STATISTICAL ESTIMATION IN MULTIVARIATE NORMAL DISTRIBUTION WITH A BLOCK OF MISSING OBSERVATIONS

by

YI LIU

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

DECEMBER 2017

To Alex and My Parents

Acknowledgements

I would like to thank my supervisor, Dr. Chien-Pai Han, for his instructing, guiding and supporting me over the years. You have set an example of excellence as a researcher, mentor, instructor and role model.

I would like to thank Dr. Shan Sun-Mitchell for her continuously encouraging and instructing. You are both a good teacher and helpful friend. I would like to thank my thesis committee members Dr. Suvra Pal and Dr. Jonghyun Yun for their discussion, ideas and feedback which are invaluable.

I would like to thank the graduate advisor, Dr. Hristo Kojouharov, for his instructing, help and patience. I would like to thank the chairman Dr. Jianzhong Su, Dr. Minerva Cordero-Epperson, Lona Donnelly, Libby Carroll and other staffs for their help.

I would like to thank my manager, Robert Schermerhorn, for his understanding, encouraging and supporting which make this happen. I would like to thank my employer Sabre and my coworkers for their support in the past two years.

I would like to thank my husband Alex for his encouraging and supporting over all these years. In particularly, I would like to thank my parents -- without the inspiration, drive and support that you have given me, I might not be the person I am today.

October 11, 2017

Abstract

STATISTICAL ESTIMATION IN MULTIVARIATE NORMAL DISTRIBUTION

WITH A BLOCK OF MISSING OBSERVATIONS

YI LIU, Ph.D.

The University of Texas at Arlington, 2017

Supervising Professor: Chien-Pai Han

Missing observations occur quite often in data analysis. We study a random sample from a multivariate normal distribution with a block of missing observations, here the observations missing is not at random. We use maximum likelihood method to obtain the estimators from such a sample. The properties of the estimators are derived. The prediction problem is considered when the response variable has missing values. The variances of the mean estimators of the response variable under with and without extra information are compared. We prove that the variance of the mean estimator of the response variable using all data is smaller than that we do not consider extra information, when the correlation between response variable and predictors meets some conditions. We derive three kinds of prediction interval for the future observation. An example of a college admission data is used to obtain the estimators for the bivariate and multivariate situations.

Table of Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Missing observations occur quite often in data analysis. The data may be missing on some variables for some observations. Generally, there are three kinds of missing: missing completely at random (MCAR), missing at random (MAR) and missing not at random(MNAR). There are a lot of research on how to minimize bias and get good estimates with the missing data. Allison (2002) discussed the strength and weakness of the conventional and novel methods to deal with the missing. The conventional methods include Listwise deletion, Pairwise deletion, dummy variable adjustment, and imputation such as replacement with the mean, regression and Hot Deck. The novel methods include maximum likelihood and multiple imputation. Other relevant papers include Anderson (1957), Rubin (1976), Chung and Han (2000), Howell (2008), Little and Zhang (2011), Han and Li (2011), Sinsomboonthong (2011), and books by Little and Rubin (2002). Though the imputation is the most popular technique to deal with the missing, it is not appropriate in some cases. For example, for the college admission data from Han and Li (2011), the TOEFL score is naturally missing for US students, and imputation does not make sense in this case.

For a random sample in multivariate normal distribution with a block of observations missing not at random, Chung and Han (2000), and Han and Li (2011) considered this situation in discriminant analysis and regression model, respectively. Anderson (1957) considered a sample from a bivariate normal distribution with missing observations. The maximum likelihood method is used to give the estimators, but the paper did not study the properties of the estimators. Allison (2002) discussed maximum likelihood method when missing is ignorable.

We use maximum likelihood method to obtain estimators from a multivariate normal distribution sample with a block of observations missing not at random. We consider all available information, do not delete any data and do not impute.

We have the following random sample with a block of missing $\boldsymbol{Y}$ values:

$$
\begin{array}{cccccccc}
X_{1,1} & X_{1,2} & \cdots & X_{1,p} & Y_{1,1} & Y_{1,2} & \cdots & Y_{1,q} \\
X_{2,1} & X_{2,2} & \cdots & X_{2,p} & Y_{2,1} & Y_{2,2} & \cdots & Y_{2,q} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
X_{m,1} & X_{m,2} & \cdots & X_{m,p} & Y_{m,1} & Y_{m,2} & \cdots & Y_{m,q} \\
X_{m+1,1} & X_{m+1,2} & \cdots & X_{m+1,p} & & & & \\
\vdots & \vdots & \vdots & \vdots & & & & \\
X_{n,1} & X_{n,1} & \cdots & X_{n,p} & & & &
\end{array}
$$

The multivariate normal probability density function (pdf) can be written as

$$f(\mathbf{x}, \boldsymbol{y}) = g(\boldsymbol{y}|\mathbf{x})h(\mathbf{x}) \tag{1 - 1}$$

where $\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{bmatrix}$ have a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

$$\boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{\mathrm{xx}} & \boldsymbol{\Sigma}_{\mathrm{xy}}\\ \boldsymbol{\Sigma}_{\mathrm{yx}} & \boldsymbol{\Sigma}_{\mathrm{yy}}\end{bmatrix} \tag{1 - 2}$$

$\boldsymbol{X}$ is a $p \times 1$ vector and $\boldsymbol{Y}$ is a $q \times 1$ vector. $g(\boldsymbol{y}|\mathbf{x})$ is the conditional pdf of $\boldsymbol{Y}$ given $\boldsymbol{X} = \mathbf{x}$, and $h(\mathbf{x})$ is the marginal pdf of $\boldsymbol{X}$.

The joint likelihood function is

$$\mathrm{L}\left(\boldsymbol{\mu}_{\mathrm{y}}, \boldsymbol{\mu}_{\mathrm{x}}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathrm{xx}}, \boldsymbol{\Sigma}_{\mathrm{e}}\right) = \prod_{\mathrm{j}=1}^{\mathrm{m}} g_{Y|X}\left(\mathbf{y}_j|\mathbf{x}_j; \boldsymbol{\mu}_{\mathrm{y}}, \boldsymbol{\mu}_{\mathrm{x}}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathrm{e}}\right) \prod_{\mathrm{i}=1}^{\mathrm{n}} h_{\boldsymbol{X}}(\mathbf{x}_{\mathrm{i}}; \boldsymbol{\mu}_{\mathrm{x}}, \boldsymbol{\Sigma}_{\mathrm{xx}}) \tag{1 - 3}$$

The missing data happened for $\boldsymbol{Y}$ in our model. We first consider a bivariate normal distribution sample (Chapter 2), i.e., $p, q = 1$. The maximum likelihood estimators are derived. We prove that they are consistent and asymptotically efficient. Their distributions are obtained. We show that the unconditional variance of $\hat{\mu}_y$ is smaller when all data is considered than that we do not consider extra information under a correlation condition. We also consider the prediction problem with missing observations. Three kinds of the prediction intervals for a future observation (X_0, Y_0) are derived. They are

1) Usual prediction interval for Y_0 – conditioning on $X = x\ and\ X_0 = x_0$
2) Prediction interval for Y_0 – unconditional on X, but conditioning on $X_0 = x_0$
3) Unconditional prediction interval for Y_0

Then we extend to multiple regression model (Chapter 3), i.e., $\boldsymbol{X}$ is a $p \times 1$ vector. Again, we study the properties of the maximum likelihood estimators and derive the three kinds of prediction intervals. Finally, we extend to multivariate regression model (Chapter 4), i.e., $\boldsymbol{X}$ is a $p \times 1$ vector and $\boldsymbol{Y}$ is a $q \times 1$ vector. We obtain the maximum likelihood estimators, study their properties and derive the three kinds of prediction interval for each response variable which follows the multiple regression model.

The comparison of the unconditional variance of $\hat{\mu}_y$ with and without extra information, and density plots for $\hat{\beta}$ are simulated by R Studio. An example of the college admission data from Han and Li (2011) (Table 1 -1) is used to obtain the estimators for the bivariate and multivariate situations. In this data set, TOEFL scores are required for students whose native language is not English, but missing for students whose native language is English such as US students. The missing values should not be imputed, as these values do not exist.

Table 1-1 College Admission Data

Obs	GRE Verbal	GRE Quantitative	GRE Analytic	TOEFL		Obs	GRE Verbal	GRE Quantitative	GRE Analytic	TOEFL
1	420	800	600	497		21	250	730	460	513
2	330	710	380	563		22	320	760	610	560
3	270	700	340	510		23	360	720	525	540
4	400	710	600	563		24	370	780	500	500
5	280	800	450	543		25	300	630	380	507
6	310	660	425	507		26	390	580	370	587
7	360	620	590	537		27	380	770	500	520
8	220	530	340	543		28	370	640	200	520
9	350	770	560	580		29	340	800	540	517
10	360	750	440	577		30	460	750	560	597
11	440	700	630	NA		31	630	540	600	NA
12	640	520	610	NA		32	350	690	620	NA
13	480	550	560	NA		33	480	610	480	NA
14	550	630	630	NA		34	630	410	530	NA
15	450	660	630	NA		35	550	450	500	NA
16	410	410	340	NA		36	510	690	730	NA
17	460	610	560	NA		37	640	720	520	NA
18	580	580	610	NA		38	440	580	620	NA
19	450	540	570	NA		39	350	430	480	NA
20	420	630	660	NA		40	480	700	670	NA

Chapter 2

STATISTICAL ESTIMATION IN BIVARIATE NORMAL DISTRIBUTION WITH A BLOCK OF MISSING OBSERVATIONS

Let $\begin{bmatrix} X \\ Y \end{bmatrix}$ have a bivariate normal distribution with mean vector $\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$ and covariance matrix $\begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$.

Suppose the following random sample with a block of missing Y values are obtained:

$$\begin{matrix} X_1 & Y_1 \\ X_2 & Y_2 \\ \vdots & \vdots \\ X_m & Y_m \\ X_{m+1} & \\ \vdots & \\ X_n & \end{matrix}$$

Based on the data, we would like to estimate the parameters. We can write the bivariate normal probability density function (pdf) as

$$f(x, y) = g(y|x)h(x) \qquad (2 - 1)$$

Where $g(y|x)$ is the conditional pdf of Y given $X = x$, and $h(x)$ is the marginal pdf of X.

$$g_{Y|X}\left(y_j|x_j; \mu_y, \mu_x, \beta, \sigma_e^2\right) = \frac{1}{\sqrt{2\pi}\sigma_e} exp\left\{-\frac{1}{2\sigma_e^2}\left[y_j - E\left(y_j|x_j\right)\right]^2\right\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_e} exp\left[-\frac{1}{2\sigma_e^2}\left(y_j - \mu_y - \beta\left(x_j - \mu_x\right)\right)^2\right] \tag{2 - 2}$$

$$h_X(x_i; \mu_x, \sigma_x^2) = \frac{1}{\sqrt{2\pi}\sigma_x} exp\left[-\frac{1}{2\sigma_x^2}(x_i - \mu_x)^2\right] \tag{2 - 3}$$

where $j = 1,2, \dots, m. i = 1,2, \dots, n.$

$$E(y_j|x_j) = \mu_y + \frac{\sigma_{yx}}{\sigma_x^2}\left(x_j - \mu_x\right) = \mu_y + \beta\left(x_j - \mu_x\right) \tag{2 - 4}$$

$$\beta = \frac{\sigma_{yx}}{\sigma_x^2} \tag{2 - 5}$$

$$\sigma_e^2 = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2} = \sigma_y^2 - \beta^2\sigma_x^2 \tag{2 - 6}$$

The joint likelihood function is

$$L\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right) = \prod_{j=1}^{m} g_{Y|X}\left(y_j|x_j; \mu_y, \mu_x, \beta, \sigma_e^2\right) \prod_{i=1}^{n} h_X(x_i; \mu_x, \sigma_x^2) \tag{2 - 7}$$

$$L\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right) = (2\pi)^{-\frac{n+m}{2}}\sigma_e^{-m}\sigma_x^{-n}$$

$$\cdot exp\left\{-\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left(y_j - \mu_y - \beta(x_j - \mu_x)\right)^2 - \frac{1}{2\sigma_x^2}\sum_{i=1}^{n}(x_i - \mu_x)^2\right\}$$

Since the log of the joint likelihood function is more convenient to use and there is no loss of information in using it, also maximizing the likelihood is equivalent maximizing the log likelihood since the log is a monotone increasing function. Therefore, we take the log of the joint likelihood function and denote it as

$$l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2) = ln(L(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2))$$

$$l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2) = -\frac{n+m}{2} ln(2\pi) - mln(\sigma_e) - nln(\sigma_x)$$

$$-\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left(y_j - \mu_y - \beta\left(x_j - \mu_x\right)\right)^2 - \frac{1}{2\sigma_x^2}\sum_{i=1}^{n}(x_i - \mu_x)^2 \tag{2 - 8}$$

2.1 Maximum Likelihood Estimators

We derive the maximum likelihood estimators (MLE) by taking the derivatives of the log likelihood function $(2-8)$ with respect to each parameter and setting it to zero. Solving the estimating equations simultaneously, we obtain the MLE.

Taking the derivative of $(2-8)$ with respective to μ_y results in

$$\frac{\partial l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \mu_y} = \frac{1}{\sigma_e^2} \sum_{j=1}^{m} \left(y_j - \mu_y - \beta(x_j - \mu_x) \right)$$

Set

$$\frac{\partial l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \mu_y} = 0$$

Then we have the estimating equation

$$\sum_{j=1}^{m} \left[y_j - \mu_y - \beta\left(x_j - \mu_x\right) \right] = 0 \qquad (2 - 9)$$

Taking the derivative of $(2-8)$ with respective to μ_x results in

$$\frac{\partial l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \mu_x} = -\frac{\beta}{\sigma_e^2} \sum_{j=1}^{m} \left[y_j - \mu_y - \beta\left(x_j - \mu_x\right) \right] + \frac{1}{\sigma_x^2} \sum_{i=1}^{n} (x_i - \mu_x)$$

Set

$$\frac{\partial l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \mu_x} = 0$$

Then

$$-\frac{\beta}{\sigma_e^2} \sum_{j=1}^{m} \left[y_j - \mu_y - \beta\left(x_j - \mu_x\right) \right] + \frac{1}{\sigma_x^2} \sum_{i=1}^{n} (x_i - \mu_x) = 0 \qquad (2 - 10)$$

Taking the derivative of $(2-8)$ with respective to β results in

$$\frac{\partial l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \beta} = \frac{1}{\sigma_e^2} \sum_{j=1}^{m} \left(y_j - \mu_y - \beta\left(x_j - \mu_x\right)\right)\left(x_j - \mu_x\right)$$

Set

$$\frac{\partial l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \beta} = 0$$

Then

$$\frac{1}{\sigma_e^2} \sum_{j=1}^{m} \left(y_j - \mu_y - \beta\left(x_j - \mu_x\right)\right)\left(x_j - \mu_x\right) = 0 \qquad (2 - 11)$$

Taking the derivative of $(2-8)$ with respective to σ_x^2 results in

$$\frac{\partial l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \sigma_x^2} = -\frac{n}{2\sigma_x^2} + \frac{1}{2\sigma_x^4} \sum_{i=1}^{n} (x_i - \mu_x)^2$$

Set

$$\frac{\partial l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \sigma_x^2} = 0$$

Then

$$-\frac{n}{2\sigma_x^2} + \frac{1}{2\sigma_x^4} \sum_{i=1}^{n} (x_i - \mu_x)^2 = 0 \qquad (2 - 12)$$

Taking the derivative of $(2-8)$ with respective to σ_e^2 results in

$$\frac{\partial l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \sigma_e^2} = -\frac{m}{2\sigma_e^2} + \frac{1}{2\sigma_e^4} \sum_{j=1}^{m} \left(y_j - \mu_y - \beta(x_j - \mu_x)\right)^2$$

Set

$$\frac{\partial l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \sigma_e^2} = 0$$

Then

$$-\frac{m}{2\sigma_e^2}+\frac{1}{2\sigma_e^4}\sum_{j=1}^{m}\left(y_j-\mu_y-\beta\left(x_j-\mu_x\right)\right)^2=0 \qquad (2\text{ - }13)$$

Simultaneously solve estimating equations $(2-9)$ to $(2-13)$, we obtain the following maximum likelihood estimators:

$$\hat{\mu}_x=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}_n \qquad (2\text{ - }14)$$

$$\hat{\mu}_y=\bar{Y}_m-\hat{\beta}(\bar{X}_m-\bar{X}_n) \qquad (2\text{ - }15)$$

$$\hat{\beta}=\frac{\sum_{j=1}^{m}(Y_j-\bar{Y}_m)\left(X_j-\bar{X}_m\right)}{\sum_{j=1}^{m}\left(X_j-\bar{X}_m\right)^2} \qquad (2\text{ - }16)$$

$$\hat{\sigma}_x^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 \qquad (2\text{ - }17)$$

$$\hat{\sigma}_e^2=\frac{1}{m}\sum_{j=1}^{m}[\left(Y_j-\bar{Y}_m\right)-\hat{\beta}\left(X_j-\bar{X}_m\right)]^2 \qquad (2\text{ - }18)$$

where

$$\bar{Y}_m=\frac{1}{m}\sum_{j=1}^{m}Y_j \qquad \bar{X}_m=\frac{1}{m}\sum_{j=1}^{m}X_j \qquad \bar{X}_n=\frac{1}{n}\sum_{j=1}^{n}X_j$$

Similarly, if we do not consider extra information $X_{m+1}, X_{m+2}, \dots X_n$ and only use the first m observations, we have

$$\hat{\mu}_{x\_no} = \frac{1}{m}\sum_{j=1}^{m} X_j = \bar{X}_m \tag{2 - 19}$$

$$\hat{\mu}_{y\_no} = \bar{Y}_m \tag{2 - 20}$$

$$\hat{\beta}_{no} = \frac{\sum_{j=1}^{m}(Y_j - \bar{Y}_m)(X_j - \bar{X}_m)}{\sum_{j=1}^{m}(X_j - \bar{X}_m)^2} \tag{2 - 21}$$

$$\hat{\sigma}^2_{x\_no} = \frac{1}{m}\sum_{j=1}^{m}(X_j - \bar{X}_m)^2 \tag{2 - 22}$$

$$\hat{\sigma}^2_{e\_no} = \frac{1}{m}\sum_{j=1}^{m}[(Y_j - \bar{Y}_m) - \hat{\beta}(X_j - \bar{X}_m)]^2 \tag{2 - 23}$$

2.2 Properties of the Maximum Likelihood Estimators

2.2.1 Estimator of the Mean of X

The expectation of $\hat{\mu}_x$ is

$$\mathrm{E}(\hat{\mu}_x) = \mathrm{E}(\bar{X}_n) = E\left\{\frac{1}{n}\sum_{i=1}^{n} X_i\right\} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}\cdot n\mu_x = \mu_x \tag{2 - 24}$$

So, $\hat{\mu}_x$ is an unbiased estimator.

The variance of $\hat{\mu}_x$ is

$$\mathrm{Var}(\hat{\mu}_x) = \mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left\{\frac{1}{n}\sum_{i=1}^{n} X_i\right\} = \frac{1}{n^2}\mathrm{Var}\left\{\sum_{i=1}^{n} X_i\right\} = \frac{1}{n^2}\left\{\sum_{i=1}^{n}\mathrm{Var}(X_i)\right\}$$

$$= \frac{1}{n^2} \cdot n\sigma_x^2 = \frac{1}{n}\sigma_x^2 \qquad (2 - 25)$$

Since X_i is normal, so is $\bar{X}_n$, hence

$$\hat{\mu}_x \sim N\left(\mu_x, \frac{1}{n}\sigma_x^2\right)$$

2.2.2 Estimator of the Variance of X

The expectation of $\hat{\sigma}_x^2$ is

$$E(\hat{\sigma}_x^2) = \mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right\} = \frac{1}{n}\mathrm{E}\left\{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right\} = \frac{\sigma_x^2}{n}\mathrm{E}\left\{\sum_{i=1}^{n}\frac{(X_i - \bar{X}_n)^2}{\sigma_x^2}\right\}$$

$$= \frac{\sigma_x^2}{n} \cdot (n-1) = \frac{n-1}{n}\sigma_x^2 \qquad (2 - 26)$$

where

$$\sum_{i=1}^{n}\frac{(X_i - \bar{X}_n)^2}{\sigma_x^2} \sim \chi^2(n-1) \qquad (2 - 27)$$

$\hat{\sigma}_x^2$ is a biased estimator. The bias of $\hat{\sigma}_x^2$ is

$$\mathrm{Bias}(\hat{\sigma}_x^2, \sigma_x^2) = E(\hat{\sigma}_x^2) - \sigma_x^2 = -\frac{1}{n}\sigma_x^2$$

The bias vanishes as $n \to \infty$.

If we define

$$\mathrm{S}_{\mathrm{xn}}^2 = \frac{n}{n-1}\hat{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \qquad (2 - 28)$$

Then

$$E(\mathrm{S}_{\mathrm{xn}}^2) = \frac{n}{n-1}E(\hat{\sigma}_x^2) = \sigma_x^2 \qquad (2 - 29)$$

$\mathrm{S}_{\mathrm{xn}}^2$ is an unbiased estimator for $\hat{\sigma}_x^2$.

The variance of $\hat{\sigma}_x^2$ is

$$Var(\hat{\sigma}_x^2) = \mathrm{Var}\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right\} = \frac{1}{n^2}\mathrm{Var}\left\{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right\} = \frac{\sigma_x^4}{n^2}\mathrm{Var}\left\{\sum_{i=1}^{n}\frac{(X_i - \bar{X}_n)^2}{\sigma_x^2}\right\}$$

$$= \frac{\sigma_x^4}{n^2}\cdot 2(n-1) = \frac{2(n-1)}{n^2}\sigma_x^4 \qquad (2 - 30)$$

The mean squared error for $\hat{\sigma}_x^2$ is

$$MSE(\hat{\sigma}_x^2) = E(\hat{\sigma}_x^2 - \sigma_x^2)^2 = \mathrm{Var}(\hat{\sigma}_x^2) + [\mathrm{Bias}(\hat{\sigma}_x^2, \sigma_x^2)]^2$$

$$= \frac{2(n-1)}{n^2}\sigma_x^4 + \frac{1}{n^2}\sigma_x^4 = \frac{(2n-1)}{n^2}\sigma_x^4 \qquad (2 - 31)$$

The distribution of $\hat{\sigma}_x^2$ is the Chi-Square distribution and

$$\frac{n\hat{\sigma}_x^2}{\sigma_x^2} \sim \chi^2\ (n-1)$$

2.2.3 Estimator of the Regression Coefficient $\hat{\beta}$

Since the formula for $\hat{\beta}$ involves X and Y, we will derive the conditional expectation and variance of $\hat{\beta}$ given $X = \mathrm{x}$ first, then we derive its unconditional expectation and variance.

The conditional expectation of $\hat{\beta}$ given $X = \mathrm{x}$ is

$$\mathrm{E}(\hat{\beta}|x) = \mathrm{E}\left\{\frac{\sum_{j=1}^{m}(Y_j - \bar{Y}_m)(x_j - \bar{x}_m)}{\sum_{j=1}^{m}(x_j - \bar{x}_m)^2}|x\right\} = \mathrm{E}\left\{\frac{\sum_{j=1}^{m}Y_j(x_j - \bar{x}_m)}{\sum_{j=1}^{m}(x_j - \bar{x}_m)^2}|x\right\}$$

$$= \frac{\mathrm{E}\{\sum_{j=1}^{m}Y_j(x_j - \bar{x}_m)\,|x\}}{\sum_{j=1}^{m}(x_j - \bar{x}_m)^2}$$

$$= \frac{\beta\sum_{j=1}^{m}(x_j - \bar{x}_m)^2}{\sum_{j=1}^{m}(x_j - \bar{x}_m)^2} = \beta \qquad (2 - 32)$$

where

$$\mathrm{E}(Y_j|x_j) = \mu_y + \beta(x_j - \mu_x) \qquad (2 - 33)$$

$$\mathrm{E}\{\textstyle\sum_{j=1}^{m} Y_j(x_j-\bar{x}_m)\,|x\} = \sum_{j=1}^{m} \mathrm{E}[Y_j(x_j-\bar{x}_m)\,|x] = \sum_{j=1}^{m} [\mathrm{E}(Y_j\,|x)(x_j-\bar{x}_m)]$$

$$= \sum_{j=1}^{m} \{[\mu_y + \beta(x_j-\mu_x)\,](x_j-\bar{x}_m)\}$$

$$= \mu_y \sum_{j=1}^{m} (x_j-\bar{x}_m) + \beta \sum_{j=1}^{m} x_j(x_j-\bar{x}_m) - \beta\mu_x \sum_{j=1}^{m}\,(x_j-\bar{x}_m)$$

$$= \beta \sum_{j=1}^{m} x_j(x_j-\bar{x}_m)$$

$$= \beta \sum_{j=1}^{m} (x_j-\bar{x}_m)^2$$

The unconditional expectation of $\hat{\beta}$ is

$$E(\hat{\beta}) = E[E(\hat{\beta}|X)] = \beta \qquad (2 - 34)$$

$\hat{\beta}$ is an unbiased estimator.

Similarly, the conditional variance of $\hat{\beta}$ given $X = \mathrm{x}$ is

$$\mathrm{Var}(\hat{\beta}|x) = \mathrm{Var}\left\{\frac{\sum_{j=1}^{m}(Y_j-\bar{Y}_m)(x_j-\bar{x}_m)}{\sum_{j=1}^{m}(x_j-\bar{x}_m)^2}\,|x\right\} = \mathrm{Var}\left\{\frac{\sum_{j=1}^{m} Y_j(x_j-\bar{x}_m)}{\sum_{j=1}^{m}(x_j-\bar{x}_m)^2}\,|x\right\}$$

$$= \frac{\mathrm{Var}\{\sum_{j=1}^{m} Y_j(x_j-\bar{x}_m)\,|x\}}{\left[\sum_{j=1}^{m}(x_j-\bar{x}_m)^2\right]^2} = \frac{\sum_{j=1}^{m}\mathrm{Var}[Y_j(x_j-\bar{x}_m)|x\,]}{\left[\sum_{j=1}^{m}(x_j-\bar{x}_m)^2\right]^2}$$

$$= \frac{\sum_{j=1}^{m}(x_j-\bar{x}_m)^2\mathrm{Var}(Y_j|x)}{[\sum_{j=1}^{m}(x_j-\bar{x}_m)^2]^2} = \frac{\sum_{j=1}^{m}(x_j-\bar{x}_m)^2\sigma_e^2}{[\sum_{j=1}^{m}(x_j-\bar{x}_m)^2]^2}$$

$$= \frac{\sigma_e^2\sum_{j=1}^{m}(x_j-\bar{x}_m)^2}{[\sum_{j=1}^{m}(x_j-\bar{x}_m)^2]^2}$$

$$= \frac{\sigma_e^2}{\sum_{j=1}^{m}(x_j-\bar{x}_m)^2} \qquad (2 - 35)$$

To obtain the unconditional variance of $\hat{\beta}$, we use the Law of Total Variance

$$\mathrm{Var}(\hat{\beta}) = \mathrm{E}\left(\mathrm{Var}(\hat{\beta}|X)\right) + \mathrm{Var}(\mathrm{E}(\hat{\beta}|X))$$

now

$$\mathrm{Var}\left(\mathrm{E}(\hat{\beta}|X)\right) = \mathrm{Var}(\beta) = 0$$

$$E\left(\mathrm{Var}(\hat{\beta}|X)\right) = E\left(\frac{\sigma_e^2}{\sum_{j=1}^{m}\left(X_j - \bar{X}_m\right)^2}\right) = \frac{\sigma_e^2}{\sigma_x^2} E\left(\frac{\sigma_x^2}{\sum_{j=1}^{m}\left(X_j - \bar{X}_m\right)^2}\right)$$

In order to obtain $E\left(\frac{\sigma_x^2}{\sum_{j=1}^{m}\left(X_j - \bar{X}_m\right)^2}\right)$, let

$$U = \sum_{j=1}^{m} \frac{\left(X_j - \bar{X}_m\right)^2}{\sigma_x^2}$$

Then

$$U \sim \chi^2(m-1)$$

So

$$\mathrm{E}\left(\frac{1}{U}\right) = \int_0^\infty \frac{1}{u} \cdot \frac{1}{\Gamma\left(\frac{m-1}{2}\right) 2^{\frac{m-1}{2}}} u^{\frac{m-1}{2}-1} e^{-\frac{u}{2}} du$$

$$= \int_0^\infty \frac{1}{\frac{m-3}{2} \cdot 2} \cdot \frac{1}{\Gamma\left(\frac{m-3}{2}\right) 2^{\frac{m-3}{2}}} u^{\frac{m-3}{2}-1} e^{-\frac{u}{2}} du$$

$$= \frac{1}{m-3} \int_0^\infty \frac{1}{\Gamma\left(\frac{m-3}{2}\right) 2^{\frac{m-3}{2}}} u^{\frac{m-3}{2}-1} e^{-\frac{u}{2}} du$$

$$= \frac{1}{m-3} \quad for \;\; m > 3$$

Hence, we have

$$\mathrm{E}\left(\mathrm{Var}(\hat{\beta}|X)\right) = \frac{\sigma_e^2}{\sigma_x^2} \mathrm{E}\left(\frac{1}{U}\right) = \frac{\sigma_e^2}{\sigma_x^2} \cdot \frac{1}{m-3} = \frac{\sigma_e^2}{(m-3)\sigma_x^2}$$

The unconditional variance of $\hat{\beta}$ is

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma_e^2}{(m-3)\sigma_x^2} + 0 = \frac{\sigma_e^2}{(m-3)\sigma_x^2} \qquad for\ \ m > 3 \tag{2 - 36}$$

2.2.3.1 The Density Function of $\hat{\beta}$

According to Kendall and Stuart (1945) , the density function of $\hat{\beta}$ is

$$f(\hat{\beta}) = \frac{\Gamma\left(\frac{m}{2}\right)\sigma_y^{m-1}(1-\rho^2)^{\frac{m-1}{2}}}{\sqrt{\pi}\Gamma\left(\frac{m-1}{2}\right)\sigma_x^{m-1}}\left[\frac{\sigma_y^2}{\sigma_x^2}(1-\rho^2) + \left(\hat{\beta} - \rho\frac{\sigma_y}{\sigma_x}\right)^2\right]^{-\frac{m}{2}}$$

$$= \frac{\Gamma\left(\frac{m}{2}\right)}{a\sqrt{\pi}\Gamma\left(\frac{m-1}{2}\right)}\left[1 + \left(\frac{\hat{\beta}-\beta}{a}\right)^2\right]^{-\frac{m}{2}}$$

$$= \frac{1}{a\mathrm{B}\left(\frac{m-1}{2}, \frac{1}{2}\right)}\left[1 + \left(\frac{\hat{\beta}-\beta}{a}\right)^2\right]^{-\frac{m}{2}} \tag{2 - 37}$$

where

$$\beta = \rho\frac{\sigma_y}{\sigma_x}\ , \quad a = \sqrt{\frac{\sigma_y^2}{\sigma_x^2} - \beta^2}\ , \quad \mathrm{B}\left(\frac{m-1}{2}, \frac{1}{2}\right) = \frac{\Gamma\left(\frac{m-1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} \tag{2 - 38}$$

It is Pearson Type VII distribution, symmetrical about the point $\beta = \rho\frac{\sigma_y}{\sigma_x}$.

