More on linear models

Analysis of Ecological and Environmental Data

QERM 514

Mark Scheuerell

8 April 2020

Goals for today

- Understand how to represent a linear model with matrix notation
- Understand the concept, assumptions & practice of least squares estimation for linear models
- Understand the concept of identifiability

Linear models in matrix form

Simple regression

$$y_i = \alpha + \beta x_i + e_i$$
$$\Downarrow^*$$
$$y_i = \beta_0 + \beta_1 x_i + e_i$$

The i subscript indicates one of a total N observations

* The reason for this notation switch will become clear later

Linear models in matrix form

Simple regression

Let's make this general statement more specific

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

$$\Downarrow$$

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + e_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + e_2 \\ &\vdots \\ y_N &= \beta_0 + \beta_1 x_N + e_N \end{aligned}$$

Linear models in matrix form

Simple regression

Let’s now make the implicit “1” multiplier on β_0 explicit

$$
\begin{aligned}
y_1 &= \beta_0 \underline{1} + \beta_1 x_1 + e_1 \\
y_2 &= \beta_0 \underline{1} + \beta_1 x_2 + e_2 \\
&\vdots \\
y_N &= \beta_0 \underline{1} + \beta_1 x_N + e_N
\end{aligned}
$$

Linear models in matrix form

Simple regression

Let's next gather the common terms into column vectors

$$
\begin{aligned}
y_1 &= \beta_0 1 + \beta_1 x_1 + e_1 \\
y_2 &= \beta_0 1 + \beta_1 x_2 + e_2 \\
&\vdots \\
y_N &= \beta_0 1 + \beta_1 x_N + e_N
\end{aligned}
$$

Linear models in matrix form

Simple regression

Maybe something like this?

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_1 \\ \vdots \\ \beta_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

An aside on linear algebra

We refer to the dimensions of matrices in a row-by-column manner

$[rows \times columns]$

An aside on linear algebra

When adding matrices, the dimensions must match

$[m \times n] + [m \times n]$ ✓

$[m \times n] + [m \times p]$ X

An aside on linear algebra

When multiplying 2 matrices, the inner dimensions must match

$[m \times \underline{n}][\underline{n} \times p]$ ✓

$[m \times \underline{n}][\underline{p} \times n]$ X

An aside on linear algebra

When multiplying 2 matrices, the dimensions are
[rows-of-first × columns-of-second]

$[\underline{m} \times n][n \times \underline{p}] = [m \times p]$

Linear models in matrix form

Simple regression

Let's check the dimensions

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_1 \\ \vdots \\ \beta_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

Linear models in matrix form

Simple regression

Let's check the dimensions

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_1 \\ \vdots \\ \beta_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

$$[N \times 1] = \underbrace{[N \times 1][N \times 1]}_{\text{OOPS!}} + \underbrace{[N \times 1][N \times 1]}_{\text{OOPS!}} + [N \times 1]$$

An aside on linear algebra

When multiplying a scalar times a vector/matrix, it's just element-wise

$$a \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} aX \\ aY \\ aZ \end{bmatrix}$$

Linear models in matrix form

Simple regression

So this looks better

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

$$[N \times 1] = [N \times 1] + [N \times 1] + [N \times 1]$$

Linear models in matrix form

Simple regression

This is nice, but can we make β_0 and β_1 more compact?

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

Linear models in matrix form

Simple regression

What if we move β_0 & β_1 to the other side of the predictors…

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \beta_0 + \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \beta_1 + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

Linear models in matrix form

Simple regression

…and group the predictors and parameters into matrices

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}
$$

Linear models in matrix form

Simple regression

Let's check the dimensions

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

$$[N \times 1] = \underbrace{[N \times 2][1 \times 2]}_{\text{OOPS!}} + [N \times 1]$$

An aside on linear algebra

Matrix multiplication works on a row-times-column manner

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} aX + bY \\ cX + dY \end{bmatrix}$$

$$[2 \times 2][2 \times 1] = [2 \times 1]$$

Linear models in matrix form

Simple regression

Let's transpose the parameter vector $[\beta_0 \;\; \beta_1]^\top$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

Linear models in matrix form

Simple regression

and check the dimensions

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}
$$

$$
\begin{aligned}
[N \times 1] &= [N \times 2][2 \times 1] + [N \times 1] \\
&= [N \times 1] + [N \times 1] \\
&= [N \times 1]
\end{aligned}
$$

Linear models in matrix form

Simple regression

Lastly, we can write the model in a more compact notation

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

$$\Downarrow$$

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

Linear models in matrix form

Multiple regression

The matrix form is generalizaable to multiple predictors

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,2} \\ \vdots & \vdots & \\ 1 & x_{1,N} & x_{2,N} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}
$$

$$
\Downarrow
$$

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

QUESTIONS?

