Introduction to Matrix Algebra and Multiple Regression with R

February 13, 2019

Multiple regression: Model and Regression Function

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + U,$$

$$X_1, X_2, \ldots, X_K \text{ are regressors,}$$

$$\beta_0, \beta_1, \ldots, \beta_K \text{ are the population parameters and}$$

$$U \text{ is the unobservable error term, } E(U) = 0$$

If we assume that $E(U|X_1, X_2, \ldots, X_K) = E(U)$ the population regression function (line/hyperplane) is

$$E(Y|X_1, X_2, \ldots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_K$$

Interpretation of β_1 and of β_0

- β_1 gives the change in $E(Y|X_1, X_2, \ldots, X_K)$ caused by a change in X_1 when $X_2, \ldots, X_K$ are held fixed at some values, $x_2, \ldots, x_K$.
- Thus, β_1 is the marginal/partial effect of X_1 on Y when we keep $X_2, \ldots, X_K$ fixed at some arbitrary values. In short, the "effect" of X_1 "controlling" for $X_2, \ldots, X_K$.
- The intercept, β_0, is the expected value of Y when all $x_1, x_2, \ldots, x_K$ equal zero.

LS Estimation

Consider $K = 2$ **Link: regression plane**

$$E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$\min_{\beta_0,\beta_1,\beta_2} \sum (Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}))^2$$

We will not derive the estimators $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$. It just becomes a lot of algebra. Using matrix algebra for the general case is more straightforward (on coming slides). Predicted (fitted) values are

$$\hat{E}(Y_i|X_{1i}, X_{2i}) = \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

and the residuals are

$$\hat{U}_i = Y_i - \hat{Y}_i$$

Example

- Use the Prestige data again and now run a regression of the following model

$$\text{prestige} = \beta_0 + \beta_1 \mathit{income} + \beta_2 \mathit{women} + U$$

- What is the predicted prestige average ($\hat{Y}$) if income is held constant at the sample average and the proportion women is 25 %?

Henceforth

The rest of the lecture aims to use matrix algebra in R and at the same time learn about Multiple regression

Matrices and vectors

$$A_{3,2} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{pmatrix} \qquad B_{3,2} = \begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ b_{3,1} & b_{3,2} \end{pmatrix}$$

$$V_3 = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

Addition

$$A_{3,2}+B_{3,2} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{pmatrix} + \begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ b_{3,1} & b_{3,2} \end{pmatrix} = \begin{pmatrix} a_{1,1}+b_{1,1} & a_{1,2}+b_{1,2} \\ a_{2,1}+b_{2,1} & a_{2,2}+b_{2,2} \\ a_{3,1}+b_{3,1} & a_{3,2}+b_{3,2} \end{pmatrix}$$

$$A_{3,2} + V_3 = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} a_{1,1}+v_1 & a_{1,2}+v_1 \\ a_{2,1}+v_2 & a_{2,2}+v_2 \\ a_{3,1}+v_3 & a_{3,2}+v_3 \end{pmatrix}$$

Subtraction

$$A_{3,2} - B_{3,2} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{pmatrix} - \begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ b_{3,1} & b_{3,2} \end{pmatrix} = \begin{pmatrix} a_{1,1} - b_{1,1} & a_{1,2} - b_{1,2} \\ a_{2,1} - b_{2,1} & a_{2,2} - b_{2,2} \\ a_{3,1} - b_{3,1} & a_{3,2} - b_{3,2} \end{pmatrix}$$

$$A_{3,2} - V_3 = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{pmatrix} - \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} a_{1,1} - v_1 & a_{1,2} - v_1 \\ a_{2,1} - v_2 & a_{2,2} - v_2 \\ a_{3,1} - v_3 & a_{3,2} - v_3 \end{pmatrix}$$

Transpose

$$A'_{3,2} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{pmatrix}' = \begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,2} & a_{2,2} & a_{3,2} \end{pmatrix}$$

$$V' = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}' = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix}$$