By $(2-6)$, $(2-36)$, and $(2-38)$, we have

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma_e^2}{(m-3)\sigma_x^2} = \frac{\sigma_y^2 - \beta^2\sigma_x^2}{(m-3)\sigma_x^2} = \frac{1}{m-3}\left(\frac{\sigma_y^2}{\sigma_x^2} - \beta^2\right) = \frac{a^2}{m-3} \tag{2 - 39}$$

Let

$$\mathrm{k} = \frac{m}{2}, \quad \sigma^2 = \mathrm{Var}(\hat{\beta}) = \frac{a^2}{m-3} = \frac{a^2}{2k-3} \tag{2 - 40}$$

If β and σ are held constant as $\mathrm{k} \to \infty$, we have

$$\lim_{k\to\infty} f(\hat{\beta}) = \lim_{k\to\infty} \frac{1}{\sigma\sqrt{2k-3}\mathrm{B}\left(k-\frac{1}{2},\frac{1}{2}\right)}\left[1+\left(\frac{\hat{\beta}-\beta}{\sigma\sqrt{2k-3}}\right)^2\right]^{-k}$$

$$= \frac{1}{\sigma\sqrt{2}\Gamma\left(\frac{1}{2}\right)} \lim_{k\to\infty}\left[\frac{\Gamma(k)}{\sqrt{\mathrm{k}-\frac{3}{2}}\,\Gamma\left(k-\frac{1}{2}\right)}\right] \lim_{k\to\infty}\left[1+\frac{\left(\frac{\hat{\beta}-\beta}{\sigma}\right)^2}{2k-3}\right]^{-k}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot 1 \cdot \exp\left[-\frac{1}{2}\left(\frac{\hat{\beta}-\beta}{\sigma}\right)^2\right] \tag{2 - 41}$$

This is the density of a normal distribution with mean β and variance σ^2 . So, $\hat{\beta}$ has an asymptotically normal distribution when sample is large.

$$\hat{\beta} \;\dot{\sim}\; N\left(\beta, \frac{\sigma_e^2}{(m-3)\sigma_x^2}\right) \tag{2 - 42}$$

2.2.3.2 Plot of the Density Function of $\hat{\beta}$

Suppose we have the following bivariate normal distribution:

$$\begin{pmatrix}\mu_y\\ \mu_x\end{pmatrix} = \begin{pmatrix}2\\ 5\end{pmatrix}$$

$$\Sigma = \begin{pmatrix}\sigma_y^2 & \sigma_{xy}\\ \sigma_{xy} & \sigma_x^2\end{pmatrix} = \begin{pmatrix}4 & 6\\ 6 & 16\end{pmatrix} \tag{2 - 43}$$

Then

$$\beta = \rho\frac{\sigma_y}{\sigma_x} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{6}{16} = 0.375$$

$$a = \sqrt{\frac{\sigma_y^2}{\sigma_x^2} - \beta^2} = \sqrt{\frac{4}{16} - \left(\frac{6}{16}\right)^2} = \frac{\sqrt{7}}{8} = 0.3307$$

For given m, we have

$$B\left(\frac{m-1}{2},\frac{1}{2}\right) = B\left(\frac{49}{2},\frac{1}{2}\right) = 0.3599 \qquad\qquad for\ m = 50$$

$$B\left(\frac{m-1}{2},\frac{1}{2}\right) = B\left(\frac{69}{2},\frac{1}{2}\right) = 0.3029 \qquad for\ m = 70$$

$$B\left(\frac{m-1}{2},\frac{1}{2}\right) = B\left(\frac{89}{2},\frac{1}{2}\right) = 0.2664 \qquad for\ m = 90$$

So

$$f(\hat{\beta}) = \frac{1}{0.1190}\left[1 + \left(\frac{\hat{\beta} - 0.375}{0.3307}\right)^2\right]^{-25} \qquad for\ m = 50$$

$$f(\hat{\beta}) = \frac{1}{0.1002}\left[1 + \left(\frac{\hat{\beta} - 0.375}{0.3307}\right)^2\right]^{-35} \qquad for\ m = 70$$

$$f(\hat{\beta}) = \frac{1}{0.0881}\left[1 + \left(\frac{\hat{\beta} - 0.375}{0.3307}\right)^2\right]^{-45} \qquad for\ m = 90$$

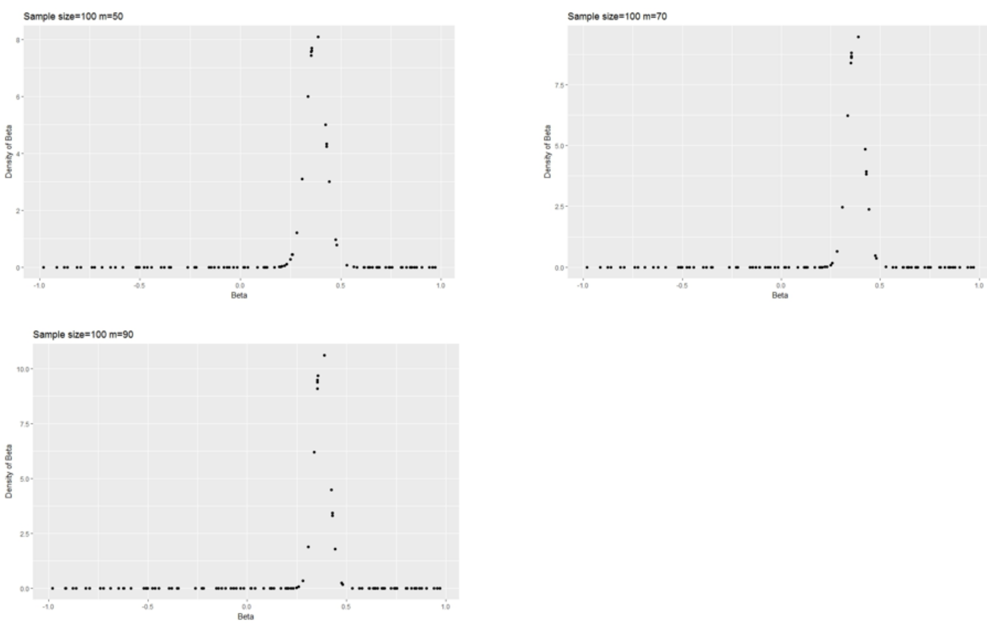

Figure 2-1 Density plots for $\hat{\beta}$ when m=50,70 and 90, respectively

2.2.4 Estimator of the Mean of Y

As we do for $\hat{\beta}$, we will derive the conditional expectation and variance of $\hat{\mu}_y$ given $X = \mathrm{x}$ first, then we derive its unconditional expectation and variance.

The conditional expectation of $\hat{\mu}_y$ given $X = \mathrm{x}$ is

$$\mathrm{E}(\hat{\mu}_y|x) = E\{[\bar{Y}_m - \hat{\beta}(\bar{x}_m - \bar{x}_n)]|x\} = E(\bar{Y}_m|x) - (\bar{x}_m - \bar{x}_n)E(\hat{\beta}|x)$$

$$= \mu_y + \beta(\bar{x}_m - \mu_x) - \beta(\bar{x}_m - \bar{x}_n)$$

$$= \mu_y + \beta(\bar{x}_n - \mu_x) \tag{2 - 44}$$

where

$$\mathrm{E}(\bar{Y}_m|x) = \mathrm{E}\left(\frac{1}{m}\sum_{j=1}^{m} Y_j\,|x_j\right) = \frac{1}{m}\sum_{j=1}^{m}\mathrm{E}(Y_j|x_j) = \frac{1}{m}\sum_{j=1}^{m}[\mu_y + \beta(x_j - \mu_x)]$$

$$= \frac{1}{m}\left[m\mu_y + \beta\sum_{j=1}^{m} x_j - m\beta\mu_x\right] = \mu_y + \beta(\bar{x}_m - \mu_x) \tag{2 - 45}$$

The unconditional expectation of $\hat{\mu}_y$ is

$$\mathrm{E}(\hat{\mu}_y) = \mathrm{E}(\mathrm{E}(\hat{\mu}_y|X)) = \mathrm{E}[\mu_y + \beta(\bar{X}_n - \mu_x)] = \mu_y + \beta[\mathrm{E}(\bar{X}_n) - \mu_x]$$

$$= \mu_y + \beta(\mu_x - \mu_x) = \mu_y \tag{2 - 46}$$

$\hat{\mu}_y$ is an unbiased estimator.

The conditional variance of $\hat{\mu}_y$ given $X = \mathrm{x}$ is

$$\mathrm{Var}(\hat{\mu}_y|x) = \mathrm{Var}\{[\bar{Y}_m - \hat{\beta}(\bar{x}_m - \bar{x}_n)]|x\}$$

$$= \mathrm{Var}(\bar{Y}_m|x) + \mathrm{Var}[(\bar{x}_m - \bar{x}_n)(\hat{\beta}|x)] - 2(\bar{x}_m - \bar{x}_n)\mathrm{Cov}[(\bar{Y}_m, \hat{\beta})|x]$$

$$= \mathrm{Var}(\bar{Y}_m|x) + (\bar{x}_m - \bar{x}_n)^2\mathrm{Var}[(\hat{\beta}|x)] - 0$$

$$= \frac{\sigma_e^2}{m} + \frac{\sigma_e^2(\bar{x}_m - \bar{x}_n)^2}{\sum_{j=1}^{m}(\mathrm{x_j} - \bar{x}_m)^2} \tag{2 - 47}$$

where

$$\mathrm{Var}(\bar{Y}_m|x) = \mathrm{Var}\left(\frac{1}{m}\sum_{j=1}^{m} Y_j|x_j\right) = \frac{1}{m^2}\sum_{j=1}^{m}\mathrm{Var}(Y_j|x_j) = \frac{1}{m^2}\cdot m\sigma_e^2 = \frac{\sigma_e^2}{m} \qquad (2 - 48)$$

$$\mathrm{Var}\left[(\bar{x}_m - \bar{x}_n)(\hat{\beta}|x)\right] = (\bar{x}_m - \bar{x}_n)^2\mathrm{Var}(\hat{\beta}|x) = \frac{\sigma_e^2(\bar{x}_m - \bar{x}_n)^2}{\sum_{j=1}^{m}\left(\mathrm{x_j} - \bar{x}_m\right)^2} \qquad (2 - 49)$$

$$\mathrm{Cov}\left[(\bar{Y}_m, \hat{\beta})\middle|x\right] = 0 \qquad (See\ Appendix\ C)$$

To obtain the unconditional variance of $\hat{\mu}_y$ we use the Law of Total Variance

$$\mathrm{Var}(\hat{\mu}_y) = \mathrm{E}\left[\mathrm{Var}(\hat{\mu}_y|X)\right] + \mathrm{Var}\left[\mathrm{E}(\hat{\mu}_y|X)\right]$$

now

$$\mathrm{Var}\left[\mathrm{E}(\hat{\mu}_y|X)\right] = \mathrm{Var}\left[\mu_y + \beta(\bar{X}_n - \mu_x)\right] = \beta^2\mathrm{Var}(\bar{X}_n) = \frac{\beta^2\sigma_x^2}{n} \qquad (2 - 50)$$

$$\mathrm{E}\left[\mathrm{Var}(\hat{\mu}_y|X)\right] = E\left(\frac{\sigma_e^2}{m} + \frac{\sigma_e^2(\bar{X}_m - \bar{X}_n)^2}{\sum_{j=1}^{m}\left(\mathrm{X_j} - \bar{X}_m\right)^2}\right) = \frac{\sigma_e^2}{m} + \sigma_e^2 E\left(\frac{(\bar{X}_m - \bar{X}_n)^2}{\sum_{j=1}^{m}\left(\mathrm{X_j} - \bar{X}_m\right)^2}\right)$$

To obtain $E\left(\frac{(\bar{X}_m-\bar{X}_n)^2}{\sum_{j=1}^{m}\left(\mathrm{X_j}-\bar{X}_m\right)^2}\right)$, we need to find out the distribution for $(\bar{X}_m - \bar{X}_n)$.

Since

$$\bar{X}_m - \bar{X}_n = \bar{X}_m - \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_m - \frac{1}{n}\left(\sum_{i=1}^{m} X_i + \sum_{j=m+1}^{n} X_j\right)$$

$$= \bar{X}_m - \frac{1}{n}[m\bar{X}_m + (n-m)\bar{X}_{n-m}] = \frac{n-m}{n}(\bar{X}_m - \bar{X}_{n-m}) \qquad (2 - 51)$$

$\bar{X}_m$ and $\bar{X}_{n-m}$ are independent and normally distributed, so we have

$$\mathrm{E}(\bar{X}_m - \bar{X}_{n-m}) = \mathrm{E}(\bar{X}_m) - \mathrm{E}(\bar{X}_{n-m}) = \mu_x - \mu_x = 0 \qquad (2 - 52)$$

$$\mathrm{Var}(\bar{X}_m - \bar{X}_{n-m}) = Var(\bar{X}_m) + Var(\bar{X}_{n-m}) = \frac{n}{m(n-m)}\sigma_x^2$$

$$= \frac{1}{m}\sigma_x^2 + \frac{1}{n-m}\sigma_x^2 = \frac{n}{m(n-m)}\sigma_x^2 \qquad (2 - 53)$$

Let

$$\mathrm{S}_{\mathrm{xm}}^2 = \frac{1}{m-1}\sum_{j=1}^{m}\left(X_j - \bar{X}_m\right)^2 \qquad (2 - 54)$$

It is known that $\bar{X}_m$ *and* $\mathrm{S}_{\mathrm{xm}}^2$ are independent, so $(\bar{X}_m - \bar{X}_{n-m})$ *and* $\mathrm{S}_{\mathrm{xm}}^2$ are independent too. Hence, we have

$$\mathrm{E}\left(\frac{(\bar{X}_m - \bar{X}_n)^2}{\sum_{j=1}^{m}\left(X_j - \bar{X}_m\right)^2}\right) = \left(\frac{n-m}{n}\right)^2 \mathrm{E}\left(\frac{(\bar{X}_m - \bar{X}_{n-m})^2}{(m-1)\mathrm{S}_{\mathrm{xm}}^2}\right)$$

$$= \left(\frac{n-m}{n}\right)^2 \cdot \frac{n}{m(n-m)} \cdot E\left\{\frac{\left[(\bar{X}_m - \bar{X}_{n-m})^2 \Big/ \frac{n}{m(n-m)}\sigma_x^2\right]}{\left[(m-1)\mathrm{S}_{\mathrm{xm}}^2 \big/ \sigma_x^2\right]}\right\}$$

where

$$U = \left[(\bar{X}_m - \bar{X}_{n-m})^2 \Big/ \frac{n}{m(n-m)}\sigma_x^2\right] \sim \chi^2(1)$$

$$V = \left[(m-1)\mathrm{S}_{\mathrm{xm}}^2 \big/ \sigma_x^2\right] \sim \chi^2(m-1)$$

Let

$$F = \frac{U/1}{V/(m-1)}$$

Then

$$F \sim F(1, m-1)$$

So

$$\mathrm{E}\left(\frac{(\bar{X}_m-\bar{X}_n)^2}{\sum_{j=1}^{m}\left(X_j-\bar{X}_m\right)^2}\right)=\frac{n-m}{mn}\cdot\frac{1}{m-1}\ \mathrm{E}(F)=\frac{n-m}{mn}\cdot\frac{1}{m-1}\cdot\frac{m-1}{m-3}$$

$$=\frac{n-m}{mn(m-3)}\qquad for\ m>3$$

Hence

$$\mathrm{E}\left(\mathrm{Var}(\hat{\mu}_y|X)\right)=\frac{\sigma_e^2}{m}+\frac{(n-m)\sigma_e^2}{mn(m-3)}=\frac{\sigma_e^2}{m}\left[1+\frac{n-m}{n(m-3)}\right]\qquad for\ m>3 \qquad (2 - 55)$$

By $(2-50)$ and $(2-55)$, the unconditional variance of $\hat{\mu}_y$ is

$$\mathrm{Var}(\hat{\mu}_y)=\mathrm{E}\left[\mathrm{Var}(\hat{\mu}_y|X)\right]+\mathrm{Var}\left[\mathrm{E}(\hat{\mu}_y|X)\right]$$

$$=\frac{\sigma_e^2}{m}\left[1+\frac{n-m}{n(m-3)}\right]+\frac{\beta^2\sigma_x^2}{n}\qquad for\ m>3 \qquad (2 - 56)$$

$\hat{\mu}_y$ has an asymptotically normal distribution when sample is large

$$\hat{\mu}_y\ \dot{\sim}\ N\left(\mu_y,\frac{\sigma_e^2}{m}\left[1+\frac{n-m}{n(m-3)}\right]+\frac{\beta^2\sigma_x^2}{n}\right) \qquad (2 - 57)$$

2.2.4.1 Comparison of the Variances

When we do not consider the extra information $X_{m+1}, X_{m+2}, \dots X_n$, and only use the first *m* observations, the unconditional variance of $\hat{\mu}_y$ is given in Appendix $(A-16)$

$$Var(\hat{\mu}_y)_{no}=\frac{\sigma_e^2}{m}+\frac{\beta^2\sigma_x^2}{m}=\frac{\sigma_y^2}{m}$$

From $(2-56)$, we have

$$\mathrm{Var}(\hat{\mu}_y)=\frac{\sigma_e^2}{m}\left[1+\frac{n-m}{n(m-3)}\right]+\frac{\beta^2\sigma_x^2}{n}+\frac{\beta^2\sigma_x^2}{m}-\frac{\beta^2\sigma_x^2}{m}$$

$$= \frac{\sigma_e^2}{m} + \frac{\beta^2 \sigma_x^2}{m} + \frac{n-m}{mn(m-3)} \sigma_e^2 - \frac{n-m}{mn} \beta^2 \sigma_x^2$$

$$= \frac{\sigma_y^2}{m} + \frac{n-m}{mn(m-3)} \sigma_e^2 - \frac{n-m}{mn} (\sigma_y^2 - \sigma_e^2)$$

$$= \frac{\sigma_y^2}{m} + \frac{n-m}{mn} \left[\left(1 + \frac{1}{m-3}\right) \sigma_e^2 - \sigma_y^2 \right]$$

$$= \frac{\sigma_y^2}{m} + \frac{n-m}{mn} \left[\left(1 + \frac{1}{m-3}\right) \sigma_y^2 (1-\rho^2) - \sigma_y^2 \right] \qquad since\ \sigma_e^2 = \sigma_y^2 (1-\rho^2)$$

$$= Var(\hat{\mu}_y)_{no} + \frac{n-m}{mn(m-3)} \sigma_y^2 [1 - (m-2)\rho^2]$$

$$= Var(\hat{\mu}_y)_{no} \left\{ 1 + \frac{n-m}{n(m-3)} [1 - (m-2)\rho^2] \right\} \tag{2 - 58}$$

Hence, if $1-(m-2)\rho^2 < 0$, then $Var(\hat{\mu}_y) < Var(\hat{\mu}_y)_{no}$, i.e.,

$$\rho^2 > \frac{1}{m-2} \tag{2 - 59}$$

Define

$$ratio = \frac{\mathrm{Var}(\hat{\mu}_y)}{Var(\hat{\mu}_y)_{no}} = 1 + \frac{n-m}{n(m-3)} [1 - (m-2)\rho^2]$$

We have the following two tables:

Table 2-1 Variance Ratio for Sample Size = 100

n	m	ρ^2	ratio		n	m	ρ^2	ratio		n	m	ρ^2	ratio
100	50	0.1	0.9596		100	70	0.1	0.9740		100	90	0.1	0.9910
		0.2	0.9085				0.2	0.9436				0.2	0.9809
		0.3	0.8574				0.3	0.9131				0.3	0.9708
		0.4	0.8064				0.4	0.8827				0.4	0.9607
		0.5	0.7553				0.5	0.8522				0.5	0.9506
		0.6	0.7043				0.6	0.8218				0.6	0.9405
		0.7	0.6532				0.7	0.7913				0.7	0.9303
		0.8	0.6021				0.8	0.7609				0.8	0.9202
		0.9	0.5511				0.9	0.7304				0.9	0.9101

Table 2-2 Variance Ratio for Sample Size = 200

n	m	ρ^2	ratio		n	m	ρ^2	ratio		n	m	ρ^2	ratio
		0.1	0.9546				0.1	0.9720				0.1	0.9905
		0.2	0.9041				0.2	0.9418				0.2	0.9805
		0.3	0.8536				0.3	0.9115				0.3	0.9704
		0.4	0.8031				0.4	0.8813				0.4	0.9603
200	100	0.5	0.7526		200	140	0.5	0.8511		200	180	0.5	0.9503
		0.6	0.7021				0.6	0.8209				0.6	0.9402
		0.7	0.6515				0.7	0.7907				0.7	0.9302
		0.8	0.6010				0.8	0.7604				0.8	0.9201
		0.9	0.5505				0.9	0.7302				0.9	0.9101

We can see the more correlated and the more missing, the smaller the ratio, i.e., $\mathrm{Var}(\hat{\mu}_y)$ is smaller than $Var(\hat{\mu}_y)_{no}$.

2.2.4.2 Simulation on Variances

From $(2-56)$, we have

$$Var(\hat{\mu}_y) = \frac{\sigma_e^2}{m}\left[1 + \frac{n-m}{n(m-3)}\right] + \frac{\beta^2\sigma_x^2}{n}$$

The estimated variance of $\hat{\mu}_y$ is

$$\widehat{Var}(\hat{\mu}_y) = \frac{\mathrm{S_e^2}}{m}\left[1 + \frac{n-m}{n(m-3)}\right] + \frac{\hat{\beta}^2\mathrm{S_{xn}^2}}{n} \qquad (2-60)$$

where

$$\mathrm{S_e^2} = \frac{1}{m-2}\sum_{j=1}^{m}[(Y_j - \bar{Y}_m) - \hat{\beta}(X_j - \bar{X}_m)]^2 \qquad (2-61)$$

and $\mathrm{S_{xn}^2}$ is given in $(2-28)$.

If we do not consider extra information $X_{m+1}, X_{m+2}, \dots, X_n$ and only use the first m observations,

$$Var(\hat{\mu}_y)_{no} = \frac{\sigma_e^2}{m} + \frac{\beta^2\sigma_x^2}{m}$$

The estimated variance is

$$\widehat{Var}(\hat{\mu}_y)_{no} = \frac{S_e^2}{m} + \frac{\hat{\beta}^2 S_{xm}^2}{m} \tag{2 - 62}$$

where S_{xm}^2 is given in $(2-54)$ and S_e^2 is given in $(2-61)$.

By R Studio and the following bivariate normal distributions - both meet $(2-59)$:

$$\begin{aligned} &\text{Low correlatio} \quad \begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2\left(\begin{pmatrix} 540 \\ 420 \end{pmatrix}, \begin{bmatrix} 970 & 600 \\ 600 & 3450 \end{bmatrix}\right) \quad (a) \\ &\text{High correlatio} \quad \begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2\left(\begin{pmatrix} 540 \\ 420 \end{pmatrix}, \begin{bmatrix} 970 & 1200 \\ 1200 & 3450 \end{bmatrix}\right) \quad (b) \end{aligned} \tag{2 - 63}$$

$$n = 40, m = 20, 28, 36$$

We have the following results by simulating 10,000 times:

Table 2-3 Simulation Results on Variance

n	m	Low Correlation				High Correlation			
		$\widehat{Var}(\hat{\mu}_y)$		$\widehat{Var}(\hat{\mu}_y)_{no}$		$\widehat{Var}(\hat{\mu}_y)$		$\widehat{Var}(\hat{\mu}_y)_{no}$	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
40	20(Miss 50%)	48.40	15.39	50.76	16.48	39.63	11.41	49.85	16.15
40	28(Miss 30%)	34.76	9.34	35.85	9.82	30.93	7.72	35.41	9.67
40	36(Miss 10%)	27.41	6.53	27.70	6.66	26.26	6.09	27.41	6.57

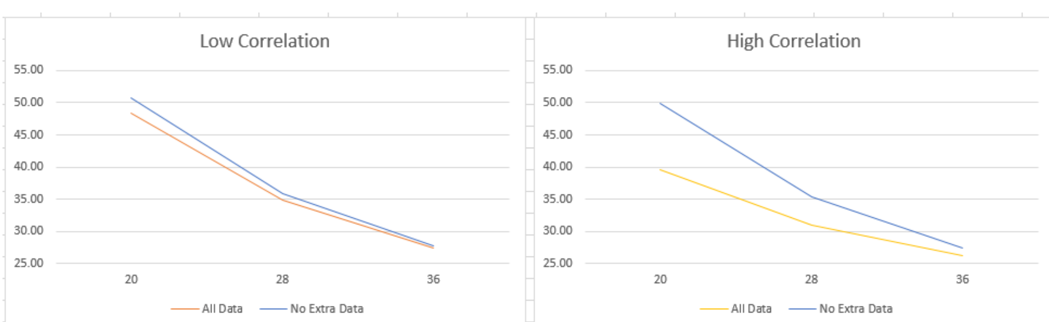

Figure 2-2 Comparison of Variance with and without Extra Information

The variance considering all information is smaller. That means the confidence interval is shorter with extra information than without extra information.

2.2.5 Estimator of the Conditional Variance of Y given x

Since

$$\hat{\sigma}_e^2 = \frac{1}{m}\sum_{j=1}^{m}[(Y_j - \bar{Y}_m) - \hat{\beta}(X_j - \bar{X}_m)]^2$$

does not involve extra information $X_{m+1}, X_{m+2}, \ldots, X_n$, it is same for both with and without extra information. So, we consider the situation for no extra information.

By $(2-4)$,

$$\mathrm{E}(Y_j|x_j) = \mu_y + \beta(x_j - \mu_x), \quad j = 1,2,\ldots,m \tag{2 - 64}$$

For given x_j, we may write $(2-64)$ as

$$Y_j = \mu_y + \beta(x_j - \mu_x) + \varepsilon_j, \quad j = 1,2,\ldots,m \tag{2 - 65}$$

where

$$\mathrm{Var}(Y_j|x_j) = \mathrm{Var}(\varepsilon_j) = \sigma_e^2$$

$$\mathrm{E}(\varepsilon_j) = \mathrm{E}(Y_j|x_j) - \mu_y - \beta(x_j - \mu_x) = 0$$

Rewrite $(2-65)$ as

$$Y_j = \beta_* + \beta(x_j - \bar{x}_m) + \varepsilon_j, \quad j = 1,2,\ldots,m \tag{2 - 66}$$

where

$$\beta_* = \mu_y + \beta(\bar{x}_m - \mu_x).$$

$(2-66)$ is the mean corrected form of the regression model.

