Chapter 3

Multivariate Linear Model

3.1 Estimation and Prediction in Multivariate Linear Model

3.1.1 Estimation of Unknown Parameters

– In multivariate linear model, for each sampling unit i, there is an observable random vector

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{im} \end{pmatrix},$$

which is assumed to follow the linear model

$$\mathbf{y}_i = \mathbf{B}'\mathbf{x}_i + \boldsymbol{\varepsilon}_i$$
$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{im} \end{pmatrix} = \begin{pmatrix} \beta_{0_1} & \beta_{1_1} & \beta_{2_1} & \dots & \beta_{p_1} \\ \beta_{0_2} & \beta_{1_2} & \beta_{2_2} & \dots & \beta_{p_2} \\ & \vdots & & & \\ \beta_{0_m} & \beta_{1_m} & \beta_{2_m} & \dots & \beta_{p_m} \end{pmatrix} \begin{pmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{im} \end{pmatrix}. \tag{3.1}$$

– In linear model, the matrix $\mathbf{B}$ is the matrix of unknown parameters, a vector $\mathbf{x}_i$ contains the values of the explanatory variables, and the random error vector $\boldsymbol{\varepsilon}_i$ is assumed to follow normal distribution $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

– Hence

$$\mathbf{y}_i \sim N(\mathbf{B}'\mathbf{x}_i, \mathbf{\Sigma}). \tag{3.2}$$

– Let $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ be independent random vectors with all following the normal distribution $\mathbf{y}_i \sim N(\mathbf{B}'\mathbf{x}_i, \mathbf{\Sigma})$.

– The multivariate linear model can be written in a form

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \tag{3.3}$$

where

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1' \\ \mathbf{y}_2' \\ \vdots \\ \mathbf{y}_n' \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix}, \qquad \mathbf{E} = \begin{pmatrix} \boldsymbol{\varepsilon}_1' \\ \boldsymbol{\varepsilon}_2' \\ \vdots \\ \boldsymbol{\varepsilon}_n' \end{pmatrix}.$$

– Estimation of B can be based on the sum of generalized least squares

$$\begin{aligned} \sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{B}'\mathbf{x}_i)'\mathbf{\Sigma}^{-1}(\mathbf{y}_i - \mathbf{B}'\mathbf{x}_i) &= \mathrm{trace}[(\mathbf{Y} - \mathbf{X}\mathbf{B})\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{B})'] \\ &= \mathrm{trace}[\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{B})'(\mathbf{Y} - \mathbf{X}\mathbf{B})] \end{aligned} \tag{3.4}$$

– It can be shown that

$$\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{3.5}$$

is the generalized least squares estimator for the $\mathbf{B}$.

– This implies that

$$\hat{\boldsymbol{\mu}}_i = \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i, \qquad \mathbf{y}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), \tag{3.6a}$$

$$\widehat{\mathbf{XB}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \tag{3.6b}$$

$$\hat{\boldsymbol{\varepsilon}}_i = \mathbf{y}_i - \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i, \tag{3.6c}$$

$$\widehat{\mathbf{E}} = \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}. \tag{3.6d}$$

– Most often used an unbiased estimator for the covariance matrix $\boldsymbol{\Sigma}$ is

$$\widehat{\boldsymbol{\Sigma}} = \frac{\widehat{\mathbf{E}}'\widehat{\mathbf{E}}}{n - \operatorname{rank}(\mathbf{X})} = \frac{\mathbf{Y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}}{n - (p+1)}. \tag{3.7}$$

– Multivariate linear model can written in general linear model form by vectorizing multivariate model

$$\operatorname{vec}(\mathbf{Y}) \sim N\left[(\mathbf{I} \otimes \mathbf{X}) \operatorname{vec}(\mathbf{B}), (\boldsymbol{\Sigma} \otimes \mathbf{I})\right] \tag{3.8}$$

– Based on linear model theory, it's easy to see that $\widehat{\mathbf{B}}$ is the maximum likelihood estimator too.

