Linear Statistical Modeling Methods with SAS

Matrix Approach to Simple Regression Analysis

Xuemao Zhang
East Stroudsburg University

February 23, 2024

Outline

- Matrix Algebra - A Prelude to Multiple Regression
- Multivariate Probability Distributions
- Matrix Approach for Multivariate Probability Distributions
- Matrix Approach to SLR

Matrix Algebra

- See the matrix algebra review document MatrixAlgebra.pdf.
- Check the videos Linear Algebra Review by Andrew Ng - Stanford University: https://youtu.be/7wpfu30FYJM?si=vtdmQfdXkDnXmI9E

Multivariate Probability Distributions

Definition. An n-dimensional **random vector** is a function from a sample space S into $\mathbb{R}^n$, n-dimensional Euclidean space.

If a random vector is 2-dimensional, the distribution is called **Bivariate** Probability Distribution.

Definition. The **joint distribution function** or **joint CDF** for any random variables X and Y is the function F defined by

$$\boxed{F(x, y) = P(X \leq x, Y \leq y)}$$

Joint probability density function. Let Y_1 and Y_2 be continuous random variables with joint distribution function $F(y_1, y_2)$. If there exists a nonnegative function $f(y_1, y_2)$, such that

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2)\, dt_2 dt_1,$$

for all $-\infty < y_1 < \infty, -\infty < y_2 < -\infty$, then Y_1 and Y_2 are said to be **jointly continuous** random variables. The function $f(y_1, y_2)$ is called the **joint probability density function**.

Multivariate Probability Distributions

If Y_1 and Y_2 are jointly continuous and $F(y_1, y_2)$ is the joint cdf, then

$$f(y_1, y_2) = \frac{\partial^2}{\partial y_1 \partial y_2} F(y_1, y_2)$$

if $F(y_1, y_2)$ is differentiable.

A pair (X, Y) of continuous random variables is described by its joint density function f: For any region R,

$$\boxed{P((X, Y) \in R) = \int \int_R f(x, y)\, dxdy}.$$

we can define a probability function (or probability density function) for the intersection of n events $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$.

The joint density function of $Y_1, Y_2, \ldots, Y_n$ is given by $f(y_1, y_2, \ldots, y_n)$ such that

$$\begin{aligned} &P(Y_1 \leq y_1, Y_2 \leq y_2, \ldots, Y_n \leq y_n) = F(y_1, y_2, \ldots, y_n) \\ &= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_n} f(t_1, t_2, \ldots, t_n) dt_n \cdots dt_1. \end{aligned}$$

Matrix Approach for Multivariate Probability Distributions

Expectation of a Random Vector: Suppose we have a n-dimensional vector, $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$.

Then the expected value of $\mathbf{Y}$, denoted by $E(\mathbf{Y})$, is defined by

$$E(\mathbf{Y}) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix}.$$

That is, the expected value of a random vector is a vector whose elements are the expected values of the random variables that are the elements of the random vector.

Expectation of a Random Matrix: Similarly, the expected value of a random matrix is defined to be a matrix whose elements are the expected values of the corresponding random variables in the original matrix.

Matrix Approach for Multivariate Probability Distributions

Variance-Covariance Matrix of a Random Vector Suppose we have a n-dimensional vector, $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$. Then the Variance-Covariance Matrix of $\mathbf{Y}$, denoted by $Var(\mathbf{Y})$, is defined by

$$Var(\mathbf{Y}) = \begin{bmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_1, Y_n) \\ Cov(Y_2, Y_1) & Var(Y_2) & \cdots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_2) & \cdots & Var(Y_n) \end{bmatrix}.$$

Note.

- The Variance-Covariance Matrix $Var(\mathbf{Y})$ is symmetric since $Cov(Y_i, Y_j) = Cov(Y_j, Y_i)$.
- $Var(\mathbf{Y}) = E\{[\mathbf{Y} - E(\mathbf{Y})][\mathbf{Y} - E(\mathbf{Y})]'\} =$
  $$E\left\{ \begin{bmatrix} Y_1 - E(Y_1) \\ Y_2 - E(Y_2) \\ \vdots \\ Y_n - E(Y_n) \end{bmatrix} [Y_1 - E(Y_1), Y_2 - E(Y_2), \cdots, Y_n - E(Y_n)] \right\}$$

Multivariate Normal Distribution

Define

$$\boxed{\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}}$$

Definition. A random vector

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

is said to have a $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution if its pdf is given by

$$f(\mathbf{y}) = f(y_1, \ldots, y_n) = \left(\frac{1}{2\pi}\right)^{n/2} \left[\frac{1}{\det \boldsymbol{\Sigma}}\right]^{1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right].$$

If $n = 2$, the distribution is called Bivariate Normal Distribution. Let Y_1 and Y_2 have a bivariate normal distribution, then

$$\boxed{\boldsymbol{\mu} = (\mu_1, \mu_2)', \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}}$$

Multivariate Normal Distribution

Theorem (Linear combination). Let $\mathbf{X} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

$$\mathbf{Y} = \mathbf{C}\mathbf{X} \sim MVN(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T),$$

where $\mathbf{C}$ is a non-singular matrix.

