What to do today ?

Part I. Introduction and Preparation

Part II. Inference under Multivariate Normal Distribution (Chp 4-7)

II.1 Multivariate Normal Distribution (Chp 4)

II.2 Inferences on Mean Vector (Chp 5)

II.3 Comparisons of Several Mean Vectors (Chp 6.1-4, 6-7)

II.4 Multivariate Linear Regression (Chp 7)

II.4.1 Classical Linear Regression (Chp7.1-3)
II.4.2 Linear Regression Based Inference (Chp7.4-5)
II.4.3 Model Checking (Chp7.6)
II.4.4 Multivariate Multiple Regression (Chp7.7)

To explore how r.v. Y depends on $X_1, \ldots, X_k$: assume

$$\begin{aligned} Y &= [\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k] + \epsilon \\ (\text{response} &= [\text{ linear function of } X_1, \ldots, X_k] + \text{error}) \end{aligned}$$

With n indpt obs on Y, where Y_i is associated with the values $x_{1i}, \ldots, x_{ki}$ of $X_1, \ldots, X_k$ for $i = 1, \ldots, n$:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 x_{11} + \ldots + \beta_k x_{1k} + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 x_{21} + \ldots + \beta_k x_{2k} + \epsilon_2 \\ \vdots & \quad \vdots \\ Y_n &= \beta_0 + \beta_1 x_{n1} + \ldots + \beta_k x_{nk} + \epsilon_n \end{aligned}$$

where (1) $\epsilon_1, \ldots, \epsilon_n$ are indpt, and (2) $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$.

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_k \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_n \end{pmatrix}$$

Classical Linear Regression Model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

with (1) $E(\boldsymbol{\epsilon}) = \mathbf{0}$, and (2) $Var(\boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{'}) = \sigma^2\mathbf{I}$.

II.4.1 Classical Linear Regression

Recalling the linear regression model, when there are n indpt obs on Y:
$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with (1) $E(\boldsymbol{\epsilon}) = \mathbf{0}$, and (2) $Var(\boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{'}) = \sigma^2\mathbf{I}$.

Least Squares Estimator.

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{'}\mathbf{X}\right)^{-1}\mathbf{X}^{'}\mathbf{Y}$$

provided $\mathbf{X}$ has full rank.

- the fitted values: $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{Y}$ with the "hat" matrix $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^{'}\mathbf{X}\right)^{-1}\mathbf{X}^{'}$.
- the residuals: $\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$, and $S(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\epsilon}}^{'}\hat{\boldsymbol{\epsilon}} = |\hat{\boldsymbol{\epsilon}}|^2$.
- $\mathbf{X}^{'}\hat{\boldsymbol{\epsilon}} = \mathbf{0}$ and $\hat{\mathbf{y}}^{'}\hat{\boldsymbol{\epsilon}} = 0$.

II.4.2A Linear Regression Based Inference

Consider $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.
Inferential procedures under the normality assumption on the error term $\boldsymbol{\epsilon}$: $\boldsymbol{\epsilon} \sim MN(\mathbf{0}, \sigma^2\mathbf{I})$

- The LSE $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \sim MN(\boldsymbol{\beta}, \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1})$.
  The same as the MLE.
- The MLE $\hat{\sigma}^2 = \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}/n$, and $n\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-[k+1]}$.
- A $100(1-\alpha)\%$ confidence region (ellipsoid) for $\boldsymbol{\beta}$ is

$$\left\{\boldsymbol{\beta} : (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq (k+1)s^2 F_{k+1,n-[k+1]}(\alpha)\right\}$$

- Simultaneous $100(1-\alpha)\%$ confidence intervals for β_j, $j = 0, 1, \ldots, k$, are

$$\hat{\beta}_j \pm \sqrt{\widehat{Var}(\hat{\beta}_j)}\sqrt{(k+1)F_{k+1,n-[k+1]}(\alpha)}$$

  Here $\widehat{Var}(\hat{\beta}_j)$ is the diagnoal element of $s^2(\mathbf{X}'\mathbf{X})^{-1}$ corresponding to $\hat{\beta}_j$.

II.4.2A Linear Regression Based Inference

Consider

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_{(1)} \\ \boldsymbol{\beta}_{(2)} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \mathbf{X}_1\boldsymbol{\beta}_{(1)} + \mathbf{X}_2\boldsymbol{\beta}_{(2)} + \boldsymbol{\epsilon}$$

To test $H_0 : \boldsymbol{\beta}_{(2)} = \mathbf{0}$ with $\boldsymbol{\beta}_{(2)}^{'} = (\beta_{q+1}, \ldots, \beta_k)$?