Ordinary least squares

In general, we have something like

$$DATA = MODEL + ERRORS$$

Ideally we have something like

$$DATA \approx MODEL$$

and hence

$$ERRORS \approx 0$$

Ordinary least squares

From this it follows that

$$\mathrm{Var}(DATA) = \mathrm{Var}(MODEL) + \mathrm{Var}(ERRORS)$$

Our hope is that

$$\mathrm{Var}(DATA) \approx \mathrm{Var}(MODEL)$$

and hence

$$\mathrm{Var}(ERRORS) \approx 0$$

Ordinary least squares

Our model for the data is

$$y_i = f(\text{predictors}_i) + e_i$$

and our estimate of y is

$$\hat{y}_i = f(\text{predictors}_i)$$

and therefore the errors (residuals) are given by

$$e_i = y_i - \hat{y}_i$$

In general, we want to minimize each of the e_i

Ordinary least squares

Specifically, we want to minimize the sum of their squares

$$\min \sum_{i=1}^{N} e_i^2 \;\Rightarrow\; \min \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Ordinary least squares

For our linear regression model, we have

$$\min \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

$$\Downarrow$$

$$\min \sum_{i=1}^{N} \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right)^2$$

An aside on linear algebra

Recall that matrix multiplication works in a row-by-column manner

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} aX + bY \\ cX + dY \end{bmatrix}$$

An aside on linear algebra

If $\mathbf{v}$ is an $[n \times 1]$ column vector & $\mathbf{v}^\top$ is its $[1 \times n]$ transpose,

multiplying $\mathbf{v}^\top \mathbf{v}$ gives the sum of the squared values in $\mathbf{v}$

$$\begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a^2 + b^2 + c^2 \end{bmatrix}$$

$$[1 \times n][n \times 1] = [1 \times 1]$$

Ordinary least squares

Writing our linear regression model in matrix form, we have

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}$$

$$\Downarrow$$

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

so the sum of squared errors is

$$\mathbf{e}^\top\mathbf{e} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Finding the minimum

For example, at what value of x does this parabola reach its minimum?

$$y = 2x^2 - 3x + 1$$

Recall from calculus that we

1. differentiate y with respect to x
2. set the result to 0
3. solve for x

Finding the minimum

For example, at what value of x does this parabola reach its minimum?

$$y = 2x^2 - 3x + 1$$
$$\Downarrow$$
$$\frac{dy}{dx} = 4x - 3$$
$$\Downarrow$$
$$4x - 3 = 0$$
$$x = \tfrac{3}{4}$$

Ordinary least squares

We want to minimize the sum of squared errors

$$\begin{aligned}\mathbf{e}^\top\mathbf{e} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}^\top\mathbf{y} - \mathbf{y}^\top\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{y} + \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}}\end{aligned}$$

and so we want

$$\frac{\partial}{\partial\hat{\boldsymbol{\beta}}}\mathbf{y}^\top\mathbf{y} - \mathbf{y}^\top\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{y} + \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}}$$

Ordinary least squares

(via several steps)

$$\frac{\partial SSE}{\partial \hat{\boldsymbol{\beta}}} = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}}$$

