Lecture7 Handout

Elizabeth Colantuoni

2/14/2021

I. Objectives

Upon completion of this session, you will be able to do the following:

- Use vector notation to specify the multiple linear regression model (*REVIEW*)
- Derive the least squares estimators using vector notation (*REVIEW*)
- Give a geometric explanation of least squares
- Derive the exact (under Gaussian model) or asymptotic distribution of the major regression results in vector notation: regression coefficients, linear functions thereof, predicted values, residuals
- Design and implement a simulation study to evaluate the properties of regression coefficients in MLR when the Gaussian model for residuals doesn't hold

II. REVIEW: MLR in vector notation

In this section, we will walk back through the derivations of $\underset{\sim}{\hat{\beta}}$ but expressing the regression models using vector and matrix notation.

Consider the following structure for our regression problem for $i = 1, ..., n$:

$$Y_i = \beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + ... + X_{pi}\beta_p + \epsilon_i$$

where ϵ_i are independently distributed as $N(0, \sigma^2)$

We can then stack each individuals data into a table structure:

$$\begin{array}{cccccccccccccc}
\underline{i} & & & & & & & & & & & & & \\
1 & Y_1 & = & 1 \times \beta_0 & + & X_{11}\beta_1 & + & X_{21}\beta_2 & + & ... & + & X_{p1}\beta_p & = & \epsilon_1 \\
2 & Y_2 & = & 1 \times \beta_0 & + & X_{12}\beta_1 & + & X_{22}\beta_2 & + & ... & + & X_{p2}\beta_p & = & \epsilon_2 \\
. & & & & & & & & & & & & & \\
. & & & & & & & & & & & & & \\
. & & & & & & & & & & & & & \\
\text{n} & Y_n & = & 1 \times \beta_0 & + & X_{1n}\beta_1 & + & X_{2n}\beta_2 & + & ... & + & X_{pn}\beta_p & = & \epsilon_n
\end{array}$$

We can then think about creating vectors that contain the same type of information for each element of our model:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_p \end{pmatrix} \begin{matrix} = \\ = \\ . \\ . \\ . \\ = \end{matrix} \begin{pmatrix} 1 \times \beta_0 \\ 1 \times \beta_0 \\ . \\ . \\ . \\ 1 \times \beta_0 \end{pmatrix} \begin{matrix} + \\ + \\ . \\ . \\ . \\ + \end{matrix} \begin{pmatrix} X_{11}\beta_1 \\ X_{12}\beta_1 \\ . \\ . \\ . \\ X_{1n}\beta_1 \end{pmatrix} \begin{matrix} + \\ + \\ . \\ . \\ . \\ + \end{matrix} \begin{pmatrix} X_{21}\beta_2 \\ X_{22}\beta_2 \\ . \\ . \\ . \\ X_{2n}\beta_2 \end{pmatrix} \begin{matrix} + & ... & + \\ + & ... & + \\ . & . & . \\ . & . & . \\ . & . & . \\ + & ... & + \end{matrix} \begin{pmatrix} X_{p1}\beta_p \\ X_{p2}\beta_p \\ . \\ . \\ . \\ X_{pn}\beta_p \end{pmatrix} \begin{matrix} = \\ = \\ . \\ . \\ . \\ = \end{matrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ . \\ \epsilon_n \end{pmatrix}$$

We can then express the system of equations in vector notation:

$$\underset{n \times 1}{\underset{\sim}{Y}} = \underset{n \times 1}{\underset{\sim}{1}\beta_0} + \underset{n \times 1}{\underset{\sim}{X}_1\beta_1} + \underset{n \times 1}{\underset{\sim}{X}_2\beta_2} + ... + \underset{n \times 1}{\underset{\sim}{X}_p\beta_p} = \underset{n \times 1}{\underset{\sim}{\epsilon}}$$