- Likelihood ratio test:

$$\frac{\max_{\boldsymbol{\beta}_{(1)}, \sigma^2} L(\boldsymbol{\beta}_{(1)}, \sigma^2)}{\max_{\boldsymbol{\beta}, \sigma^2} L(\boldsymbol{\beta}, \sigma^2)} = \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}^2}\right)^{-n/2} = \left(1 + \frac{\hat{\sigma}_1^2 - \hat{\sigma}^2}{\hat{\sigma}^2}\right)^{-n/2}$$

- It is equivalent to test H_0 : Model$_0$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_{(1)} + \boldsymbol{\epsilon}$ vs H_1 : Model$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

$$F = \frac{[SS_{res}(Model_0) - SS_{res}(Model)]/(k-q)}{SS_{res}(Model)/(n-[k+1])} \sim F_{k-q, n-[k+1]}$$

under H_0

II.4.2B Inference by the Fitted Linear Regression Model

Provided the LSE $\hat{\boldsymbol{\beta}}$ using data with size n for the model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ under the normality assumption $\boldsymbol{\epsilon} \sim MN(\mathbf{0}, \sigma^2\mathbf{I})$:
if Y_0 is the response when the indpendent variables are $x_{01}, \ldots, x_{0k}$,

- *to estimate* $E(Y_0|\mathbf{x}_0) = \beta_0 + \beta_1 x_{01} + \ldots + \beta_k x_{0k} = \mathbf{x}_0^{'}\boldsymbol{\beta}$
  - point estimate: $\mathbf{x}_0^{'}\hat{\boldsymbol{\beta}}$
  - $100(1-\alpha)\%$ confidence interval:

$$\mathbf{x}_0^{'}\hat{\boldsymbol{\beta}} \pm t_{n-[k+1]}(\alpha/2)\sqrt{[\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0]s^2}$$

- *to predict for* $Y_0 = \mathbf{x}_0^{'}\boldsymbol{\beta} + \epsilon_0$
  - point estimate: $\mathbf{x}_0^{'}\hat{\boldsymbol{\beta}}$
  - $100(1-\alpha)\%$ prediction interval:

$$\mathbf{x}_0^{'}\hat{\boldsymbol{\beta}} \pm t_{n-[k+1]}(\alpha/2)\sqrt{[\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 + 1]s^2}$$

II.4.3 Checking for Liner Regression Model (Chp7.6)

Is the model appropriate?

To examine the residuals $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \hat{\mathbf{y}} = [\mathbf{I} - \mathbf{H}]\mathbf{y}$: if the model is true, $\hat{\boldsymbol{\epsilon}} \sim MN(\mathbf{0}, \sigma^2[\mathbf{I} - \mathbf{H}])$

- plot $\hat{\epsilon}_i$ vs the predicted $\hat{y}_i$:
  - check for any dependence of $\hat{\epsilon}_i$ on $\hat{y}_i$
  - check for any pattern of nonconstant variance
- plot $\hat{\epsilon}_i$ vs x_{ig}, values of an independent variable:
  check for any systematic pattern
- Q-Q plot/histogram $\hat{\epsilon}_i$:
  check for any departure from normality assumption
- plot $\hat{\epsilon}_i$ vs index of the observations:
  check for any correlation

Some further issues.

- use the standardized residuals $\hat{\epsilon}_i^* = \hat{\epsilon}_i \big/ \sqrt{s^2(1-h_{ii})}$ for $i = 1, \ldots, n$.

- the leverage h_{ii}, the (i,i) element of the hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$:
  - a measure of how far $(y_i, \mathbf{x}_i)$ is from the other obs
  - $\hat{y}_i = h_{ii} y_i + \sum_{g \neq i} h_{ig} y_g$

- variable selection, criterion quantities
  - $R^2 = SS_{reg} \big/ SS_{total} = 1 - SS_{res} \big/ SS_{total}$
  - R^2-adjusted$= 1 - (1 - R^2)\frac{n-1}{n-[k+1]}$
  - Mallow's $C_p = SSE_p \big/ MSE_{full} - (n - 2p)$
  - Akaike's AIC: $-2 \ln L + 2K$
  - BIC: $-2 \ln L + K \ln n$

Example. (textbook p372) [Linear regression analysis with real estate data] $n = 20$

- Y = selling price (in \$1000)
- Z_1 = total dwelling size (100 square feet); Z_2 = assessed value (in \$1000)

Modeling checking by residual analysis?

