Introduction to Multivariate Linear Regression Models

Linear regression models are used to formulate relationship between response variable (commonly denoted by Y, as in 'yield') and explanatory variable (X).

Linear models provide a clear relation between input variables and the output. They are simple, interpretable, and often surprisingly effective. Linear regression is among the first steps in data analysis. Linear models are integral components in more sophisticated statistical methods.

1 Review: Classical univariate linear regression model

Linear regression models formulates relationship between response Y and explanatory variable $X = (X_1, \cdots, X_r)$, assuming that $\mathbb{E}(Y|X)$, the conditional expectation of Y given the values of X, is linear in the X_i's. The model form is

$$\mathbb{E}(Y|X) = \beta_0 + X_1\beta_1 + \cdots + X_p\beta_r.$$

Linear models encompass variable transformations and expansions.

In this section we review classical linear regression models with univariate response variables, employing their vector-matrix formulation to generalize to the multivariate versions. Model parameters are typically estimated by the method of least squares. There are a set of standard inferences. Model checking and goodness of fit diagnostic methods are well developed.

The model

In classical linear regression model, the response variable is modeled as linear combination of r explanatory variables $z_1, \cdots, z_r$, plus a random variation.

$$Y = \beta_0 + \beta_1 z_1 + \cdots + \beta_r z_r + \varepsilon, \qquad \mathbb{E}(\varepsilon) = 0,\ V(\varepsilon) = \sigma^2.$$

The parameters β_i are to be estimated, along with the variance σ^2 of the error term ε. The formula implies conditional expectation.

$$\mathbb{E}(Y) = \mathbb{E}(Y \mid Z_1 = z_1, \cdots, Z_r = z_r) = \beta_0 + \beta_1 z_1 + \cdots + \beta_r z_r$$

If $\hat{\beta}_i$'s are estimates of β_i's, then the estimated Y is

$$\hat{Y} = \widehat{\mathbb{E}}(Y \mid z_1, \cdots, z_r) = \hat{\beta}_0 + \hat{\beta}_1 z_1 + \cdots + \hat{\beta}_r z_r$$

If sample data are collected, let Y_j denote the jth measurement of the response variable Y when the values of the r explanatory variables are set at $z_{j1}, \cdots, z_{jr}$. The regression model to be fitted by the data has the form

$$Y_j = \beta_0 + \beta_1 z_{j1} + \cdots + \beta_r z_{jr} + \varepsilon_j, \quad j = 1, \cdots, n.$$

Under the assumption that the n observations are independent, the error terms have the properties

$$\mathbb{E}(\varepsilon_j) = 0, \qquad V(\varepsilon_j) = \sigma^2, \qquad cov(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j.$$

The objective is to estimate the β_i's and σ^2 based on the observed data.

The regression model to be fitted with n observed data points can be written in matrix form as

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_j \\ \vdots \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1r} \\ \vdots & \vdots & & \vdots \\ 1 & z_{j1} & \cdots & z_{jr} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & z_{n1} & \cdots & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_j \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

In matrix notations, the above n model equations can be written as

$$\boldsymbol{Y}_{n\times 1} = \boldsymbol{Z}_{n\times(r+1)}\boldsymbol{\beta}_{(r+1)\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$$

or simply

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Matrix $\boldsymbol{Z}$ is the **design matrix** of the model. The independence, mean zero, and common variance assumptions on the error terms of the observations can be written in vector-matrix form as

$$\mathbb{E}(\boldsymbol{\varepsilon}) = 0, \qquad Cov(\boldsymbol{\varepsilon}) = \mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \boldsymbol{I}_n.$$

Least squares estimation of model parameters

When $n > r$, classical linear regression model parameters are commonly estimated by the method of least squares (LS). The approach is to minimize the sum of squares of errors (error = observed - estimated).

$$S(\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}) = \sum_{j=1}^{n}(Y_j - \beta_0 - \beta_1 z_{j1} - \cdots - \beta_r z_{jr})^2$$

when $\boldsymbol{Z}$ is of full rank, $rank(\boldsymbol{Z}) = r + 1$, the LS estimate of $\boldsymbol{\beta}$ that minimizes the sum of squares of errors $S(\boldsymbol{\beta})$ is

$$\hat{\boldsymbol{\beta}}_{LS} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$$

(Notes: When $rank(\boldsymbol{Z}) < r + 1$, $(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$ is replace by a generalized inverse of $\boldsymbol{Z}'\boldsymbol{Z}$.)