– The covariance matrix for the $\hat{\boldsymbol{\mu}}_i$ is

$$\begin{aligned}\mathrm{Cov}(\hat{\boldsymbol{\mu}}_i) &= \mathrm{Cov}(\widehat{\mathbf{B}}'\mathbf{x}_i) = \mathrm{Cov}(\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i) \\ &= \mathrm{Cov}(\mathrm{vec}(\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i)) = \mathrm{Cov}((\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{I})\,\mathrm{vec}(\mathbf{Y}')) \\ &= (\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{I})\,\mathrm{Cov}(\mathrm{vec}(\mathbf{Y}'))(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \otimes \mathbf{I}) \\ &= (\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \mathbf{I})(\mathbf{I} \otimes \boldsymbol{\Sigma})(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \otimes \mathbf{I}) \\ &= (\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \otimes \boldsymbol{\Sigma})(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \otimes \mathbf{I}) \\ &= (\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \otimes \boldsymbol{\Sigma}) = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\boldsymbol{\Sigma}. \end{aligned} \tag{3.9}$$

3.1.2 Prediction of New Observations

– Consider predicting the value of new random vector $\mathbf{y}_f$, and let us assume that $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ and $\mathbf{y}_f$ are independent from each others.

– The best empirical predictor for the new random vector $\mathbf{y}_f$ is

$$\widehat{\mathrm{BP}}(\mathbf{y}_f|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n) = \widehat{\mathrm{BP}}(\mathbf{y}_f|\mathbf{Y}) = \widehat{\mathbf{B}}'\mathbf{x}_f = \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f. \tag{3.10}$$

– The covariance matrix of the prediction error $\mathbf{y}_f - \widehat{\mathrm{BP}}(\mathbf{y}_f|\mathbf{Y})$ is

$$\begin{aligned}\mathrm{Cov}(\mathbf{y}_f - \widehat{\mathrm{BP}}(\mathbf{y}_f|\mathbf{Y})) &= \mathrm{Cov}(\mathbf{y}_f) + \mathrm{Cov}(\widehat{\mathrm{BP}}(\mathbf{y}_f|\mathbf{Y})) \qquad (3.11) \\ &= \boldsymbol{\Sigma} + \mathbf{x}_f'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(1 + \mathbf{x}_f'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f),\end{aligned}$$

and hence

$$\mathbf{y}_f - \widehat{\mathrm{BP}}(\mathbf{y}_f|\mathbf{Y}) \sim N(\mathbf{0}, \mathbf{\Sigma}(1 + \mathbf{x}_f'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f)). \tag{3.12}$$

– Simultaneous $100(1-\alpha)\%$ prediction intervals for each element z_{fk} of the random vector $\mathbf{y}_f$ are

$$\left[\widehat{\mathrm{BP}}(\mathbf{y}_f|\mathbf{Y}) \pm z_{\alpha/2}\,\mathbf{diag}(\widehat{\mathbf{\Sigma}})^{1/2}\sqrt{(1 + \mathbf{x}_f'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_f)}\right] \tag{3.13}$$

– Consider the prediction in a partitioned random vector $\mathbf{y}_f$ situation

$$\mathbf{y}_f = \begin{pmatrix} \mathbf{y}_{f_1} \\ \mathbf{y}_{f_2} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1' \\ \mathbf{B}_2' \end{pmatrix} \mathbf{x}_f + \begin{pmatrix} \boldsymbol{\varepsilon}_{f_1} \\ \boldsymbol{\varepsilon}_{f_2} \end{pmatrix}. \tag{3.14}$$

– Let the problem be the prediction of $\mathbf{y}_{f_2}$ given $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ and $\mathbf{y}_{f_1}$.