II.4.4A Multivariate Multiple Regression: modeling

- To explore how r.v. $Y_1, \ldots, Y_m$ depends on $X_1, \ldots, X_k$? That is, to explore how r.v. $\mathbf{Y}$ depends on $X_1, \ldots, X_k$?

Recall in the univariate situations, assume

$$\begin{aligned} Y &= [\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k] + \epsilon \\ (\text{response} &= [\text{ linear function of } X_1, \ldots, X_k] + \text{error}) \end{aligned}$$

With n indpt obs on Y, where Y_i is associated with the values $x_{1i}, \ldots, x_{ki}$ of $X_1, \ldots, X_k$ for $i = 1, \ldots, n$:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 x_{11} + \ldots + \beta_k x_{1k} + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 x_{21} + \ldots + \beta_k x_{2k} + \epsilon_2 \\ \vdots \quad & \quad \vdots \\ Y_n &= \beta_0 + \beta_1 x_{n1} + \ldots + \beta_k x_{nk} + \epsilon_n \end{aligned}$$

where (1) $\epsilon_1, \ldots, \epsilon_n$ are indpt, and (2) $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$.

II.4.4A Multivariate Multiple Regression: modeling

- To explore how r.v. $Y_1, \ldots, Y_m$ depends on $X_1, \ldots, X_k$?
- That is, to explore how r.v. $\mathbf{Y}$ depends on $X_1, \ldots, X_k$?

Assume

$$\begin{array}{rcl} \mathbf{Y} & = & [\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 X_1 + \ldots + \boldsymbol{\beta}_k X_k] + \boldsymbol{\epsilon} \\ (\text{response} & = & [\text{ linear function of } X_1, \ldots, X_k] + \text{error}) \end{array}$$

With n indpt obs on $\mathbf{Y}$, where $\mathbf{Y}_i$ is associated with the values $x_{1i}, \ldots, x_{ki}$ of $X_1, \ldots, X_k$ for $i = 1, \ldots, n$:

$$\begin{array}{rcl} \mathbf{Y}_1 & = & \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{11} + \ldots + \boldsymbol{\beta}_k x_{1k} + \boldsymbol{\epsilon}_1 \\ \mathbf{Y}_2 & = & \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{21} + \ldots + \boldsymbol{\beta}_k x_{2k} + \boldsymbol{\epsilon}_2 \\ \vdots & & \vdots \\ \mathbf{Y}_n & = & \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{n1} + \ldots + \boldsymbol{\beta}_k x_{nk} + \boldsymbol{\epsilon}_n \end{array}$$

where (1) $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ are indpt, and (2) $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$, $Var(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$.

II.4.4A Multivariate Multiple Regression: modeling

$$\begin{pmatrix} \mathbf{Y}_1' \\ \mathbf{Y}_2' \\ \vdots \\ \mathbf{Y}_n' \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_0' \\ \boldsymbol{\beta}_1' \\ \vdots \\ \boldsymbol{\beta}_k' \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1' \\ \boldsymbol{\epsilon}_2' \\ \vdots \\ \boldsymbol{\epsilon}_n' \end{pmatrix}$$

Multivariate Linear Regression Model

$$\mathbf{Y}_{n\times m} = \mathbf{X}_{n\times(k+1)}\boldsymbol{B}_{(k+1)\times m} + \boldsymbol{E}_{n\times m}$$

with (1) $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$, and (2) $Cov(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_l) = \boldsymbol{\Sigma}$ if $i = l$; $\mathbf{0}_{m\times m}$ if $i \neq l$.

The jth component $\mathbf{Y}_{(j)}$ follows

$$\mathbf{Y}_{(j)} = \mathbf{X}_{n\times(k+1)}\boldsymbol{\beta}_{(j)} + \boldsymbol{\epsilon}_{(j)} \quad j = 1, \dots, m$$

with $Var(\boldsymbol{\epsilon}_{(j)}) = \sigma_{jj}\mathbf{I}$.