Proof. Denote $\boldsymbol{\beta}_* = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$.

$$\begin{aligned} S(\boldsymbol{\beta}) &= (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}) \\ &= [\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_* + \boldsymbol{Z}(\boldsymbol{\beta}_* - \boldsymbol{\beta})]'[\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_* + \boldsymbol{Z}(\boldsymbol{\beta}_* - \boldsymbol{\beta})] \\ &= (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_*)'(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_*) + (\boldsymbol{\beta}_* - \boldsymbol{\beta})'\boldsymbol{Z}'\boldsymbol{Z}(\boldsymbol{\beta}_* - \boldsymbol{\beta}) + 2(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_*)'\boldsymbol{Z}(\boldsymbol{\beta}_* - \boldsymbol{\beta}) \end{aligned}$$

The last term is a zero vector, because

$$\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_*) = \boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y} = \boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{Y} = \boldsymbol{0}_{(r+1)\times 1}$$

Thus

$$(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_*)'\boldsymbol{Z} = \boldsymbol{0}_{1\times(r+1)} \qquad \Rightarrow \qquad (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_*)'\boldsymbol{Z}(\boldsymbol{\beta}_* - \boldsymbol{\beta}) = 0$$

Therefore,

$$\begin{aligned} S(\boldsymbol{\beta}) &= (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_*)'(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_*) + (\boldsymbol{\beta}_* - \boldsymbol{\beta})'\boldsymbol{Z}'\boldsymbol{Z}(\boldsymbol{\beta}_* - \boldsymbol{\beta}) \\ &= \|\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_*\|^2 + \|\boldsymbol{Z}(\boldsymbol{\beta}_* - \boldsymbol{\beta})\|^2 \\ &\geq \|\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}_*\|^2 \end{aligned}$$

The equality holds if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}_* = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$. In other words,

$$S(\boldsymbol{\beta}^*) = \min_{\boldsymbol{\beta}} S(\boldsymbol{\beta})$$

which means $\boldsymbol{\beta}_* = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y} = \hat{\boldsymbol{\beta}}_{LS}$, the least squares estimator of β in the linear regression model. □

Remarks on Least Squares Estimator (LSE)

- The above least squares estimation results can also be stated using matrix calculus, by taking derivative of $s(\boldsymbol{\beta})$ with respect to vector β setting $S'(\boldsymbol{\beta}) = 0_{(r+1),1}$, and solve for $\boldsymbol{\beta}$.
- $\boldsymbol{H} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ is the **hat matrix**, a symmetric projection matrix satisfying $\boldsymbol{H}^2 = \boldsymbol{H}$ (idempotent).
- For observed data $\boldsymbol{y}$, $\hat{\boldsymbol{y}} = \boldsymbol{Z}\hat{\boldsymbol{\beta}} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} = \boldsymbol{H}\boldsymbol{y}$ are the model fitted values of $\boldsymbol{y}$.
- $\hat{\varepsilon} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$ are the residuals.
- Both $\hat{\boldsymbol{\beta}}$ and $\hat{\varepsilon}$ can obtained from the design matrix $\boldsymbol{Z}$ and observed response variable $\boldsymbol{Y}$.
- Both $\hat{\boldsymbol{\beta}}$ and $\hat{\varepsilon}$ can be expressed as linear combinations of observed response variables $Y_1, \cdots, Y_n$.

Geometry of Least Squares

- Minimizing $S(\boldsymbol{\beta})$ is equivalent to finding minimal error norm in projecting $\boldsymbol{Y}$ to $C(\boldsymbol{Z})$,the column space of $\boldsymbol{Z}$. The optimal is achieved by orthogonal projection.
- $\hat{\boldsymbol{Y}} = \boldsymbol{H}\boldsymbol{Y}$ is the result of the orthogonal projection of $\boldsymbol{Y}$ to $C(\boldsymbol{Z})$.
- $\boldsymbol{Z}'\hat{\varepsilon} = 0$, the residual vector $\hat{\varepsilon} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$ is orthogonal to $C(\boldsymbol{Z})$, a result of the orthogonal projection, can be derived via $\boldsymbol{Z}'\hat{\varepsilon} = \boldsymbol{Z}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} = \boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y} = \boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{Y} = 0$.
- $\hat{\boldsymbol{Y}}'\hat{\varepsilon} = 0$, the residual vector $\hat{\varepsilon}$ is orthogonal to the projected, fitted value $\hat{\boldsymbol{Y}}$, another result of the orthogonal projection, can be obtained from the derivation $\hat{\boldsymbol{Y}}'\hat{\varepsilon} = (\boldsymbol{H}\boldsymbol{Y})'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} = \boldsymbol{Y}'(\boldsymbol{H} - \boldsymbol{H}^2)\boldsymbol{Y} = 0$, by $\boldsymbol{H}^2 = \boldsymbol{H}$.