Theorem (Marginal distributions). Let $\mathbf{X} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The marginal distribution of any set of component $\mathbf{X}$ is multivariate normal with means, variance and covariance obtained by taking the corresponding components of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively.

Theorem (Conditional distributions). Let $\mathbf{X}$ be a n-dimensional random vector and $\mathbf{Y}$ be an m-dimensional random vector. Suppose

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim MVN_{n+m}\left(\begin{pmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) \quad \text{with} \quad \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^T,$$

then

$$\mathbf{X} | \mathbf{Y} = \mathbf{y} \sim MVN_n(\boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}}), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

Bivariate Normal Distribution

Theorem (Marginal distributions). Let $\mathbf{Y}_1$ and Y_2 have a bivariate normal distribution. Then

(a). The marginal distribution of Y_1 is normal with mean μ_1 and variance σ_1^2.

(b). The marginal distribution of Y_2 is normal with mean μ_2 and variance σ_2^2.

Theorem (Conditional distributions). Let Y_1 and Y_2 have a bivariate normal distribution. Then the conditional distribution of Y_1 given that $Y_2 = y_2$ is a normal distribution with mean

$$\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y_2 - \mu_2)$$

and variance

$$\sigma_1^2(1 - \rho^2).$$

Matrix Approach to SLR

The General Linear Model

Consider the SLR model

$$Y_i|_{X=x_i} = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where x_i is the ith observation of X, β_0 and β_1 are unknown parameters, ε_i's are i.i.d random variables with 0 mean and common variance σ^2, $i = 1, \ldots, n$.

We now define the following matrices:

$$\boxed{\begin{array}{ll} \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, & \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{pmatrix} \\ \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, & \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \end{array}}$$

Matrix Approach to SLR

The General Linear Model

Then the SLR model can be written as

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

where ε has a multivariate distribution with mean $\mathbf{0}$ and variance-covariance matrix $\sigma^2 I_n$, and I_n is a n-dimensional identity matrix. Note that the variance-covariance matrix is given by

$$\begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}.$$

And hence

$$E(\mathbf{Y}) = \mathbf{X}\beta$$

since $E(\varepsilon) = \mathbf{0}$.

Matrix Approach to SLR

The General Linear Model

Least-Squares Equations and Solutions for a General Linear Model:
Equations:$(\mathbf{X}^{'}\mathbf{X})\widehat{\beta} = \mathbf{X}^{'}\mathbf{Y}$
Solutions:$\widehat{\beta} = (\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}\mathbf{Y}$

Note that $\mathbf{X}^{'}\mathbf{X} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$ and $\mathbf{X}^{'}\mathbf{Y} = \begin{bmatrix} \sum Y_i \\ \sum x_i Y_i \end{bmatrix}$

Furthermore,

$$\boxed{SSE = \mathbf{Y}^{'}\mathbf{Y} - \widehat{\beta}^{'}\mathbf{X}^{'}\mathbf{Y}}.$$

Fitted Values and Residuals

Fitted Values In matrix notation, the vector of the fitted values, denoted by $\widehat{\mathbf{Y}}$ is

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\beta},$$

where $\widehat{\beta} = (\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}\mathbf{Y}$.

Hat Matrix. The hat matrix is defined by

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}.$$

The vector of fitted values can be written as

$$\widehat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}.$$

Properties of the Hat matrix

- $\mathbf{H}$ is symmetric.
- Idempotency: $\mathbf{H}\ \mathbf{H} = \mathbf{H}$.

Proof.

Fitted Values and Residuals

Residuals The vector of the residuals is denoted by $\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$. We have

$$\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\widehat{\beta} = \mathbf{Y} - \mathbf{H}\mathbf{Y} = (\mathbf{I} - \mathbf{H})\mathbf{Y}.$$

Properties of the $\mathbf{I} - \mathbf{H}$

- $\mathbf{I} - \mathbf{H}$ is symmetric.
- Idempotency: $(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H}) = (\mathbf{I} - \mathbf{H})$.

Proof.

Variance-Covariance Matrix of Residuals:

$$Var(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{H})$$

and it is estimated by

$$MSE(\mathbf{I} - \mathbf{H})$$

since the common variance σ^2 is estimated by MSE.

Proof.

Analysis of Variance

$$\boxed{SST = \mathbf{Y}^{'}\mathbf{Y} - \frac{1}{n}\mathbf{Y}^{'}\mathbf{J}\mathbf{Y},}$$

where $\mathbf{J}$ is the matrix of 1s, $\mathbf{J} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$.

$$\boxed{\begin{aligned} SSE = \quad & \mathbf{e}^{'}\mathbf{e} = (\mathbf{Y} - \mathbf{X}\widehat{\beta})^{'}(\mathbf{Y} - \mathbf{X}\widehat{\beta}) \\ = \quad & \mathbf{Y}^{'}\mathbf{Y} - \widehat{\beta}^{'}\mathbf{X}^{'}\mathbf{Y} \end{aligned}}$$

$$\boxed{SSR = SST - SSE = \widehat{\beta}^{'}\mathbf{X}^{'}\mathbf{Y} - \frac{1}{n}\mathbf{Y}^{'}\mathbf{J}\mathbf{Y}.}$$

Proof.

License