Let

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} \qquad \boldsymbol{X}_c = \begin{bmatrix} 1 & x_1 - \bar{x}_m \\ 1 & x_2 - \bar{x}_m \\ & \vdots \\ 1 & x_m - \bar{x}_m \end{bmatrix} \qquad \boldsymbol{\beta_c} = \begin{bmatrix} \beta_* \\ \cdots \\ \beta \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

$(2-66)$ can be written as

$$\boldsymbol{Y} = \boldsymbol{X}_c\boldsymbol{\beta_c} + \boldsymbol{\varepsilon} \qquad (2 - 67)$$

By Results 7.1, Results 7.2 and Results 7,4 in Johnson and Wichern (1998), the least square estimators are

$$\widehat{\boldsymbol{\beta}}_c = (\boldsymbol{X}_c'\boldsymbol{X}_c)^{-\mathbf{1}}\boldsymbol{X}_c'\boldsymbol{Y} \qquad (2 - 68)$$

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = [\boldsymbol{I} - \boldsymbol{X}_c(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-\mathbf{1}}\boldsymbol{X}_c']\boldsymbol{Y} \qquad (2 - 69)$$

and

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} \sim \sigma_e^2\chi_{m-2}^2 \qquad (2 - 70)$$

From $(2-69)$, we have

$$\hat{\boldsymbol{\varepsilon}} = [\boldsymbol{I} - \boldsymbol{X}_c(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-\mathbf{1}}\boldsymbol{X}_c']\boldsymbol{Y}$$

$$= \left\{\boldsymbol{I} - [\mathbf{1} \vdots \boldsymbol{X}_{c2}]\begin{bmatrix}\frac{1}{m} & \mathbf{0}' \\ \mathbf{0} & \frac{1}{\sum_{j=1}^{m}\left(x_j - \bar{x}_m\right)^2}\end{bmatrix}\begin{bmatrix}\mathbf{1}' \\ \boldsymbol{X}_{c2}'\end{bmatrix}\right\}\boldsymbol{Y}$$

$$= \left\{\boldsymbol{I} - \frac{\mathbf{11}'}{\boldsymbol{m}} - \frac{\boldsymbol{X}_{c2}\boldsymbol{X}_{c2}'}{\sum_{j=1}^{m}\left(x_j - \bar{x}_m\right)^2}\right\}\boldsymbol{Y}$$

$$= \begin{bmatrix} Y_1 \\ Y_2 \\ \dots\dots \\ Y_m \end{bmatrix} - \begin{bmatrix} \bar{Y}_m \\ \bar{Y}_m \\ \dots\dots \\ \bar{Y}_m \end{bmatrix} - \hat{\beta}\begin{bmatrix} x_1 - \bar{x}_m \\ x_2 - \bar{x}_m \\ \dots\dots \\ x_m - \bar{x}_m \end{bmatrix}$$

$$= \begin{bmatrix} Y_1 - \bar{Y}_m - \hat{\beta}(x_1 - \bar{x}_m) \\ Y_2 - \bar{Y}_m - \hat{\beta}(x_2 - \bar{x}_m) \\ \dots \dots \\ Y_m - \bar{Y}_m - \hat{\beta}(x_m - \bar{x}_m) \end{bmatrix} \quad (2 - 71)$$

So,

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \sum_{j=1}^{m} \left[\left(Y_j - \bar{Y}_m\right) - \hat{\beta}\left(x_j - \bar{x}_m\right) \right]^2 | x \quad (2 - 72)$$

The conditional expectation of $\hat{\sigma}_e^2$ given $X = \mathrm{x}$ is

$$E(\hat{\sigma}_e^2 | x) = \frac{1}{m} E\left[\sum_{j=1}^{m} \left[\left(Y_j - \bar{Y}_m\right) - \hat{\beta}\left(X_j - \bar{X}_m\right) \right]^2 | x \right] = \frac{1}{m} E(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}})$$

$$= \frac{1}{m} \cdot (m-2)\sigma_e^2 = \frac{m-2}{m} \sigma_e^2 \quad (2 - 73)$$

Then the unconditional expectation of $\hat{\sigma}_e^2$ is

$$\mathrm{E}(\hat{\sigma}_e^2) = \mathrm{E}[\mathrm{E}(\hat{\sigma}_e^2 | X)] = \mathrm{E}[\frac{m-2}{m} \sigma_e^2] = \frac{m-2}{m} \sigma_e^2 \quad (2 - 74)$$

$\hat{\sigma}_e^2$ is a biased estimator. The bias of $\hat{\sigma}_e^2$ is

$$Bias(\hat{\sigma}_e^2) = E(\hat{\sigma}_e^2) - \sigma_e^2 = E\left(-\frac{2}{m}\sigma_e^2\right) = -\frac{2}{m}\sigma_e^2$$

The bias vanishes as $m \to \infty$.

The conditional variance of $\hat{\sigma}_e^2$ given $X = \mathrm{x}$ is

$$\mathrm{Var}(\hat{\sigma}_e^2 | x) = \frac{1}{m^2} \mathrm{Var}(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}) = \frac{1}{m^2} \cdot 2(m-2)\sigma_e^4 = \frac{2(m-2)}{m^2} \sigma_e^4 \quad (2 - 75)$$

By the Law of Total Variance, we have the unconditional variance of $\hat{\sigma}_e^2$

$$\mathrm{Var}(\hat{\sigma}_e^2) = \mathrm{E}\left(\mathrm{Var}(\hat{\sigma}_e^2 | X)\right) + \mathrm{Var}(\mathrm{E}(\hat{\sigma}_e^2 | X))$$

$$= \mathrm{E}\left[\frac{2(m-2)\sigma_e^4}{m^2}\right] + 0 = \frac{2(m-2)\sigma_e^4}{m^2} \qquad (2 - 76)$$

The mean squared error for $\hat{\sigma}_e^2$ is

$$MSE(\hat{\sigma}_e^2) = E(\hat{\sigma}_e^2 - \sigma_e^2)^2 = \mathrm{Var}(\hat{\sigma}_e^2) + [Bias(\hat{\sigma}_e^2)]^2 = \mathrm{Var}(\hat{\sigma}_e^2) + E\left(-\frac{2}{m}\sigma_e^2\right)^2$$

$$= \frac{2(m-2)\sigma_e^4}{m^2} + \frac{4\sigma_e^4}{m^2} = \frac{2\sigma_e^4}{m} \qquad (2 - 77)$$

Since the conditional distribution

$$\frac{m\hat{\sigma}_e^2}{\sigma_e^2}|x \sim \chi_{m-2}^2$$

does not depend on x, so the unconditional distribution of $\frac{m\hat{\sigma}_e^2}{\sigma_e^2}$ is also χ_{m-2}^2, i.e.,

$$\frac{m\hat{\sigma}_e^2}{\sigma_e^2} \sim \chi_{m-2}^2 \qquad (2 - 78)$$

2.3 Fisher Information Matrix

Upon taking the negative expectation of the second partial derivatives with respect to parameters in $(2-8)$ (See Appendix B), we obtain the following Fisher Information Matrix

$$\boldsymbol{I}(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2) = \begin{bmatrix} I_{\mu_y\mu_y} & I_{\mu_y\mu_x} & I_{\mu_y\beta} & I_{\mu_y\sigma_x^2} & I_{\mu_y\sigma_e^2} \\ I_{\mu_x\mu_y} & I_{\mu_x\mu_x} & I_{\mu_x\beta} & I_{\mu_x\sigma_x^2} & I_{\mu_x\sigma_e^2} \\ I_{\beta\mu_y} & I_{\beta\mu_x} & I_{\beta\beta} & I_{\beta\sigma_x^2} & I_{\beta\sigma_e^2} \\ I_{\sigma_x^2\mu_y} & I_{\sigma_x^2\mu_x} & I_{\sigma_x^2\beta} & I_{\sigma_{xx}\sigma_x^2} & I_{\sigma_x^2\sigma_e^2} \\ I_{\sigma_e^2\mu_y} & I_{\sigma_e^2\mu_x} & I_{\sigma_e^2\beta} & I_{\sigma_e^2\sigma_x^2} & I_{\sigma_e^2\sigma_e^2} \end{bmatrix}$$

$$
= \begin{bmatrix} \frac{m}{\sigma_e^2} & -\frac{m\beta}{\sigma_e^2} & 0 & 0 & 0 \\ -\frac{m\beta}{\sigma_e^2} & \frac{m\beta^2}{\sigma_e^2}+\frac{n}{\sigma_x^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{m\sigma_x^2}{\sigma_e^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{n}{2\sigma_x^4} & 0 \\ 0 & 0 & 0 & 0 & \frac{m}{2\sigma_e^4} \end{bmatrix} \tag{2 - 79}
$$

The inverse of the Fisher Information Matrix is

$$
\boldsymbol{I}^{-1}\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right) = \begin{bmatrix} \frac{\sigma_e^2}{m}+\frac{\beta^2\sigma_x^2}{n} & \frac{\beta\sigma_x^2}{n} & 0 & 0 & 0 \\ \frac{\beta\sigma_x^2}{n} & \frac{\sigma_x^2}{n} & 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma_e^2}{m\sigma_x^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{2\sigma_x^4}{n} & 0 \\ 0 & 0 & 0 & 0 & \frac{2\sigma_e^4}{m} \end{bmatrix} \tag{2 - 80}
$$

Denote the elements on the diagonal of $(2-80)$ as

$$\mathrm{Var}(\hat{\mu}_y)_{Fisher} = \frac{\sigma_e^2}{m}+\frac{\beta^2\sigma_x^2}{n}$$

$$\mathrm{Var}(\hat{\mu}_x)_{Fisher} = \frac{\sigma_x^2}{n}$$

$$\mathrm{Var}(\hat{\beta})_{Fisher} = \frac{\sigma_e^2}{m\sigma_x^2}$$

$$Var(\hat{\sigma}_x^2)_{Fisher} = \frac{2\sigma_x^4}{n}$$

$$\mathrm{Var}(\hat{\sigma}_e^2)_{Fisher} = \frac{2\sigma_e^4}{m}$$

respectively, we will compare them with the variance of each parameter.

1) $\mathrm{Var}(\hat{\mu}_y)$ vs $\mathrm{Var}(\hat{\mu}_y)_{Fisher}$

Since

$$\frac{\mathrm{Var}(\hat{\mu}_y)}{\mathrm{Var}(\hat{\mu}_y)_{Fisher}} = \frac{\frac{\sigma_e^2}{m}\left[1 + \frac{n-m}{n(m-3)}\right] + \frac{\beta^2\sigma_x^2}{n}}{\frac{\sigma_e^2}{m} + \frac{\beta^2\sigma_x^2}{n}} = 1 + \frac{\frac{1-\delta}{m-3}\sigma_e^2}{\sigma_e^2 + \delta\beta^2\sigma_x^2} \to 1 \; as \; m \to \infty$$

where $\delta = \frac{m}{n} = constant$. So $\hat{\mu}_\mathrm{y}$ is an asymptotically efficient estimator.

2) $\mathrm{Var}(\hat{\mu}_x)$ vs $\mathrm{Var}(\hat{\mu}_x)_{Fisher}$

Since

$$\frac{\mathrm{Var}(\hat{\mu}_x)}{\mathrm{Var}(\hat{\mu}_x)_{Fisher}} = \frac{\frac{\sigma_x^2}{n}}{\frac{\sigma_x^2}{n}} = 1$$

So $\hat{\mu}_\mathrm{x}$ is an efficient estimator.

3) $\mathrm{Var}(\hat{\beta})$ vs $\mathrm{Var}(\hat{\beta})_{Fisher}$

Since

$$\frac{\mathrm{Var}(\hat{\beta})}{\mathrm{Var}(\hat{\beta})_{Fisher}} = \frac{\frac{\sigma_e^2}{(m-3)\sigma_x^2}}{\frac{\sigma_e^2}{m\sigma_x^2}} = \frac{m}{m-3} \to 1 \; as \; m \to \infty$$

So $\hat{\beta}$ is an asymptotically efficient estimator.

4) $Var(\hat{\sigma}_x^2)$ vs $\left[\frac{dE(\hat{\sigma}_x^2)}{d\sigma_x^2}\right]^2 Var(\hat{\sigma}_x^2)_{Fisher}$

The expectation of $\hat{\sigma}_x^2$ is

$$E(\hat{\sigma}_x^2) = \frac{n-1}{n}\sigma_x^2$$

and its derivative on σ_x^2 is

$$\frac{dE(\hat{\sigma}_x^2)}{d\sigma_x^2} = \frac{n-1}{n}$$

We have

$$\frac{Var(\hat{\sigma}_x^2)}{\left[\frac{dE(\hat{\sigma}_x^2)}{d\sigma_x^2}\right]^2 Var(\hat{\sigma}_x^2)_{Fisher}} = \frac{\frac{2(n-1)}{n^2}\sigma_x^4}{\left(\frac{n-1}{n}\right)^2 \frac{2\sigma_x^4}{n}} = \frac{n}{n-1} \to 1 \; as \; n \to \infty$$

So, $\hat{\sigma}_x^2$ is an asymptotically efficient estimator.

5) $\mathrm{Var}(\hat{\sigma}_e^2)$ vs $\left[\frac{d\mathrm{E}(\hat{\sigma}_e^2)}{d\sigma_e^2}\right]^2 \mathrm{Var}(\hat{\sigma}_e^2)_{Fisher}$

Since the expectation of $\hat{\sigma}_e^2$ is

$$\mathrm{E}(\hat{\sigma}_e^2) = \frac{m-2}{m}\sigma_e^2$$

and its derivative on σ_e^2 is

$$\frac{d\mathrm{E}(\hat{\sigma}_e^2)}{d\sigma_e^2} = \frac{m-2}{m}$$

We have

$$\frac{\mathrm{Var}(\hat{\sigma}_e^2)}{\left[\frac{d\mathrm{E}(\hat{\sigma}_e^2)}{d\sigma_e^2}\right]^2 \mathrm{Var}(\hat{\sigma}_e^2)_{Fisher}} = \frac{\frac{2(m-2)\sigma_e^4}{m^2}}{\left(\frac{m-2}{m}\right)^2 \frac{2\sigma_e^4}{m}} = \frac{m}{m-2} \to 1 \; as \; m \to \infty$$

So, $\hat{\sigma}_e^2$ is an asymptotically efficient estimator.

2.4 Prediction

Suppose we have a future observation (X_0, Y_0) with a bivariate normal distribution

$$\begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} \sim N_2\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}\right) \quad (2 - 81)$$

We have the following three kinds of prediction interval for Y_0:

1) Usual prediction interval for Y_0 – conditioning on $X = x \text{ and } X_0 = x_0$

2) Prediction interval for Y_0 – unconditional on X, but conditioning on $X_0 = x_0$

3) Unconditional prediction interval for Y_0

2.4.1Usual prediction interval

– Conditioning on $X = x \text{ and } X_0 = x_0$

The prediction value of Y_0 given $\mathrm{X} = x \text{ and } \mathrm{X}_0 = x_0$ is

$$\hat{Y}_0|x, x_0 = \hat{\mu}_y + \hat{\beta}(x_0 - \bar{x}_n) \tag{2 - 82}$$

By $(2-32)$, $(2-35)$ and our assumption $(2-81)$, the distribution of $(\hat{\beta}|x)$ is

$$(\hat{\beta}|x) \sim N\left(\beta, \frac{\sigma_e^2}{\sum_{j=1}^{m}(x_j - \bar{x}_m)^2}\right) \tag{2 - 83}$$

By $(2-44)$, $(2-47)$ and our assumption $(2-81)$, the distribution of $(\hat{\mu}_y|x)$ is

$$(\hat{\mu}_y|x) \sim N\left(\mu_y + \beta(\bar{x}_n - \mu_x), \frac{\sigma_e^2}{m} + \frac{\sigma_e^2(\bar{x}_m - \bar{x}_n)^2}{\sum_{j=1}^{m}(\mathrm{x_j} - \bar{x}_m)^2}\right) \tag{2 - 84}$$

and $(\hat{\mu}_y, \hat{\beta})$ are independent of (X_0, Y_0).

So, by $(2-83)$ and $(2-84)$, the expectation of $\hat{Y}_0|x, x_0$ is

$$E(\hat{Y}_0|x, x_0) = E(\hat{\mu}_y|x, x_0) + E(\hat{\beta}|x, x_0)(x_0 - \bar{x}_n)$$

$$= E(\hat{\mu}_y|x) + E(\hat{\beta}|x)(x_0 - \bar{x}_n)$$

$$= \mu_y + \beta(\bar{x}_n - \mu_x) + \beta(x_0 - \bar{x}_n) = \mu_y + \beta(x_0 - \mu_x) \tag{2 - 85}$$

The variance of $\hat{Y}_0|x, x_0$ is

$$Var(\hat{Y}_0|x,x_0) = Var(\hat{\mu}_y|x) + (x_0 - \bar{x}_n)^2 Var(\hat{\beta}|x) + 2(x_0 - \bar{x}_n)Cov[(\hat{\mu}_y, \hat{\beta})|x]$$

$$= \frac{\sigma_e^2}{m} + \frac{\sigma_e^2(\bar{x}_m - \bar{x}_n)^2}{\sum_{j=1}^m (\mathrm{x_j} - \bar{x}_m)^2} + \frac{\sigma_e^2(x_0 - \bar{x}_n)^2}{\sum_{j=1}^m (x_j - \bar{x}_m)^2} - \frac{2\sigma_e^2(x_0 - \bar{x}_n)(\bar{x}_m - \bar{x}_n)}{\sum_{j=1}^m (x_j - \bar{x}_m)^2}$$

$$= \frac{\sigma_e^2}{m} + \frac{\sigma_e^2[(x_0 - \bar{x}_n) - (\bar{x}_m - \bar{x}_n)]^2}{\sum_{j=1}^m (\mathrm{x_j} - \bar{x}_m)^2}$$

$$= \frac{\sigma_e^2}{m} + \frac{\sigma_e^2(x_0 - \bar{x}_m)^2}{\sum_{j=1}^m (\mathrm{x_j} - \bar{x}_m)^2} \qquad (2 - 86)$$

where

$$Cov[(\hat{\mu}_y, \hat{\beta})|x] = Cov[(\bar{Y}_m - \hat{\beta}(\bar{x}_m - \bar{x}_n), \hat{\beta})|x]$$

$$= Cov(\bar{Y}_m, \hat{\beta})|x - (\bar{x}_m - \bar{x}_n)Cov(\hat{\beta}, \hat{\beta})|x$$

$$= -\frac{\sigma_e^2(\bar{x}_m - \bar{x}_n)}{\sum_{j=1}^m (x_j - \bar{x}_m)^2} \qquad (2 - 87)$$

By assumption, we have

$$Y_0|x, x_0 \sim N(\mu_y + \beta(x_0 - \mu_x), \sigma_e^2) \qquad (2 - 88)$$

By $(2-85)$ and $(2-88)$, the expectation of $(Y_0 - \hat{Y}_0)|x, x_0$ is

$$E(Y_0 - \hat{Y}_0)|x, x_0 = E(Y_0|x, x_0) - E(\hat{Y}_0|x, x_0)$$

$$= \mu_y + \beta(x_0 - \mu_x) - [\mu_y + \beta(x_0 - \mu_x)] = 0 \qquad (2 - 89)$$

By $(2-86)$ and $(2-88)$, the variance of $(Y_0 - \hat{Y}_0)|x, x_0$ is

$$Var(Y_0 - \hat{Y}_0)|x, x_0 = Var(Y_0)|x, x_0 + Var(\hat{Y}_0)|x, x_0 - 2Cov(\hat{Y}_0, Y_0)|x, x_0$$

$$= \sigma_e^2 + \frac{\sigma_e^2}{m} + \frac{\sigma_e^2(x_0 - \bar{x}_m)^2}{\sum_{j=1}^m (\mathrm{x_j} - \bar{x}_m)^2} - 2 \cdot 0$$

$$= \sigma_e^2 \left[1 + \frac{1}{m} + \frac{(x_0 - \bar{x}_m)^2}{\sum_{j=1}^m (\mathrm{x_j} - \bar{x}_m)^2}\right] \qquad (2 - 90)$$

where

$$\mathrm{Cov}\left(\hat{Y}_0, Y_0\right)\big|x, x_0 = \mathrm{Cov}\left[\hat{\mu}_y + \hat{\beta}(x_0 - \bar{x}_n), Y_0\right]\big|x, x_0 = 0$$

Since Y_0 *and* $\hat{Y}_0$ are normal, so $Y_0 - \hat{Y}_0$ is normal, then

$$Z = \frac{\left(Y_0 - \hat{Y}_0\right)|x, x_0 - E\left(Y_0 - \hat{Y}_0\right)|x, x_0}{\sqrt{\sigma_e^2\left[1 + \frac{1}{m} + \frac{(x_0 - \bar{x}_m)^2}{\sum_{j=1}^{m}\left(x_j - \bar{x}_m\right)^2}\right]}} \sim N(0,1)$$

By $(2 - 78)$,

$$U = \frac{m\hat{\sigma}_e^2}{\sigma_e^2} \sim \chi_{m-2}^2$$

and Z *and* U are independent, so

$$T = \frac{Z}{\sqrt{U/(m-2)}} \sim t_{m-2}$$

i.e.,

$$T = \frac{\dfrac{\left(Y_0 - \hat{Y}_0\right)|x, x_0 - 0}{\sqrt{\sigma_e^2\left[1 + \frac{1}{m} + \frac{(x_0 - \bar{x}_m)^2}{\sum_{j=1}^{m}\left(\mathrm{x_j} - \bar{x}_m\right)^2}\right]}}}{\sqrt{\dfrac{m\hat{\sigma}_e^2}{(m-2)\sigma_e^2}}} = \frac{\left(Y_0 - \hat{Y}_0\right)|x, x_0}{\sqrt{\frac{m\hat{\sigma}_e^2}{m-2}\left[1 + \frac{1}{m} + \frac{(x_0 - \bar{x}_m)^2}{\sum_{j=1}^{m}\left(\mathrm{x_j} - \bar{x}_m\right)^2}\right]}} \sim t_{m-2}$$

The 95% prediction interval for Y_0 $given\ X = x\ and\ \ X_0 = x_0$ is

$$\hat{Y}_0|x, x_0 \pm t_{0.025,m-2}\sqrt{\frac{m\hat{\sigma}_e^2}{m-2}\left[1 + \frac{1}{m} + \frac{(x_0 - \bar{x}_m)^2}{\sum_{j=1}^{m}\left(\mathrm{x_j} - \bar{x}_m\right)^2}\right]} \qquad (2 - 91)$$

2.4.2 Prediction interval

– Unconditional on X, but conditioning on $X_0 = x_0$

The prediction value of Y_0 given $X_0 = x_0$ is

$$\hat{Y}_0|x_0 = \hat{\mu}_y + \hat{\beta}(x_0 - \hat{\mu}_x) \tag{2 - 92}$$

By $(2-34)$ and $(2-46)$, the expectation of $\hat{Y}_0|x_0$ is

$$E(\hat{Y}_0|x_0) = E(\hat{\mu}_y|x_0) + E[\hat{\beta}(x_0 - \hat{\mu}_x)|x_0]$$

$$= E(\hat{\mu}_y) + x_0 E(\hat{\beta}) - E(\hat{\beta}\hat{\mu}_x)$$

$$= \mu_y + \beta(x_0 - \mu_x) \tag{2 - 93}$$

where

$$E(\hat{\beta}\hat{\mu}_x) = E[E(\hat{\beta}\hat{\mu}_x)|x] = E[\hat{\mu}_x E(\hat{\beta})|x] = \beta E(\hat{\mu}_x) = \beta\mu_x \tag{2 - 94}$$

$\hat{\mu}_x$ and $\hat{\beta}$ are not correlated, and by assumptions $(\hat{\mu}_x, \hat{\beta})$ are independent of (X_0, Y_0).

We derive the variance of $\hat{Y}_0|x_0$ by the Δ-Method. Let

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = \begin{bmatrix} \hat{\mu}_y \\ \hat{\mu}_x \\ \hat{\beta} \end{bmatrix} \tag{2 - 95}$$

$\boldsymbol{Z}$ *and* X_0 are independent. The expectation and covariance of $\boldsymbol{Z}$ given $X_0 = x_0$ are

$$\mathrm{E}(\boldsymbol{Z}|x_0) = \mathrm{E}(\boldsymbol{Z}) = \boldsymbol{\mu}_Z = \begin{bmatrix} E(Z_1) \\ E(Z_2) \\ E(Z_3) \end{bmatrix} = \begin{bmatrix} E(\hat{\mu}_y) \\ E(\hat{\mu}_x) \\ E(\hat{\beta}) \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_x \\ \beta \end{bmatrix} \tag{2 - 96}$$

and

$$\mathrm{Cov}(\boldsymbol{Z}|x_0) = \mathrm{Cov}(\boldsymbol{Z}) = E[\boldsymbol{Z} - \mathrm{E}(\boldsymbol{Z})][\boldsymbol{Z} - \mathrm{E}(\boldsymbol{Z})]' = \begin{bmatrix} Var(Z_1) & Cov(Z_1, Z_2) & Cov(Z_1, Z_3) \\ Cov(Z_1, Z_2) & Var(Z_2) & Cov(Z_2, Z_3) \\ Cov(Z_1, Z_3) & Cov(Z_2, Z_3) & Var(Z_3) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma_e^2}{m}\left[1 + \frac{n-m}{n(m-3)}\right] + \frac{\beta^2\sigma_x^2}{n} & \frac{\beta\sigma_x^2}{n} & 0 \\ \frac{\beta\sigma_x^2}{n} & \frac{\sigma_x^2}{n} & 0 \\ 0 & 0 & \frac{\sigma_e^2}{(m-3)\sigma_x^2} \end{bmatrix} \tag{2 - 97}$$

where, by $(2-56)$, we have

$$Var(Z_1) = \mathrm{Var}(\hat{\mu}_y) = \frac{\sigma_e^2}{m}\left[1 + \frac{n-m}{n(m-3)}\right] + \frac{\beta^2\sigma_x^2}{n}$$

by $(2-25)$,

$$Var(Z_2) = \mathrm{Var}(\hat{\mu}_x) = \frac{\sigma_x^2}{n}$$

by $(2-36)$,

$$Var(Z_3) = \mathrm{Var}(\hat{\beta}) = \frac{\sigma_e^2}{(m-3)\sigma_x^2}$$

By Appendix C, we have

$$Cov(Z_1, Z_2) = Cov(\hat{\mu}_y, \hat{\mu}_x) = \frac{1}{n}\beta\sigma_x^2$$

$$\mathrm{Cov}(Z_1, Z_3) = \mathrm{Cov}(\hat{\mu}_y, \hat{\beta}) = 0$$

$$\mathrm{Cov}(Z_2, Z_3) = \mathrm{Cov}(\hat{\mu}_x, \hat{\beta}) = 0$$

In terms of $\boldsymbol{z}$, we have

$$\hat{y}_0|x_0 = z_1 + z_3\,(x_o - z_2\,) \tag{2 - 98}$$

By the Δ-method,

$$\hat{y}_0|x_0 \approx \mu_y + \beta(x_0 - \mu_x) + \sum_{j=1}^{3}(\hat{y}'_{0j}|x_0)(\,z_j - \mu_{zj}) \tag{2 - 99}$$

where

$$\hat{y}'{}_0|x_0 = \begin{bmatrix} \hat{y}'_{01}|x_0 \\ \hat{y}'_{02}|x_0 \\ \hat{y}'_{03}|x_0 \end{bmatrix} = \begin{bmatrix} \dfrac{\partial y_0}{\partial z_1} \\ \dfrac{\partial y_0}{\partial z_2} \\ \dfrac{\partial y_0}{\partial z_3} \end{bmatrix}_{z=\boldsymbol{\mu}_z} = \begin{bmatrix} 1 \\ -\beta \\ x_0 - \mu_x \end{bmatrix} \tag{2 - 100}$$

Hence, the expectation of $\hat{Y}_0|x_0$ is

$$E(\hat{Y}_0|x_0) \approx \mu_y + \beta(x_0 - \mu_x) + \sum_{j=1}^{3}(\hat{y}'_{0j}|x_0)E(Z_j - \mu_{Zj}) = \mu_y + \beta(x_0 - \mu_x) \qquad (2 - 101)$$

The variance of $\hat{Y}_0|x_0$ is

$$\mathrm{Var}(\hat{Y}_0|x_0) = E\left[\hat{Y}_0|x_0 - E(\hat{Y}_0|x_0)\right]^2 \approx E\left[\sum_{j=1}^{3}(\hat{y}'_{0j}|x_0)(Z_j - \mu_{Zj})\right]^2$$

$$= \sum_{j=1}^{3}\left[(\hat{y}'_{0j}|x_0)\right]^2 Var(Z_j) + 2\sum_{i=1}^{3}\sum_{j=1,i\neq j}^{3}(\hat{y}'_{0i}|x_0)(\hat{y}'_{0j}|x_0)Cov(Z_i, Z_j)$$

$$= 1^2 Var(Z_1) + (-\beta)^2 Var(Z_2) + (x_0 - \mu_x)^2 Var(Z_3) + 2 \cdot 1 \cdot (-\beta)Cov(Z_1, Z_2)$$

$$= 1^2 Var(\hat{\mu}_y) + (-\beta)^2 Var(\hat{\mu}_x) + (x_0 - \mu_x)^2 Var(\hat{\beta}) + 2 \cdot 1 \cdot (-\beta)Cov(\hat{\mu}_y, \hat{\mu}_x)$$

$$= \frac{\sigma_e^2}{m}\left[1 + \frac{n-m}{n(m-3)}\right] + \frac{\beta^2\sigma_x^2}{n} + \frac{\beta^2\sigma_x^2}{n} + (x_0 - \mu_x)^2\frac{\sigma_e^2}{(m-3)\sigma_{xx}} - 2\frac{\beta^2\sigma_x^2}{n}$$

$$= \sigma_e^2\left[\frac{1}{m} + \frac{n-m}{mn(m-3)} + \frac{(x_0 - \mu_x)^2}{(m-3)\sigma_x^2}\right] \qquad (2 - 102)$$

By assumptions,

$$Y_0|x_0 \sim N\left(\mu_y + \beta(x_0 - \mu_x), \sigma_e^2\right) \qquad (2 - 103)$$

So, the expectation of $(Y_0 - \hat{Y}_0)|x_0$ is

$$E(Y_0 - \hat{Y}_0)|x_0 = E(Y_0|x_0) - E(\hat{Y}_0|x_0)$$

$$= \mu_y + \beta(x_0 - \mu_x) - \left[\mu_y + \beta(x_0 - \mu_x)\right] = 0 \qquad (2 - 104)$$

The variance of $\left(Y_0 - \hat{Y}_0\right)|x_0$ is

$$Var\left(Y_0 - \hat{Y}_0\right)|x_0 = Var(Y_0)|x_0 + Var\left(\hat{Y}_0\right)|x_0 - 2Cov(\hat{Y}_0, Y_0)|x_0$$

$$= \sigma_e^2 + \sigma_e^2\left[\frac{1}{m} + \frac{n-m}{mn(m-3)} + \frac{(x_0-\mu_x)^2}{(m-3)\sigma_x^2}\right] - 2\cdot 0$$

$$= \sigma_e^2\left[1 + \frac{1}{m} + \frac{n-m}{mn(m-3)} + \frac{(x_0-\mu_x)^2}{(m-3)\sigma_x^2}\right] \qquad (2 - 105)$$

where

$$\mathrm{Cov}\left(\hat{Y}_0, Y_0\right)\big|x_0 = \mathrm{Cov}\left[\hat{\mu}_y + \hat{\beta}(x_0 - \hat{\mu}_x), Y_0\right]\big|x_0 = 0$$

When sample is large,

$$Z = \frac{\left(Y_0 - \hat{Y}_0\right)|x_0 - E\left(Y_0 - \hat{Y}_0\right)|x_0}{\sqrt{\sigma_e^2\left[1 + \frac{1}{m} + \frac{n-m}{mn(m-3)} + \frac{(x_0-\mu_x)^2}{(m-3)\sigma_x^2}\right]}} \dot{\sim} N(0,1)$$

The 95% prediction interval for Y_0 given $X_0 = x_0$ is

$$\hat{Y}_0|x_0 \pm z_{0.025}\sqrt{\mathrm{S}_e^2\left[1 + \frac{1}{m} + \frac{n-m}{mn(m-3)} + \frac{(x_0-\bar{X}_n)^2}{(m-3)\mathrm{S}_{\mathrm{xn}}^2}\right]} \qquad (2 - 106)$$

where $\mathrm{S}_{\mathrm{xn}}^2$ and S_e^2 are given in $(2-28)$ and $(2-61)$, respectively.

