Multivariate: Linear regression

Table of contents

1 Goals

1.1 Goals

1.1.1 Goals of this lecture

- Fully transition to **matrix form** for linear regression
- Describe **matrix solution** to least squares estimation

2 Matrices in multiple regression

2.1 Matrices in multiple regression

2.1.1 Matrices in multiple regression

Data matrix

$$\underset{(n,p)}{\mathbf{X}} = \begin{bmatrix} X_{11} & X_{12} & ... & X_{1p} \\ X_{21} & X_{22} & ... & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & ... & X_{np} \end{bmatrix}$$

2.1.2 Matrices in multiple regression

Outcome variable

$$\underset{(n,1)}{\underline{y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

2.1.3 Matrices in multiple regression

Predicted outcome variable

$$\underset{(n,1)}{\underline{\hat{y}}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$$

2.1.4 Regression equation in matrix form

$$\underset{(n,1)}{\underline{\hat{y}}} = \underset{(n,p)}{\mathbf{X}} \underset{(p,1)}{\underline{b}} + \underset{(n,1)}{\underline{b}_0}$$

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & ... & X_{1p} \\ X_{21} & X_{22} & ... & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & ... & X_{np} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} + \begin{bmatrix} b_0 \\ b_0 \\ \vdots \\ b_0 \end{bmatrix}$$

2.2 Covariation, covariance, and correlation matrices

2.2.1 Covariation matrix P

We talked about the partitioned variation covariation matrix in general before

$$\mathbf{P}_{XX,YY} = \mathbf{M}'\,\mathbf{M} - \frac{1}{n}\mathbf{M}'\,\mathbf{E}\,\mathbf{M} = \left[\begin{array}{c|c} \mathbf{P}_{XX} & \mathbf{P}_{XY} \\ \hline \mathbf{P}_{YX} & \mathbf{P}_{YY} \end{array}\right]$$

2.2.2 Covariation matrix P

In linear regression, the variation covariation matrix becomes:

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{P}_{XX} & \underline{p}_{XY} \\ \hline \underline{p}_{YX} & SS_Y \end{array}\right] = \left[\begin{array}{cccc|c} SS_{x1} & SP_{x1,x2} & \dots & SP_{x1,xp} & SP_{x1,y} \\ SP_{x2,x1} & SS_{x2} & \dots & SP_{x2,xp} & SP_{x2,y} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ SP_{xp,x1} & SP_{xp,x2} & \dots & SS_{xp} & SP_{xp,y} \\ \hline SP_{y,x1} & SP_{y,x2} & \dots & SP_{y,xp} & SS_y \end{array}\right]$$

2.2.3 Covariation matrix P

- $\mathbf{P}_{XX}$: covariation matrix of the predictors
  - $p \times p$ matrix
- $\underline{p}_{XY}$: vector of covariations of each predictor with the outcome Y
  - $p \times 1$ vector
  - Its transpose, $\underline{p}_{YX}$, is a $1 \times p$ vector
- SS_Y: variation in the outcome
  - 1×1 or a scalar

2.2.4 Covariance matrix S

We talked about the partitioned variance covariance matrix in general before

$$\mathbf{S}_{XX,YY} = \frac{1}{(n-1)}\left(\mathbf{M}'\,\mathbf{M} - \frac{1}{n}\mathbf{M}'\,\mathbf{E}\,\mathbf{M}\right) = \left[\begin{array}{c|c} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \hline \mathbf{S}_{YX} & \mathbf{S}_{YY} \end{array}\right]$$

2.2.5 Covariance matrix S

In linear regression, the variance covariance matrix becomes:

$$\mathbf{S} = \frac{1}{n-1}\,\mathbf{P} = \left[\begin{array}{c|c} \mathbf{S}_{XX} & \underline{s}_{XY} \\ \hline \underline{s}_{YX} & s_y^2 \end{array}\right] = \left[\begin{array}{cccc|c} s_{x1}^2 & s_{x1,x2} & \cdots & s_{x1,xp} & s_{x1,y} \\ s_{x2,x1} & s_{x2}^2 & \cdots & s_{x2,xp} & s_{x2,y} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{xp,x1} & s_{xp,x2} & \cdots & s_{xp}^2 & s_{xp,y} \\ \hline s_{y,x1} & s_{y,x2} & \cdots & s_{y,xp} & s_y^2 \end{array}\right]$$

