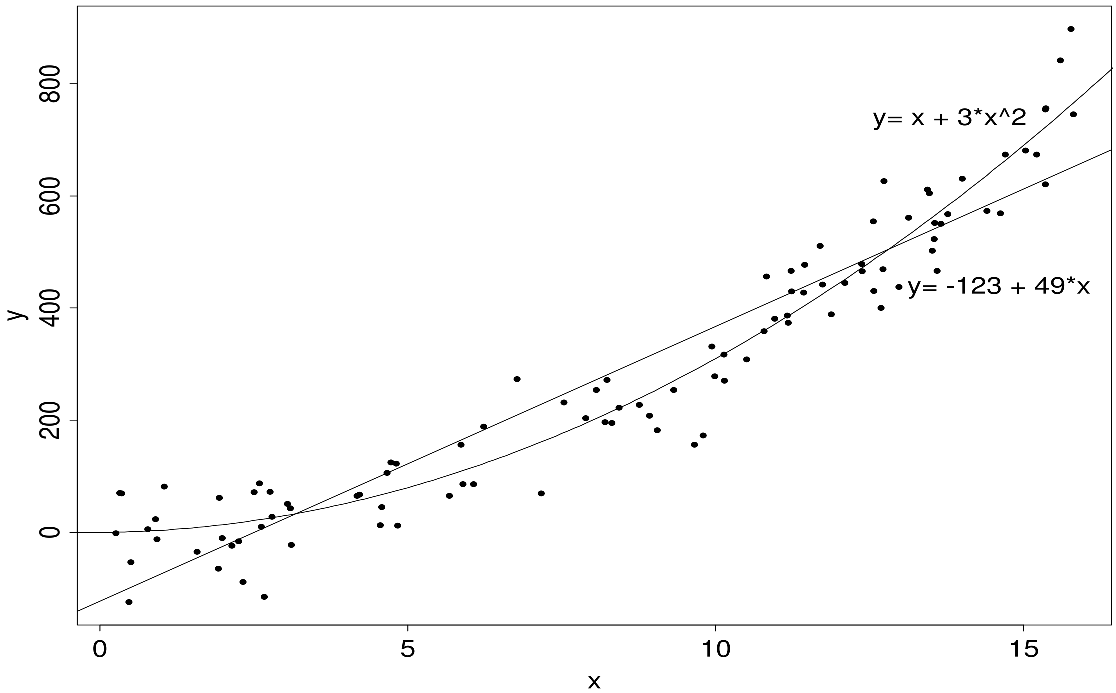
   MODEL:

   $$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

   where

   - Y_i is the *response* variable,
   - x_i is a regressor variable,
   - β_0, β_1 are regression coefficients – unknown model parameters to be estimated,
   - ε_i is an error term.

Figure 1



2. EXAMPLE 2. *Polynomial regression.*

OBJECTIVE: Same as EXAMPLE 1.

MODEL:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$

Is this still a linear model?

3. EXAMPLE 3. *Multiple linear regression.*

OBJECTIVE: Relate blood pressure to weight and age.

For the i-th patient, $x_{i1} =$ is weight and $x_{i2} =$ is age.

MODEL:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

where $x_{i3} = x_{i1} \times x_{i2}$ is the weight-by-age interaction term.

4. EXAMPLE 4. *Data Transformations.*

Seber & Lee, Example 1.2–The Law of Gravity

The Inverse Square Law states that the force of gravity F between two bodies a distance D apart is given by

$$F = \frac{c}{D^\beta}.$$

Question: By transforming variables, how can this be viewed as a linear regression model for the paramter β?

MODEL:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

- where $Y_i = log(F_i),$
- $x_i = -log(D_i),$
- $\beta_0 = log(c),$
- $\beta_1 = \beta,$
- ε_i is an error term.

Seber states "... and from experimental data we can estimate β and test whether $\beta = 2$."

MATRIX REPRESENTATION OF LINEAR MODELS

The general linear model in matrix form:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n0} & x_{n1} & x_{n2} & \cdots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Equivalent shorthand form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$\mathbf{Y}$ $(n \times 1)$ is the response vector

$\mathbf{X}$ $(n \times p)$ is the design (or model or regression) matrix

$\boldsymbol{\beta}$ $(p \times 1)$ is the vector of regression coefficients (model parameters)

$\boldsymbol{\varepsilon}$ $(n \times 1)$ is the error vector (mean $\mathbf{0}$)

(a) Example 1

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

(b) Example 2

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

(c) Example 3

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

(d) Example 4

$$\begin{pmatrix} log(F_1) \\ log(F_2) \\ \vdots \\ log(F_n) \end{pmatrix} = \begin{pmatrix} 1 & -log(D_1) \\ 1 & -log(D_2) \\ \vdots & \vdots \\ 1 & -log(D_n) \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

(e) EXAMPLE 5. *One-way analysis of variance.*

OBJECTIVE: Compare two treatments for blood pressure.

Consider random samples of J individuals taking one of two blood pressure medications. Y_{ij} is the blood pressure for individual j from treatment group i.