2.4.3 Unconditional prediction interval

By assumptions we have

$$Y_0 \sim N\left(\mu_y, \sigma_y^2\right) \qquad (2 - 107)$$

$$X_0 \sim N(\mu_x, \sigma_x^2) \qquad (2 - 108)$$

The prediction value of Y_0 is

$$\hat{Y}_0 = \hat{\mu}_y + \hat{\beta}(X_0 - \hat{\mu}_x) \tag{2 - 109}$$

By $(2-34)$ and $(2-46)$, the expectation of $\hat{Y}_0$ is

$$E(\hat{Y}_0) = E(\hat{\mu}_y) + E\left[\hat{\beta}(X_0 - \hat{\mu}_x)\right] = \mu_y \tag{2 - 110}$$

where

$$E\left[\hat{\beta}(X_0 - \hat{\mu}_x)\right] = E(\hat{\beta}X_0) - E(\hat{\beta}\hat{\mu}_x) = E(\hat{\beta})E(X_0) - E(\hat{\beta})E(\hat{\mu}_x) = \beta\mu_x - \beta\mu_x = 0$$

$\hat{\mu}_x$ and $\hat{\beta}$ are not correlated by $(2-94)$, and by assumptions $(\hat{\mu}_x, \hat{\beta})$ are independent of (X_0, Y_0).

We derive the variance of $\hat{Y}_0$ by the Δ-Method. Let

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix} = \begin{bmatrix} \hat{\mu}_y \\ \hat{\mu}_x \\ \hat{\beta} \\ X_0 \end{bmatrix} \tag{2 - 111}$$

The expectation and covariance of $\boldsymbol{Z}$ are

$$\mathrm{E}(\boldsymbol{Z}) = \boldsymbol{\mu}_Z = \begin{bmatrix} E(Z_1) \\ E(Z_2) \\ E(Z_3) \\ E(Z_4) \end{bmatrix} = \begin{bmatrix} E(\hat{\mu}_y) \\ E(\hat{\mu}_x) \\ E(\hat{\beta}) \\ E(X_0) \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_x \\ \beta \\ \mu_x \end{bmatrix} \tag{2 - 112}$$

and

$$\mathrm{Cov}(\boldsymbol{Z}) = E[\boldsymbol{Z} - \mathrm{E}(\boldsymbol{Z})][\boldsymbol{Z} - \mathrm{E}(\boldsymbol{Z})]' = \begin{bmatrix} Var(Z_1) & Cov(Z_1, Z_2) & Cov(Z_1, Z_3) & Cov(Z_1, Z_4) \\ Cov(Z_1, Z_2) & Var(Z_2) & Cov(Z_2, Z_3) & Cov(Z_2, Z_4) \\ Cov(Z_1, Z_3) & Cov(Z_2, Z_3) & Var(Z_3) & Cov(Z_3, Z_4) \\ Cov(Z_1, Z_4) & Cov(Z_2, Z_4) & Cov(Z_3, Z_4) & Var(Z_4) \end{bmatrix}$$

$$
= \begin{bmatrix} \frac{\sigma_e^2}{m}\left[1+\frac{n-m}{n(m-3)}\right]+\frac{\beta^2\sigma_x^2}{n} & \frac{\beta\sigma_x^2}{n} & 0 & 0 \\ \frac{\beta\sigma_x^2}{n} & \frac{\sigma_x^2}{n} & 0 & 0 \\ 0 & 0 & \frac{\sigma_e^2}{(m-3)\sigma_x^2} & 0 \\ 0 & 0 & 0 & \sigma_x^2 \end{bmatrix} \quad (2 - 113)
$$

In terms of $\boldsymbol{z}$, we have

$$
\hat{y}_0 = z_1 + z_3\,(z_4 - z_2) \quad (2 - 114)
$$

By the Δ-method,

$$
\hat{y}_0 \approx \mu_y + \sum_{j=1}^{4} \hat{y}'_{0j}(\boldsymbol{\mu}_Z)(z_j - \mu_{Zj}) \quad (2 - 115)
$$

where

$$
\hat{y}'_0(\boldsymbol{\mu}_Z) = \begin{bmatrix} \hat{y}'_{01}(\boldsymbol{\mu}_Z) \\ \hat{y}'_{02}(\boldsymbol{\mu}_Z) \\ \hat{y}'_{03}(\boldsymbol{\mu}_Z) \\ \hat{y}'_{04}(\boldsymbol{\mu}_Z) \end{bmatrix} = \begin{bmatrix} \frac{\partial y_0}{\partial z_1} \\ \frac{\partial y_0}{\partial z_2} \\ \frac{\partial y_0}{\partial z_3} \\ \frac{\partial y_0}{\partial z_4} \end{bmatrix}_{z=\boldsymbol{\mu}_Z} = \begin{bmatrix} 1 \\ -\beta \\ 0 \\ \beta \end{bmatrix} \quad (2 - 116)
$$

so

$$
E(\hat{Y}_0) \approx \mu_y + \sum_{j=1}^{4} \hat{y}'_{0j}(\boldsymbol{\mu}_Z)E(Z_j - \mu_{Zj}) = \mu_y \quad (2 - 117)
$$

The variance of $\hat{Y}_0$ is

$$
\mathrm{Var}(\hat{Y}_0) = \mathrm{E}\left[\hat{Y}_0 - E(\hat{Y}_0)\right]^2 \approx \mathrm{E}\left[\sum_{j=1}^{4} \hat{y}'_{0j}(\boldsymbol{\mu}_Z)E(Z_j - \mu_{Zj})\right]^2
$$

$$
= \sum_{j=1}^{4}\left[\hat{y}'_{0j}(\boldsymbol{\mu}_Z)\right]^2 Var(Z_j) + 2\sum_{i=1}^{4}\sum_{j=1, i\neq j}^{4} \hat{y}'_{0i}(\boldsymbol{\mu}_Z)\hat{y}'_{0j}(\boldsymbol{\mu}_Z)Cov(Z_i, Z_j)
$$

$$= 1^2 Var(Z_1) + (-\beta)^2 Var(Z_2) + 0^2 \cdot Var(Z_3) + \beta^2 Var(Z_4) + 2 \cdot 1 \cdot (-\beta) Cov(Z_1, Z_2)$$

$$= 1^2 Var(\hat{\mu}_y) + (-\beta)^2 Var(\hat{\mu}_x) + 0^2 \cdot Var(\hat{\beta}) + \beta^2 Var(X_0) + 2 \cdot 1 \cdot (-\beta) Cov(\hat{\mu}_y, \hat{\mu}_x)$$

$$= \frac{\sigma_e^2}{m}\left[1 + \frac{n-m}{n(m-3)}\right] + \frac{\beta^2\sigma_x^2}{n} + \frac{\beta^2\sigma_x^2}{n} + 0 + \beta^2\sigma_x^2 - 2\frac{\beta^2\sigma_x^2}{n}$$

$$= \frac{\sigma_e^2}{m}\left[1 + \frac{n-m}{n(m-3)}\right] + \beta^2\sigma_x^2 \qquad (2 - 118)$$

The expectation and variance of $(Y_0 - \hat{Y}_0)$ are

$$E(Y_0 - \hat{Y}_0) = E(Y_0) - E(\hat{Y}_0) = \mu_y - \mu_y = 0 \qquad (2 - 119)$$

$$Var(Y_0 - \hat{Y}_0) = Var(Y_0) + Var(\hat{Y}_0) - 2Cov(\hat{Y}_0, Y_0)$$

$$= \sigma_y^2 + \frac{\sigma_e^2}{m}\left[1 + \frac{n-m}{n(m-3)}\right] + \beta^2\sigma_x^2 - 2 \cdot \beta^2\sigma_x^2$$

$$= \sigma_y^2 + \frac{\sigma_e^2}{m}\left[1 + \frac{n-m}{n(m-3)}\right] - \beta^2\sigma_x^2$$

$$= \sigma_e^2\left[1 + \frac{1}{m} + \frac{n-m}{mn(m-3)}\right] \qquad (2 - 120)$$

where

$$\mathrm{Cov}(\hat{Y}_0, Y_0) = \mathrm{Cov}[\hat{\mu}_y + \hat{\beta}(\mathrm{X}_0 - \hat{\mu}_x), Y_0]$$

$$= \mathrm{Cov}(\hat{\mu}_y, Y_0) + \mathrm{Cov}(\hat{\beta}\mathrm{X}_0, Y_0) - \mathrm{Cov}(\hat{\beta}\hat{\mu}_x, Y_0)$$

$$= 0 + E(\hat{\beta}\mathrm{X}_0 Y_0) - E(\hat{\beta}\mathrm{X}_0)E(Y_0) - 0$$

$$= E(\hat{\beta})\mathrm{E}(\mathrm{X}_0 Y_0) - E(\hat{\beta})E(\mathrm{X}_0)E(Y_0)$$

$$= \beta[E(\mathrm{X}_0 Y_0) - E(\mathrm{X}_0)E(Y_0)] = \beta\sigma_{xy} = \beta^2\sigma_x^2 \qquad (2 - 121)$$

When sample is large,

$$Z = \frac{(Y_0 - \hat{Y}_0) - E(Y_0 - \hat{Y}_0)}{\sqrt{\sigma_e^2\left[1 + \frac{1}{m} + \frac{n-m}{mn(m-3)}\right]}} \dot{\sim} N(0,1)$$

The 95% prediction interval for Y_0 is

$$\hat{Y}_0 \pm z_{0.025}\sqrt{S_e^2\left[1+\frac{1}{m}+\frac{n-m}{mn(m-3)}\right]} \qquad (2 - 122)$$

where S_e^2 is given in $(2-61)$.

This is unconditional (not dependent on X_0) prediction interval for Y_0 of a future observation.

2.5 An Example for the Bivariate Situation

College Admission data (Table 1-1) is from Han and Li (2011). In this example, we take TOEFL score as Y and GRE Verbal, Quantitative and Analytic as X, respectively to estimate above 5 maximum likelihood estimators. Normality test shows that Y and X are normally distributed.

If we consider all information $X_1, X_2, \dots, X_n$, the five estimators are

Table 2-4 Estimators with All Information

Considering the regression of Y on GRE Verbal					
X	$\hat{\mu}_x$	$\hat{\mu}_y$	$\hat{\beta}$	$\hat{\sigma}_x^2$	$\hat{\sigma}_e^2$
Verbal	419.50	552.44	0.1727	11614.75	826.19
Considering the regression of Y on GRE Quantitative					
X	$\hat{\mu}_x$	$\hat{\mu}_y$	$\hat{\beta}$	$\hat{\sigma}_x^2$	$\hat{\sigma}_e^2$
Quantitative	646.50	540.98	-0.0301	11707.75	918.78
Considering the regression of Y on GRE Analytic					
X	$\hat{\mu}_x$	$\hat{\mu}_y$	$\hat{\beta}$	$\hat{\sigma}_x^2$	$\hat{\sigma}_e^2$
Analytic	523.00	541.62	0.0471	12239.75	898.47

If we do not consider extra information $X_{m+1}, X_{m+2}, \dots, X_n$ and only consider first m observations, the five estimators are

Table 2-5 Estimators without Extra Information

Considering the regression of Y on GRE Verbal					
X	$\hat{\mu}_{x\_no}$	$\hat{\mu}_{y\_no}$	$\hat{\beta}_{no}$	$\hat{\sigma}^2_{x\_no}$	$\hat{\sigma}^2_{e\_no}$
Verbal	342.00	539.05	0.1727	3276.00	826.19
Considering the regression of Y on GRE Quantitative					
X	$\hat{\mu}_{x\_no}$	$\hat{\mu}_{y\_no}$	$\hat{\beta}_{no}$	$\hat{\sigma}^2_{x\_no}$	$\hat{\sigma}^2_{e\_no}$
Quantitative	710.50	539.05	-0.0301	5694.75	918.78
Considering the regression of Y on GRE Analytic					
X	$\hat{\mu}_{x\_no}$	$\hat{\mu}_{y\_no}$	$\hat{\beta}_{no}$	$\hat{\sigma}^2_{x\_no}$	$\hat{\sigma}^2_{e\_no}$
Analytic	468.50	539.05	0.0471	11500.25	898.47

Chapter 3

STATISTICAL ESTIMATION IN MULTIPLE REGRESSION MODEL

WITH A BLOCK OF MISSING OBSERVATIONS

Let $\begin{bmatrix} \boldsymbol{X} \\ Y \end{bmatrix} = [X_1, X_2, \dots, X_p, Y]^T$ have a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu_x} \\ \mu_y \end{bmatrix} = [\mu_{x1}, \mu_{x2}, \dots, \mu_{xp}, \mu_y]^T, \ \ \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma_{xx}} & \boldsymbol{\Sigma_{xy}} \\ \boldsymbol{\Sigma_{yx}} & \sigma_y^2 \end{bmatrix}$$

Suppose we have the following random sample with a block of missing Y values:

$$\begin{matrix} X_{1,1} & X_{1,2} & \cdots & X_{1,p} & Y_1 \\ X_{2,1} & X_{2,2} & \cdots & X_{2,p} & Y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{m,1} & X_{m,2} & \cdots & X_{m,p} & Y_m \\ X_{m+1,1} & X_{m+1,2} & \cdots & X_{m+1,p} & \\ \vdots & \vdots & \vdots & \vdots & \\ X_{n,1} & X_{n,1} & \cdots & X_{n,p} & \end{matrix}$$

Based on the data, we want to estimate the parameters. We can write the multivariate normal probability density function (pdf) as

$$f(\mathbf{x}, y) = g(y|\mathbf{x})h(\mathbf{x}) \qquad (3 - 1)$$

where $g(y|\mathbf{x})$ is the conditional pdf of Y given $\boldsymbol{X} = \mathbf{x}$, and $h(\mathbf{x})$ is the marginal pdf of $\boldsymbol{X}$.

$$g_{Y|X}(\mathrm{y_j}|\mathbf{x_j}; \mu_\mathrm{y}, \boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\beta}, \sigma_\mathrm{e}^2) = \frac{1}{\sqrt{2\pi}\sigma_\mathrm{e}} \exp\left\{-\frac{1}{2\sigma_\mathrm{e}^2}\left[\mathrm{y_j} - E(\mathrm{y_j}|\mathbf{x_j})\right]^2\right\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_\mathrm{e}} \exp\left\{-\frac{1}{2\sigma_\mathrm{e}^2}\left[\mathrm{y_j} - \mu_y - \boldsymbol{\beta}^T(\mathbf{x_j} - \boldsymbol{\mu}_\mathbf{x})\right]^2\right\} \qquad j = 1,2,\dots,m \qquad (3 - 2)$$

$$h_X(\mathbf{x_i};\boldsymbol{\mu}_\mathbf{x},\boldsymbol{\Sigma}_\mathbf{xx}) = \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}_\mathbf{xx}|}}\exp\left\{-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_\mathbf{xx}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_x)\right\} \quad i = 1,2,\dots,n \tag{3 - 3}$$

where

$$E(\mathrm{y_j}|\mathbf{x_j}) = \mu_y + \boldsymbol{\Sigma}_\mathbf{yx}\boldsymbol{\Sigma}_\mathbf{xx}^{-1}(\boldsymbol{x_j}-\boldsymbol{\mu}_x) = \mu_y + \boldsymbol{\beta}^{\boldsymbol{T}}(\boldsymbol{x_j}-\boldsymbol{\mu}_x) \tag{3 - 4}$$

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_\mathbf{xx}^{-1}\boldsymbol{\Sigma}_\mathbf{xy} \tag{3 - 5}$$

$$\sigma_e^2 = \sigma_y^2 - \boldsymbol{\Sigma}_\mathbf{yx}\boldsymbol{\Sigma}_\mathbf{xx}^{-1}\boldsymbol{\Sigma}_\mathbf{xy} = \sigma_y^2 - \boldsymbol{\beta}^T\boldsymbol{\Sigma}_\mathbf{xx}\boldsymbol{\beta} \tag{3 - 6}$$

The joint likelihood function is

$$\mathrm{L}(\mu_y,\boldsymbol{\mu}_\mathbf{x},\boldsymbol{\beta},\boldsymbol{\Sigma}_\mathbf{xx},\sigma_e^2) = \prod_{j=1}^{m} g_{Y|X}(\mathrm{y_j}|\mathbf{x_j};\mu_y,\boldsymbol{\mu}_\mathbf{x},\boldsymbol{\beta},\sigma_e^2)\prod_{i=1}^{n} h_X(\mathbf{x_i};\boldsymbol{\mu}_\mathbf{x},\boldsymbol{\Sigma}_\mathbf{xx})$$

$$\mathrm{L}(\mu_y,\boldsymbol{\mu}_\mathbf{x},\boldsymbol{\beta},\boldsymbol{\Sigma}_\mathbf{xx},\sigma_e^2) = (2\pi)^{-\frac{n+m}{2}}\sigma_e^{-m}exp\left\{-\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left[\mathrm{y_j}-\mu_y-\boldsymbol{\beta}^T(\mathbf{x_j}-\boldsymbol{\mu}_\mathbf{x})\right]^2\right\}$$

$$\cdot|\boldsymbol{\Sigma}_\mathbf{xx}|^{-\frac{n}{2}}exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_\mathbf{xx}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_x)\right\} \tag{3 - 7}$$

We will derive the maximum likelihood estimators by maximizing the likelihood function in the following section.

3.1 Maximum Likelihood Estimators

To obtain maximum likelihood estimators, we need to maximize following $(3-8)$ and $(3-9)$ simultaneously

$$|\boldsymbol{\Sigma}_\mathbf{xx}|^{-\frac{n}{2}}exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_\mathbf{xx}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_x)\right\} \tag{3 - 8}$$

$$\sigma_e^{-m}\, exp\left\{-\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left[y_j-\mu_y-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\boldsymbol{\mu}_\mathbf{x}\right)\right]^2\right\} \tag{3 - 9}$$

Let us consider the exponent and find the MLE of μ_y , $\boldsymbol{\mu}_\mathbf{x}$ and $\boldsymbol{\beta}$ to minimize

$$\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left[y_j-\mu_y-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\boldsymbol{\mu}_\mathbf{x}\right)\right]^2+\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_x) \tag{3 - 10}$$

Since the sum of trace of the matrix is equal to the trace of sum of the matrix, we have

$$\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_x)=\frac{1}{2}\sum_{i=1}^{n}tr[(\mathbf{x}_i-\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_x)]$$

$$=\frac{1}{2}\sum_{i=1}^{n}tr[\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_x)(\mathbf{x}_i-\boldsymbol{\mu}_x)^T]$$

$$=\frac{1}{2}tr\left\{\sum_{i=1}^{n}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_x)(\mathbf{x}_i-\boldsymbol{\mu}_x)^T\right\}$$

$$=\frac{1}{2}tr\left\{\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\left[\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu}_x)(\mathbf{x}_i-\boldsymbol{\mu}_x)^T\right]\right\} \tag{3 - 11}$$

let

$$\bar{\mathbf{x}}_n=\frac{1}{n}\left[\sum_{i=1}^{n}x_{i1},\sum_{i=1}^{n}x_{i2},\dots,\quad\sum_{i=1}^{n}x_{ip}\right]^T=\left[\bar{\mathbf{x}}_{n1},\bar{\mathbf{x}}_{n2},\dots,\bar{\mathbf{x}}_{np}\right]^T$$

and rewrite each $(\mathbf{x}_i-\boldsymbol{\mu}_x)$ as

$$\mathbf{x}_i-\boldsymbol{\mu}_x=(\mathbf{x}_i-\bar{\mathbf{x}}_n)+(\bar{\mathbf{x}}_n-\boldsymbol{\mu}_x)$$

We have

$$\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu}_x)(\mathbf{x}_i-\boldsymbol{\mu}_x)^T=\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}}_n+\bar{\mathbf{x}}_n-\boldsymbol{\mu}_x)(\mathbf{x}_i-\bar{\mathbf{x}}_n+\bar{\mathbf{x}}_n-\boldsymbol{\mu}_x)^T$$

$$= \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)^T + n(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)^T \qquad (3 - 12)$$

where the cross-product terms

$$\sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)^T = \left[\sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)\right](\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)^T = \mathbf{0}$$

and

$$\sum_{i=1}^{n} (\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)(\mathbf{x}_i - \bar{\mathbf{x}}_n)^T = (\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)\left[\sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)^T\right] = \mathbf{0}$$

Replace $\sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu}_x)(\mathbf{x}_i - \boldsymbol{\mu}_x)^T$ in $(3 - 11)$ with $(3 - 12)$, we obtain

$$\frac{1}{2}\sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu}_x)^T \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_x) = \frac{1}{2} tr\left\{\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\left[\sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)^T + n(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)^T\right]\right\}$$

$$= \frac{1}{2} tr\left\{\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\left[\sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)^T\right]\right\} + \frac{n}{2}(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x)^T \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} (\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x) \qquad (3 - 13)$$

Similarly, let

$$\bar{y}_m = \frac{1}{m}\sum_{j=1}^{m} y_j$$

$$\bar{\mathbf{x}}_m = \frac{1}{m}\left[\sum_{i=1}^{m} x_{i1}, \sum_{i=1}^{m} x_{i2}, \dots, \sum_{i=1}^{m} x_{ip}\right]^T = \left[\bar{\mathbf{x}}_{m1}, \bar{\mathbf{x}}_{m2}, \dots, \bar{\mathbf{x}}_{mp}\right]^T$$

Each $\left[y_j - \mu_y - \boldsymbol{\beta}^T(\mathbf{x}_\mathrm{j} - \boldsymbol{\mu}_x)\right]$ can be written as

$$y_j - \mu_y - \boldsymbol{\beta}^T(\mathbf{x}_\mathrm{j} - \boldsymbol{\mu}_x) = \left[y_j - \bar{y}_m - \boldsymbol{\beta}^T(\mathbf{x}_\mathrm{j} - \bar{\mathbf{x}}_m)\right] + \left[\bar{y}_m - \mu_y - \boldsymbol{\beta}^T(\bar{\mathbf{x}}_m - \boldsymbol{\mu}_x)\right]$$

Then we get

$$\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left[\mathrm{y_j}-\mu_y-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\boldsymbol{\mu_x}\right)\right]^2$$

$$=\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left\{\left[\mathrm{y}_j-\bar{y}_m-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\bar{\mathbf{x}}_m\right)\right]+\left[\bar{y}_m-\mu_y-\boldsymbol{\beta}^T(\bar{\mathbf{x}}_m-\boldsymbol{\mu_x})\right]\right\}^2$$

$$=\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left[\mathrm{y}_j-\bar{y}_m-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\bar{\mathbf{x}}_m\right)\right]^2+\frac{m}{2\sigma_e^2}\left[\bar{y}_m-\mu_y-\boldsymbol{\beta}^T(\bar{\mathbf{x}}_m-\boldsymbol{\mu_x})\right]^2 \qquad (3 - 14)$$

where the cross-product term

$$\sum_{j=1}^{m}\left[\mathrm{y}_j-\bar{y}_m-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\bar{\mathbf{x}}_m\right)\right]\left[\bar{y}_m-\mu_y-\boldsymbol{\beta}^T(\bar{\mathbf{x}}_m-\boldsymbol{\mu_x})\right]=0$$

So, if we minimize $(3-13)$ and $(3-14)$ simultaneously, $(3-10)$ will be minimized.