– Under normality

$$\begin{pmatrix} \mathbf{y}_{f_1} \\ \mathbf{y}_{f_2 2} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{B}_1'\mathbf{x}_f \\ \mathbf{B}_2'\mathbf{x}_f \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}\right), \tag{3.15}$$

the best predictor for the $\mathbf{y}_{f_2}$ is

$$\mathrm{BP}(\mathbf{y}_{f_2}|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n, \mathbf{y}_{f_1}) = \mathrm{BP}(\mathbf{y}_{f_2}|\mathbf{Y}, \mathbf{y}_{f_1}) = \mathbf{B}_2'\mathbf{x}_f + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{y}_{f_1} - \mathbf{B}_1'\mathbf{x}_f) \tag{3.16}$$

with the covariance matrix of the prediction error being

$$\mathrm{Cov}(\mathbf{y}_{f_2} - \mathrm{BP}(\mathbf{y}_{f_2}|\mathbf{Y}, \mathbf{y}_{f_1})) = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}. \tag{3.17}$$

– Thus the best empirical predictor is

$$\widehat{\mathrm{BP}}(\mathbf{y}_{f_2}|\mathbf{Y},\mathbf{y}_{f_1}) = \widehat{\mathbf{B}}_2'\mathbf{x}_f + \widehat{\boldsymbol{\Sigma}}_{21}\widehat{\boldsymbol{\Sigma}}_{11}^{-1}(\mathbf{y}_{f_1} - \widehat{\mathbf{B}}_1'\mathbf{x}_f) \tag{3.18}$$

with the (asymptotic) estimated covariance matrix of the prediction error being

$$\widehat{\mathrm{Cov}}(\mathbf{y}_{f_2} - \widehat{\mathrm{BP}}(\mathbf{y}_{f_2}|\mathbf{Y},\mathbf{y}_{f_1})) = \widehat{\boldsymbol{\Sigma}}_{22} - \widehat{\boldsymbol{\Sigma}}_{21}\widehat{\boldsymbol{\Sigma}}_{11}^{-1}\widehat{\boldsymbol{\Sigma}}_{12}. \tag{3.19}$$

– Simultaneous $100(1-\alpha)\%$ prediction intervals for each element of the random vector $\mathbf{y}_{f_2}$ are

$$\left[\widehat{\mathrm{BP}}(\mathbf{y}_{f_2}|\mathbf{Y},\mathbf{y}_{f_1}) \pm z_{\alpha/2}\,\mathbf{diag}(\widehat{\boldsymbol{\Sigma}}_{22} - \widehat{\boldsymbol{\Sigma}}_{21}\widehat{\boldsymbol{\Sigma}}_{11}^{-1}\widehat{\boldsymbol{\Sigma}}_{12})^{1/2}\right] \tag{3.20}$$

Example 3.1.

Consider the following dataset:

```
> library(MASS)
> data(Cars93)
> data<-Cars93[,c(4:6,13,25)]
> head(data)
  Min.Price Price Max.Price Horsepower Weight
1      12.9  15.9      18.8        140   2705
2      29.2  33.9      38.7        200   3560
3      25.9  29.1      32.3        172   3375
4      30.8  37.7      44.6        172   3405
5      23.7  30.0      36.2        208   3640
6      14.2  15.7      17.3        110   2880
```

Model the variables $y_1 = \mathsf{Min.Price}$, $y_2 = \mathsf{Price}$ and $y_3 = \mathsf{Max.Price}$ by the multivariate linear model

$$\begin{aligned} y_{i1} &= \beta_{0_1} + \beta_{1_1} x_{i1} + \beta_{2_1} x_{i2} + \varepsilon_{i_1} \\ y_{i2} &= \beta_{0_2} + \beta_{1_2} x_{i1} + \beta_{2_2} x_{i2} + \varepsilon_{i_2} \\ y_{i3} &= \beta_{0_3} + \beta_{1_3} x_{i1} + \beta_{2_3} x_{i2} + \varepsilon_{i_3}, \end{aligned}$$

where $x_1 = \mathsf{Horsepower}$, $x_2 = \mathsf{Weight}$.

(a) Let us write the model for whole data as $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$, where, for each row of $\mathbf{E}$, it is assumed $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Calculate the maximum likelihood estimate $\widehat{\mathbf{B}}$ for the parameters $\mathbf{B}$.

(b) Find the unbiased estimate $\widehat{\boldsymbol{\Sigma}}$ for the covariance matrix $\boldsymbol{\Sigma}$.

(c) Predict the value of the new observation $\mathbf{y}_f$, when $x_1 = \mathsf{Horsepower} = 200$ and $x_2 = \mathsf{Weight} = 3500$. Create also 80 % simultaneous prediction intervals for elements of the random vector $\mathbf{y}_f$.