- Addition Rule: $(A + B)' = A' + B'$

Matrix multiplication

$$A'_{3,2}B_{3,2} = \begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,2} & a_{2,2} & a_{3,2} \end{pmatrix} \times \begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ b_{3,1} & b_{3,2} \end{pmatrix} =$$

$$= \begin{pmatrix} \sum_{i=1}^{3} a_{i,1}b_{i,1} & \sum_{i=1}^{3} a_{i,1}b_{i,2} \\ \sum_{i=1}^{3} b_{i,2}a_{i,1} & \sum_{i=1}^{3} a_{i,2}b_{i,2} \end{pmatrix}$$

$$A'_{3,2}V = \begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,2} & a_{2,2} & a_{3,2} \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{3} a_{i,1}v_i \\ \sum_{i=1}^{3} a_{i,2}v_i \end{pmatrix}$$

Rules:

- A(B+C)=AB+AC
- (A+B)C=AC+BC
- $(AB)' = B'A'$

In R, AB is written $A\% * \%B$.

Exersice

Consider

$$X_i = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$$

and

$$\beta = \begin{pmatrix} 0.5 \\ 1 \\ 2 \end{pmatrix}$$

what is

$$X_i \beta \ ?$$

Exersice

Consider

$$X_i = \begin{pmatrix} 1 & X_{1i} & X_{2i} \end{pmatrix}$$

and

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

what is

$$X_i \beta \text{ ?}$$

Extend on the exercise

Consider

$$\mathbf{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{K1} \\ 1 & X_{12} & \cdots & X_{K2} \\ \vdots & \vdots & \ddots & \ldots \\ 1 & X_{1n} & \cdots & X_{Kn} \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \mathbf{U} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}$$

then we can write the regression models for each observation combined like

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}$$

which is equivalent to

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}$$

The inverse and the identity Matrix

Consider a **square** matrix, e.g.

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}.$$

If

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

then, by definition, B is the inverse of A denoted A^{-1}. The matrix on the right-hand side is called the identity matrix denoted

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example

The function *solve(A)* give the inverse of matrix A in R. Find the inverse of

$$\mathbf{X} = \begin{pmatrix} 5 & 4 \\ 3 & 7 \end{pmatrix}$$

and control

$$\mathbf{X}\mathbf{X}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Also compute $\mathbf{X}I$ and $I\beta$ where

$$\beta = \begin{pmatrix} 2.5 \\ 1.5 \end{pmatrix}$$

Symmetric and diagonal matrices

A **symmetric matrix** A has elements $a_{i,j} = a_{j,i}$, e.g.

$$\mathbf{A} = \begin{pmatrix} 5 & 3 & 1 \\ 3 & 7 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

A **diagonal matrix** B is a symmetric matrix but all $b_{i,j} = 0,\ i \neq j$, the identity matrix is one example and a different example is this matrix:

$$\mathbf{B} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Rule:

- If A is symmetric then A=A'

Matrix operations in R

Consider two $n \times m$ matrices A and B and a vector V of length n

Opearation	In R
Addition	$A + B$ and $A + V$
Subtraction	$A - B$ and $A - V$
transpose	$t(A)$
Multiplication	$t(A)$ %*%B and t(A)%*%V
Hadamard product	$A * B$ and $A * V$
Inverse of a square matrix C	solve(C)
Diagonal of a square matrix C	diag(C)

Gradient

A gradient is a vector including all first order partial derivatives of a multivariate function $f(x_1, x_2 \dots, x_K) = f(\mathbf{x})$:

$$g = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_k} \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_k \end{bmatrix}$$

Two Rules:

- $\frac{\partial V'\mathbf{x}}{\partial \mathbf{x}} = V$ where both $\mathbf{x}$ and V are vectors
- $\frac{\partial \mathbf{x}' A\mathbf{x}}{\partial \mathbf{x}} = 2A\mathbf{x}$ where A is a symmetric matrix

Link: Examples of the two rules

Rules useful for LS-derivations

1. $(A + B)' = A' + B'$
2. $A(B + C) = AB + AC$
3. $(A + B)C = AC + BC$
4. $(AB)' = B'A'$
5. $(A^{-1})' = (A')^{-1}$
6. If A is symmetric then A=A'
7. $V_1'V_2 = V_2'V_1$, if V_1 and V_2 are two vectors of the same length
8. $\frac{\partial V'\mathbf{x}}{\partial \mathbf{x}} = V$ where both $\mathbf{x}$ and V are vectors
9. $\frac{\partial \mathbf{x}'A\mathbf{x}}{\partial \mathbf{x}} = 2A\mathbf{x}$ where A is a symmetric matrix
10. $cAB = AcB = ABc$ where A and B are matrices or vectors and c is a single value (scalar)

