- **Univariate multiple linear regression model**

  - **Population-level model:**

    $$Y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon,$$

    where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)'$, and $\mathbf{x} = (1, X_1, \ldots, X_q)'$.

  - **Sample-level model:**

    Based on the sample $\{y_i, \mathbf{x}_i,\ i = 1, \ldots, n\}$, where $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{iq})'$, we have

    $$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

    where $\mathbf{y} = (y_1, \ldots, y_n)'$, $\mathbf{X}_{n\times(q+1)} = (\mathbf{x}_1', \ldots, \mathbf{x}_n')'$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$.

  - **Assumptions for statistical inference:**

    $$\varepsilon_i \text{ i.i.d. } N(0, \sigma^2),\ i = 1, \ldots, n.$$

    (Note that the normality is needed only when statistical inferences are of interest, i.e., tests, confidence intervals, etc. It's not a necessity for estimation stage.) Then show the following:

    1. y_i independently follows $N(\boldsymbol{\beta}'\mathbf{x}_i, \sigma^2)$, $i = 1, \ldots, n$.
       (This implies that y_i's are only independent, but not identically distributed, since they have distinct means due to different x-information. In addition, the randomness of Y purely comes from the error term, as we here follow the classical regression setup which assumes $\mathbf{x}$ is fixed. Or, we may work on the conditional mean, where the population-level model is written as a regression equation: $E(Y|\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$.)
    2. $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$.
       (I'd like to mark here, this $\mathbf{y}$ is the n-dimensional sample/observation vector of Y, not the p-dimensional random vector appearing in previous chapters.)
    3. $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_q)' \sim N_{q+1}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, where $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.
       (Then $\hat{\beta}_j \sim N(\beta_j, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}_{jj})$, which can be used in the inference for β_j.)

- **Multivariate multiple linear regression model**

  - **Population-level model:**

    $$\mathbf{y} = \mathbf{B}'\mathbf{x} + \boldsymbol{\varepsilon},$$

    where $\mathbf{x}$ is defined as the univariate case, but $\mathbf{y} = (Y_1, \ldots, Y_p)'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)'$. In addition, $\mathbf{B}_{(q+1)\times p} = (\boldsymbol{\beta}_{(1)}, \ldots, \boldsymbol{\beta}_{(p)})$, where the jth column $\boldsymbol{\beta}_{(j)} = (\beta_{0j}, \beta_{1j}, \ldots, \beta_{qj})'$. (Therefore, β_{lj} depicts the predictive effect of X_l to Y_j.)

– **Sample-level model:**

Based on the sample $\{\mathbf{y}_i, \mathbf{x}_i,\ i = 1, \ldots, n\}$, where $\mathbf{x}_i$ is defined as before, but $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})'$, we have

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\Xi}.$$

Here $\mathbf{X}$ is defined as the univariate case, but

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n_2} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{y}'_1 \\ \vdots \\ \mathbf{y}'_n \end{pmatrix} = (\mathbf{y}_{(1)}, \ldots, \mathbf{y}_{(p)}),$$

where

* $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})'$ denotes the ith row of $\mathbf{Y}$. It is the ith realization/observation of the random vector $\mathbf{y}$;
* $\mathbf{y}_{(j)} = (y_{1j}, \ldots, y_{nj})'$ denotes the jth column of $\mathbf{Y}$. It consists of n realizations of the jth random variable Y_j in $\mathbf{y}$.

$\boldsymbol{\Xi} = (\boldsymbol{\varepsilon}'_1, \ldots, \boldsymbol{\varepsilon}'_n)' = (\boldsymbol{\varepsilon}_{(1)}, \ldots, \boldsymbol{\varepsilon}_{(p)})$ is defined following the same fashion as $\mathbf{Y}$.

– **Assumptions for statistical inference:**

$$\boldsymbol{\varepsilon}_i \text{ i.i.d. } N_p(0, \boldsymbol{\Sigma}),\ i = 1, \ldots, n,$$

where $\boldsymbol{\Sigma}$ is the $p \times p$ covariance matrix with element σ_{jk}.

Then show the following:

1. $\mathbf{y}_i$ independently follows $N_p(\mathbf{B}'\mathbf{x}_i, \boldsymbol{\Sigma})$, $i = 1, \ldots, n$.
2. $\mathbf{y}_{(j)} \sim N_n(\mathbf{X}\boldsymbol{\beta}_{(j)}, \sigma_{jj}\mathbf{I}_{n\times n})$, $j = 1, \ldots, p$.
3. The covariance matrix between $\mathbf{y}_{(j)}$ and $\mathbf{y}_{(k)}$ is $\sigma_{jk}\mathbf{I}_{n\times n}$.
   (This means $\mathbf{y}_{(j)}$ and $\mathbf{y}_{(k)}$ are NOT independent!)
4. $\widehat{\boldsymbol{\beta}}_{(j)} = (\hat{\beta}_{0j}, \hat{\beta}_{1j}, \ldots, \hat{\beta}_{qj})' \sim N_{q+1}(\boldsymbol{\beta}_{(j)}, \sigma_{jj}(\mathbf{X}'\mathbf{X})^{-1})$, where $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.
5. The covariance matrix between $\widehat{\boldsymbol{\beta}}_{(j)}$ and $\widehat{\boldsymbol{\beta}}_{(k)}$ is $\sigma_{jk}(\mathbf{X}'\mathbf{X})^{-1}$.
   (That is the advantage of multivariate regression over separately univariate regression - the covariance structure between columns of $\widehat{\mathbf{B}}$ can be taken into account.)