First, let us consider $(3-13)$. Since $\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}$ is positive definite, each term in $(3-13)$ is greater than or equal to zero. The second term $n(\bar{\mathbf{x}}_n-\boldsymbol{\mu_x})^T\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\bar{\mathbf{x}}_n-\boldsymbol{\mu_x})/2$ can be minimized if we set $\boldsymbol{\mu}_\mathbf{x}=\bar{\mathbf{x}}_\mathrm{n}$, so we have MLE for $\boldsymbol{\mu}_\mathbf{x}$

$$\widehat{\boldsymbol{\mu}}_\mathbf{x}=\bar{\mathbf{x}}_\mathrm{n} \qquad (3 - 15)$$

Second, let us consider $(3-14)$. Both terms in it are non-negative, to minimize the first term in $(3-14)$,i.e.,

$$min\sum_{j=1}^{m}\left[\mathrm{y}_j-\bar{y}_m-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\bar{\mathbf{x}}_m\right)\right]^2$$

We take derivative with respect to $\boldsymbol{\beta}$ first, then set the derivative to zero, and obtain the MLE for $\boldsymbol{\beta}$ which makes the above minimum. By method in Petersen and Pedersen (2012),

$$\frac{\partial\sum_{j=1}^{m}\left[\mathrm{y}_j-\bar{y}_m-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\bar{\mathbf{x}}_m\right)\right]^2}{\partial\boldsymbol{\beta}}=-2\sum_{j=1}^{m}\left(\mathrm{y}_j-\bar{y}_m-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\bar{\mathbf{x}}_m\right)\right)\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\boldsymbol{\beta}^T\left(\mathbf{x_j}-\bar{\mathbf{x}}_m\right)\right\}$$

$$= -2\sum_{j=1}^{m}\left(y_j - \bar{y}_m - \boldsymbol{\beta}^T\left(\mathbf{x_j} - \bar{\mathbf{x}}_m\right)\right)(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})$$

and set above to **0**, we have

$$\sum_{j=1}^{m}\left(y_j - \bar{y}_m - \boldsymbol{\beta}^T\left(\mathbf{x_j} - \bar{\mathbf{x}}_m\right)\right)(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}) = \mathbf{0}$$

Solve above equation, we have

$$\widehat{\boldsymbol{\beta}} = \mathbf{S}_{\mathrm{xxm}}^{-\mathbf{1}}\mathbf{S}_{\mathrm{xym}} \tag{3 - 16}$$

where

$$\mathbf{S_{xxm}} = \sum_{j=1}^{m}(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})\left((\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)^T \tag{3 - 17}$$

$$\mathbf{S_{xym}} = \sum_{j=1}^{m}(\mathrm{y_j} - \bar{y}_m)(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}) \tag{3 - 18}$$

By minimizing the second term in $(3 - 14)$ to give

$$\left[\bar{y}_m - \mu_y - \boldsymbol{\beta}^T(\bar{\mathbf{x}}_m - \boldsymbol{\mu}_x)\right]^2 = 0$$

We have MLE for μ_y by solving above equation:

$$\hat{\mu}_y = \bar{y}_m - \widehat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_m - \widehat{\boldsymbol{\mu}}_\mathbf{x}) \tag{3 - 19}$$

Now back to maximize $(3 - 8)$ and $(3 - 9)$ simultaneously. When $\widehat{\boldsymbol{\mu}}_\mathbf{x} = \bar{\mathbf{x}}_\mathrm{n}$, $(3 - 8)$ is reduced to

$$|\boldsymbol{\Sigma}_\mathbf{xx}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_{\boldsymbol{i}} - \boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_\mathbf{xx}^{-1}(\mathbf{x}_{\boldsymbol{i}} - \boldsymbol{\mu}_x)\right\}$$

$$= |\boldsymbol{\Sigma}_\mathbf{xx}|^{-\frac{n}{2}} exp\left\{-\frac{1}{2}tr\left\{\boldsymbol{\Sigma}_\mathbf{xx}^{-1}\left[\sum_{i=1}^{n}(\mathbf{x}_{\boldsymbol{i}} - \bar{\mathbf{x}}_n)(\mathbf{x}_{\boldsymbol{i}} - \bar{\mathbf{x}}_n)^T\right]\right\}\right\} \tag{3 - 20}$$

By Results 4.10 in Johnson and Wichern (1998), $(3 - 20)$ reaches maximum when

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{xx}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_{\boldsymbol{i}} - \bar{\mathbf{x}}_n)(\mathbf{x}_{\boldsymbol{i}} - \bar{\mathbf{x}}_n)^T \tag{3 - 21}$$

Similarly, when $\hat{\mu}_y = \bar{y}_m - \widehat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_m - \widehat{\boldsymbol{\mu}}_{\mathbf{x}})$ and $\widehat{\boldsymbol{\mu}}_{\mathbf{x}} = \bar{\mathbf{x}}_{\mathrm{n}}$, $(3-9)$ is reduced to

$$\sigma_e^{-m} exp\left\{-\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left(y_j - \mu_y - \boldsymbol{\beta}^T\left(\mathbf{x}_{\mathbf{j}} - \boldsymbol{\mu}_x\right)\right)^2\right\}$$

$$= \sigma_e^{-m} exp\left\{-\frac{1}{2\sigma_e^2} tr\sum_{j=1}^{m}\left[\mathrm{y}_j - \bar{y}_m - \widehat{\boldsymbol{\beta}}^T\left(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_m\right)\right]^2\right\} \tag{3 - 22}$$

So, by Results 4.10 in Johnson and Wichern (1998), $(3-22)$ reaches maximum when

$$\hat{\sigma}_e^2 = \frac{1}{m}\sum_{j=1}^{m}\left[\mathrm{y}_j - \bar{y}_m - \widehat{\boldsymbol{\beta}}^T\left(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_m\right)\right]^2 \tag{3 - 23}$$

In summary, we have the following 5 maximum likelihood estimators:

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}} = \overline{\boldsymbol{X}}_n \tag{3 - 24}$$

$$\hat{\mu}_y = \bar{Y}_m - \widehat{\boldsymbol{\beta}}^T(\overline{\boldsymbol{X}}_m - \overline{\boldsymbol{X}}_n) \tag{3 - 25}$$

$$\widehat{\boldsymbol{\beta}} = \mathbf{S}_{\mathrm{xxm}}^{\mathbf{-1}}\mathbf{S}_{\mathrm{xym}} \tag{3 - 26}$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{xx}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_{\boldsymbol{i}} - \overline{\boldsymbol{X}}_n)(\boldsymbol{X}_{\boldsymbol{i}} - \overline{\boldsymbol{X}}_n)^T \tag{3 - 27}$$

$$\hat{\sigma}_e^2 = \frac{1}{m}\sum_{j=1}^{m}\left[Y_j - \bar{Y}_m - \widehat{\boldsymbol{\beta}}^T\left(\boldsymbol{X}_{\boldsymbol{j}} - \overline{\boldsymbol{X}}_m\right)\right]^2 \tag{3 - 28}$$

Similarly, if we do not consider extra information $\boldsymbol{X}_{\mathrm{m+1,}}\boldsymbol{X}_{\mathrm{m+2,}} \dots, \boldsymbol{X}_{\mathrm{n}}$ and only use the first m observations, we have

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{x}\_no} = \overline{\mathbf{X}}_{\mathrm{m}} \tag{3 - 29}$$

$$\hat{\mu}_{y\_no} = \bar{Y}_{\mathrm{m}} \tag{3 - 30}$$

$$\widehat{\boldsymbol{\beta}}_{no} = \mathbf{S}_{\mathrm{xxm}}^{-1}\mathbf{S}_{\mathrm{xym}} \tag{3 - 31}$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{xx}\_no} = \frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{X}_{\boldsymbol{i}} - \overline{\boldsymbol{X}}_{m})(\boldsymbol{X}_{\boldsymbol{i}} - \overline{\boldsymbol{X}}_{m})^{T} \tag{3 - 32}$$

$$\widehat{\sigma}_{\mathrm{e}\_no}^{2} = \frac{1}{m}\sum_{j=1}^{m}\left[Y_j - \bar{Y}_m - \widehat{\boldsymbol{\beta}}^{T}\left(\boldsymbol{X}_{\boldsymbol{j}} - \overline{\boldsymbol{X}}_{m}\right)\right]^{2} \tag{3 - 33}$$

3.2 Properties of the Maximum Likelihood Estimators

3.2.1 Estimator of the Mean Vector of $\boldsymbol{X}$

The expectation of $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}}$ is

$$E(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}}) = E(\overline{\boldsymbol{X}}_{\boldsymbol{n}}) = E\left[\bar{X}_{n1}, \bar{X}_{n2}, \dots, \bar{X}_{np}\right]^{T} = \left[E(\bar{X}_{n1}), E(\bar{X}_{n2}), \dots, E(\bar{X}_{np})\right]^{T}$$

$$= \left[\mu_{x1}, \mu_{x2}, \dots \mu_{xp}\right]^{T} = \boldsymbol{\mu}_{\boldsymbol{x}} \tag{3 - 34}$$

So $\hat{\boldsymbol{\mu}}_{\mathbf{x}}$ is an unbiased estimator. The covariance of $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}}$ is

$$\mathrm{Cov}(\hat{\boldsymbol{\mu}}_{\mathbf{x}}) = \mathrm{Cov}(\overline{\mathbf{X}}_{\mathbf{n}}) = \mathrm{E}(\overline{\mathbf{X}}_{\mathbf{n}} - \boldsymbol{\mu}_{\mathbf{x}})(\overline{\mathbf{X}}_{\mathbf{n}} - \boldsymbol{\mu}_{\mathbf{x}})^{T} = \frac{1}{n^{2}}\left\{\sum_{j=1}^{n}\sum_{l=1}^{n}\mathrm{E}(\mathbf{X}_{\mathbf{j}} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{X}_{\mathbf{l}} - \boldsymbol{\mu}_{\mathbf{x}})^{T}\right\}$$

$$= \frac{1}{n^2}\sum_{j=1}^{n} \mathrm{E}\big(\mathbf{X_j} - \boldsymbol{\mu}_\mathbf{x}\big)\big(\mathbf{X_j} - \boldsymbol{\mu}_\mathbf{x}\big)^T = \frac{1}{n}\boldsymbol{\Sigma}_\mathbf{xx} \qquad (3 - 35)$$

By assumptions, $\boldsymbol{X} \sim N_{\boldsymbol{p}}(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\Sigma}_{\mathrm{xx}})$, so $\widehat{\boldsymbol{\mu}}_\mathbf{x}$ has a p-variate normal distribution too, i.e.,

$$\widehat{\boldsymbol{\mu}}_\mathbf{x} \sim N_p(\boldsymbol{\mu}_\mathbf{x}, \frac{1}{n}\boldsymbol{\Sigma}_{\mathrm{xx}})$$

3.2.2 Estimator of the Covariance Matrix of $\boldsymbol{X}$

Since $\boldsymbol{X}_{1,}\boldsymbol{X}_{2,} \dots, \boldsymbol{X}_{\mathrm{n}}$ is a random sample of size n from a p-variate normal distribution with mean $\boldsymbol{\mu}_\mathbf{x}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathrm{xx}}$, so

$$\sum_{i=1}^{n}(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_n)(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_n)^T \sim W_p(\boldsymbol{\Sigma}_{\mathrm{xx}}, n-1)$$

where $W_p(\boldsymbol{\Sigma}_\mathbf{xx}, n-1)$ is Wishart distribution with $(n-1)$ degree of freedom.

We have

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{xx}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_n)(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_n)^T$$

So

$$n\widehat{\boldsymbol{\Sigma}}_{\mathrm{xx}} \sim W_p(\boldsymbol{\Sigma}_{\mathrm{xx}}, n-1)$$

By Nydick (2012), we have

$$E(n\widehat{\boldsymbol{\Sigma}}_{\mathrm{xx}}) = (n-1)\boldsymbol{\Sigma}_{\mathrm{xx}}$$

$$Var\big(n\widehat{\Sigma}_{ij}\big) = (n-1)\big(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}\big), \qquad i,j = 1,2,\dots,p$$

So, the expectation of $\widehat{\boldsymbol{\Sigma}}_{\mathrm{xx}}$ is

$$E(\widehat{\boldsymbol{\Sigma}}_{\mathrm{xx}}) = \frac{n-1}{n}\boldsymbol{\Sigma}_{\mathrm{xx}} \qquad (3 - 36)$$

$\widehat{\boldsymbol{\Sigma}}_\mathbf{xx}$ is a biased estimator.

$$Var(\hat{\Sigma}_{ij}) = \frac{n-1}{n^2}(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}) \tag{3 - 37}$$

If we define

$$\mathbf{S}_{\mathrm{xn}} = \frac{n}{n-1}\widehat{\mathbf{\Sigma}}_{\mathrm{xx}} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_n)(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_n)^T \tag{3 - 38}$$

Then we have

$$E(\mathbf{S}_{\mathrm{xn}}) = \frac{n}{n-1}E(\widehat{\mathbf{\Sigma}}_{\mathrm{xx}}) = \mathbf{\Sigma}_{\mathrm{xx}}$$

$\mathbf{S}_{\mathrm{xn}}$ is an unbiased estimator for $\widehat{\mathbf{\Sigma}}_{\mathrm{xx}}$.

3.2.3 Estimator of the Regression Coefficient Vector $\widehat{\boldsymbol{\beta}}$

In this section, we will derive the conditional expectation and covariance matrix of $\widehat{\boldsymbol{\beta}}$ given $\boldsymbol{X} = \mathbf{x}$ first, then derive the unconditional expectation and covariance matrix of the estimator.

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}|\mathbf{x}) = \mathrm{E}(\mathbf{S}_{\mathrm{xxm}}^{-1}\mathbf{S}_{\mathrm{xym}}|\mathbf{x}) = \mathbf{S}_{\mathrm{xxm}}^{-1}\mathrm{E}(\mathbf{S}_{\mathrm{xym}}|\mathbf{x}) = \mathbf{S}_{\mathrm{xxm}}^{-1}\mathrm{E}\left\{\sum_{j=1}^{m}(\mathbf{x_j} - \bar{\mathbf{x}}_{\mathbf{m}})(Y_j - \bar{Y}_m)|\mathbf{x}\right\}$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\sum_{j=1}^{m}(\mathbf{x_j} - \bar{\mathbf{x}}_{\mathbf{m}})\mathrm{E}(Y_j\,|\mathbf{x}) = \mathbf{S}_{\mathrm{xxm}}^{-1}\sum_{j=1}^{m}(\mathbf{x_j} - \bar{\mathbf{x}}_{\mathbf{m}})\left[\mu_y + (\mathbf{x_j} - \boldsymbol{\mu_x})^T\boldsymbol{\beta}\right]$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\sum_{j=1}^{m}(\mathbf{x_j} - \bar{\mathbf{x}}_{\mathbf{m}})\,\mathbf{x}_j^T\boldsymbol{\beta} = \mathbf{S}_{\mathrm{xxm}}^{-1}\sum_{\boldsymbol{j=1}}^{\boldsymbol{m}}(\mathbf{x_j} - \bar{\mathbf{x}}_{\mathbf{m}})(\mathbf{x_j} - \bar{\mathbf{x}}_{\mathbf{m}})^T\boldsymbol{\beta}$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\mathbf{S}_{\mathrm{xxm}}\boldsymbol{\beta} = \boldsymbol{\beta} \tag{3 - 39}$$

So the unconditional expectation of $\widehat{\boldsymbol{\beta}}$ is

$$E(\widehat{\boldsymbol{\beta}}) = E[E(\widehat{\boldsymbol{\beta}}|\mathbf{x})] = E(\boldsymbol{\beta}) = \boldsymbol{\beta} \tag{3 - 40}$$

$\widehat{\boldsymbol{\beta}}$ is an unbiased estimator.

Similarly, we derive the conditional covariance first, then by the Law of Total Covariance to obtain the unconditional covariance.

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}|\mathbf{x}) = \mathrm{Cov}(\mathbf{S}_{\mathrm{xxm}}^{-1}\mathbf{S}_{\mathrm{xym}}|\mathbf{x}) = \mathbf{S}_{\mathrm{xxm}}^{-1}\mathrm{Cov}(\mathbf{S}_{\mathrm{xym}}|\mathbf{x})(\mathbf{S}_{\mathrm{xxm}}^{-1})^T$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\mathrm{Cov}\left\{\sum_{j=1}^{m}(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})(Y_j - \bar{Y}_m)|\mathbf{x}\right\}\mathbf{S}_{\mathrm{xxm}}^{-1}$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\left\{\sum_{j=1}^{m}(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})Var[(Y_j)|\mathbf{x}]\,(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T\right\}\mathbf{S}_{\mathrm{xxm}}^{-1}$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\left\{\sum_{j=1}^{m}(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})\sigma_e^2\,(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T\right\}\mathbf{S}_{\mathrm{xxm}}^{-1}$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\sigma_e^2\left\{\sum_{j=1}^{m}(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})\,(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T\right\}\mathbf{S}_{\mathrm{xxm}}^{-1} = \sigma_e^2\mathbf{S}_{\mathrm{xxm}}^{-1} \qquad (3 - 41)$$

By the Law of Total Covariance and by Nydick (2012), we have

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = E\,[\mathrm{Cov}(\hat{\boldsymbol{\beta}}|\mathbf{x})] + \mathrm{Cov}[E(\hat{\boldsymbol{\beta}}|\mathbf{x})] = E[\sigma_e^2\mathbf{S}_{\mathrm{xxm}}^{-1}] + \mathrm{Cov}(\boldsymbol{\beta})$$

$$= \sigma_e^2 E[\mathbf{S}_{\mathrm{xxm}}^{-1}] = \frac{\sigma_e^2}{m - p - 2}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} \qquad (3 - 42)$$

where $\mathbf{S}_{\mathrm{xxm}}^{-1}$ has an inverse Wishart distribution

$$\mathbf{S}_{\mathrm{xxm}}^{-1} \sim InvW_p(\,\boldsymbol{\Sigma}_{\mathbf{xx}}, m - 1) \qquad (3 - 43)$$

When sample is large, $\hat{\boldsymbol{\beta}}$ has an asymptotically p-variate normal distribution.

3.2.4 Estimator of the Mean of Y

As we do in 3.2.3, in this section, we will derive the conditional expectation and variance of $\hat{\mu}_y$ given $\boldsymbol{X} = \mathbf{x}$ first, then derive the unconditional expectation and variance of the estimator.

$$E(\hat{\mu}_y|\mathbf{x}) = E\{[\bar{Y}_m - \widehat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)]|\mathbf{x}\} = E(\bar{Y}_\mathrm{m}|\mathbf{x}) - E(\widehat{\boldsymbol{\beta}}^T|\mathbf{x})(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)$$

$$= \mu_y + \boldsymbol{\beta}^T(\bar{\mathbf{x}}_m - \boldsymbol{\mu}_x) - \boldsymbol{\beta}^T(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n) = \mu_y + \boldsymbol{\beta}^T(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x) \qquad (3 - 44)$$

where

$$E(\bar{Y}_\mathrm{m}|\mathbf{x}) = E\left(\frac{1}{m}\sum_{j=1}^{m} Y_j\,|\mathbf{x}_j\right) = \frac{1}{m}\sum_{j=1}^{m} E(Y_j|\mathbf{x}_j) = \frac{1}{m}\sum_{j=1}^{m}[\mu_y + \boldsymbol{\beta}^T(\mathbf{x}_\mathbf{j} - \boldsymbol{\mu}_x)]$$

$$= \frac{1}{m}\left[m\mu_y + \boldsymbol{\beta}^T\left(\sum_{j=1}^{m}\mathbf{x}_\mathbf{j} - m\boldsymbol{\mu}_x\right)\right] = \mu_y + \boldsymbol{\beta}^T(\bar{\mathbf{x}}_{\boldsymbol{m}} - \boldsymbol{\mu}_x) \qquad (3 - 45)$$

So, we have

$$E(\hat{\mu}_y) = E(E(\hat{\mu}_y|\mathbf{X})) = E[\mu_y + \boldsymbol{\beta}^T(\bar{\mathbf{X}}_{\boldsymbol{n}} - \boldsymbol{\mu}_x)] = \mu_y + \boldsymbol{\beta}^T(\boldsymbol{\mu}_x - \boldsymbol{\mu}_x) = \mu_y \qquad (3 - 46)$$

$\hat{\mu}_y$ is an unbiased estimator.

Since

$$Var(\bar{Y}_\mathrm{m}|\mathbf{x}) = Var\left(\frac{1}{m}\sum_{j=1}^{m} Y_j\,|\mathbf{x}_j\right) = \frac{1}{m^2}\sum_{j=1}^{m} Var(Y_j|\mathbf{x}_j) = \frac{\sigma_e^2}{m} \qquad (3 - 47)$$

$$Var[\widehat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)|\mathbf{x}] = Var[(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T\widehat{\boldsymbol{\beta}}|\mathbf{x}] = (\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T Var(\widehat{\boldsymbol{\beta}}|\mathbf{x})(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)$$

$$= \sigma_e^2(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T\mathbf{S}_\mathrm{xxm}^{-\mathbf{1}}(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n) \qquad (3 - 48)$$

$$\mathrm{Cov}[\bar{Y}_\mathrm{m}, \widehat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)]|\mathbf{x} = \mathrm{Cov}[(\bar{Y}_\mathrm{m}, \widehat{\boldsymbol{\beta}}^T)|\mathbf{x}](\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T = [\mathrm{Cov}(\bar{Y}_\mathrm{m}, \hat{\beta}_j^T)|\mathbf{x}](\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T = 0$$

So, the conditional variance of $\hat{\mu}_y$ is

$$Var(\hat{\mu}_y|\mathbf{x}) = Var\{[\bar{Y}_m - \widehat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)]|\mathbf{x}\}$$

$$= Var(\bar{Y}_\mathrm{m}|\mathbf{x}) + Var[\widehat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)|\mathbf{x}] - 2Cov[\bar{Y}_\mathrm{m}, \widehat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)]|\mathbf{x}$$

$$= \frac{\sigma_e^2}{m} + \sigma_e^2 (\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T \mathbf{S}_{\mathrm{xxm}}^{-1} (\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n) \qquad (3 - 49)$$

To obtain the unconditional variance of $\hat{\mu}_y$, we use the Law of Total Variance

$$Var(\hat{\mu}_y) = Var\left[E(\hat{\mu}_y|\mathbf{x})\right] + E\left[Var(\hat{\mu}_y|\mathbf{x})\right]$$

now

$$Var\left[E(\hat{\mu}_y|\mathbf{x})\right] = Var\left[\mu_y + \boldsymbol{\beta}^T(\bar{\mathbf{X}}_n - \boldsymbol{\mu}_x)\right] = \boldsymbol{\beta}^T Var(\bar{\mathbf{X}}_n)\boldsymbol{\beta} = \frac{1}{n}\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\beta} \qquad (3 - 50)$$

To obtain $E\left[Var(\hat{\mu}_y|\mathbf{x})\right]$, we need to find out the distribution of $(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)$ first.

$$\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_n = \bar{\mathbf{X}}_m - \frac{1}{n}(m\bar{\mathbf{X}}_m + (n-m)\bar{\mathbf{X}}_{n-m}) = \frac{n-m}{n}(\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m}) \qquad (3 - 51)$$

$\bar{\mathbf{X}}_m$ and $\bar{\mathbf{X}}_{n-m}$ are independent and normally distributed, and since

$$E(\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m}) = E(\bar{\mathbf{X}}_m) - E(\bar{\mathbf{X}}_{n-m}) = \boldsymbol{\mu}_x - \boldsymbol{\mu}_x = \mathbf{0} \qquad (3 - 52)$$

$$\mathrm{Cov}(\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m}) = \mathrm{Cov}(\bar{\mathbf{X}}_m) + \mathrm{Cov}(\bar{\mathbf{X}}_{n-m}) = \frac{n}{m(n-m)}\boldsymbol{\Sigma}_{\mathbf{xx}} \qquad (3 - 53)$$

So, we have

$$\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m} \sim N_p\left(\mathbf{0}, \frac{n}{m(n-m)}\boldsymbol{\Sigma}_{\mathbf{xx}}\right)$$

Replace $(\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_n)$ with $(\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m})$,

$$(\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_n)^T \mathbf{S}_{\mathrm{xxm}}^{-1} (\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_n) = \left(\frac{n-m}{n}\right)^2 (\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m})^T \mathbf{S}_{\mathrm{xxm}}^{-1} (\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m})$$

$$= \left(\frac{n-m}{n}\right)^2 \cdot \frac{n}{m(n-m)} \cdot \frac{1}{m-1} \frac{(\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m})^T}{\sqrt{\frac{n}{m(n-m)}}} \left(\frac{\mathbf{S}_{\mathrm{xxm}}}{m-1}\right)^{-1} \frac{(\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m})}{\sqrt{\frac{n}{m(n-m)}}}$$

$$= \frac{n-m}{mn(m-1)} T_{p,m-1}^2 \qquad (3 - 54)$$

where

$$T_{p,m-1}^2 = \frac{(\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m})^T}{\sqrt{\frac{n}{m(n-m)}}} \left(\frac{\mathbf{S}_{\mathrm{xxm}}}{m-1}\right)^{-1} \frac{(\bar{\mathbf{X}}_m - \bar{\mathbf{X}}_{n-m})}{\sqrt{\frac{n}{m(n-m)}}}$$

Hence,

$$E\left[Var(\hat{\mu}_y|\mathbf{x})\right] = E\left[\frac{\sigma_e^2}{m} + \sigma_e^2(\overline{\mathbf{X}}_m - \overline{\mathbf{X}}_n)^T \mathbf{S}_{\text{xxm}}^{-1}(\overline{\mathbf{X}}_m - \overline{\mathbf{X}}_n)\right]$$

$$= \frac{\sigma_e^2}{m} + \frac{(n-m)\sigma_e^2}{mn(m-1)} E(T_{p,m-1}^2)$$

$$= \frac{\sigma_e^2}{m} + \frac{(n-m)\sigma_e^2}{mn(m-1)} \cdot \frac{(m-1)p}{m-p-2}$$

$$= \frac{\sigma_e^2}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right] \qquad (3 - 55)$$

where

$$E\left(T_{p,m-1}^2\right) = \frac{(m-1)p}{m-p} E\left(F_{p,m-p}\right) = \frac{(m-1)p}{m-p} \cdot \frac{m-p}{m-p-2} = \frac{(m-1)p}{m-p-2}$$

Using $(3-50)$ and $(3-55)$, we have the unconditional variance of $\hat{\mu}_y$

$$Var(\hat{\mu}_y) = \frac{\sigma_e^2}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right] + \frac{1}{n}\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}} \boldsymbol{\beta} \qquad (3 - 56)$$

$\hat{\mu}_y$ has an asymptotically normal distribution when sample is large.

*3.2.5 Estimator of the Conditional Variance of Y given **x***

We use similar idea for the bivariate normal distribution. Since

$$\mathrm{E}\left(Y_j|\mathbf{x}_j\right) = \mu_y + \boldsymbol{\beta}^T\left(\mathbf{x}_\mathrm{j} - \boldsymbol{\mu}_x\right), \qquad j = 1,2,\ldots,m$$

For given $\mathbf{x}_j$, we may write

$$\mathrm{Y}_j = \mu_y + \boldsymbol{\beta}^T\left(\mathbf{x}_\mathrm{j} - \boldsymbol{\mu}_x\right) + \varepsilon_j, \quad j = 1,2,\ldots,m \qquad (3 - 57)$$

where

$$E\left(\varepsilon_j\right) = E\left(\mathrm{Y}_j|\mathbf{x}_\mathbf{j}\right) - \mu_y - \boldsymbol{\beta}^T\left(\mathbf{x}_\mathbf{j} - \boldsymbol{\mu}_x\right) = 0$$

$$\mathrm{Var}(\mathrm{Y}_\mathrm{j}|\mathbf{x}_\mathrm{j}) = Var\left(\varepsilon_j\right) = \sigma_e^2$$

Then we have

$$\hat{\varepsilon}_j = Y_j - \hat{\mu}_y - \widehat{\boldsymbol{\beta}}^T(\boldsymbol{X}_j - \widehat{\boldsymbol{\mu}}_{\boldsymbol{x}})$$

$$= Y_j - [\bar{Y}_m - \boldsymbol{\beta}^T(\overline{\boldsymbol{X}}_m - \overline{\boldsymbol{X}}_n)] - \boldsymbol{\beta}^T(\boldsymbol{X}_j - \overline{\boldsymbol{X}}_n)$$

$$= Y_j - \bar{Y}_m - \boldsymbol{\beta}^T(\boldsymbol{X}_j - \overline{\boldsymbol{X}}_m)$$

Hence

$$\hat{\sigma}_e^2|\mathbf{x} = \frac{1}{m}\sum_{j=1}^{m}\left\{Y_j - \bar{Y}_m - \boldsymbol{\beta}^T(\boldsymbol{X}_j - \overline{\boldsymbol{X}}_m)\right\}^2|\mathbf{x} = \frac{1}{m}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} \qquad (3 - 58)$$

We may rewrite $(3-57)$ as

$$Y_j = \beta_* + \boldsymbol{\beta}^T(\mathbf{x}_\mathrm{j} - \bar{\mathbf{x}}_m) + \varepsilon_j, \qquad j = 1,2,\dots,m \qquad (3 - 59)$$

where

$$\beta_* = \mu_y + \boldsymbol{\beta}^T(\bar{\mathbf{x}}_m - \boldsymbol{\mu}_{\boldsymbol{x}})$$

Equation $(3-59)$ is the mean corrected form of the multiple regression model.

By Results 7.2 and Results 7,4 in Johnson and Wichern (1998),

$$\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} \sim \sigma_e^2\chi_{m-p-1}^2$$

So

$$E(\hat{\sigma}_e^2|\mathbf{x}) = \frac{m-p-1}{m}\sigma_e^2, \qquad Var(\hat{\sigma}_e^2|\mathbf{x}) = \frac{2(m-p-1)}{m^2}\sigma_e^4$$

Hence, we have

$$E(\hat{\sigma}_e^2) = E[E(\hat{\sigma}_e^2|\mathbf{x})] = \frac{m-p-1}{m}\sigma_e^2 \qquad (3 - 60)$$

$\hat{\sigma}_e^2$ is a biased estimator. The bias of $\hat{\sigma}_e^2$ is

$$Bias(\hat{\sigma}_e^2, \sigma_e^2) = E(\hat{\sigma}_e^2) - \sigma_e^2 = -\frac{p+1}{m}\sigma_e^2$$

The bias vanishes as $\mathrm{m} \to \infty$.