2.2.6 Covariance matrix S

- $\mathbf{S}_{XX}$: covariance matrix of the predictors
  - $p \times p$ matrix
- $\underline{s}_{XY}$: vector of covariances of each predictor with the outcome Y
  - $p \times 1$ vector
  - Its transpose, $\underline{s}_{YX}$, is a $1 \times p$ vector
- s_y^2 is the variance in the outcome
  - 1×1 or a scalar

2.2.7 Correlation matrix R

We talked about the partitioned correlation matrix in general before

$$\mathbf{R}_{XX,YY} = \left[\begin{array}{c|c} \mathbf{R}_{XX} & \mathbf{R}_{XY} \\ \hline \mathbf{R}_{YX} & \mathbf{R}_{YY} \end{array}\right]$$

2.2.8 Correlation matrix R

In linear regression, the correlation matrix becomes:

$$\mathbf{R} = \left[\begin{array}{c|c} \mathbf{R}_{XX} & \underline{r}_{XY} \\ \hline \underline{r}_{YX} & 1 \end{array}\right] = \left[\begin{array}{cccc|c} 1 & r_{x1,x2} & \cdots & r_{x1,xp} & r_{x1,y} \\ r_{x2,x1} & 1 & \cdots & r_{x2,xp} & r_{x2,y} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{xp,x1} & r_{xp,x2} & \cdots & 1 & r_{xp,y} \\ \hline r_{y,x1} & r_{y,x2} & \cdots & r_{y,xp} & 1 \end{array}\right]$$

2.2.9 Correlation matrix R

- $\mathbf{R}_{XX}$: correlation matrix of the predictors
  - $p \times p$ matrix
- $\underline{r}_{XY}$: vector of correlations of each predictor with the outcome Y
  - $p \times 1$ vector
  - Its transpose, $\underline{r}_{YX}$, is a $1 \times p$ vector
- 1 (in the bottom right): correlation of the outcome with itself
  - 1×1 or a scalar

3 Linear regression solution: Matrix!

3.1 Least squares solution

3.1.1 From last time...

Last time, we went through the **least squares solution** and the *normal equations* to solve for the regression coefficients in a model with a single predictor

$$b_1 = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2} = \frac{SP_{XY}}{SS_X} = \frac{s_{XY}}{s_X^2}$$

The regression coefficient b_1 is equal to *either*:

- **Covariation** between X and Y, divided by **variation** of X
- **Covariance** between X and Y, divided by **variance** of X

3.2 Regression solution in matrix form

3.2.1 General solution for linear regression

In the non-matrix approach, we could solve for coefficients in terms of **covariation**, **covariance**, or **correlation** (standardized solution)

There are several equivalent **matrix formulations** for solving for regression coefficients

1. In terms of **covariation** (unstandardized solution)
2. In terms of the **covariance** (unstandardized solution)
3. In terms of the **correlation** (standardized solution)

3.2.2 General solution (in terms of covariation)

In matrix form, the solution for **unstandardized** coefficients is:

$$\underline{b} = \mathbf{P}_{XX}^{-1}\ \underline{p}_{XY}$$

- $\underline{b}$: vector of regression coefficients
  - $p \times 1$ vector – does **not** include the **intercept**
- $\mathbf{P}_{XX}^{-1}$: **inverse** of the **covariation** matrix of the predictors
  - $p \times p$ matrix, just like the covariation matrix
- $\underline{p}_{XY}$: vector of **covariations** of each predictor with the outcome Y
  - $p \times 1$ vector

3.2.3 General solution (in terms of covariance)

In matrix form, the solution for **unstandardized** coefficients is:

$$\underline{b} = \mathbf{S}_{XX}^{-1}\ \underline{s}_{XY}$$

- $\underline{b}$: vector of regression coefficients
  - $p \times 1$ vector – does **not** include the **intercept**
- $\mathbf{S}_{XX}^{-1}$: **inverse** of the **covariance** matrix of the predictors
  - $p \times p$ matrix, just like the covariance matrix
- $\underline{s}_{XY}$: vector of **covariances** of each predictor with the outcome Y
  - $p \times 1$ vector