Sum of squares decomposition

The total response sum of squares

$$\sum_{i=1}^{n} Y_i^2 = \boldsymbol{Y}'\boldsymbol{Y}$$

By $\hat{\mathbf{Y}}'\hat{\varepsilon} = 0$, we obtain the decomposition

$$\boldsymbol{Y}'\boldsymbol{Y} = (\hat{\boldsymbol{Y}} + \hat{\varepsilon})'(\hat{\boldsymbol{Y}} + \hat{\varepsilon}) = \hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}} + \hat{\varepsilon}'\hat{\varepsilon}$$

To consider the centralized sum of squares,

$$\boldsymbol{Y}'\boldsymbol{Y} - n\bar{Y}^2 = \hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}} - n\bar{Y}^2 + \hat{\varepsilon}'\hat{\varepsilon}.$$

Using $\bar{Y} = \bar{\hat{Y}}$, the above equation can be written in terms of sums of squares. which can be stated as

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 + \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

The decomposition can be described as

$$\sum(\text{total sum of squares about mean})^2 = \sum(\text{regression sum of squares})^2 + \sum(\text{residual sum of squares})^2$$

On simplified notations,

$$SS_{total} = SS_{reg} + SS_{err}$$

Coefficient of determination

The sum of squares decomposition provides a measure of the quality of the regression model, in terms of the closeness of the fitted model to the observed data.

The measure is called coefficient of determination, also commonly known as R-square.

$$R^2 = \frac{SS_{reg}}{SS_{total}} = 1 - \frac{SS_{err}}{SS_{total}}$$

The coefficient of determination measures the proportion of the total variation "explained" by the fitted regression model, or by the explanatory variables. The variation is in terms of sums of squares about the mean.

Checking normal assumptions

To have statistical inference on the model fitting and parameter estimation, normality assumption is needed.

To check if observations $Y_1, \cdot, Y_n$ are indeed from a normal distribution, Q-Q plot is a quick diagnostic tool.

Normal Q-Q plot is a graphical method to check the normality assumption of the ε_i's.

The method of normal Q-Q plot

Let $\varepsilon_{(1)} \le \cdots \le \varepsilon_{(n)}$ be the ordered values of observed $\varepsilon_1, \cdots, \varepsilon_n$, called order statistics of $\varepsilon_1, \cdots, \varepsilon_n$.

If ε_i's are from independent random variables $\sim N(\mu, \sigma^2)$, then there is a linear relationship

$$\varepsilon_{(i)} = \mu + \sigma Z_{(i)}$$

where

$$Z_{(1)} \le \cdots \le Z_{(n)}$$

denote the order statistics of n independent observations $Z_1, \cdots, Z_n$, with

$$Z_i \sim N(0,1), \qquad i = 1, \cdots, n.$$

The so called Q-Q plot (quantile-quantile plot) for the ε_i's is the plot of quantile points

$$\left(\Phi^{(-1)}\Big(\frac{i}{n+1}\Big),\ \varepsilon_{(i)}\right), \qquad i =, 1 \cdots, n,$$

where

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

is the cumulative distribution function of $N(0,1)$, and (proof is based on order statistics, as in Mathematical Statistics Methods II)

$$\Phi^{(-1)}\left(\frac{i}{n+1}\right) \approx \mathbb{E}[Z_{(i)}]$$

The plot should resemble a straight line, if the ε_i's are indeed i.i.d. observations from a normal distribution.