MODEL:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

- μ = overall mean blood pressure,
- α_i = effect on blood pressure for treatment i $(i = 1, 2)$,
- ε_{ij} = error term for subject j receiving treatment i.

ALTERNATE MODEL REPRESENTATION:

$$Y_{ij} = \mu + \alpha_1 I_1 + \alpha_2 I_2 + \varepsilon_{ij}$$

where I_1 is an indicator variable for membership in treatment group 1 and I_2 is an indicator variable for membership in treatment group 2.

(f) EXAMPLE 6. *Two-way analysis of variance.*

OBJECTIVE: Same as EXAMPLE 5, but we now also consider a patient's sex.

MODEL:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

- μ = overall mean blood pressure,
- α_i = effect on blood pressure for treatment i $(i = 1, 2)$,
- β_j = effect on blood pressure for sex j $(j = 1, 2)$,
- ε_{ijk} = error term for subject k of sex j receiving tmnt i.

ALTERNATE MODEL REPRESENTATION:

$$Y_{ij} = \mu + \alpha_1 I_1 + \alpha_2 I_2 + +\beta_1 J_1 + \beta_2 J_2 + \varepsilon_{ijk}$$

where I_1 is an indicator variable for membership in treatment group 1, I_2 is an indicator variable for membership in treatment group 2, J_1 indicates sex 1, and J_2 indicates sex 2.

(g) EXAMPLE 7. *Analysis of covariance.*

OBJECTIVE: Same as EXAMPLE 5, controlling for age.

MODEL:

$$Y_{ij} = \mu + \alpha_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij}$$

- μ = overall mean blood pressure,
- α_i = effect on blood pressure for treatment i ($i = 1, 2$),
- β = slope parameter,
- x_{ij} = age of subject j receiving treatment i,
- $\bar{x}_{..}$ = overall mean age,
- ε_{ij} = error term for subject j receiving treatment i.

NOTE:

The alternative model representations for these ANOVA and ANCOVA models make it clear that these are linear models. Let's continue with matrix representation of these models.

5. Example 5

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1J} \\ Y_{21} \\ \vdots \\ Y_{2J} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1J} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2J} \end{pmatrix}$$

6. Example 6

$$\begin{pmatrix} Y_{111} \\ \vdots \\ Y_{11K} \\ \hline Y_{121} \\ \vdots \\ Y_{12K} \\ \hline Y_{211} \\ \vdots \\ Y_{21K} \\ \hline Y_{221} \\ \vdots \\ Y_{22K} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 1 & 0 \\ \hline 1 & 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 1 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{111} \\ \vdots \\ \varepsilon_{11K} \\ \hline \varepsilon_{121} \\ \vdots \\ \varepsilon_{12K} \\ \hline \varepsilon_{211} \\ \vdots \\ \varepsilon_{21K} \\ \hline \varepsilon_{221} \\ \vdots \\ \varepsilon_{22K} \end{pmatrix}$$

7. Example 7

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1J} \\ Y_{21} \\ \vdots \\ Y_{2J} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & (x_{11} - \bar{x}_{..}) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & (x_{1J} - \bar{x}_{..}) \\ 1 & 0 & 1 & (x_{21} - \bar{x}_{..}) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & (x_{2J} - \bar{x}_{..}) \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1J} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2J} \end{pmatrix}$$

In Summary: Linear models have the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{Y}^{n\times 1}$ response vector, $\mathbf{X}^{n\times p}$ model matrix, $\boldsymbol{\beta}^{p\times 1}$ vector of unknown regression parameters, $\boldsymbol{\varepsilon}^{n\times 1}$ mean zero random error vector.

Notes:

1. Usually $x_{i0} = 1$ for all i. That is, usually there is an *intercept* β_0 in the model and the first column of the design matrix X is all 1's.
2. $x_{i0}, x_{i1}, \ldots, x_{i,p-1}$ are called the *predictor* variables or *regressor* variables or the *covariates*. They are the data.
3. "Linear Model" means the model is *linear* in the unknown regression coefficients $\beta_0, \beta_1, \ldots, \beta_{p-1}$.
4. Instead of matrices, we can write the model in terms of vectors:

$$\mathbf{Y} = \sum_{j=0}^{p-1} \beta_j \mathbf{x}_j + \boldsymbol{\varepsilon},$$

where $\mathbf{x}_j = (x_{1j}, x_{2j}, \ldots, x_{nj})'$. (Also note that we will use the convention in this course that a vector is a column vector.)