(d) Predict the value of the random variable $y_3 = \mathsf{Max.Price}$ when it is known that $x_1 = \mathsf{Horsepower} = 200$ and $x_2 = \mathsf{Weight} = 3500$, and further $y_1 = \mathsf{Min.Price} = 25.00$ and $y_2 = \mathsf{Price} = 29.50$. Create also 80 % simultaneous (asymptotic) prediction intervals for $y_3 = \mathsf{Max.Price}$.

3.2 Testing in Multivariate Linear Model

3.2.1 General Linear Hypotheses

– Consider two competing multivariate linear models

$$\begin{aligned}\mathcal{M}_1 &: \ \mathbf{Y} = \mathbf{X}_1\mathbf{B}_1 + \mathbf{E},\\ \mathcal{M}_2 &: \ \mathbf{Y} = \mathbf{X}_1\mathbf{B}_1 + \mathbf{X}_2\mathbf{B}_2 + \mathbf{E}.\end{aligned}$$

– Let us test the hypotheses

$$\begin{aligned}H_0 &: \ \mathcal{M}_1 \text{ is the true model},\\ H_1 &: \ \mathcal{M}_2 \text{ is the true model}.\end{aligned}$$

– Let

$$\widehat{\mathbf{E}}_{\mathcal{M}_1} = \mathbf{Y} - \mathbf{X}_1\widehat{\mathbf{B}}_1, \tag{3.21}$$

$$\widehat{\mathbf{E}}_{\mathcal{M}_2} = \mathbf{Y} - \mathbf{X}_1\widehat{\mathbf{B}}_1 - \mathbf{X}_2\widehat{\mathbf{B}}_2, \tag{3.22}$$

and furthermore

$$\mathbf{S}_{\mathcal{M}_1} = \widehat{\mathbf{E}}'_{\mathcal{M}_1}\widehat{\mathbf{E}}_{\mathcal{M}_1}, \tag{3.23}$$

$$\mathbf{S}_{\mathcal{M}_2} = \widehat{\mathbf{E}}'_{\mathcal{M}_2}\widehat{\mathbf{E}}_{\mathcal{M}_2}. \tag{3.24}$$

– Then testing can be based on either of these statistic

(a) Wilks Λ: $\frac{|\mathbf{S}_{\mathcal{M}_2}|}{|\mathbf{S}_{\mathcal{M}_1}|}$,

(b) Hotelling–Lawley trace: $\mathrm{trace}(\mathbf{S}_{\mathcal{M}_2}^{-1}\mathbf{S}_{\mathcal{M}_1} - \mathbf{I})$,

(c) Pillai's trace: $\mathrm{trace}(\mathbf{I} - (\mathbf{S}_{\mathcal{M}_2}\mathbf{S}_{\mathcal{M}_1}^{-1})^{-1})$.

– The transformation for the Pillai's trace V

$$F = \frac{n-r-q+s}{|q-d|+s} \cdot \frac{V}{s-V} \tag{3.25}$$

is following approximately F distribution with $df_1 = s(|q-d|+s)$ and $df_2 = s(n-r-q+s)$ degrees of freedom, where

$$\begin{aligned} n &= \text{the sample size}, \\ r &= \mathrm{rank}(\mathbf{X}) \\ d &= \mathrm{rank}(\mathbf{X}_2) \\ q &= \text{amount of columns in } \mathbf{B}, \\ s &= \min(d, q). \end{aligned}$$

– A general linear hypotheses

$$\begin{aligned} H_0 &: \ \mathbf{KB} = \mathbf{A}, \\ H_1 &: \ \mathbf{KB} \neq \mathbf{A}, \end{aligned}$$

means that under H_0 hypothesis $\mathbf{B}$ can have a form

$$\mathbf{B} = \mathbf{K}^{+}\mathbf{A} + (\mathbf{I} - \mathbf{K}^{+}\mathbf{K})\mathbf{C}, \tag{3.26}$$

where $\mathbf{C}$ is free to wary.