The general least squares estimator

To obtain the least squares estimator one minimise the sum of the squared residuals also called Residual sum of squares (RSS). Note

$$RSS = \sum_{i=1}^{n} U_i^2 = \mathbf{U}'\mathbf{U} = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

and with some matrix algebra **Link: here**

$$\begin{gathered}(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) = \\ \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta\end{gathered}$$

Now take the first order condition for minimising RSS w.r.t. β

$$\frac{\partial \mathbf{U}'\mathbf{U}}{\partial \beta} = -2(\mathbf{Y}'\mathbf{X})' + 2\mathbf{X}'\mathbf{X}\beta = \mathbf{0}$$

Link: Derivations

The general least squares estimator

$$
\begin{aligned}
&-2(\mathbf{Y}'\mathbf{X})' + 2\mathbf{X}'\mathbf{X}\beta = \mathbf{0} \Leftrightarrow \\
&\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{Y} \Leftrightarrow \\
&\beta = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}
\end{aligned}
$$

thus

$$\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \tag{1}$$

is the least squares estimator.

Exercise

- Use the Prestige data in the car package and compute the Least squares estimator given by (1), with help of matrix algebra operators in R.

$$\mathrm{prestige} = \beta_0 + \beta_1 \mathit{income} + \beta_2 \mathit{women} + U$$

Hypothesis testing

The lm() function in R tests the following hypotheses for each single coefficient

$$H_0 : \beta_k = 0,\ H_a : \beta_k \neq 0$$

For example for the model:

$$\text{prestige} = \beta_0 + \beta_1 income + \beta_2 women + U$$

$$H_0 : \beta_1 = 0,\ H_a : \beta_1 \neq 0$$

means that we test

- the null-hypothesis (H_0): *income* has no effect on *prestige*
- against the alternative hypothesis (H_a): *income* has an effect on *prestige*

The LS estimator: Small sample inference

If

1. $U|X \sim N(0, \sigma^2)$
2. And we sample (X_i, Y_i), $i = 1, \ldots, n$ such that $U_1, U_2, \ldots, U_n$ are iid link about iid, then

$$T = \frac{\hat{\beta}_k - \beta_k}{\sqrt{S^2 \left((\mathbf{X}'\mathbf{X})^{-1}\right)_{(k+1)(k+1)}}} \sim t_{n-K-1},$$

where $S^2 = \sum_{i=1}^{n} \hat{U}_i^2/(n-K-1)$, K is the number of regressors (X-variables) and $\left((\mathbf{X}'\mathbf{X})^{-1}\right)_{(k+1)(k+1)}$ is the $((k+1),(k+1))$ element of $(\mathbf{X}'\mathbf{X})^{-1}$

R and many other statistical packages are using this T-statistic as the default way to test:

$$H_0 : \beta_k = 0, \; H_a : \beta_k \neq 0$$

The LS estimator: Small sample inference

The assumptions

1. $U|X \sim N(0, \sigma^2)$
2. Sampling $(X_i, Y_i),\ i = 1, \ldots, n$ such that $U_1, U_2, \ldots, U_n$ are iid

imply

- U is normally distributed, i.e. we assume we know the population distribution of the unexplained part of Y!?
- Homoskedasticity: $V(U|X) = V(U) = \sigma^2$
  - For example if, Y is wage and X is age, then we assume that the wage-dispersion is constant across different age groups
- The sampling is independent
  - This more or less rules out time-series data, for which it is difficult to assume that the random errors for the same unit or individual are not dependent over time. $Cov(U_1, U_2) \neq 0$

The LS estimator: Large sample inference

If we have a large sample size, large n, we don't have to assume:

- Normal U
- Homoskedasticity
- Independence $Cov(U_1, U_2) = 0$

Next Lecture

- Large sample T-testing
- More about the concepts of homo- and heteroskedastic random error U
- Joint hypothesis testing
- Non-linear regression curves