In linear regression case, the true residual ε_i's are not observed. If the normality assumption on the Y_i's and $varepsilon_i$'s are appropriate, the fitted residual $\hat{\varepsilon}_i$'s, though not exactly, but are close to i.i.d. $\sim N(0, \hat{\sigma}^2)$. Common normality checks such as the Normal Q-Q plot are applied to the fitted residuals $\hat{\varepsilon}_i$'s.

2 Multivariate Linear Regression (non-tensor version)

Recall the regression model with univariate response random variable Y,

$$Y = \beta_0 + \beta_1 z_1 + \cdots + z_r\beta_r + \varepsilon = [1\ z_1\ \cdots\ z_r]\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_r\end{bmatrix} + \varepsilon = \boldsymbol{z}'\boldsymbol{\beta} + \varepsilon, \quad \varepsilon \sim N(0,\sigma^2).$$

Assume, for $i = 1, \cdots, n$, the observed independent outcomes of Y are y_j conditioned on z variables to be set at $(z_{j1}, \cdots, z_{jr})$, so

$$y_j = \beta_0 + \beta_1 z_{j1} + \cdots + \beta_r z_{jr} + \varepsilon_j, \qquad j = 1, \cdots, n.$$

In element-wise vector-matrix form, the model to be fitted with the observed data can be written as

$$\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_j\\ \vdots\\ y_n\end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1r}\\ 1 & z_{21} & \cdots & z_{21p}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & z_{j1} & \cdots & z_{jp}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & z_{n1} & \cdots & z_{np}\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_r\end{bmatrix} + \begin{bmatrix}\varepsilon_1\\ \epsilon_2\\ \vdots\\ \epsilon_j\\ \vdots\\ \epsilon_n\end{bmatrix}$$

The abbreviated matrix notation for the to-be-fitted model with observed data is

$$\boldsymbol{Y}_{n\times 1} = \boldsymbol{Z}_{n\times(1+r)}\boldsymbol{\beta}_{(1+r)\times 1} + \boldsymbol{\varepsilon}_{n\times 1}, \qquad \boldsymbol{\varepsilon} \sim N(0, \sigma^2\boldsymbol{I}_n)$$

In a multivariate regression model, the response

$$\boldsymbol{Y} = \begin{bmatrix} Y_1\\ \vdots\\ Y_p\end{bmatrix}$$

is a p-dimensional random vector, the explanatory variables are of the same structure as univariate case — for each component of the response vector.

Multivariate population regression model

The multivariate regression model can also be written in the form

$$\boldsymbol{Y} = \boldsymbol{\beta}_0 + z_1\boldsymbol{\beta}_1 + \cdots + z_r\boldsymbol{\beta}_r + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N_p(\boldsymbol{0}_p, \Sigma_{p\times p}),$$

which represents a joint model of p-univariate equations:

$$\begin{bmatrix} Y_1\\ Y_2\\ \vdots\\ Y_p\end{bmatrix} = \begin{bmatrix}\beta_{01} + \beta_{11}z_1 + \cdots + z_r\beta_{r1} + \varepsilon_1\\ \beta_{02} + \beta_{12}z_1 + \cdots + z_r\beta_{r2} + \varepsilon_2\\ \vdots\\ \beta_{0p} + \beta_{1p}z_1 + \cdots + z_r\beta_{rp} + \varepsilon_p\end{bmatrix}$$

This form shows that the multivariate regression consists of p univariate regression models sharing the same predictors or explanatory variables $z_1, \cdots, z_r$.

Equivalently,

$$\boldsymbol{Y} = \begin{bmatrix}\beta_{01}\\ \beta_{02}\\ \vdots\\ \beta_{0p}\end{bmatrix} + z_1\begin{bmatrix}\beta_{11}\\ \beta_{12}\\ \vdots\\ \beta_{1p}\end{bmatrix} + \cdots + z_r\begin{bmatrix}\beta_{r1}\\ \beta_{r2}\\ \vdots\\ \beta_{rp}\end{bmatrix} + \begin{bmatrix}\varepsilon_1\\ \epsilon_2\\ \vdots\\ \epsilon_p\end{bmatrix}$$

Notice the parameter coefficients for each predictor z_i,

$$\boldsymbol{\beta}_i = \begin{bmatrix}\beta_{i1}\\ \beta_{i2}\\ \vdots\\ \beta_{ip}\end{bmatrix}, \qquad i = 0, 1, \cdots, r$$

are p-vectors.