Further, we can append the column vectors represented by $\underset{\sim}{1}$, $\underset{\sim}{X}_1$, $\ldots$, $\underset{\sim}{X}_p$ into a matrix and multiply this matrix with the vector of regression coefficients:

$$\underset{\sim}{Y} = \left(\underset{\sim}{1}, \underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_p\right) \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_p \end{pmatrix} + \underset{\sim}{\epsilon}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & ... & X_{p1} \\ 1 & X_{12} & X_{22} & ... & X_{p2} \\ . & . & . & ... & . \\ . & . & . & ... & . \\ . & . & . & ... & . \\ 1 & X_{1n} & X_{2n} & ... & X_{pn} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ . \\ \epsilon_n \end{pmatrix}$$

This organization of the model leaves us with the following matrix representation of the MLR:

$$\underset{n\times 1}{\underset{\sim}{Y}} = \underset{n\times(p+1)}{X} \quad \underset{(p+1)\times 1}{\underset{\sim}{\beta}} + \underset{n\times 1}{\underset{\sim}{\epsilon}}$$

How do we express the distribution of $\underset{\sim}{\epsilon}$?

A. Multivariate Gaussian distribution

The multivariate Gaussian disribution describes the marginal and joint distribution of 2 or more Gaussian random variables. In matrix notation, we can define the multivariate Gaussian distribution as:

$$\underset{\sim}{Y} \sim MVN(\underset{\sim}{\mu}, V)$$

where,

$$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix}, \underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ . \\ \mu_n \end{pmatrix}, \text{ and } V = \begin{bmatrix} v_{11} & v_{12} & . & . & . & v_{1n} \\ v_{21} & v_{22} & . & . & . & v_{2n} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ v_{n1} & v_{n2} & . & . & . & v_{nn} \end{bmatrix}$$

with $v_{ii} = Var(Y_i)$ and $v_{ij} = Cov(Y_i, Y_j)$.

Given our MLR under the assumption that $\epsilon_i iid N(0, \sigma^2)$ then we have:

- $E(Y_i) = \mu_i = X_i \underset{\sim}{\beta}$ where $X_i = (1, X_{1i}, X_{2i}, ..., X_{pi})$ (the ith row of X)
- $Var(Y_i) = \sigma^2$ and $Cov(Y_i, Y_j) = 0$

And we can express the multivariate normal distribution for $\underset{\sim}{\epsilon}$ as $\underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \sigma^2 I)$ where I is the identity matrix with $1s$ on the diagonal elements and $0s$ on the off-diagonal elements.

B. Maximum likelihood estimation = least squares using vector notation

Using vector notation, our MLR is:

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}, \underset{\sim}{\epsilon} \sim MVN(\underset{\sim}{0}, \sigma^2 I)$$

In the remainder of this section, I will drop the $\underset{\sim}{}$ so you should assume Y and β are $n \times 1$ and $(p+1) \times 1$ vectors, respectively, and X is the $n \times (p+1)$ design matrix.

Our goal is to select estimates of β and σ^2 to minimize $\sum_{i=1}^{n} r_i(\hat{\beta}) = \sum_{i=1}^{n} (y_i - X_i\beta)^2$.

Using the vector notation, we can express the sums of squared residuals as:

$$\sum_{i=1}^{n} (y_i - X_i\beta)^2 = (Y - X\beta)'(Y - X\beta) = ||Y - X\beta||^2$$

where Y' is the transpose of Y: if Y is a $n \times 1$ vector, then Y' is a $1 \times n$ vector. If X is a $n \times (p+1)$ matrix, then X' is a $(p+1) \times n$ matrix.

The score equations for β can be written as:

$$U_\beta(\beta) = \frac{\partial}{\partial \beta}(Y - X\beta)'(Y - X\beta) = X'(Y - X\beta) = 0$$

Solving the score equations for β, we have:

$$\begin{aligned} X'Y - X'X\beta &= 0 \\ X'Y &= (X'X)\beta \\ (X'X)^{-1}X'Y &= \hat{\beta} \end{aligned}$$

C. Major regression results in vector notation with distributions

Now that we have an expression for the MLE of β in the MLR expressed using vector notation. In this section, we will provide definitions of key regression results in vector notation and derive distributions of key results under the Gaussian assumption.