3.2.4 Obtaining the intercept

- For the solutions based on the covariation or the covariance:
  - **Intercept is not included in the vector of regression coefficients**

$$b_0 = \overline{Y} - \underline{\overline{X}}\ \underline{b}$$

$$= \overline{Y} - (b_1\overline{X}_1 + b_2\overline{X}_2 + \cdots + b_p\overline{X}_p)$$

3.2.5 General solution (in terms of correlation)

The matrix solution for **standardized** regression coefficients:

$$\underline{b} = \mathbf{R}_{XX}^{-1}\, \underline{r}_{XY}$$

- $\underline{b}$: vector of regression coefficients
  - $p \times 1$ vector – **no intercept for standardized solution**
- $\mathbf{R}_{XX}^{-1}$: **inverse** of the **correlation** matrix of the predictors
  - $p \times p$ matrix, just like the correlation matrix
- $\underline{r}_{XY}$: vector of **correlations** of each predictor with the outcome Y
  - $p \times 1$ vector

3.3 Least squares solution with augmented data matrix

3.3.1 Least squares solution with augmented data matrix

An alternative form of the solution uses the **augmented data matrix**

$$\underset{(n,p+1)}{\mathbf{X}_A} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

Note: I use $\mathbf{X}_A$ but there is no standard notation for raw data matrix vs augmented data matrix. Count the columns!

3.3.2 Regression with augmented data matrix

$$\underset{(n,1)}{\underline{\hat{y}}} = \underset{(n,p+1)}{\mathbf{X}_A}\ \underset{(p+1,1)}{\underline{b}}$$

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$$

3.3.3 Augmented vector of regression coefficients

Adds the intercept (b_0) to the vector of regression coefficients

Vector of regression coefficients becomes: $\underset{(p+1,1)}{\underline{b}} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$

3.3.4 Augmented data matrix

Augmented data matrix ($\mathbf{X}_A$) has a column of 1s as the first column of the matrix

The solution to OLS regression using the augmented data matrix:

$$\underline{b} = \left(\mathbf{X}'_A \mathbf{X}_A\right)^{-1} \mathbf{X}'_A \, \underline{y}$$

where $\underline{b}$ is the $(p+1) \times 1$ matrix of regression coefficients

Remember: this version includes the intercept in the vector of coefficients

3.4 Hat matrix

3.4.1 Regression diagnostics

- Regression diagnostics are measures of the extent to which deviant cases affect the outcome of the regression analysis
  - **Leverage**: Extreme cases in the predictor space
    - Most X values between 1 and 10, but one person has a value of 20
  - **Discrepancy**: Extreme cases in terms of residuals
    - How far is an observed point from its predicted value?
  - **Influence**: Cases that change the coefficients
    - Need to have high leverage and high discrepancy

3.4.2 Regression diagnostics: Leverage

- There are several measures of **leverage** and some slight differences between them depending on the software package you're using
  - They're all based on the **hat matrix**
  - The hat matrix is an $n \times n$ matrix
  - The values on the diagonal (one for each of the n subjects) are the **leverage** statistics

3.4.3 Hat matrix

- Using the augmented data matrix solution:
  - Predicted scores are given by: $\underline{\hat{y}} = \mathbf{X}_A \underline{b}$
- From a few slides ago: $\underline{b} = \left(\mathbf{X}_A' \mathbf{X}_A\right)^{-1} \mathbf{X}_A' \, \underline{y}$

Substitution:

$\underline{\hat{y}} = \mathbf{X}_A \left(\mathbf{X}_A' \mathbf{X}_A\right)^{-1} \mathbf{X}_A' \, \underline{y}$

3.4.4 Hat matrix

$\underline{\hat{y}} = \mathbf{X}_A \left(\mathbf{X}_A' \mathbf{X}_A\right)^{-1} \mathbf{X}_A' \underline{y}$

- **Hat matrix**
  - Everything highlighted in blue
  - Everything on the right side before $\underline{y}$
- **Why is it called that???**
  - It's how you go from Y (observed) to $\hat{Y}$ (predicted)
    - It puts the **hats** on the Ys