Since the dependence among the Y_i's is of main interests, **residuals** $\varepsilon_i, i = 1, \cdots, p$**, usually are not independent.** Therefore the covariance matrix of ε_i's, Σ, is usually not a diagonal matrix.

The above is the population model.

Fitting multivariate regression model with data

When data are collected to fit a multivariate regression model, $j = 1, \cdots, n$ observations are obtained.

For the jth explanatory variable setting $\{z_{j1}, \cdots, z_{jr}\}$, the observed jth outcome is a p-variate vector $[y_{j1}\ \cdots\ y_{jp}]'$, a realization of a common random vector $\boldsymbol{Y} = [Y_1\ \cdots\ Y_p]'$.

We may use the convention that each observation is represented in a row.

The multivariate regression model to be fitted with n observed data can be written as

$$\begin{bmatrix} y_{11} & \cdots & y_{1p}\\ y_{21} & \cdots & y_{2p}\\ \vdots & \cdots & \vdots\\ y_{j1} & \cdots & y_{jp}\\ \vdots & \cdots & \vdots\\ y_{n1} & \cdots & y_{np}\end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1r}\\ 1 & z_{21} & \cdots & z_{2r}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & z_{j1} & \cdots & z_{jr}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & z_{n1} & \cdots & z_{nr}\end{bmatrix}\begin{bmatrix}\beta_{01} & \cdots & \beta_{0p}\\ \beta_{11} & \cdots & \beta_{1p}\\ \vdots & \vdots & \vdots\\ \beta_{r1} & \cdots & \beta_{rp}\end{bmatrix} + \begin{bmatrix}\epsilon_{11} & \cdots & \epsilon_{1p}\\ \epsilon_{21} & \cdots & \epsilon_{2p}\\ \vdots & \cdots & \vdots\\ \epsilon_{j1} & \cdots & \epsilon_{jp}\\ \vdots & \cdots & \vdots\\ \epsilon_{n1} & \cdots & \epsilon_{np}\end{bmatrix} \tag{1}$$

The corresponding matrix notation for the multivariate regression model with n observations is

$$\boldsymbol{Y}_{n\times p} = \boldsymbol{Z}_{n\times(1+r)}\boldsymbol{\beta}_{(1+r)\times p} + \boldsymbol{\epsilon}_{n\times p}$$

The kth column on the left hand side response matrix corresponds to observations of the k-component of the response vector, that is, the observed outcomes of the $k$the random variable Y_k.

On the right hand side, only the kth column of the $\boldsymbol{\beta}$ matrix and kth column of the ε matrix, along with the $\boldsymbol{Z}$ matrix, are related to the kth response variable.

$$\boldsymbol{Y}_{(k)} = \begin{bmatrix} y_{1k}\\ y_{2k}\\ \vdots\\ y_{jk}\\ \vdots\\ y_{nk}\end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1r}\\ 1 & z_{21} & \cdots & z_{2r}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & z_{j1} & \cdots & z_{jr}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & z_{n1} & \cdots & z_{nr}\end{bmatrix}\begin{bmatrix}\beta_{0k}\\ \beta_{1k}\\ \vdots\\ \beta_{rk}\end{bmatrix} + \begin{bmatrix}\epsilon_{1k}\\ \epsilon_{2k}\\ \vdots\\ \epsilon_{jk}\\ \vdots\\ \epsilon_{nk}\end{bmatrix} = \boldsymbol{Z}\boldsymbol{\beta}_{(k)} + \boldsymbol{\varepsilon}_{(k)}$$

Therefore, by looking at the part of the multivariate model corresponding to a univariate component, we are back to the familiar univariate setting.

Under the assumption that the n observations are independent, the rows are independent. Therefore,

$$\mathbb{E}(\boldsymbol{\varepsilon}_{(k)}) = \mathbf{0}_n, \qquad Cov(\boldsymbol{\varepsilon}_{(k)}) = \sigma_{kk}\boldsymbol{I}_n$$

The model assumptions corresponding to each univariate component response variable is just as in the univariate regression model. The constant σ_{kk} is the kth diagonal element of Σ, the population covariance matrix of the population random p-vector $\boldsymbol{Y}$ and random p-variate error term $\boldsymbol{\varepsilon}$.