– Hence the general linear hypotheses are equivalent to the hypotheses

$$\begin{aligned} H_0 &: \mathbf{Y} - \mathbf{X}\mathbf{K}^{+}\mathbf{A} = \mathbf{X}(\mathbf{I} - \mathbf{K}^{+}\mathbf{K})\mathbf{C} + \mathbf{E}, \\ H_1 &: \mathbf{Y} = \mathbf{XB} + \mathbf{E}. \end{aligned}$$

– Testing can then be based on matrices $\widehat{\mathbf{E}}_{H_0}$ and $\widehat{\mathbf{E}}_{H_1}$.

3.2.2 Testing Additional Information

– Consider the partitioned random vector $\mathbf{y}_i$

$$\mathbf{y}_i = \begin{pmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1' \\ \mathbf{B}_2' \end{pmatrix} \mathbf{x}_i + \begin{pmatrix} \boldsymbol{\varepsilon}_{i1} \\ \boldsymbol{\varepsilon}_{i2} \end{pmatrix}. \tag{3.27}$$

– Then under normality

$$\mathbf{y}_{i2}|\mathbf{y}_{i1} \sim N\left[\mathbf{B}_2'\mathbf{x}_i + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_{i1} - \mathbf{B}_1'\mathbf{x}_i), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right]. \tag{3.28}$$

– Hypotheses testing

$$\begin{aligned} H_0 &: \mathbf{y}_{i2}|\mathbf{y}_{i1} \sim N\left[\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}_{i1}, \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right], \\ H_1 &: \mathbf{y}_{i2}|\mathbf{y}_{i1} \sim N\left[\mathbf{B}_2'\mathbf{x}_i + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_{i1} - \mathbf{B}_1'\mathbf{x}_i), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right], \end{aligned}$$

is interpreted as testing whether $\mathbf{y}_{i2}$ contains any additional information about $\mathbf{B}$ beyond that available in $\mathbf{y}_{i1}$.

– Conditional distributions $\mathbf{y}_{i2}|\mathbf{y}_{i1}$ can be written as a model

$$\begin{aligned} \mathbf{Y}_2 &= \mathbf{X}\mathbf{B}_2 + (\mathbf{Y}_1 - \mathbf{X}\mathbf{B}_1)\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} + \mathbf{E}_{2\cdot 1} \\ &= \mathbf{X}(\mathbf{B}_2 - \mathbf{B}_1\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}) + \mathbf{Y}_1\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} + \mathbf{E}_{2\cdot 1} \\ &= \mathbf{X}\mathbf{C}_1 + \mathbf{Y}_1\mathbf{C}_2 + \mathbf{E}_{2\cdot 1}, \end{aligned} \tag{3.29}$$

where $\mathbf{C}_1 = \mathbf{B}_2 - \mathbf{B}_1\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ and $\mathbf{C}_2 = \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

– Hence testing the additional information is same as testing the hypotheses

$$H_0 : \mathbf{Y}_2 = \mathbf{Y}_1\mathbf{C}_2 + \mathbf{E}_{2\cdot 1},$$
$$H_1 : \mathbf{Y}_2 = \mathbf{X}\mathbf{C}_1 + \mathbf{Y}_1\mathbf{C}_2 + \mathbf{E}_{2\cdot 1}.$$

Example 3.2.

Consider the following dataset:

```
> library(MASS)
> data(Cars93)
> data<-Cars93[,c(4:6,13,25)]
> head(data)
  Min.Price Price Max.Price Horsepower Weight
1      12.9  15.9      18.8        140   2705
2      29.2  33.9      38.7        200   3560
3      25.9  29.1      32.3        172   3375
4      30.8  37.7      44.6        172   3405
5      23.7  30.0      36.2        208   3640
```

Model the variables $Y_1 = \mathsf{Min.Price}$, $Y_2 = \mathsf{Price}$ and $Y_3 = \mathsf{Max.Price}$ by the multivariate linear model

$$y_{i1} = \beta_{0_1} + \beta_{1_1}x_{i1} + \beta_{2_1}x_{i2} + \varepsilon_{i_1}$$
$$y_{i2} = \beta_{0_2} + \beta_{1_2}x_{i1} + \beta_{2_2}x_{i2} + \varepsilon_{i_2}$$
$$y_{i3} = \beta_{0_3} + \beta_{1_3}x_{i1} + \beta_{2_3}x_{i2} + \varepsilon_{i_3},$$

where $x_1 = \mathsf{Horsepower}$, $x_2 = \mathsf{Weight}$.