1. Distribution of $\hat{\beta}$

The MLE of β is given by: $\hat{\beta} = (X'X)^{-1}X'Y = AY$ where $A = (X'X)^{-1}X'$.

We know that $Y \sim MVN(X\beta, \sigma^2 I)$, so $AY \sim MVN(AX\beta, \sigma^2 AIA')$.

Plugging in $A = (X'X)^{-1}X'$, we have:

$$\begin{aligned} AX\beta &= (X'X)^{-1}X'X\beta \\ &= \beta \end{aligned}$$

$$\begin{aligned} AIA' &= (X'X)^{-1}X'((X'X)^{-1}X')' \\ &= (X'X)^{-1}X'X((X'X)^{-1})' \\ &= (X'X)^{-1}(X'X)(X'X)^{-1} \\ &= (X'X)^{-1} \end{aligned}$$

So that: $\hat{\beta} \sim MVN(\beta, \sigma^2(X'X)^{-1})$

- Does this look familiar?
- NOTE: $X'X$ and $(X'X)^{-1}$ are symmetric and $X'X(X'X)^{-1} = (X'X)^{-1}X'X = I_{n\times n}$

2. Predicted values

In vector notation, the predicted values are:

$$
\begin{aligned}
\hat{Y} &= X\hat{\beta} \\
&= X(X'X)^{-1}X'Y \\
&= \left[X(X'X)^{-1}X'\right]Y \\
&= HY
\end{aligned}
$$

The matrix H is referred to as the *hat-matrix*.

To obtain the predicted value for Y_i, you multiple the ith row of H with the vector Y, such that the predicted value for each i is a weighted average of the observed y with the weights defined by the ith row of H. **See HW1**

The distribution of $\hat{Y}$ is multivariate normal with mean $E(\hat{Y}) = E(X\hat{\beta}) = X\beta$ and variance:

$$
\begin{aligned}
Var(\hat{Y}) &= Var(X\hat{\beta}) \\
&= XVar(\hat{\beta})X' \\
&= \sigma^2 X(X'X)^{-1}X' \\
&= \sigma^2 H
\end{aligned}
$$

i. Properties of the Hat matrix

The Hat matrix has some unique properties.

1. H is symmetric:

$$
\begin{aligned}
H' &= \left[X(X'X)^{-1}X'\right]' \\
&= X(X'X)^{-1'}X' \\
&= X(X'X)^{-1}X' \\
&= H
\end{aligned}
$$

2. H is idempotent, i.e. $HH = H$

$$
\begin{aligned}
HH &= \left[X(X'X)^{-1}X'\right]\left[X(X'X)^{-1}X'\right] \\
&= X(X'X)^{-1}X'X(X'X)^{-1}X' \\
&= X(X'X)^{-1}X' \\
&= H
\end{aligned}
$$

3. Residuals

In vector notation, the residuals are: $\hat{R} = Y - \hat{Y} = Y - HY = (I - H)Y$.

The distribution of the residuals will be multivariate normal with mean: $E(Y - \hat{Y}) = X\beta - X\beta = 0$ with variance: $Var(Y - \hat{Y}) = Var((I - H)Y) = \sigma^2(I - H)$. Can you derive the $Var(\hat{R})$ yourself?

4. Relationship between $\hat{Y}$ and R

We can show that the predicted values and residuals are independent. Does this make sense?

$$\begin{aligned} Cov(\hat{Y}, \hat{R}) &= E\left[HY\left((I-H)Y\right)'\right] \\ &= HE\left[YY'\right](I-H) \\ &= H\sigma^2 I(I-H) \\ &= \sigma^2 H(I-H) = 0 \end{aligned}$$

D. Geometry of least squares

See lecture slides for motivation of geometry of least squares.

III. Normality of regression coefficients

See extra Rmarkdown file for the simulation set up and your group exercise.