II.4.4B Multivariate Multiple Regression: estimation

Least Squares Estimator (LSE).

$$\hat{\beta}_{(j)} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}_{(j)}$$

provided $\mathbf{X}$ has full rank.

$$\widehat{\boldsymbol{B}} = \left(\hat{\beta}_{(1)}\hat{\beta}_{(2)}\dots\hat{\beta}_{(m)}\right) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{Y}_{(1)}\mathbf{Y}_{(2)}\dots\mathbf{Y}_{(m)}\right)$$

That is

$$\widehat{\boldsymbol{B}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

- the fitted values: $\hat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{B}} = \mathbf{H}\mathbf{Y}$ with the “hat” matrix $\mathbf{H} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$.
- the residuals: $\hat{\boldsymbol{E}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$, and $\mathbf{X}'\hat{\boldsymbol{E}} = \mathbf{0}$
- $\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\boldsymbol{E}}'\hat{\boldsymbol{E}}$

II.4.4B Multivariate Multiple Regression: estimation

- $\widehat{\boldsymbol{B}} = \left(\hat{\beta}_{(1)} \dots \hat{\beta}_{(m)}\right)$ is unbiased: $E(\widehat{\boldsymbol{B}}) = \boldsymbol{B}$
- $Cov(\hat{\beta}_{(j)}, \hat{\beta}_{(l)}) = \sigma_{jl}\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ for $j, l = 1, \dots, k$.
- $E(\hat{\boldsymbol{E}}) = \mathbf{0}$ and $E(\hat{\boldsymbol{E}}^{'}\hat{\boldsymbol{E}}) = (n - [k+1])\boldsymbol{\Sigma}$

$$\Longrightarrow \qquad \widehat{\boldsymbol{\Sigma}} = \frac{\hat{\boldsymbol{E}}^{'}\hat{\boldsymbol{E}}}{n - [k+1]}.$$

- $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{E}}$ are uncorrelated.

II.4.4C Multivariate Multiple Regression: inference

Consider $\mathbf{Y} = \mathbf{X}\boldsymbol{B} + \boldsymbol{E}$.
Inferential procedures under the normality assumption on the error terms ϵ_i, $i = 1, \ldots, n$: $\epsilon_i \sim MN(\mathbf{0}, \mathbf{\Sigma})$ iid

- The LSE $\hat{\beta}_{(j)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{(j)} \sim MN(\beta_{(j)}, \sigma_{jj}(\mathbf{X}'\mathbf{X})^{-1})$.

  The same as the MLE.

- The MLE $\hat{\mathbf{\Sigma}} = \hat{\boldsymbol{E}}'\hat{\boldsymbol{E}}/n$

II.4.4C Multivariate Multiple Regression: inference

Consider

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{(1)} \\ \boldsymbol{B}_{(2)} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{B} + \boldsymbol{E} = \mathbf{X}_1\boldsymbol{B}_{(1)} + \mathbf{X}_2\boldsymbol{B}_{(2)} + \boldsymbol{E}$$

To test $H_0 : \boldsymbol{B}_{(2)} = \mathbf{0}$?

- Likelihood ratio test:

$$\Lambda = \frac{\max_{\boldsymbol{\beta}_{(1)}, \boldsymbol{\Sigma}} L(\boldsymbol{\beta}_{(1)}, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\beta}, \boldsymbol{\Sigma}} L(\boldsymbol{\beta}, \boldsymbol{\Sigma})} = \left(\frac{|\hat{\boldsymbol{\Sigma}}_1|}{|\hat{\boldsymbol{\Sigma}}|} \right)^{-n/2}$$

  Wilks' lambda statistic: $\Lambda^{2/n} = \frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_1|}$

- It is equivalent to test H_0 : Model$_0$ $\mathbf{Y} = \mathbf{X}_1\boldsymbol{B}_{(1)} + \boldsymbol{E}$ vs H_1 : Model $\mathbf{Y} = \mathbf{X}\boldsymbol{B} + \boldsymbol{E}$ for model selection.

II.4.4C Multivariate Multiple Regression: inference

Provided the LSE $\hat{\boldsymbol{B}}$ using data with size n for the model $\mathbf{Y} = \mathbf{X}\boldsymbol{B} + \boldsymbol{E}$ under the normality assumption $\epsilon_i \sim MN(\mathbf{0}, \mathbf{\Sigma})$ iid:
if $\mathbf{Y}_0$ is the response when the indpendent variables are $x_{01}, \ldots, x_{0k}$,