(a) Test at 5% significance level, is the explanatory variable $X_2 = \mathsf{Weight}$ statistically significant variable in the model.

(b) Test the hypotheses

$$H_0 : \beta_{2_1} = 0 \text{ and } \beta_{2_2} = 0,$$
$$H_1 : \beta_{2_1} \neq 0 \text{ or } \beta_{2_2} \neq 0.$$

(c) Test does the variables $Y_2 = \mathsf{Price}$ and $Y_3 = \mathsf{Max.Price}$ contains any additional information about the parameters $\mathbf{B}$ beyond that is available in variables Y_1.

3.2.3 Repeated Measures Model for Two Population

– If the elements of the random vector

$$\mathbf{y}_i = \begin{pmatrix} y_{it_1} \\ y_{ir_2} \\ \vdots \\ y_{it_m} \end{pmatrix}$$

are random variables measuring the outcome of the same random phenomenon in time points $t_1, t_2, \ldots, t_m$ for each sampling unit i, then the model

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

can be called as a repeated measures model.

– A repeated measures model can be written as

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \tag{3.30}$$

but the hypotheses are often a form

$$H_0: \ \mathbf{KBL} = \mathbf{0},$$
$$H_1: \ \mathbf{KBL} \neq \mathbf{0}.$$

– Above hypotheses can be tested by first transforming the original model to the model

$$\begin{aligned} \mathbf{YL} &= \mathbf{XBL} + \mathbf{EL} \\ \mathbf{Y}_* &= \mathbf{XB}_* + \mathbf{E}_*, \end{aligned} \tag{3.31}$$

and then testing the hypotheses

$$\begin{aligned} H_0 &: \ \mathbf{KB}_* = \mathbf{0}, \\ H_1 &: \ \mathbf{KB}_* \neq \mathbf{0}. \end{aligned}$$

– For example, two population $(j = 1, 2)$ repeated measures model

$$\begin{aligned} \mu_{jt_1} &= \beta_{0_1} + \beta j_1 \\ \mu_{jt_2} &= \beta_{0_2} + \beta j_2 \\ &\vdots \\ \mu_{jt_m} &= \beta_{0_m} + \beta j_m \end{aligned}$$

can be written as $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$, where (since $j = 1$ is baseline category)

$$\mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} \beta_{0_1} & \beta_{0_2} & \ldots \beta_{0_m} \\ \beta_{2_1} & \beta_{2_2} & \ldots \beta_{2_m} \end{pmatrix}.$$

– For example, average equality hypotheses

$$
\begin{aligned}
H_0 &: \frac{\mu_{1t_1} + \mu_{1t_2} + \cdots + \mu_{1t_m}}{m} = \frac{\mu_{2t_1} + \mu_{2t_2} + \cdots + \mu_{2t_m}}{m}, \\
H_1 &: \frac{\mu_{1t_1} + \mu_{1t_2} + \cdots + \mu_{1t_m}}{m} \neq \frac{\mu_{2t_1} + \mu_{2t_2} + \cdots + \mu_{2t_m}}{m},
\end{aligned}
$$

is tested by the matrices

$$
\mathbf{K} = \begin{pmatrix} 0 & 1 \end{pmatrix}, \qquad \mathbf{L} = \frac{1}{m} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.
$$

– And horizontal sub-population hypotheses

$$
\begin{aligned}
H_0 &: \mu_{2t_1} = \mu_{2t_2} = \cdots = \mu_{2t_m}, \\
H_1 &: \mu_{2t_1} \neq \mu_{2t_2} \neq \cdots \neq \mu_{2t_m},
\end{aligned}
$$

can be tested by the matrices

$$
\mathbf{K} = \begin{pmatrix} 1 & 1 \end{pmatrix}, \qquad \mathbf{L}' = \begin{pmatrix} 1 & -1 & 0 & 0 \ldots & 0 & 0 \\ 0 & 1 & -1 & 0 \ldots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 \ldots & 1 & -1 \end{pmatrix}.
$$