$$\Downarrow$$

$$-2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

$$\mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}$$

$$\Downarrow$$

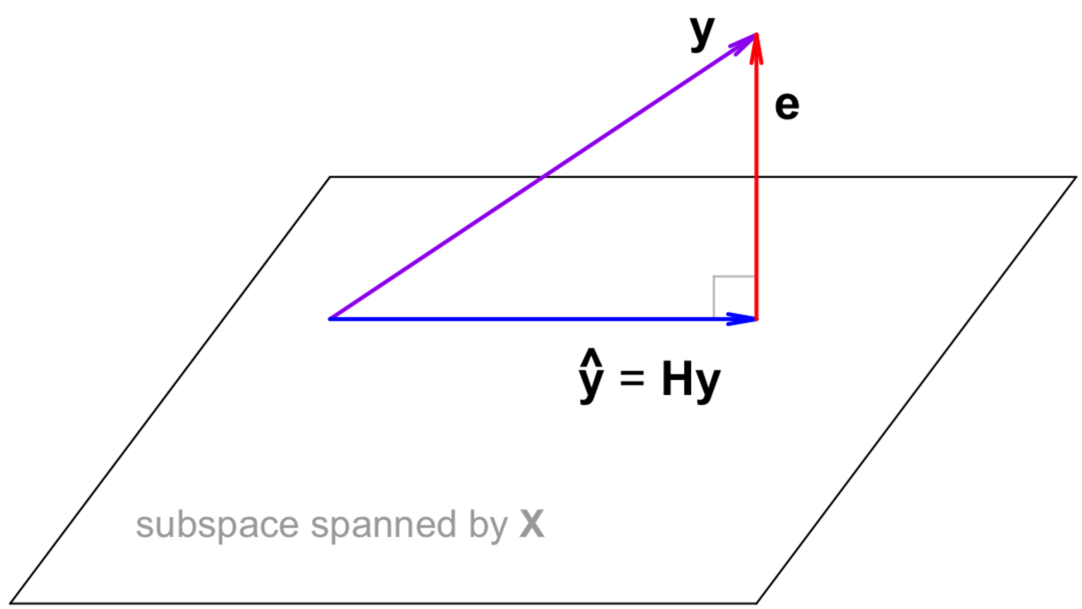
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$$

Ordinary least squares

Returning to our estimate of the data, we have

$$\begin{aligned}
\hat{\mathbf{y}} &= \mathbf{X}\hat{\boldsymbol{\beta}} \\
&= \mathbf{X}\left((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}\right) \\
&= \underbrace{\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top}_{\mathbf{H}}\mathbf{y} \\
&= \mathbf{H}\mathbf{y}
\end{aligned}$$

Ordinary least squares



$\mathbf{H}$ is called the “hat matrix” because it maps $\mathbf{y}$ onto $\hat{\mathbf{y}}$ (“y-hat”)

Ordinary least squares

Consider for a moment what it means if

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

Ordinary least squares

Consider for a moment what it means if

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$$

We can estimate the data without any model parameters!

Ordinary least squares

Key assumptions

- Model is linear in parameters*
- Observations y_i are a random sample from the population
- The predictor(s) is/are known without measurement error
- The predictor(s) is/are independent of the response
- If 2+ predictors, they are independent of each other
- Errors are IID: $e_i \sim \text{N}(0, \sigma^2); \text{Cov}(e_i, e_j) = 0$

*parameters are not multiplied or divided by other parameters, nor do they appear as an exponent

Independent & identically distributed

How do we know if our errors are IID?

- Knowledge of the problem/design
- Examine residual plots
- Tests of model fits

We will discuss this more in later lectures

Ordinary least squares

What can we say about $\hat{\boldsymbol{\beta}}$ when estimated this way?

1. It's the *maximum likelihood estimate* (MLE)
2. It's the *best linear unbiased estimate* (BLUE)

NOTE: these propoerties only hold if the errors (e_i) are *independent and identically distributed* (IID)

Identifiability

Recall the solution for $\hat{\boldsymbol{\beta}}$ where

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

If the quantity $\mathbf{X}^\top \mathbf{X}$ is not invertible, then $\hat{\boldsymbol{\beta}}$ is partially unidentifiable.

This occurs when the columns of $\mathbf{X}$ are not independent
(ie, $\mathbf{X}$ is not of “full rank”)

Lack of identifiability

When does it arise?

- analysis of designed experiments (more later)
- two predictors are perfectly correlated (eg, temperature entered in both degrees F & degrees C)
- predictors are subsets of one another (eg, counts of trees in 3 categories: DBH $\geq$ 10 cm, DBH $\geq$ 20 cm, DBH $\geq$ 30 cm)
- number of parameters equals or exceeds the observations
  $p = n$: model is *saturated*
  $p \geq n$: model is *supersaturated*