The unconditional variance of $\hat{\sigma}_e^2$ is

$$Var(\hat{\sigma}_e^2) = Var[E(\hat{\sigma}_e^2|\mathbf{x})] + E[Var(\hat{\sigma}_e^2|\mathbf{x})] = \frac{2(m-p-1)}{m^2}\sigma_e^4 \qquad (3 - 61)$$

Since $\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}/\sigma_e^2$ does not depend on x, so the unconditional $\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}/\sigma_e^2$ also has $\chi^2_{\mathrm{m-p-1}}$ distribution, i.e.,

$$\frac{m\hat{\sigma}_e^2}{\sigma_e^2} \sim \chi^2_{m-p-1}$$

The mean square error for $\hat{\sigma}_e^2$ is

$$MSE(\hat{\sigma}_e^2) = E(\hat{\sigma}_e^2 - \sigma_e^2)^2 = \mathrm{Var}(\hat{\sigma}_e^2) + [Bias(\hat{\sigma}_e^2, \sigma_e^2)]^2$$

$$= \frac{2(m-p-1)\sigma_e^4}{m^2} + \frac{(p+1)^2\sigma_e^4}{m^2} = \frac{2m + (p+1)(p-1)}{m^2}\sigma_e^4 \qquad (3 - 62)$$

3.3 Prediction

Suppose we have a future observation $(X_{01}, X_{02}, \dots, X_{0p}, Y_0)$,

$$\begin{pmatrix} \boldsymbol{X}_0 \\ Y_0 \end{pmatrix} \sim N_{p+1}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}} \\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \sigma_y^2 \end{bmatrix}\right) \qquad (3 - 63)$$

Where $\boldsymbol{X}_0$ is a $p-$ dimensional vector.

We have following three kinds of prediction interval for Y_0:

1) Usual prediction interval for Y_0– conditioning on $\mathbf{X} = \mathbf{x}\ and\ \mathbf{X}_0 = \mathbf{x}_0$

2) Prediction interval for Y_0– unconditional on $\mathbf{X}$, but conditioning on $\mathbf{X}_0 = \mathbf{x}_0$

3) Unconditional prediction interval for Y_0

3.3.1 Usual prediction interval

– Conditioning on $\boldsymbol{X} = \boldsymbol{x}\ and\ \boldsymbol{X}_0 = \boldsymbol{x}_0$

The prediction value of Y_0 given $\mathbf{X} = \mathbf{x}\ and\ \ \mathbf{X}_0 = \mathbf{x}_0$ is

$$\hat{Y}_0|\mathbf{x}, \mathbf{x}_0 = \hat{\mu}_y + \widehat{\boldsymbol{\beta}}^T(\mathbf{x}_0 - \bar{\mathbf{x}}_n) \quad (3 - 64)$$

By our assumption, $(\hat{\mu}_y, \widehat{\boldsymbol{\beta}})$ are independent of $(\boldsymbol{X}_0, Y_0)$. By equation $(3-26), (3-39), (3-41)$ and since $(Y|\mathbf{x})$ is normal, the distribution of $(\widehat{\boldsymbol{\beta}}|\mathbf{x})$ is

$$(\widehat{\boldsymbol{\beta}}|\mathbf{x}) \sim N_p(\boldsymbol{\beta}, \sigma_e^2 \mathbf{S}_{\mathrm{xxm}}^{-1}) \quad (3 - 65)$$

By equation $(3-25), (3-44), (3-49)$ and since $(Y|\mathbf{x})$ and $(\widehat{\boldsymbol{\beta}}|\mathbf{x})$ are normal, the distribution of $(\hat{\mu}_y|\mathbf{x})$ is

$$(\hat{\mu}_y|\mathbf{x}) \sim N\left(\mu_y + \boldsymbol{\beta}^T(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x), \frac{\sigma_e^2}{m} + \sigma_e^2(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T \mathbf{S}_{\mathrm{xxm}}^{-1}(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)\right) \quad (3 - 66)$$

So, we have

$$\begin{aligned}
E(\hat{Y}_0|\mathbf{x}, \mathbf{x}_0) &= E(\hat{\mu}_y|\mathbf{x}, \mathbf{x}_0) + E(\widehat{\boldsymbol{\beta}}^T|\mathbf{x}, \mathbf{x}_0)(\mathbf{x}_0 - \bar{\mathbf{x}}_n) \\
&= E(\hat{\mu}_y|\mathbf{x}) + E(\widehat{\boldsymbol{\beta}}^T|\mathbf{x})(\mathbf{x}_0 - \bar{\mathbf{x}}_n) \\
&= \mu_y + \boldsymbol{\beta}^T(\bar{\mathbf{x}}_n - \boldsymbol{\mu}_x) + \boldsymbol{\beta}^T(\mathbf{x}_0 - \bar{\mathbf{x}}_n) \\
&= \mu_y + \boldsymbol{\beta}^T(\mathbf{x}_0 - \boldsymbol{\mu}_x) \quad (3 - 67)
\end{aligned}$$

By our assumption,

$$Y_0|\mathbf{x}, \mathbf{x}_0 \sim N(\mu_y + \boldsymbol{\beta}^T(\mathbf{x}_0 - \boldsymbol{\mu}_x), \sigma_e^2) \quad (3 - 68)$$

Then we have conditional variance of $\hat{Y}_0$ as follows:

$$\begin{aligned}
Var(\hat{Y}_0|\mathbf{x}, \mathbf{x}_0) &= Var(\hat{\mu}_y|\mathbf{x}) + (\mathbf{x}_0 - \bar{\mathbf{x}}_n)^T \mathrm{Cov}(\widehat{\boldsymbol{\beta}}|\mathbf{x})(\mathbf{x}_0 - \bar{\mathbf{x}}_n) + 2Cov[(\hat{\mu}_y, \widehat{\boldsymbol{\beta}}^T)|\mathbf{x}](\mathbf{x}_0 - \bar{\mathbf{x}}_n) \\
&= \frac{\sigma_e^2}{m} + \sigma_e^2(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T \mathbf{S}_{\mathrm{xxm}}^{-1}(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n) + \sigma_e^2(\mathbf{x}_0 - \bar{\mathbf{x}}_n)^T \mathbf{S}_{\mathrm{xxm}}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_n) \\
&\quad - 2\sigma_e^2(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T \mathbf{S}_{\mathrm{xxm}}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_n) \\
&= \frac{\sigma_e^2}{m} + \sigma_e^2(\mathbf{x}_0 - \bar{\mathbf{x}}_m)^T \mathbf{S}_{\mathrm{xxm}}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_m) \quad (3 - 69)
\end{aligned}$$

where

$$Cov[(\hat{\mu}_y, \widehat{\boldsymbol{\beta}}^T)|\mathbf{x}] = Cov[(\bar{Y}_m - (\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^T)|\mathbf{x}]$$

$$= -(\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T \mathrm{Cov}(\widehat{\boldsymbol{\beta}}|\mathbf{x}) = -\sigma_e^2 (\bar{\mathbf{x}}_m - \bar{\mathbf{x}}_n)^T \mathbf{S}_{\mathrm{xxm}}^{-1} \quad (3 - 70)$$

By (3 – 67) and (3 – 68), the expectation $(Y_0 - \hat{Y}_0)$ given $\mathbf{X} = \mathbf{x}$ *and* $\mathbf{X}_0 = \mathbf{x}_0$ is

$$E(Y_0 - \hat{Y}_0)|\mathbf{x}, \mathbf{x}_0 = E(Y_0|\mathbf{x}, \mathbf{x}_0) - E(\hat{Y}_0|\mathbf{x}, \mathbf{x}_0) = 0 \quad (3 - 71)$$

By (3 – 68) and (3 – 69), the variance $(Y_0 - \hat{Y}_0)$ given $\mathbf{X} = \mathbf{x}$ *and* $\mathbf{X}_0 = \mathbf{x}_0$ is

$$Var(Y_0 - \hat{Y}_0)|\mathbf{x}, \mathbf{x}_0 = Var(Y_0)|\mathbf{x}, \mathbf{x}_0 + Var(\hat{Y}_0)|\mathbf{x}, \mathbf{x}_0 - 2Cov(\hat{Y}_0, Y_0)|\mathbf{x}, \mathbf{x}_0$$

$$= \sigma_e^2 + \frac{\sigma_e^2}{m} + \sigma_e^2 (\mathbf{x}_0 - \bar{\mathbf{x}}_m)^T \mathbf{S}_{\mathrm{xxm}}^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_m) - 2 \cdot 0$$

$$= \sigma_e^2 \left[1 + \frac{1}{m} + (\mathbf{x}_0 - \bar{\mathbf{x}}_m)^T \mathbf{S}_{\mathrm{xxm}}^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_m)\right] \quad (3 - 72)$$

where

$$\mathrm{Cov}(\hat{Y}_0, Y_0)|\mathbf{x}, \mathbf{x}_0 = \mathrm{Cov}[\hat{\mu}_y + \widehat{\boldsymbol{\beta}}^T (\mathbf{x}_0 - \bar{\mathbf{x}}_n), Y_0]|\mathbf{x}, \mathbf{x}_0 = 0$$

Since Y_0 *and* $\hat{Y}_0$ are normal, so $Y_0 - \hat{Y}_0$ is normal, then

$$Z = \frac{(Y_0 - \hat{Y}_0)|\mathbf{x}, \mathbf{x}_0 - E(Y_0 - \hat{Y}_0)|\mathbf{x}, \mathbf{x}_0}{\sqrt{Var(Y_0 - \hat{Y}_0)|\mathbf{x}, \mathbf{x}_0}} = \frac{(Y_0 - \hat{Y}_0)|\mathbf{x}, \mathbf{x}_0 - 0}{\sqrt{\sigma_e^2 \left[1 + \frac{1}{m} + (\mathbf{x}_0 - \bar{\mathbf{x}}_m)^T \mathbf{S}_{\mathrm{xxm}}^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_m)\right]}} \sim N(0,1)$$

Since

$$U = \frac{m\hat{\sigma}_e^2}{\sigma_e^2} \sim \chi_{m-p-1}^2$$

and Z *and* U are independent, so

$$T = \frac{Z}{\sqrt{U/(m-p-1)}} \sim t_{m-p-1}$$

i.e.,

$$T = \frac{\dfrac{\left(Y_0 - \hat{Y}_0\right)|\mathbf{x}, \mathbf{x}_0 - 0}{\sqrt{\sigma_e^2\left[1 + \dfrac{1}{m} + (\mathbf{x}_0 - \bar{\mathbf{x}}_m)^T \mathbf{S}_{\mathrm{xxm}}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_m)\right]}}}{\sqrt{\dfrac{m\hat{\sigma}_e^2}{(m-p-1)\sigma_e^2}}}$$

$$= \frac{\left(Y_0 - \hat{Y}_0\right)|\mathbf{x}, \mathbf{x}_0}{\sqrt{\dfrac{m\hat{\sigma}_e^2}{m-p-1}\left[1 + \dfrac{1}{m} + (\mathbf{x}_0 - \bar{\mathbf{x}}_m)^T \mathbf{S}_{\mathrm{xxm}}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_m)\right]}} \sim t_{m-p-1}$$

Hence, the 95% prediction interval for $Y_0 \; given \; \mathbf{X} = \mathbf{x} \; and \; \mathbf{X}_0 = \mathbf{x}_0$ is

$$\hat{Y}_0|\mathbf{x}, \mathbf{x}_0 \pm t_{0.025,m-p-1}\sqrt{\frac{m\hat{\sigma}_e^2}{m-p-1}\left[1 + \frac{1}{m} + (\mathbf{x}_0 - \bar{\mathbf{x}}_m)^T \mathbf{S}_{\mathrm{xxm}}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_m)\right]} \qquad (3 - 73)$$

3.3.2 Prediction interval

– Unconditional on $\boldsymbol{X}$, but conditioning on $\boldsymbol{X}_0 = \boldsymbol{x}_0$

In this situation, the prediction value of Y_0 given $\mathbf{X}_0 = \mathbf{x}_0$ is

$$\hat{Y}_0|\mathbf{x}_0 = \hat{\mu}_y + \hat{\boldsymbol{\beta}}^T(\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_x) \qquad (3 - 74)$$

By $(3-42)$ and $(3-46)$, the expectation of Y_0 given $\mathbf{X}_0 = \mathbf{x}_0$ is

$$E\left(\hat{Y}_0\middle|\mathbf{x}_0\right) = E\left(\hat{\mu}_y|\mathbf{x}_0\right) + E\left[\hat{\boldsymbol{\beta}}^T(\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_x)|\mathbf{x}_0\right]$$

$$= E\left(\hat{\mu}_y\right) + E(\hat{\boldsymbol{\beta}}^T)\mathbf{x}_0 - E(\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\mu}}_x)$$

$$= \mu_y + \boldsymbol{\beta}^T(\mathbf{x}_0 - \boldsymbol{\mu}_x) \qquad (3 - 75)$$

where

$$E\left(\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\mu}}_x\right) = E\left[E\left(\boldsymbol{\mu}_x^T\hat{\boldsymbol{\beta}}\right)|\mathbf{x}\right] = E\left[\boldsymbol{\mu}_x^T E\left(\hat{\boldsymbol{\beta}}\right)|\mathbf{x}\right] = \boldsymbol{\beta}^T E(\hat{\boldsymbol{\mu}}_x) = \boldsymbol{\beta}^T\boldsymbol{\mu}_x \qquad (3 - 76)$$

$\hat{\boldsymbol{\mu}}_x \; and \; \hat{\boldsymbol{\beta}}^T$ are not correlated. By our assumption,

$$Y_0|\mathbf{x}_0 \sim N\left(\mu_y + \boldsymbol{\beta}^T(\mathbf{x}_0 - \boldsymbol{\mu}_x), \sigma_e^2\right) \tag{3 - 77}$$

We derive the variance of $\hat{Y}_0|\mathbf{x}_0$ by the Δ-Method. Let

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = \begin{bmatrix} \hat{\mu}_y \\ \hat{\boldsymbol{\mu}}_x \\ \hat{\boldsymbol{\beta}} \end{bmatrix} \tag{3 - 78}$$

$\boldsymbol{Z}$ *and* $\boldsymbol{X}_0$ are independent. So

$$\mathrm{E}(\boldsymbol{Z}|\mathbf{x}_0) = \mathrm{E}(\boldsymbol{Z}) = \boldsymbol{\mu}_Z = \begin{bmatrix} E(Z_1) \\ E(Z_2) \\ E(Z_3) \end{bmatrix} = \begin{bmatrix} E(\hat{\mu}_y) \\ E(\hat{\boldsymbol{\mu}}_x) \\ E(\hat{\boldsymbol{\beta}}) \end{bmatrix} = \begin{bmatrix} \mu_y \\ \boldsymbol{\mu}_x \\ \boldsymbol{\beta} \end{bmatrix} \tag{3 - 79}$$

$$\mathrm{Cov}(\boldsymbol{Z}|\mathbf{x}_0) = \mathrm{Cov}(\boldsymbol{Z}) = E[\boldsymbol{Z} - \mathrm{E}(\boldsymbol{Z})][\boldsymbol{Z} - \mathrm{E}(\boldsymbol{Z})]' = \begin{bmatrix} Var(Z_1) & Cov(Z_1, Z_2) & Cov(Z_1, Z_3) \\ Cov(Z_2, Z_1) & Var(Z_2) & Cov(Z_2, Z_3) \\ Cov(Z_3, Z_1) & Cov(Z_3, Z_2) & Var(Z_3) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma_e^2}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right] + \frac{1}{n}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\beta} & \frac{1}{n}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\mathbf{xx}} & \mathbf{0} \\ \frac{1}{n}\boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\beta} & \frac{1}{n}\boldsymbol{\Sigma}_{\mathbf{xx}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{\sigma_e^2}{m-p-2}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-\mathbf{1}} \end{bmatrix} \tag{3 - 80}$$

where, by $(3-56)$, we have

$$Var(Z_1) = \mathrm{Var}(\hat{\mu}_y) = \frac{\sigma_e^2}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right] + \frac{1}{n}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\beta}$$

By $(3-35)$, we have

$$Var(Z_2) = \mathrm{Cov}(\hat{\boldsymbol{\mu}}_x) = \frac{1}{n}\boldsymbol{\Sigma}_{\mathbf{xx}}$$

By $(3-42)$, we have

$$Var(Z_3) = \mathrm{Cov}\left(\hat{\boldsymbol{\beta}}\right) = \frac{\sigma_e^2}{m-p-2}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-\mathbf{1}}$$

$$Cov(Z_1, Z_2) = Cov(\hat{\mu}_y, \hat{\boldsymbol{\mu}}_{\boldsymbol{x}}) = Cov[\bar{Y}_m - \hat{\boldsymbol{\beta}}^T(\bar{\boldsymbol{X}}_{\boldsymbol{m}} - \bar{\boldsymbol{X}}_{\boldsymbol{n}}), \bar{\boldsymbol{X}}_{\boldsymbol{n}}]$$

$$= Cov(\bar{Y}_m, \bar{\boldsymbol{X}}_{\boldsymbol{n}}) - Cov[\hat{\boldsymbol{\beta}}^T(\bar{\boldsymbol{X}}_{\boldsymbol{m}} - \bar{\boldsymbol{X}}_{\boldsymbol{n}}), \bar{\boldsymbol{X}}_{\boldsymbol{n}}]$$

$$= E(\bar{Y}_m\bar{\boldsymbol{X}}_n^T) - E(\bar{Y}_m)E(\bar{\boldsymbol{X}}_n^T) - E[\hat{\boldsymbol{\beta}}^T(\bar{\boldsymbol{X}}_{\boldsymbol{m}} - \bar{\boldsymbol{X}}_{\boldsymbol{n}})\bar{\boldsymbol{X}}_n^T] + E[\hat{\boldsymbol{\beta}}^T(\bar{\boldsymbol{X}}_{\boldsymbol{m}} - \bar{\boldsymbol{X}}_{\boldsymbol{n}})]E(\bar{\boldsymbol{X}}_n^T)$$

$$= E\{[\mu_y + \boldsymbol{\beta}^T(\bar{\boldsymbol{X}}_{\boldsymbol{m}} - \boldsymbol{\mu}_{\boldsymbol{x}})]\bar{\boldsymbol{X}}_n^T\} - \mu_y\boldsymbol{\mu}_{\boldsymbol{x}}^T - \boldsymbol{\beta}^T E[(\bar{\boldsymbol{X}}_{\boldsymbol{m}} - \bar{\boldsymbol{X}}_{\boldsymbol{n}})\bar{\boldsymbol{X}}_n^T] + \mathbf{0}$$

$$= -\boldsymbol{\beta}^T\boldsymbol{\mu}_{\boldsymbol{x}}\boldsymbol{\mu}_{\boldsymbol{x}}^T + \boldsymbol{\beta}^T E(\bar{\boldsymbol{X}}_{\boldsymbol{n}}\bar{\boldsymbol{X}}_n^T) = -\boldsymbol{\beta}^T\boldsymbol{\mu}_{\boldsymbol{x}}\boldsymbol{\mu}_{\boldsymbol{x}}^T + \boldsymbol{\beta}^T[Var(\bar{\boldsymbol{X}}_{\boldsymbol{n}}) + \boldsymbol{\mu}_{\boldsymbol{x}}\boldsymbol{\mu}_{\boldsymbol{x}}^T] = \frac{1}{n}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\boldsymbol{xx}}$$

$$\mathrm{Cov}(\hat{\mu}_y, \hat{\boldsymbol{\beta}}) = \mathbf{0} \quad \rightarrow \mathrm{Cov}(Z_1, Z_3) = \mathrm{Cov}(\hat{\mu}_y, \hat{\boldsymbol{\beta}}^T) = \mathbf{0} \quad (1 \times p)$$

$$\mathrm{Cov}(\hat{\boldsymbol{\mu}}_x, \hat{\boldsymbol{\beta}}) = \mathbf{0} \quad \rightarrow \mathrm{Cov}(Z_2, Z_3) = \mathrm{Cov}(\hat{\boldsymbol{\mu}}_x, \hat{\boldsymbol{\beta}}^T) = \mathbf{0} \quad (p \times p)$$

In terms of $\boldsymbol{z}$, we have

$$\hat{y}_0|\mathbf{x}_0 = z_1 + \mathbf{z}_3^T(\mathbf{x}_0 - \boldsymbol{z}_2) \qquad (3 - 81)$$

By the Delta-method,

$$\hat{y}_0|\mathbf{x}_0 \approx \mu_y + \boldsymbol{\beta}^T(\mathbf{x}_0 - \boldsymbol{\mu}_x) + \sum_{j=1}^{3}[\hat{y}'_{0j}(\boldsymbol{\mu}_Z|\mathbf{x}_0)](z_j - \mu_{Zj})$$

Where

$$\hat{y}'_0(\boldsymbol{\mu}_Z|\mathbf{x}_0) = \begin{bmatrix} \hat{y}'_{01}(\boldsymbol{\mu}_Z|\mathbf{x}_0) \\ \hat{y}'_{02}(\boldsymbol{\mu}_Z|\mathbf{x}_0) \\ \hat{y}'_{03}(\boldsymbol{\mu}_Z|\mathbf{x}_0) \end{bmatrix} = \begin{bmatrix} \frac{\partial y_0}{\partial z_1} \\ \frac{\partial y_0}{\partial \boldsymbol{z}_2} \\ \frac{\partial y_0}{\partial \boldsymbol{z}_3} \end{bmatrix}_{\boldsymbol{z}=\boldsymbol{\mu}_Z} = \begin{bmatrix} 1 \\ -\boldsymbol{\beta} \\ \mathbf{x}_0 - \boldsymbol{\mu}_x \end{bmatrix} \qquad (3 - 82)$$

Hence the expectation of $\hat{Y}_0|\mathbf{x}_0$ is

$$\mathrm{E}(\hat{Y}_0|\mathbf{x}_0) \approx \mu_y + \boldsymbol{\beta}^T(\mathbf{x}_0 - \boldsymbol{\mu}_x) + \sum_{j=1}^{3}[\hat{y}'_{0j}(\boldsymbol{\mu}_Z|\mathbf{x}_0)]E(Z_j - \mu_{Zj}) = \mu_y + \boldsymbol{\beta}^T(\mathbf{x}_0 - \boldsymbol{\mu}_x) \qquad (3 - 83)$$

The variance of $\hat{Y}_0|\mathbf{x}_0$ is

$$\mathrm{Var}(\hat{Y}_0|\mathbf{x}_0) = \mathrm{E}\left[\hat{Y}_0|\mathbf{x}_0 - E\left(\hat{Y}_0|\mathbf{x}_0\right)\right]^2$$

$$\approx \mathrm{E}\left[\sum_{j=1}^{3}\left[\hat{y}'_{0j}(\boldsymbol{\mu}_Z|\mathbf{x}_0)\right](Z_j - \mu_{Zj})\right]^2$$

$$= \sum_{j=1}^{3}\left[\hat{y}'_{0j}(\boldsymbol{\mu}_Z|\mathbf{x}_0)\right]^T Var(Z_j)\left[\hat{y}'_{0j}(\boldsymbol{\mu}_Z|\mathbf{x}_0)\right] + 2\sum_{i=1}^{3}\sum_{j=1,i\neq j}^{3}[\hat{y}'_{0i}(\boldsymbol{\mu}_Z|\mathbf{x}_0)]^T Cov(Z_i, Z_j)\left[\hat{y}'_{0j}(\boldsymbol{\mu}_Z|\mathbf{x}_0)\right]$$

$$= Var(Z_1) + (-\boldsymbol{\beta}^T)Var(Z_2)(-\boldsymbol{\beta}) + (\mathbf{x}_0 - \boldsymbol{\mu}_{\boldsymbol{x}})^T Var(Z_3)(\mathbf{x}_0 - \boldsymbol{\mu}_{\boldsymbol{x}}) + 2\cdot 1 Cov(Z_1, Z_2)(-\boldsymbol{\beta})$$

$$= Var\left(\hat{\mu}_y\right) + (-\boldsymbol{\beta}^T)Cov(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}})(-\boldsymbol{\beta}) + (\mathbf{x}_0 - \boldsymbol{\mu}_{\boldsymbol{x}})^T \mathrm{Cov}\left(\hat{\boldsymbol{\beta}}\right)(\mathbf{x}_0 - \boldsymbol{\mu}_{\boldsymbol{x}}) + 2\cdot 1\cdot Cov\left(\hat{\mu}_y, \hat{\boldsymbol{\mu}}_{\boldsymbol{x}}\right)(-\boldsymbol{\beta})$$

$$= \frac{\sigma_e^2}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right] + \frac{1}{n}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\mathrm{xx}}\boldsymbol{\beta} + \frac{1}{n}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\mathrm{xx}}\boldsymbol{\beta} + (\mathbf{x}_0 - \boldsymbol{\mu}_{\boldsymbol{x}})^T\frac{\sigma_e^2}{m-p-2}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-\mathbf{1}}(\mathbf{x}_0 - \boldsymbol{\mu}_{\boldsymbol{x}}) - \frac{2}{n}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\mathrm{xx}}\boldsymbol{\beta}$$

$$= \sigma_e^2\left[\frac{1}{m} + \frac{(n-m)p}{mn(m-p-2)} + \frac{(\mathbf{x}_0 - \boldsymbol{\mu}_{\boldsymbol{x}})^T\boldsymbol{\Sigma}_{\mathbf{xx}}^{-\mathbf{1}}(\mathbf{x}_0 - \boldsymbol{\mu}_{\boldsymbol{x}})}{m-p-2}\right] \qquad (3 - 84)$$

By $(3-77)$ and $(3-83)$, the expectation of $\left(Y_0 - \hat{Y}_0\right)$ given $\mathbf{X}_0 = \mathbf{x}_0$ is

$$E\left(Y_0 - \hat{Y}_0\right)|\mathbf{x}_0 = E(Y_0|\mathbf{x}_0) - E\left(\hat{Y}_0\middle|\mathbf{x}_0\right) = 0 \qquad (3 - 85)$$

The variance of $\left(Y_0 - \hat{Y}_0\right)$ given $\mathbf{X}_0 = \mathbf{x}_0$ is

$$Var\left(Y_0 - \hat{Y}_0\right)|\mathbf{x}_0 = Var(Y_0)|\mathbf{x}_0 + Var\left(\hat{Y}_0\right)|\mathbf{x}_0 - 2Cov(\hat{Y}_0, Y_0)|\mathbf{x}_0$$

$$= \sigma_e^2 + \sigma_e^2\left[\frac{1}{m} + \frac{(n-m)p}{mn(m-p-2)} + \frac{(\mathbf{x}_0 - \boldsymbol{\mu}_{\boldsymbol{x}})^T\boldsymbol{\Sigma}_{\mathbf{xx}}^{-\mathbf{1}}(\mathbf{x}_0 - \boldsymbol{\mu}_{\boldsymbol{x}})}{m-p-2}\right] - 2\cdot 0$$

$$= \sigma_e^2 \left[1 + \frac{1}{m} + \frac{(n-m)p}{mn(m-p-2)} + \frac{(\mathbf{x}_0 - \boldsymbol{\mu}_x)^T \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}_x)}{m-p-2} \right] \qquad (3 - 86)$$

where

$$\text{Cov}(\hat{Y}_0, Y_0)|\mathbf{x}_0 = \text{Cov}[\hat{\mu}_y + \hat{\boldsymbol{\beta}}^T(\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_x), Y_0]|\mathbf{x}_0 = 0$$

When sample is large and by $(3-85)$ and $(3-86)$

$$Z = \frac{(Y_0 - \hat{Y}_0)|\mathbf{x}_0 - 0}{\sqrt{\sigma_e^2 \left[1 + \frac{1}{m} + \frac{(n-m)p}{mn(m-p-2)} + \frac{(\mathbf{x}_0 - \boldsymbol{\mu}_x)^T \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}_x)}{m-p-2} \right]}} \dot{\sim} N(0,1)$$

The 95% prediction interval for $Y_0 \; given \; \mathbf{X}_0 = \mathbf{x}_0$ is

$$\hat{Y}_0|\mathbf{x}_0 \pm z_{0.025} \sqrt{S_e^2 \left[1 + \frac{1}{m} + \frac{(n-m)p}{mn(m-p-2)} + \frac{(\mathbf{x}_0 - \bar{\mathbf{x}}_n)^T \mathbf{S}_{\mathbf{xn}}^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_n)}{m-p-2} \right]} \qquad (3 - 87)$$