– Furthermore, parallel hypotheses

$$H_0 : \mu_{1t_1} - \mu_{2t_1} = \mu_{1t_2} - \mu_{2t_2} = \cdots = \mu_{1t_m} - \mu_{2t_m},$$
$$H_1 : \mu_{1t_1} - \mu_{2t_1} \neq \mu_{1t_2} - \mu_{2t_2} \neq \cdots \neq \mu_{1t_m} - \mu_{2t_m},$$

can be tested by the matrices

$$\mathbf{K} = \begin{pmatrix} 0 & 1 \end{pmatrix}, \qquad \mathbf{L}' = \begin{pmatrix} 1 & -1 & 0 & 0 \dots & 0 & 0 \\ 0 & 1 & -1 & 0 \dots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 \dots & 1 & -1 \end{pmatrix}.$$

Example 3.3.

Consider the dataset knee2.txt:

	N	Th	Age	Sex	R1	R2	R3	R4
1	1	1	28	1	4	4	4	4
2	2	1	32	1	4	4	4	4
3	3	1	41	1	3	3	3	3
4	4	2	21	1	4	3	3	2
5	5	2	34	1	4	3	3	2
6	6	1	24	1	3	3	3	2
7	7	2	28	1	4	3	3	2
8	8	2	40	1	3	2	2	2

```
A data frame with 127 observations on the following 8 variables.
N- Patient's number
Th - Therapy ( placebo = 1, treatment = 2)
Age - Age in years
Sex -Gender (male = 0, female = 1)
R1 -Pain before treatment (no pain = 1, severe pain = 5)
R2 -Pain after three days of treatment
R3 -Pain after seven days of treatment
R4 -Pain after ten days of treatment
```

.

```
In a clinical study n=127 patients with sport related injuries have been treated with
two different therapies (chosen by random design).
After 3,7 and 10 days of treatment the pain occuring during knee movement was observed.
```

Model the data by the repeated measures model

$$\mu_{jt_1} = \beta_{0_1} + \beta j_1,$$
$$\mu_{jt_2} = \beta_{0_2} + \beta j_2,$$
$$\mu_{jt_3} = \beta_{0_3} + \beta j_3,$$
$$\mu_{jt_4} = \beta_{0_4} + \beta j_4,$$

where $X = \mathsf{Th}$.

(a) Test at 5% significance level, is the explanatory variable $X = \mathsf{Th}$ statistically significant variable in the model.

(b) Test the hypotheses

$$
\begin{aligned}
H_0 &: \frac{\mu_{1t_1} + \mu_{1t_2} + \mu_{1t_3} + \mu_{1t_4}}{4} = \frac{\mu_{2t_1} + \mu_{2t_2} + \mu_{2t_3} + \mu_{2t_m}}{4}, \\
H_1 &: \frac{\mu_{1t_1} + \mu_{1t_2} + \mu_{1t_3} + \mu_{1t_4}}{4} \neq \frac{\mu_{2t_1} + \mu_{2t_2} + \mu_{2t_3} + \mu_{2t_m}}{4}.
\end{aligned}
$$

(c) Test the hypotheses

$$
\begin{aligned}
H_0 &: \mu_{2t_1} = \mu_{2t_2} = \mu_{2t_3} = \mu_{2t_4}, \\
H_1 &: \mu_{2t_1} \neq \mu_{2t_2} \neq \mu_{2t_3} \neq \mu_{2t_4}.
\end{aligned}
$$

(d) Test the hypotheses

$$
\begin{aligned}
H_0 &: \mu_{1t_1} - \mu_{2t_1} = \mu_{1t_2} - \mu_{2t_2} = \mu_{1t_3} - \mu_{2t_3} = \mu_{1t_4} - \mu_{2t_4}, \\
H_1 &: \mu_{1t_1} - \mu_{2t_1} \neq \mu_{1t_2} - \mu_{2t_2} \neq \mu_{1t_3} - \mu_{2t_3} \neq \mu_{1t_4} - \mu_{2t_4}.
\end{aligned}
$$