Therefore, using the observations $\boldsymbol{Y}_{(k)}$ of the kth component response variable, we obtain the least squares solution for the corresponding unknown parameter $\boldsymbol{\beta}_{(k)}$ is

$$\hat{\boldsymbol{\beta}}_{(k)} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}\boldsymbol{Y}_{(k)}$$

Analogously, the formulation for the ith variable is

$$\boldsymbol{Y}_{(i)} = \begin{bmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{ji} \\ \vdots \\ y_{ni} \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1r} \\ 1 & z_{21} & \cdots & z_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & z_{j1} & \cdots & z_{jr} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & z_{n1} & \cdots & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_{0i} \\ \beta_{1i} \\ \vdots \\ \beta_{ri} \end{bmatrix} + \begin{bmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \vdots \\ \epsilon_{ji} \\ \vdots \\ \epsilon_{ni} \end{bmatrix} = \boldsymbol{Z}\boldsymbol{\beta}_{(i)} + \boldsymbol{\varepsilon}_{(i)},$$

with

$$\mathbb{E}(\boldsymbol{\varepsilon}_{(i)}) = \mathbf{0}_n, \qquad Cov(\boldsymbol{\varepsilon}_{(i)}) = \sigma_{ii}\boldsymbol{I}_n.$$

We now combine the solution vectors for the linear regression for each component variable,

$$\hat{\boldsymbol{\beta}}_{(i)} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}\boldsymbol{Y}_{(i)}, \qquad i = 1, \cdots, p.$$

Recall matrix $\boldsymbol{\beta} = [\boldsymbol{\beta}_{(1)}, \cdots, \boldsymbol{\beta}_{(r)}]$, we have the least squares solutions for the multivariate regression model,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}\boldsymbol{Y}$$

Adding the dimensions,

$$\hat{\boldsymbol{\beta}}_{(r+1)\times p} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}\boldsymbol{Y}_{n\times p}$$

where $\boldsymbol{Z}$ is the same $n \times (r+1)$ design matrix.

Dependence in component response variables

A single observation, say, jth observation with $z_{j1}, \cdots, z_{jr}$ and $y_{j1}, \cdots, y_{jp}$, corresponds to a row in (1).

$$\begin{bmatrix} y_{j1} & \cdots & y_{jp} \end{bmatrix} = \begin{bmatrix} 1 & z_{j1} & \cdots & z_{jr} \end{bmatrix} \begin{bmatrix} \beta_{01} & \cdots & \beta_{0p} \\ \beta_{11} & \cdots & \beta_{1p} \\ \vdots & \vdots & \vdots \\ \beta_{r1} & \cdots & \beta_{rp} \end{bmatrix} + \begin{bmatrix} \epsilon_{j1} & \cdots & \epsilon_{jp} \end{bmatrix}$$

where $\begin{bmatrix} y_{j1} & \cdots & y_{jp} \end{bmatrix}$ an realization (independent in j) of a common random vector $\boldsymbol{Y} = [Y_1 \ \cdots \ Y_p]'$, with covariance matrix $\Sigma = [\sigma_{ij}]_{p\times p}$.

If the ith and kth component variables Y_i and Y_k are correlated, that is, if the ith and kth components of the same observation are correlated,

$$cov(\epsilon_{ji}, \epsilon_{jk}) = \sigma_{ik} \neq 0, \qquad j = 1, \cdots, n.$$

The covariance matrix between the residuals of two variables is

$$Cov\left(\boldsymbol{\varepsilon}_{(i)}, \boldsymbol{\varepsilon}_{(k)}\right) = \sigma_{ik}\boldsymbol{I}_n.$$

Multivariate variation decomposition*

In multivariate regression model, we still have the relation

$$\boldsymbol{Y}'\boldsymbol{Y} = \hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}} + \hat{\varepsilon}'\hat{\varepsilon}$$

However now the equation is on decomposition in terms of $p \times p$ matrices, they are sums of squares on the diagonals, cross products on the off-diagonal entries. Thus the relation can be stated as

(Total sum of squares and cross products)

$\|$

(Predicted sum of squares and cross product) $+$ (Residual sum of squares and product)

Regression techniques for multivariate

Many useful regression techniques can be generalized to multivariate setting. For example, the maximum likelihood ratio test is used in multiple places in this course.

Note: Relevant chapter in Johnson and Wichern: Chapter 7.