- *to estimate* $E(\mathbf{Y}_0|\mathbf{x}_0) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{01} + \ldots + \boldsymbol{\beta}_k x_{0k}$, *which is* $E\big(\mathbf{Y}_0|\mathbf{x}_0\big) = \boldsymbol{B}^{'}\mathbf{x}_0$
    - point estimate: $\hat{\boldsymbol{B}}^{'}\mathbf{x}_0 \sim MN\big(E(\mathbf{Y}_0|\mathbf{x}_0), \mathbf{x}_0^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{x}_0\mathbf{\Sigma}\big)$.
    - $100(1-\alpha)\%$ simultaneous confidence interval for $E(Y_{0j}) = \mathbf{x}_0^{'}\boldsymbol{\beta}_{(j)}$

$$\mathbf{x}_0^{'}\hat{\boldsymbol{\beta}}_{(j)} \pm \sqrt{\frac{m(n-[k+1])}{n-k-m}F_{m,n-k-m}(\alpha)}\sqrt{[\mathbf{x}_0^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{x}_0]\frac{n}{n-[k+1]}\hat{\sigma}_{jj}}$$

- *to predict for* $\mathbf{Y}_0 = \boldsymbol{B}^{'}\mathbf{x}_0 + \epsilon_0$
    - point estimate: $\hat{\boldsymbol{B}}^{'}\mathbf{x}_0 \sim MN\big(E(\mathbf{Y}_0|\mathbf{x}_0), \mathbf{x}_0^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{x}_0\mathbf{\Sigma}\big)$.
    - $100(1-\alpha)\%$ prediction interval for Y_{0j}:

$$\mathbf{x}_0^{'}\hat{\boldsymbol{\beta}}_{(j)} \pm \sqrt{\frac{m(n-[k+1])}{n-k-m}F_{m,n-k-m}(\alpha)}\sqrt{[1+\mathbf{x}_0^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{x}_0]\frac{n}{n-[k+1]}\hat{\sigma}_{jj}}$$

II.4.4C Multivariate Multiple Regression: inference

Provided the LSE $\hat{\boldsymbol{B}}$ using data with size n for the model $\mathbf{Y} = \mathbf{X}\boldsymbol{B} + \boldsymbol{E}$ under the normality assumption $\boldsymbol{\epsilon}_i \sim MN(\mathbf{0}, \boldsymbol{\Sigma})$ iid:
if $\mathbf{Y}_0$ is the response when the indpendent variables are $x_{01}, \ldots, x_{0k}$,

- *to estimate* $E(\mathbf{Y}_0|\mathbf{x}_0) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{01} + \ldots + \boldsymbol{\beta}_k x_{0k}$, *which is* $E(\mathbf{Y}_0|\mathbf{x}_0) = \boldsymbol{B}^{'}\mathbf{x}_0$
  - point estimate: $\hat{\boldsymbol{B}}^{'}\mathbf{x}_0 \sim MN\big(E(\mathbf{Y}_0|\mathbf{x}_0), \mathbf{x}_0^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{x}_0\boldsymbol{\Sigma}\big)$.
  - $100(1-\alpha)\%$ simultaneous confidence interval for $E(Y_{0j}) = \mathbf{x}_0^{'}\boldsymbol{\beta}_{(j)}$

$$\mathbf{x}_0^{'}\hat{\boldsymbol{\beta}}_{(j)} \pm \sqrt{\frac{m(n-[k+1])}{n-k-m}F_{m,n-k-m}(\alpha)}\sqrt{[\mathbf{x}_0^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{x}_0]\frac{n}{n-[k+1]}\hat{\sigma}_{jj}}$$

- *to predict for* $\mathbf{Y}_0 = \boldsymbol{B}^{'}\mathbf{x}_0 + \boldsymbol{\epsilon}_0$
  - point estimate: $\hat{\boldsymbol{B}}^{'}\mathbf{x}_0 \sim MN\big(E(\mathbf{Y}_0|\mathbf{x}_0), \mathbf{x}_0^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{x}_0\boldsymbol{\Sigma}\big)$.
  - $100(1-\alpha)\%$ prediction ellipsoid for $\mathbf{Y}_0$:

$$\Big\{\mathbf{y} : (\mathbf{y}-\hat{\boldsymbol{B}}^{'}\mathbf{x}_0)^{'}\Big(\frac{n}{n-[k+1]}\hat{\boldsymbol{\Sigma}}\Big)^{-1}(\mathbf{y}-\hat{\boldsymbol{B}}^{'}\mathbf{x}_0) \leq \frac{m(n-[k+1])}{n-k-m}F_{m,n-k-m}(\alpha)[1+\mathbf{x}_0^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{x}_0$$

- There will be more information on midterm be posted (range, practice questions)