Where

$$S_e^2 = \frac{1}{m-p-1} \sum_{j=1}^{m} \left\{ Y_j - \bar{Y}_{\text{m}} - \hat{\boldsymbol{\beta}}^T (\mathbf{X}_{\text{j}} - \bar{\mathbf{X}}_{\boldsymbol{m}}) \right\}^2 \qquad (3 - 88)$$

$\mathbf{S}_{\text{xn}}$ is given in $(3-38)$ as

$$\mathbf{S}_{\text{xn}} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)^T$$

3.3.3 Unconditional prediction interval

By assumptions in $(3-63)$, we have

$$Y_0 \sim N\left(\mu_y, \sigma_y^2\right) \quad (3 - 89)$$

$$\boldsymbol{X}_0 \sim N_p(\boldsymbol{\mu}_{\mathbf{x}}, \Sigma_{\mathbf{xx}}) \quad (3 - 90)$$

The prediction value of Y_0 is

$$\hat{Y}_0 = \hat{\mu}_y + \widehat{\boldsymbol{\beta}}^T(\boldsymbol{X}_0 - \widehat{\boldsymbol{\mu}}_{\mathbf{x}}) \quad (3 - 91)$$

Let

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix} = \begin{bmatrix} \hat{\mu}_y \\ \widehat{\boldsymbol{\mu}}_{\mathbf{x}} \\ \widehat{\boldsymbol{\beta}} \\ \boldsymbol{X}_0 \end{bmatrix} \quad (3 - 92)$$

We have

$$\mathrm{E}(\boldsymbol{Z}) = \boldsymbol{\mu}_Z = \begin{bmatrix} E(Z_1) \\ E(Z_2) \\ E(Z_3) \\ E(Z_4) \end{bmatrix} = \begin{bmatrix} E(\hat{\mu}_y) \\ E(\widehat{\boldsymbol{\mu}}_{\mathbf{x}}) \\ E(\widehat{\boldsymbol{\beta}}) \\ E(\boldsymbol{X}_0) \end{bmatrix} = \begin{bmatrix} \mu_y \\ \boldsymbol{\mu}_{\mathbf{x}} \\ \boldsymbol{\beta} \\ \boldsymbol{\mu}_{\mathbf{x}} \end{bmatrix} \quad (3 - 93)$$

$$\mathrm{Cov}(\boldsymbol{Z}) = E[\boldsymbol{Z} - \mathrm{E}(\boldsymbol{Z})][\boldsymbol{Z} - \mathrm{E}(\boldsymbol{Z})]^T = \begin{bmatrix} Var(Z_1) & Cov(Z_1, Z_2) & Cov(Z_1, Z_3) & Cov(Z_1, Z_4) \\ Cov(Z_2, Z_1) & Var(Z_2) & Cov(Z_2, Z_3) & Cov(Z_2, Z_4) \\ Cov(Z_3, Z_1) & Cov(Z_3, Z_2) & Var(Z_3) & Cov(Z_3, Z_4) \\ Cov(Z_4, Z_1) & Cov(Z_4, Z_2) & Cov(Z_4, Z_3) & Var(Z_4) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma_e^2}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right] + \frac{1}{n}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\beta} & \frac{1}{n}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\mathbf{xx}} & 0 & 0 \\ \frac{1}{n}\boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\beta} & \frac{1}{n}\boldsymbol{\Sigma}_{\mathbf{xx}} & 0 & 0 \\ 0 & 0 & \frac{\sigma_e^2}{m-p-2}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-\mathbf{1}} & 0 \\ 0 & 0 & 0 & \boldsymbol{\Sigma}_{\mathbf{xx}} \end{bmatrix} \quad (3 - 94)$$

In terms of $\boldsymbol{z}$, we have

$$\hat{y}_0 = z_1 + \mathbf{z}_3^T(\boldsymbol{z}_4 - \boldsymbol{z}_2) \quad (3 - 95)$$

By the Delta-method,

$$\hat{y}_0 \approx \mu_y + \sum_{j=1}^{4} \hat{y}'_{0j}(\boldsymbol{\mu}_Z)(\, z_j - \mu_{Zj})$$

where

$$\hat{y}'_0(\boldsymbol{\mu}_Z) = \begin{bmatrix} \hat{y}'_{01}(\boldsymbol{\mu}_Z) \\ \hat{y}'_{02}(\boldsymbol{\mu}_Z) \\ \hat{y}'_{03}(\boldsymbol{\mu}_Z) \\ \hat{y}'_{04}(\boldsymbol{\mu}_Z) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial y_0}{\partial z_1} \\ \dfrac{\partial y_0}{\partial \boldsymbol{z}_2} \\ \dfrac{\partial y_0}{\partial \boldsymbol{z}_3} \\ \dfrac{\partial y_0}{\partial \boldsymbol{z}_4} \end{bmatrix}_{\boldsymbol{z}=\boldsymbol{\mu}_Z} = \begin{bmatrix} 1 \\ -\boldsymbol{\beta} \\ \boldsymbol{0} \\ \boldsymbol{\beta} \end{bmatrix} \tag{3 - 96}$$

so, the expectation of $\hat{Y}_0$ is

$$\mathrm{E}(\hat{Y}_0) \approx \mu_y + \sum_{j=1}^{4} \hat{y}'_{0j}(\boldsymbol{\mu}_Z) E(\, Z_j - \mu_{Zj}) = \mu_y \tag{3 - 97}$$

The variance of $\hat{Y}_0$ is

$$\mathrm{Var}(\hat{Y}_0) = \mathrm{E}[\hat{Y}_0 - E(\hat{Y}_0)]^2$$

$$\approx \mathrm{E}\left[\sum_{j=1}^{4} \hat{y}'_{0j}(\boldsymbol{\mu}_Z) E(\, Z_j - \mu_{Zj})\right]^2$$

$$= \sum_{j=1}^{4} [\hat{y}'_{0j}(\boldsymbol{\mu}_Z|x_0)]^T Var(Z_j)\, [\hat{y}'_{0j}(\boldsymbol{\mu}_Z|x_0)] + 2\sum_{i=1}^{4} \sum_{j=1, i\neq j}^{4} [\hat{y}'_{0i}(\boldsymbol{\mu}_Z|x_0)]^T Cov(Z_i, Z_j)\, [\hat{y}'_{0j}(\boldsymbol{\mu}_Z|x_0)]$$

$$= Var(Z_1) + (-\boldsymbol{\beta}^T) Var(Z_2)(-\boldsymbol{\beta}) + \boldsymbol{0} \cdot Var(Z_3) + \boldsymbol{\beta}^T Var(Z_4)\boldsymbol{\beta} + 2 \cdot 1 \cdot Cov(Z_1, Z_2)(-\boldsymbol{\beta})$$

$$= Var(\hat{\mu}_y) + (-\boldsymbol{\beta}^T) Cov(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}})(-\boldsymbol{\beta}) + \boldsymbol{0} \cdot Cov(\hat{\boldsymbol{\beta}}) + \boldsymbol{\beta}^T Cov(\boldsymbol{X}_0)\boldsymbol{\beta} + 2 \cdot 1 \cdot Cov(\hat{\mu}_y, \hat{\boldsymbol{\mu}}_{\boldsymbol{x}})(-\boldsymbol{\beta})$$

$$= \frac{\sigma_e^2}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right] + \frac{1}{n}\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\beta} + \frac{1}{n}\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\beta} + 0 + \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\beta} - \frac{2}{n}\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\beta}$$

$$= \frac{\sigma_e^2}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right] + \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}} \boldsymbol{\beta} \tag{3 - 98}$$

By $(3-89)$ and $(3-97)$, the expectation of $\left(Y_0 - \hat{Y}_0\right)$ is

$$E\left(Y_0 - \hat{Y}_0\right) = E(Y_0) - E\left(\hat{Y}_0\right) = 0 \tag{3 - 99}$$

By $(3-6), (3-63), (3-89)$ and $(3-98)$, the variance of $\left(Y_0 - \hat{Y}_0\right)$ is

$$Var\left(Y_0 - \hat{Y}_0\right) = Var(Y_0) + Var\left(\hat{Y}_0\right) - 2Cov(\hat{Y}_0, Y_0)$$

$$= \sigma_y^2 + \frac{\sigma_e^2}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right] + \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}} \boldsymbol{\beta} - 2 \cdot \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}} \boldsymbol{\beta}$$

$$= \sigma_y^2 + \frac{\sigma_e^2}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right] - \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}} \boldsymbol{\beta}$$

$$= \sigma_e^2\left[1 + \frac{1}{m} + \frac{(n-m)p}{nm(m-p-2)}\right] \tag{3 - 100}$$

Where

$$\mathrm{Cov}\left(\hat{Y}_0, Y_0\right) = \mathrm{Cov}\left[\hat{\mu}_y + \hat{\boldsymbol{\beta}}^T(\boldsymbol{X}_0 - \hat{\boldsymbol{\mu}}_{\mathbf{x}}), Y_0\right]$$

$$= \mathrm{Cov}\left(\hat{\mu}_y, Y_0\right) + \mathrm{Cov}\left(\hat{\boldsymbol{\beta}}^T \mathbf{X}_0, Y_0\right) - \mathrm{Cov}\left(\hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\mu}}_{\mathbf{x}}, Y_0\right)$$

$$= 0 + E\left(\hat{\boldsymbol{\beta}}^T \mathbf{X}_0 Y_0\right) - E\left(\hat{\boldsymbol{\beta}}^T \mathbf{X}_0\right) E(Y_0) - 0$$

$$= E\left(\hat{\boldsymbol{\beta}}^T\right) \mathrm{E}(\mathbf{X}_0 Y_0) - E\left(\hat{\boldsymbol{\beta}}^T\right) E\left(\mathbf{X}_0\right) E(Y_0)$$

$$= \boldsymbol{\beta}^T [E(\mathbf{X}_0 Y_0) - E(\mathbf{X}_0) E(Y_0)] = \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{x}y} = \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\mathbf{xx}} \boldsymbol{\beta} \tag{3 - 101}$$

When sample is large,

$$Z = \frac{\left(Y_0 - \hat{Y}_0\right) - 0}{\sqrt{\sigma_e^2\left[1 + \frac{1}{m} + \frac{(n-m)p}{nm(m-p-2)}\right]}} \dot{\sim} N(0,1)$$

The 95% prediction interval for Y_0 is

$$\hat{Y}_0 \pm z_{0.025} \sqrt{\mathrm{S}_e^2\left[1 + \frac{1}{m} + \frac{(n-m)p}{nm(m-p-2)}\right]} \tag{3 - 102}$$

where S_e^2 is given in $(3-88)$ as

$$S_e^2 = \frac{1}{m-p-1}\sum_{j=1}^{m}\left\{Y_j - \bar{Y}_m - \widehat{\boldsymbol{\beta}}^T\left(\mathbf{X}_j - \bar{\mathbf{X}}_{\boldsymbol{m}}\right)\right\}^2$$

This is unconditional (not dependent on $\boldsymbol{X}_0$) prediction interval for Y_0 of a future observation.

3.4 An example for multiple regression model

We use the data from Han and Li (2011) to estimate those 5 MLE estimators (data in Table 1 – 1). In this example, we take TOEFL score as Y, GRE Verbal, GRE Quantitative and GRE Analytic as X_1, X_2 *and* X_3, respectively.

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{GRE\ Verbal} \\ \boldsymbol{GRE\ Quantitative} \\ \boldsymbol{GRE\ Analytic} \end{bmatrix} \qquad y = \boldsymbol{TOEFL}$$

Normality test shows that Y and $\mathbf{x} = [x_1, x_2, x_3]^T$ are normally distributed.

The five estimators are:

$$\widehat{\boldsymbol{\mu}}_{\mathbf{x}} = \begin{bmatrix} 419.5 \\ 646.5 \\ 523 \end{bmatrix} \qquad \hat{\mu}_y = 563 \qquad \widehat{\boldsymbol{\beta}} = \begin{bmatrix} 0.1776 \\ -0.1122 \\ 0.0513 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{xx}} = \begin{bmatrix} 11614.75 & -4656.75 & 5825.25 \\ -4656.75 & 11707.75 & 1510.5 \\ 5825.25 & 1510.5 & 12239.75 \end{bmatrix} \qquad \hat{\sigma}_e^2 = 776$$

If we do not consider extra information, we have

$$\widehat{\boldsymbol{\mu}}_{\mathbf{x}} = \begin{bmatrix} 342 \\ 710.5 \\ 468.5 \end{bmatrix} \qquad \hat{\mu}_y = 539.05 \qquad \widehat{\boldsymbol{\beta}} = \begin{bmatrix} 0.1776 \\ -0.1122 \\ 0.0513 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{xx}} = \begin{bmatrix} 3276 & 1324 & 2585.5 \\ 2585.5 & 5694.75 & 4525.75 \\ 5825.25 & 4525.75 & 11500.25 \end{bmatrix} \qquad \hat{\sigma}_e^2 = 776$$

Chapter 4

STATISTICAL ESTIMATION IN MULTIVARIATE REGRESSION MODEL WITH A BLOCK OF MISSING OBSERVATIONS

Let $\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{bmatrix}$ have a multivariate normal distribution with mean vector $\begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix}$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}}\\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \boldsymbol{\Sigma}_{\mathbf{yy}}\end{bmatrix}$$

where $\boldsymbol{X}$ is a $p \times 1$ vector and $\boldsymbol{Y}$ is a $q \times 1$ vector.

Suppose the following random sample with a block of missing $\boldsymbol{Y}$ values are obtained:

$$\begin{array}{cccccccc}
X_{1,1} & X_{1,2} & \cdots & X_{1,p} & Y_{1,1} & Y_{1,2} & \cdots & Y_{1,q}\\
X_{2,1} & X_{2,2} & \cdots & X_{2,p} & Y_{2,1} & Y_{2,2} & \cdots & Y_{2,q}\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
X_{m,1} & X_{m,2} & \cdots & X_{m,p} & Y_{m,1} & Y_{m,2} & \cdots & Y_{m,q}\\
X_{m+1,1} & X_{m+1,2} & \cdots & X_{m+1,p} & & & & \\
\vdots & \vdots & \vdots & \vdots & & & & \\
X_{n,1} & X_{n,1} & \cdots & X_{n,p} & & & &
\end{array}$$

Based on the data, We want to estimate the parameters. We can write the multivariate normal probability density function (pdf) as

$$f(\mathbf{x}, \boldsymbol{y}) = g(\boldsymbol{y}|\mathbf{x})h(\mathbf{x}) \tag{4- 1}$$

where $g(\boldsymbol{y}|\mathbf{x})$ is the conditional pdf of $\boldsymbol{Y}$ given $\boldsymbol{X} = \mathbf{x}$, and $h(\mathbf{x})$ is the marginal pdf of $\boldsymbol{X}$.

$$g_{Y|X}(\mathbf{y}_j|\mathbf{x}_j;\boldsymbol{\mu}_y,\boldsymbol{\mu}_x,\boldsymbol{\beta},\boldsymbol{\Sigma}_e) = \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}_e|}}\exp\left\{-\frac{1}{2}\left[\mathbf{y}_j - E(\mathbf{y}_j|\mathbf{x}_j)\right]^T\boldsymbol{\Sigma}_e^{-1}\left[\mathbf{y}_j - E(\mathbf{y}_j|\mathbf{x}_j)\right]\right\}$$

$$= \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}_e|}}\exp\left\{-\frac{1}{2}\left[\boldsymbol{y}_j - \boldsymbol{\mu}_y - \boldsymbol{\beta}^T(\boldsymbol{x}_j - \boldsymbol{\mu}_x)\right]^T\boldsymbol{\Sigma}_e^{-1}\left[\boldsymbol{y}_j - \boldsymbol{\mu}_y - \boldsymbol{\beta}^T(\boldsymbol{x}_j - \boldsymbol{\mu}_x)\right]\right\} \quad (4\text{- }2)$$

$$h_X(\mathbf{x}_i;\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_{xx}) = \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}_{xx}|}}\exp\left\{-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_x)\right\} \quad (4\text{- }3)$$

where

$$E(\mathbf{y}_j|\mathbf{x}_j) = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_x) = \boldsymbol{\mu}_y + \boldsymbol{\beta}^T(\boldsymbol{x}_j - \boldsymbol{\mu}_x) \quad (4 - 4)$$

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \quad (4 - 5)$$

$$\boldsymbol{\Sigma}_e = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\beta}^T\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} \quad (4 - 6)$$

$$i = 1,2,\dots,n. \; j = 1,2,\dots,m.$$

The joint likelihood function is

$$\mathrm{L}(\boldsymbol{\mu}_y,\boldsymbol{\mu}_x,\boldsymbol{\beta},\boldsymbol{\Sigma}_{xx},\boldsymbol{\Sigma}_e) = \prod_{j=1}^{m} g_{Y|X}(\mathbf{y}_j|\mathbf{x}_j;\boldsymbol{\mu}_y,\boldsymbol{\mu}_x,\boldsymbol{\beta},\boldsymbol{\Sigma}_e)\prod_{i=1}^{n} h_X(\mathbf{x}_i;\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_{xx})$$

$$\mathrm{L}(\boldsymbol{\mu}_y,\boldsymbol{\mu}_x,\boldsymbol{\beta},\boldsymbol{\Sigma}_{xx},\boldsymbol{\Sigma}_e)$$

$$= (2\pi)^{-\frac{n+m}{2}}|\boldsymbol{\Sigma}_e|^{-\frac{m}{2}}\exp\left\{-\frac{1}{2}\sum_{j=1}^{m}\left[\boldsymbol{y}_j - \boldsymbol{\mu}_y - \boldsymbol{\beta}^T(\boldsymbol{x}_j - \boldsymbol{\mu}_x)\right]^T\boldsymbol{\Sigma}_e^{-1}\left[\boldsymbol{y}_j - \boldsymbol{\mu}_y - \boldsymbol{\beta}^T(\boldsymbol{x}_j - \boldsymbol{\mu}_x)\right]\right\}$$

$$\cdot|\boldsymbol{\Sigma}_{xx}|^{-\frac{n}{2}}exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_x)\right\} \quad (4 - 7)$$

4.1 Maximum Likelihood Estimators

To obtain maximum likelihood estimators, we need to maximize the following $(4-8)$ and $(4-9)$ simultaneously.

$$|\mathbf{\Sigma}_{\mathrm{xx}}|^{-\frac{n}{2}} exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x_i}-\boldsymbol{\mu_x})^T\mathbf{\Sigma}_{\mathrm{xx}}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu_x})\right\} \quad (4 - 8)$$

$$|\mathbf{\Sigma}_{\mathrm{e}}|^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{m}\left[\boldsymbol{y_j}-\boldsymbol{\mu_y}-\boldsymbol{\beta}^T\left(\boldsymbol{x_j}-\boldsymbol{\mu_x}\right)\right]^T\mathbf{\Sigma}_{\mathrm{e}}^{-1}\left[\boldsymbol{y_j}-\boldsymbol{\mu_y}-\boldsymbol{\beta}^T\left(\boldsymbol{x_j}-\boldsymbol{\mu_x}\right)\right]\right\} \quad (4 - 9)$$

Let us consider the exponent first and find the MLE of $\boldsymbol{\mu_y}$, $\boldsymbol{\mu_x}$ and $\boldsymbol{\beta}$ to minimize

$$\frac{1}{2}\sum_{j=1}^{m}\left[\boldsymbol{y_j}-\boldsymbol{\mu_y}-\boldsymbol{\beta}^T\left(\boldsymbol{x_j}-\boldsymbol{\mu_x}\right)\right]^T\mathbf{\Sigma}_{\mathrm{e}}^{-1}\left[\boldsymbol{y_j}-\boldsymbol{\mu_y}-\boldsymbol{\beta}^T\left(\boldsymbol{x_j}-\boldsymbol{\mu_x}\right)\right]$$
$$+\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x_i}-\boldsymbol{\mu_x})^T\mathbf{\Sigma}_{\mathrm{xx}}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu_x}) \quad (4 - 10)$$

Since the sum of trace of the matrix is equal to the trace of sum of the matrix, we have

$$\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x_i}-\boldsymbol{\mu_x})^T\mathbf{\Sigma}_{\mathrm{xx}}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu_x}) = \frac{1}{2}\sum_{i=1}^{n} tr[(\boldsymbol{x_i}-\boldsymbol{\mu_x})^T\mathbf{\Sigma}_{\mathrm{xx}}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu_x})]$$

$$= \frac{1}{2} tr\left[\sum_{i=1}^{n}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu_x})(\boldsymbol{x_i}-\boldsymbol{\mu_x})^T\right]$$

$$= \frac{1}{2} tr\left[\sum_{i=1}^{n}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}[(\boldsymbol{x_i}-\overline{\boldsymbol{x}}_{\boldsymbol{n}})+(\overline{\boldsymbol{x}}_{\boldsymbol{n}}-\boldsymbol{\mu_x})][(\boldsymbol{x_i}-\overline{\boldsymbol{x}}_{\boldsymbol{n}})+(\overline{\boldsymbol{x}}_{\boldsymbol{n}}-\boldsymbol{\mu_x})]^T\right]$$

$$= \frac{1}{2} tr\left[\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\sum_{i=1}^{n}(\boldsymbol{x_i}-\overline{\boldsymbol{x}}_{\boldsymbol{n}})(\boldsymbol{x_i}-\overline{\boldsymbol{x}}_{\boldsymbol{n}})^T\right] + \frac{n}{2}(\overline{\boldsymbol{x}}_{\boldsymbol{n}}-\boldsymbol{\mu_x})^T\mathbf{\Sigma}_{\mathbf{xx}}^{-1}(\overline{\boldsymbol{x}}_{\boldsymbol{n}}-\boldsymbol{\mu_x}) \quad (4 - 11)$$

where

$$\overline{\boldsymbol{x}}_{\boldsymbol{n}} = \frac{1}{n}\left[\sum_{i=1}^{n}x_{i1}, \sum_{i=1}^{n}x_{i2}, \dots, \quad \sum_{i=1}^{n}x_{ip}\right]^T = \left[\overline{\boldsymbol{x}}_{n1}, \overline{\boldsymbol{x}}_{n2}, \dots, \overline{\boldsymbol{x}}_{np}\right]^T \quad (4 - 12)$$

and cross-product terms

$$\sum_{i=1}^{n}(\mathbf{x}_{\boldsymbol{i}}-\overline{\mathbf{x}}_n)\left(\overline{\mathbf{x}}_n-\boldsymbol{\mu_x}\right)^T = \left[\sum_{i=1}^{n}(\mathbf{x}_{\boldsymbol{i}}-\overline{\mathbf{x}}_n)\right]\left(\overline{\mathbf{x}}_n-\boldsymbol{\mu_x}\right)^T = \mathbf{0}$$

$$\sum_{i=1}^{n}(\bar{\mathbf{x}}_n-\boldsymbol{\mu}_x)(\mathbf{x}_i-\bar{\mathbf{x}}_n)^T=(\bar{\mathbf{x}}_n-\boldsymbol{\mu}_x)\left[\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}}_n)^T\right]=\mathbf{0}$$

Similarly, we have

$$\frac{1}{2}\sum_{j=1}^{m}\left[\boldsymbol{y}_j-\boldsymbol{\mu}_y-\boldsymbol{\beta}^T(\boldsymbol{x}_j-\boldsymbol{\mu}_x)\right]^T\boldsymbol{\Sigma}_e^{-1}\left[\boldsymbol{y}_j-\boldsymbol{\mu}_y-\boldsymbol{\beta}^T(\boldsymbol{x}_j-\boldsymbol{\mu}_x)\right]$$

$$=\frac{1}{2}tr\left\{\sum_{j=1}^{m}\boldsymbol{\Sigma}_e^{-1}\left[\boldsymbol{y}_j-\boldsymbol{\mu}_y-\boldsymbol{\beta}^T(\boldsymbol{x}_j-\boldsymbol{\mu}_x)\right]\left[\boldsymbol{y}_j-\boldsymbol{\mu}_y-\boldsymbol{\beta}^T(\boldsymbol{x}_j-\boldsymbol{\mu}_x)\right]^T\right\}$$

$$=\frac{1}{2}tr\left\{\sum_{j=1}^{m}\boldsymbol{\Sigma}_e^{-1}\left[\boldsymbol{y}_j-\bar{\boldsymbol{y}}_m-\boldsymbol{\beta}^T(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_m)\right]\left[\boldsymbol{y}_j-\bar{\boldsymbol{y}}_m-\boldsymbol{\beta}^T(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_m)\right]^T\right\}$$

$$+\frac{m}{2}\left[\bar{\boldsymbol{y}}_m-\boldsymbol{\mu}_y-\boldsymbol{\beta}^T(\bar{\boldsymbol{x}}_m-\boldsymbol{\mu}_x)\right]^T\boldsymbol{\Sigma}_e^{-1}\left[\bar{\boldsymbol{y}}_m-\boldsymbol{\mu}_y-\boldsymbol{\beta}^T(\bar{\mathbf{x}}_m-\boldsymbol{\mu}_x)\right] \qquad (4 - 13)$$

where

$$\bar{\mathbf{x}}_m=\frac{1}{m}\left[\sum_{i=1}^{m}x_{i1},\sum_{i=1}^{m}x_{i2},\dots,\quad\sum_{i=1}^{m}x_{ip}\right]^T=\left[\bar{\mathbf{x}}_{m1},\bar{\mathbf{x}}_{m2},\dots,\bar{\mathbf{x}}_{mp}\right]^T \qquad (4 - 14)$$

$$\bar{\boldsymbol{y}}_m=\frac{1}{m}\left[\sum_{j=1}^{m}y_{j1},\sum_{j=1}^{m}y_{j2},\dots,\quad\sum_{j=1}^{m}y_{jq}\right]^T=\left[\bar{\mathbf{y}}_{m1},\bar{\mathbf{y}}_{m2},\dots,\bar{\mathbf{y}}_{mq}\right]^T \qquad (4 - 15)$$

$$\sum_{j=1}^{m}\left[\mathbf{y}_j-\bar{\boldsymbol{y}}_m-\boldsymbol{\beta}^T(\mathbf{x}_j-\bar{\mathbf{x}}_m)\right]=0$$

Hence, minimizing $(4-11)$ and $(4-13)$ simultaneously will minimize $(4-10)$.