5. $\boldsymbol{\varepsilon}$ is the random part of the model. Right now we just assume that $E(\varepsilon) = 0$. Later, we will make more assumptions about the distribution of $\boldsymbol{\varepsilon}$.
6. $\mathbf{Y}$ is random because $\boldsymbol{\varepsilon}$ is random. $\mathbf{Y}$ "inherits" randomness from $\boldsymbol{\varepsilon}$.
7. We can thus evaluate

$$E[\mathbf{Y}] = E[\sum_{j=0}^{p-1} \beta_j \mathbf{x}_j + \boldsymbol{\varepsilon}] = E[\sum_{j=0}^{p-1} \beta_j \mathbf{x}_j] + E[\boldsymbol{\varepsilon}] = \sum_{j=0}^{p-1} \beta_j \mathbf{x}_j.$$

8. The vector $E[\mathbf{Y}]$ is a linear combination of the $\mathbf{x}_j$.
9. $E[\mathbf{Y}] \in span(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{p-1}) \equiv \Omega,$

FACT: We obtain least squares estimators (LSE's) of the β_j, denoted $\hat{\beta}_j$, by projecting $\mathbf{Y}$ onto Ω. Note that $\mathbf{Y}$ is n-dimensional and Ω has dimension $\leq p$. The projection of $\mathbf{Y}$ onto Ω is denoted $\hat{\mathbf{Y}}$.

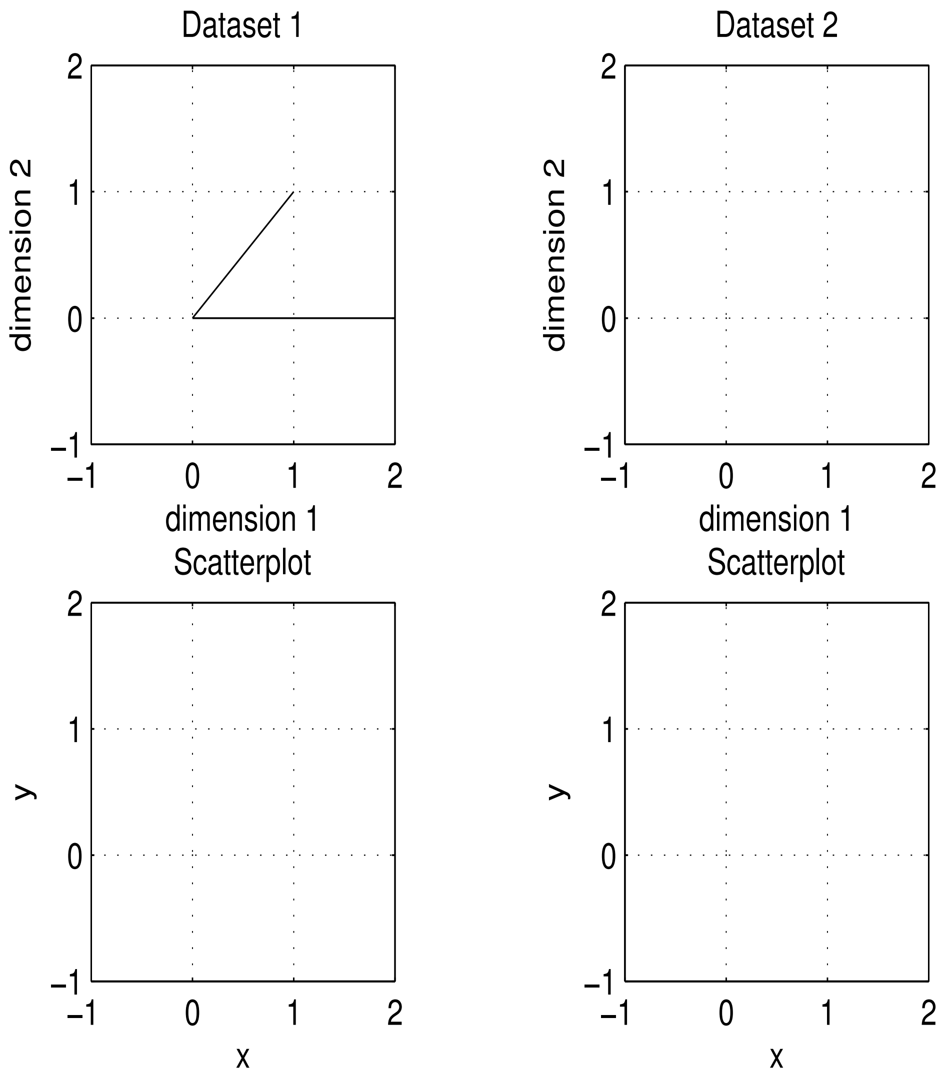
In Class Exercise: Model $Y_i = \beta x_i + \varepsilon_i$ (linear regression through the origin)

Two datasets:

1. $\{(x, y) = (2, 1), (0, 1)\}$
2. $\{(x, y) = (1, 2), (1/2, 2)\}$

For each dataset:

1. Write out the model with vectors and matrices (using the data).
2. Show the vectors $\mathbf{y}$ and $\mathbf{x}$ on a graph.
3. Identify (on your graph) Ω,
4. Plot $\hat{\mathbf{y}} = Proj_{\Omega}(\mathbf{y})$,
5. Plot $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}}$, identify $\Omega^{\perp}$
6. What is the dimension of Ω? What is the dimension of $\Omega^{\perp}$?
7. Also make a scatterplot of the data and sketch the least squares line.
8. Alternatively, if we fit the model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, how do your answers to the previous questions change?


Dataset 1
dimension 2
dimension 1
Dataset 2
dimension 2
dimension 1
Scatterplot
y
x
Scatterplot
y
x