First, let us consider $(4-11)$. Since $\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}$ is positive definite, each term in $(4-11)$ is greater than or equal to zero. The second term $n(\bar{\mathbf{x}}_n-\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\bar{\mathbf{x}}_n-\boldsymbol{\mu}_x)/2$ can be minimized if we set

$$\frac{n}{2}(\bar{\boldsymbol{x}}_n-\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\bar{\boldsymbol{x}}_n-\boldsymbol{\mu}_x)=0 \qquad (4 - 16)$$

Second, let us consider $(4-13)$. Similarly, since $\boldsymbol{\Sigma}_{\mathbf{e}}^{-1}$ is positive definite, each term in $(4-13)$ is greater than or equal to zero. To minimize the first term in $(4-13)$, i.e.,

$$Min\left\{\frac{1}{2}tr\left\{\sum_{j=1}^{m}\boldsymbol{\Sigma}_e^{-1}\left[\boldsymbol{y_j}-\overline{\boldsymbol{y}}_m-\boldsymbol{\beta}^T\left(\boldsymbol{x_j}-\overline{\mathbf{x}}_{\boldsymbol{m}}\right)\right]\left[\boldsymbol{y_j}-\overline{\boldsymbol{y}}_m-\boldsymbol{\beta}^T\left(\boldsymbol{x_j}-\overline{\mathbf{x}}_{\boldsymbol{m}}\right)\right]^T\right\}\right\}$$

We take derivative with respect to $\boldsymbol{\beta}$ first, then set the derivative to zero, and obtain the MLE for $\boldsymbol{\beta}$ which makes the above minimum. We used the derivatives of trace for the first and second order in Petersen and Pedersen (2012),

$$\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\frac{1}{2}trace\boldsymbol{\Sigma}_e^{-1}\sum_{j=1}^{m}\left[\mathbf{y_j}-\overline{\boldsymbol{y}}_m-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\right]\left[\mathbf{y_j}-\overline{\boldsymbol{y}}_m-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\right]^T\right\}$$

$$=\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\beta}}\left\{trace\sum_{j=1}^{m}\boldsymbol{\Sigma}_e^{-1}\left[\mathbf{y_j}-\overline{\boldsymbol{y}}_m-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\right]\left[\mathbf{y_j}-\overline{\boldsymbol{y}}_m-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\right]^T\right\}$$

$$=\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\beta}}\left\{trace\sum_{j=1}^{m}\boldsymbol{\Sigma}_e^{-1}\left[\left(\mathbf{y_j}-\overline{\boldsymbol{y}}_m\right)\left(\mathbf{y_j}-\overline{\boldsymbol{y}}_m\right)^T-\boldsymbol{\beta}^T\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\left(\mathbf{y_j}-\overline{\boldsymbol{y}}_m\right)^T\right.\right.$$

$$\left.\left.-\left(\mathbf{y_j}-\overline{\boldsymbol{y}}_m\right)\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)^T\boldsymbol{\beta}+\boldsymbol{\beta}^T\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)^T\boldsymbol{\beta}\right]\right\}$$

$$=\frac{1}{2}\sum_{J=1}^{m}\left\{0-\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\left(\mathbf{y_j}-\overline{\boldsymbol{y}}_m\right)^T\boldsymbol{\Sigma}_e^{-1}-\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\left(\mathbf{y_j}-\overline{\boldsymbol{y}}_m\right)^T\boldsymbol{\Sigma}_e^{-1}+\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)^T\boldsymbol{\beta}\boldsymbol{\Sigma}_e^{-1}\right.$$

$$\left.+\left[\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)^T\right]^T\boldsymbol{\beta}[\boldsymbol{\Sigma}_e^{-1}]^T\right\}$$

$$=\sum_{J=1}^{m}-\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\left(\mathbf{y_j}-\overline{\boldsymbol{y}}_m\right)^T\boldsymbol{\Sigma}_e^{-1}+\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)^T\boldsymbol{\beta}\boldsymbol{\Sigma}_e^{-1} \qquad (4-17)$$

By equations (102) and (117) in Petersen and Pedersen (2012),

$$\frac{\partial}{\partial \boldsymbol{\beta}}\left\{trace\left[\boldsymbol{\Sigma}_e^{-1}\boldsymbol{\beta}^T(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{y_j}-\bar{\boldsymbol{y}}_m)^T\right]\right\} = (\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{y_j}-\bar{\boldsymbol{y}}_m)^T\boldsymbol{\Sigma}_e^{-1}$$

$$\frac{\partial}{\partial \boldsymbol{\beta}}\left\{trace\left[\boldsymbol{\Sigma}_e^{-1}(\mathbf{y_j}-\bar{\boldsymbol{y}}_m)(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})^T\boldsymbol{\beta}\right]\right\}$$

$$= \frac{\partial}{\partial \boldsymbol{\beta}}\left\{trace\left[\boldsymbol{\Sigma}_e^{-1}(\mathbf{y_j}-\bar{\boldsymbol{y}}_m)(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})^T\boldsymbol{\beta}\right]^T\right\} \qquad tr(\boldsymbol{A}) = tr(\boldsymbol{A}^T)$$

$$= \frac{\partial}{\partial \boldsymbol{\beta}}\left\{trace\left[\boldsymbol{\beta}^T(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{y_j}-\bar{\boldsymbol{y}}_m)^T\boldsymbol{\Sigma}_e^{-1}\right]\right\}$$

$$\frac{\partial}{\partial \boldsymbol{\beta}}\left\{trace\left[\boldsymbol{\Sigma}_e^{-1}\boldsymbol{\beta}^T(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{y_j}-\bar{\boldsymbol{y}}_m)^T\right]\right\}$$

$$= (\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{y_j}-\bar{\boldsymbol{y}}_m)^T\boldsymbol{\Sigma}_e^{-1}$$

$$\frac{\partial}{\partial \boldsymbol{\beta}}\left\{trace\left[\boldsymbol{\Sigma}_e^{-1}\boldsymbol{\beta}^T(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})^T\boldsymbol{\beta}\right]\right\}$$

$$= \frac{\partial}{\partial \boldsymbol{\beta}}\left\{trace\left[\boldsymbol{\beta}^T(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})^T\boldsymbol{\beta}\boldsymbol{\Sigma}_e^{-1}\right]\right\}$$

$$= (\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})^T\boldsymbol{\beta}\boldsymbol{\Sigma}_e^{-1} + \left[(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})^T\right]^T\boldsymbol{\beta}[\boldsymbol{\Sigma}_e^{-1}]^T$$

$$= 2(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})^T\boldsymbol{\beta}\boldsymbol{\Sigma}_e^{-1}$$

Set $(4-17) = \mathbf{0}$ to give

$$\sum_{J=1}^{m} -(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{y_j}-\bar{\boldsymbol{y}}_m)^T\boldsymbol{\Sigma}_e^{-1} + (\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})(\mathbf{x_j}-\bar{\mathbf{x}}_\mathbf{m})^T\boldsymbol{\beta}\boldsymbol{\Sigma}_e^{-1} = \mathbf{0} \qquad (4 - 18)$$

The second term $m[\bar{\boldsymbol{y}}_m - \boldsymbol{\mu_y} - \boldsymbol{\beta}^T(\bar{\boldsymbol{x}}_{\boldsymbol{m}} - \boldsymbol{\mu_x})]^T\boldsymbol{\Sigma}_e^{-1}[\bar{\boldsymbol{y}}_m - \boldsymbol{\mu_y} - \boldsymbol{\beta}^T(\bar{\mathbf{x}}_\mathbf{m} - \boldsymbol{\mu}_\mathbf{x})]/2$

can be minimized if we set

$$\frac{m}{2}[\bar{\boldsymbol{y}}_m - \boldsymbol{\mu_y} - \boldsymbol{\beta}^T(\bar{\boldsymbol{x}}_{\boldsymbol{m}} - \boldsymbol{\mu_x})]^T\boldsymbol{\Sigma}_e^{-1}[\bar{\boldsymbol{y}}_m - \boldsymbol{\mu_y} - \boldsymbol{\beta}^T(\bar{\boldsymbol{x}}_{\boldsymbol{m}} - \boldsymbol{\mu_x})] = 0 \qquad (4 - 19)$$

Simultaneously solving $(4-16)$, $(4-18)$ and $(4-19)$, we obtain the MLE for $\boldsymbol{\mu}_y$, $\boldsymbol{\mu}_x$ and $\boldsymbol{\beta}$ as follows:

$$\widehat{\boldsymbol{\mu}}_x = \overline{\boldsymbol{X}}_n \qquad (4 - 20)$$

$$\widehat{\boldsymbol{\mu}}_y = \overline{\boldsymbol{Y}}_m - \widehat{\boldsymbol{\beta}}^T(\overline{\boldsymbol{X}}_m - \overline{\boldsymbol{X}}_n) \qquad (4 - 21)$$

$$\widehat{\boldsymbol{\beta}} = \mathbf{S}_{\mathrm{xxm}}^{-1}\mathbf{S}_{\mathrm{xym}} \qquad (4 - 22)$$

where

$$\mathbf{S}_{\mathbf{xxm}} = \sum_{\mathbf{j=1}}^{\mathbf{m}}(\mathbf{X_j} - \overline{\mathbf{X}}_{\mathbf{m}})\big((\mathbf{X_j} - \overline{\mathbf{X}}_{\mathbf{m}}\big)^T \qquad (4 - 23)$$

$$\mathbf{S}_{\mathbf{xym}} = \sum_{j=1}^{m}\big(\mathbf{X_j} - \overline{\mathbf{X}}_{\mathbf{m}}\big)\,(\mathbf{Y}_{\mathrm{j}} - \overline{\mathbf{Y}}_{\mathbf{m}})^T \qquad (4 - 24)$$

Now back to maximize $(4-8)$ and $(4-9)$ simultaneously. Since when $\widehat{\boldsymbol{\mu}}_x = \overline{\boldsymbol{X}}_n$, $(4-8)$ is reduced to

$$|\boldsymbol{\Sigma}_{\mathbf{xx}}|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x_i} - \boldsymbol{\mu}_{\mathbf{x}})^T\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x_i} - \boldsymbol{\mu}_{\mathbf{x}})\right\}$$

$$= |\boldsymbol{\Sigma}_{\mathbf{xx}}|^{-\frac{n}{2}}exp\left\{-\frac{1}{2}tr(\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\sum_{i=1}^{n}(\mathbf{x_i} - \overline{\mathbf{x}}_{\mathbf{n}})(\mathbf{x_i} - \overline{\mathbf{x}}_{\mathbf{n}})^T\right\} \qquad (4 - 25)$$

By Results 4.10 in Johnson and Wichern (1998), $(4-25)$ reaches maximum when

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{xx}} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(\mathbf{X_i} - \overline{\mathbf{X}}_{\mathbf{n}})(\mathbf{X_i} - \overline{\mathbf{X}}_{\mathbf{n}})^T \qquad (4 - 26)$$

Similarly, when $\widehat{\boldsymbol{\mu}}_y = \overline{\boldsymbol{Y}}_m - \widehat{\boldsymbol{\beta}}^T(\overline{\boldsymbol{X}}_m - \overline{\boldsymbol{X}}_n)$, $\widehat{\boldsymbol{\mu}}_x = \overline{\boldsymbol{X}}_n$ and $\widehat{\boldsymbol{\beta}} = \mathbf{S}_{\mathrm{xxm}}^{-1}\mathbf{S}_{\mathrm{xym}}$, $(4-9)$ is reduced to

$$|\boldsymbol{\Sigma}_{\mathrm{e}}|^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{m}\left[\boldsymbol{y_j}-\boldsymbol{\mu_y}-\boldsymbol{\beta}^T\left(\boldsymbol{x_j}-\boldsymbol{\mu_x}\right)\right]^T\boldsymbol{\Sigma}_{\mathrm{e}}^{-1}\left[\boldsymbol{y_j}-\boldsymbol{\mu_y}-\boldsymbol{\beta}^T\left(\boldsymbol{x_j}-\boldsymbol{\mu_x}\right)\right]\right\}$$

$$= |\boldsymbol{\Sigma}_{\mathrm{e}}|^{-\frac{m}{2}} exp\left\{-\frac{1}{2} tr\boldsymbol{\Sigma}_e^{-1}\sum_{j=1}^{m}\left[\mathbf{y_j}-\overline{\boldsymbol{y}}_m-\widehat{\boldsymbol{\beta}}^T\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\right]\left[\mathbf{y_j}-\overline{\boldsymbol{y}}_m-\widehat{\boldsymbol{\beta}}^T\left(\mathbf{x_j}-\overline{\mathbf{x}}_{\mathbf{m}}\right)\right]^T\right\} \quad (4 - 27)$$

Again, by Results 4.10 in Johnson and Wichern (1998), $(4-27)$ reaches maximum when

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{e}} = \frac{1}{m}\sum_{j=1}^{m}\left[\boldsymbol{Y_j}-\overline{\boldsymbol{Y}}_m-\widehat{\boldsymbol{\beta}}^T\left(\boldsymbol{X_j}-\overline{\boldsymbol{X}}_{\boldsymbol{m}}\right)\right]\left[\boldsymbol{Y_j}-\overline{\boldsymbol{Y}}_m-\widehat{\boldsymbol{\beta}}^T\left(\boldsymbol{X_j}-\overline{\boldsymbol{X}}_{\boldsymbol{m}}\right)\right]^T \quad (4 - 28)$$

In summary, we have following 5 maximum likelihood estimators:

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}} = \overline{\boldsymbol{X}}_{\boldsymbol{n}} \quad (4 - 29)$$

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}} = \overline{\boldsymbol{Y}}_m - \widehat{\boldsymbol{\beta}}^T(\overline{\boldsymbol{X}}_{\boldsymbol{m}} - \overline{\boldsymbol{X}}_{\boldsymbol{n}}) \quad (4 - 30)$$

$$\widehat{\boldsymbol{\beta}} = \mathbf{S}_{\mathrm{xxm}}^{-\mathbf{1}}\mathbf{S}_{\mathrm{xym}} \quad (4 - 31)$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{xx}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X_i}-\overline{\mathbf{X}}_{\boldsymbol{n}})\,(\mathbf{X_i}-\overline{\mathbf{X}}_{\boldsymbol{n}})^T \quad (4 - 32)$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{e}} = \frac{1}{m}\sum_{j=1}^{m}\left[\boldsymbol{Y_j}-\overline{\boldsymbol{Y}}_m-\widehat{\boldsymbol{\beta}}^T\left(\boldsymbol{X_j}-\overline{\boldsymbol{X}}_{\boldsymbol{m}}\right)\right]\left[\boldsymbol{Y_j}-\overline{\boldsymbol{Y}}_m-\widehat{\boldsymbol{\beta}}^T\left(\boldsymbol{X_j}-\overline{\boldsymbol{X}}_{\boldsymbol{m}}\right)\right]^T \quad (4 - 33)$$

Similarly, if we do not consider extra information $\boldsymbol{X}_{\mathrm{m+1,}}\boldsymbol{X}_{\mathrm{m+2,}}\dots,\boldsymbol{X}_{\mathrm{n}}$ and only use the first m observations, we have

$$\widehat{\boldsymbol{\mu}}_{x\_no} = \overline{\boldsymbol{X}}_{\boldsymbol{m}} \quad (4 - 34)$$

$$\widehat{\boldsymbol{\mu}}_{y\_no} = \overline{\boldsymbol{Y}}_{\mathbf{m}} \tag{4 - 35}$$

$$\widehat{\boldsymbol{\beta}}_{no} = \mathbf{S}_{\mathrm{xxm}}^{-1}\mathbf{S}_{\mathrm{xym}} \tag{4 - 36}$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{xx}\_no} = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{X_i} - \overline{\mathbf{X}}_{\boldsymbol{m}})(\mathbf{X_i} - \overline{\mathbf{X}}_{\boldsymbol{m}})^T \tag{4 - 37}$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{e}\_no} = \frac{1}{m}\sum_{j=1}^{m}\left[\boldsymbol{Y_j} - \overline{\boldsymbol{Y}}_m - \widehat{\boldsymbol{\beta}}^T\left(\boldsymbol{X_j} - \overline{\mathbf{X}}_{\boldsymbol{m}}\right)\right]\left[\boldsymbol{Y_j} - \overline{\boldsymbol{Y}}_m - \widehat{\boldsymbol{\beta}}^T\left(\boldsymbol{X_j} - \overline{\mathbf{X}}_{\boldsymbol{m}}\right)\right]^T \tag{4 - 38}$$

4.2 Properties of the Maximum Likelihood Estimators

4.2.1 Estimator of the Mean Vector of $\boldsymbol{X}$

The expectation of $\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}}$ is

$$E(\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}}) = E(\overline{\boldsymbol{X}}_{\boldsymbol{n}}) = E\left[\bar{X}_{n1}, \bar{X}_{n2}, \dots, \bar{X}_{np}\right]^T = \left[E(\bar{X}_{n1}), E(\bar{X}_{n2}), \dots, E(\bar{X}_{np})\right]^T$$

$$= \left[\mu_{x1}, \mu_{x2}, \dots \mu_{xp}\right]^T = \boldsymbol{\mu}_{\boldsymbol{x}} \tag{4 - 39}$$

So $\widehat{\boldsymbol{\mu}}_{\mathbf{x}}$ is an unbiased estimator. The covariance of $\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}}$ is

$$\mathrm{Cov}(\widehat{\boldsymbol{\mu}}_{\mathbf{x}}) = \mathrm{Cov}(\overline{\mathbf{X}}_{\mathbf{n}}) = \mathrm{E}(\overline{\mathbf{X}}_{\mathbf{n}} - \boldsymbol{\mu}_{\mathbf{x}})(\overline{\mathbf{X}}_{\mathbf{n}} - \boldsymbol{\mu}_{\mathbf{x}})^T = \frac{1}{n^2}\left\{\sum_{j=1}^{n}\sum_{l=1}^{n}\mathrm{E}(\mathbf{X_j} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{X_l} - \boldsymbol{\mu}_{\mathbf{x}})^T\right\}$$

$$= \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{E}\left(\mathbf{X_j} - \boldsymbol{\mu}_{\mathbf{x}}\right)\left(\mathbf{X_j} - \boldsymbol{\mu}_{\mathbf{x}}\right)^T = \frac{1}{n}\boldsymbol{\Sigma}_{\mathrm{xx}} \tag{4 - 40}$$

By our assumptions,

$$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\mathrm{xx}})$$

Hence $\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}} = \overline{\boldsymbol{X}}_{\boldsymbol{n}}$ is distributed as

$$\hat{\boldsymbol{\mu}}_x \sim N_p(\boldsymbol{\mu}_x, \frac{1}{n}\boldsymbol{\Sigma}_{\mathrm{xx}})$$

4.2.2 Estimator of the Covariance Matrix of $\boldsymbol{X}$

Since $\boldsymbol{X}_{1,}\boldsymbol{X}_{2,} \dots, \boldsymbol{X}_{\mathrm{n}}$ is a random sample of size n from a p-variate normal distribution with mean $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{\Sigma}_{\mathrm{xx}}$, so

$$\sum_{i=1}^{n}(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_n)(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_n)^T \sim W_p(\boldsymbol{\Sigma}_{\mathrm{xx}}, n-1)$$

where $W_p(\boldsymbol{\Sigma}_{\mathbf{xx}}, n-1)$ is Wishart distribution with $(n-1)$ degree of freedom.

We have

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{xx}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_n)(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_n)^T$$

So

$$n\hat{\boldsymbol{\Sigma}}_{\mathrm{xx}} \sim W_p(\boldsymbol{\Sigma}_{\mathrm{xx}}, n-1)$$

Then by Nydick (2012), we have

$$E(n\hat{\boldsymbol{\Sigma}}_{\mathrm{xx}}) = (n-1)\boldsymbol{\Sigma}_{\mathrm{xx}}$$

$$Var\left(n\hat{\Sigma}_{ij}\right) = (n-1)(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj})$$

The expectation of $\hat{\boldsymbol{\Sigma}}_{\mathrm{xx}}$ is

$$E(\hat{\boldsymbol{\Sigma}}_{\mathrm{xx}}) = \frac{n-1}{n}\boldsymbol{\Sigma}_{\mathrm{xx}} \qquad (4 - 41)$$

So $\hat{\boldsymbol{\Sigma}}_{\mathbf{xx}}$ is a biased estimator.

$$Var\left(\hat{\Sigma}_{ij}\right) = \frac{n-1}{n^2}\left(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}\right) \qquad (4 - 42)$$

If we define

$$\mathbf{S}_{\mathrm{xn}} = \frac{n}{n-1}\widehat{\mathbf{\Sigma}}_{\mathrm{xx}} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_{\boldsymbol{i}} - \overline{\boldsymbol{X}}_n)(\boldsymbol{X}_{\boldsymbol{i}} - \overline{\boldsymbol{X}}_n)^T \tag{4 - 43}$$

Then we have

$$E(\mathbf{S}_{\mathrm{xn}}) = \frac{n}{n-1}E\left(\widehat{\mathbf{\Sigma}}_{\mathrm{xx}}\right) = \mathbf{\Sigma}_{\mathrm{xx}}$$

$\mathbf{S}_{\mathrm{xn}}$ is an unbiased estimator for $\mathbf{\Sigma}_{\mathrm{xx}}$.

4.2.3 Estimator of the Regression Coefficient Matrix

As we do in Chapter 3, we will derive the conditional expectation and covariance matrix of $\widehat{\boldsymbol{\beta}}$ given $\boldsymbol{X} = \mathbf{x}$ first, then derive the unconditional expectation and covariance matrix of the estimator.

The conditional expectation of $\widehat{\boldsymbol{\beta}}$ given $\boldsymbol{X} = \mathbf{x}$ is

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}|\mathbf{x}) = \mathrm{E}(\mathbf{S}_{\mathrm{xxm}}^{-1}\mathbf{S}_{\mathrm{xym}}|\mathbf{x}) = \mathbf{S}_{\mathrm{xxm}}^{-1}\mathrm{E}(\mathbf{S}_{\mathrm{xym}}|\mathbf{x}) = \mathbf{S}_{\mathrm{xxm}}^{-1}\mathrm{E}\left\{\sum_{j=1}^{m}\left(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_{\mathbf{m}}\right)\left(\mathbf{Y}_{\mathrm{j}} - \overline{\mathbf{Y}}_{\mathbf{m}}\right)^T|\mathbf{x}\right\}$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\sum_{j=1}^{m}\left(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_{\mathbf{m}}\right)\mathrm{E}(\mathbf{Y}_j^T\,|\mathbf{x}) = \mathbf{S}_{\mathrm{xxm}}^{-1}\sum_{j=1}^{m}\left(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_{\mathbf{m}}\right)\left[\boldsymbol{\mu}_{\mathbf{y}} + \boldsymbol{\beta}^T\left(\mathbf{x}_{\boldsymbol{j}} - \boldsymbol{\mu}_{\mathbf{x}}\right)\right]^T$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\sum_{j=1}^{m}\left(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_{\mathbf{m}}\right)\left[\boldsymbol{\mu}_{\boldsymbol{y}}^T + \left(\mathbf{x}_{\boldsymbol{j}} - \boldsymbol{\mu}_{\mathbf{x}}\right)^T\boldsymbol{\beta}\right]$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\sum_{j=1}^{m}\left(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_{\mathbf{m}}\right)\mathbf{x}_{\boldsymbol{j}}^T\boldsymbol{\beta} = \mathbf{S}_{\mathrm{xxm}}^{-1}\sum_{\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{m}}(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_{\mathbf{m}})\left(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_{\mathbf{m}}\right)^T\boldsymbol{\beta}$$

$$= \mathbf{S}_{\mathrm{xxm}}^{-1}\mathbf{S}_{\mathrm{xxm}}\boldsymbol{\beta} = \boldsymbol{\beta} \tag{4 - 44}$$

So we have the unconditional expectation of $\widehat{\boldsymbol{\beta}}$

$$E\left(\widehat{\boldsymbol{\beta}}\right) = E\left[E\left(\widehat{\boldsymbol{\beta}}\middle|\mathbf{x}\right)\right] = E(\boldsymbol{\beta}) = \boldsymbol{\beta} \tag{4 - 45}$$

$\widehat{\boldsymbol{\beta}}$ is an unbiased estimator.

We use **vec**-operator to obtain the conditional covariance matrix of $\widehat{\boldsymbol{\beta}}$ given $\boldsymbol{X} = \mathbf{x}$. Since

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{p1} & \beta_{p2} & \cdots & \beta_{pq} \end{bmatrix} = \left[\boldsymbol{\beta}_{(1)} \vdots \boldsymbol{\beta}_{(2)} \vdots \cdots \vdots \boldsymbol{\beta}_{(q)}\right] \qquad (4 - 46)$$

so

$$\mathbf{vec}(\boldsymbol{\beta}) = \left[\beta_{11}, \beta_{21}, \dots, \beta_{p1}, \beta_{12}, \dots, \beta_{p2}, \dots, \beta_{1q}, \dots, \beta_{pq}\right]^T = [\boldsymbol{\beta}_{(1)}, \boldsymbol{\beta}_{(2)}, \dots, \boldsymbol{\beta}_{(q)}]^T \qquad (4 - 47)$$

Then by Loan (2009), we have

$$\mathrm{Cov}\left[\mathbf{vec}\left(\widehat{\boldsymbol{\beta}}|\mathbf{x}\right)\right] = \mathrm{Cov}\left[\mathbf{vec}(\mathbf{S}_{\mathrm{xxm}}^{-1}\mathbf{S}_{\mathrm{xym}}|\mathbf{x})\right]$$

$$= \mathrm{Cov}\left\{\mathbf{vec}\sum_{j=1}^{m} \mathbf{S}_{\mathrm{xxm}}^{-1}\left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)\left[\left(\mathbf{Y}_\mathrm{j} - \bar{\mathbf{Y}}_\mathbf{m}\right)^T|\mathbf{x}\right]\right\}$$

$$= \mathrm{Cov}\left\{\mathbf{vec}\sum_{j=1}^{m} \mathbf{S}_{\mathrm{xxm}}^{-1}\left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)(\mathbf{Y}_\mathrm{j}{}^T|\mathbf{x})\right\}$$

$$= \mathrm{Cov}\left\{\sum_{j=1}^{m}\left[\mathbf{I} \otimes \mathbf{S}_{\mathrm{xxm}}^{-1}\left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)\right]\mathbf{vec}(\mathbf{Y}_\mathrm{j}{}^T|\mathbf{x})\right\}$$

$$= \sum_{j=1}^{m}\left[\mathbf{I} \otimes \mathbf{S}_{\mathrm{xxm}}^{-1}\left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)\right]\mathrm{Cov}\left[\mathbf{vec}(\mathbf{Y}_\mathrm{j}{}^T|\mathbf{x})\right]\left[\mathbf{I} \otimes \mathbf{S}_{\mathrm{xxm}}^{-1}\left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)\right]^T \quad \textit{independent of } \boldsymbol{Y}_j$$

$$= \sum_{j=1}^{m}\left[\mathbf{I} \otimes \mathbf{S}_{\mathrm{xxm}}^{-1}\left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)\right]\mathrm{Cov}\left(\mathbf{Y}_j|\mathbf{x}\right)\left[\mathbf{I} \otimes \left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)^T\mathbf{S}_{\mathrm{xxm}}^{-1}\right]$$

$$= \sum_{j=1}^{m}\left[\mathbf{I} \otimes \mathbf{S}_{\mathrm{xxm}}^{-1}\left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)\right]\boldsymbol{\Sigma}_e\left[\mathbf{I} \otimes \left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)^T\mathbf{S}_{\mathrm{xxm}}^{-1}\right]$$

$$= \sum_{j=1}^{m}\left[\mathbf{I} \otimes \mathbf{S}_{\mathrm{xxm}}^{-1}\left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)\right]\left[\boldsymbol{\Sigma}_e \otimes \left(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}\right)^T\mathbf{S}_{\mathrm{xxm}}^{-1}\right]$$

$$= \sum_{j=1}^{m} \boldsymbol{\Sigma}_e \otimes \mathbf{S}_{\mathrm{xxm}}^{-1} (\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}) (\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T \mathbf{S}_{\mathrm{xxm}}^{-1}$$

$$= \boldsymbol{\Sigma}_e \otimes \sum_{j=1}^{m} \mathbf{S}_{\mathrm{xxm}}^{-1} (\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}) (\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T \mathbf{S}_{\mathrm{xxm}}^{-1}$$

$$= \boldsymbol{\Sigma}_e \otimes \mathbf{S}_{\mathrm{xxm}}^{-1} \left[\sum_{j=1}^{m} (\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}) (\mathbf{x} \quad - \bar{\mathbf{x}}_\mathbf{m})^T \right] \mathbf{S}_{\mathrm{xxm}}^{-1}$$

$$= \boldsymbol{\Sigma}_e \otimes \mathbf{S}_{\mathrm{xxm}}^{-1} \qquad (4 - 48)$$

where $\otimes$ stands for Kronecker Product.

By the Law of Total Covariance and by Nydick (2012), the unconditional covariance matrix of $\mathbf{vec}(\hat{\boldsymbol{\beta}})$ is

$$\mathrm{Cov}[\mathbf{vec}(\hat{\boldsymbol{\beta}})] = E\{\mathrm{Cov}\,[\mathbf{vec}(\hat{\boldsymbol{\beta}}|\mathbf{x})]\} + \mathrm{Cov}\{E[\mathbf{vec}(\hat{\boldsymbol{\beta}}|\mathbf{x})]\}$$

$$= E[\boldsymbol{\Sigma}_e \otimes \mathbf{S}_{\mathrm{xxm}}^{-1}] + \mathrm{Cov}[\mathbf{vec}(\boldsymbol{\beta})]$$

$$= \boldsymbol{\Sigma}_e \otimes \mathrm{E}(\mathbf{S}_{\mathrm{xxm}}^{-1}\,) + \mathbf{0}$$

$$= \frac{1}{m-p-2} \boldsymbol{\Sigma}_e \otimes \boldsymbol{\Sigma}_{\mathrm{xx}}^{-1} \qquad (4 - 49)$$

Then we have

$$\mathrm{Cov}(\boldsymbol{\beta}_{(i)}, \boldsymbol{\beta}_{(j)}) = \frac{\boldsymbol{\Sigma}_{e(i,j)}}{m-p-2} \boldsymbol{\Sigma}_{\mathrm{xx}}^{-1} \qquad (4 - 50)$$

$$i, j = 1,2, \dots, q$$

When sample is large, $\hat{\boldsymbol{\beta}}$ is asymptotically normally distributed.

4.2.4 Estimator of the Mean Vector of $\boldsymbol{Y}$

As we do in 4.2.3, first we will derive the conditional expectation and covariance matrix of $\hat{\boldsymbol{\mu}}_y$ given $\boldsymbol{X} = \mathbf{x}$, then we derive the unconditional expectation and covariance matrix of the estimator. The conditional expectation of $\hat{\boldsymbol{\mu}}_y$ is

$$E(\hat{\boldsymbol{\mu}}_y|\mathbf{x}) = E\{[\overline{\boldsymbol{Y}}_{\mathbf{m}} - \hat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_{\boldsymbol{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})]|\mathbf{x}\} = E(\overline{\boldsymbol{Y}}_{\mathbf{m}}|\mathbf{x}) - E(\hat{\boldsymbol{\beta}}^T|\mathbf{x})(\bar{\mathbf{x}}_{\boldsymbol{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})$$

$$= \boldsymbol{\mu}_y + \boldsymbol{\beta}^T(\bar{\mathbf{x}}_{\boldsymbol{m}} - \boldsymbol{\mu}_x) - \boldsymbol{\beta}^T(\bar{\mathbf{x}}_{\boldsymbol{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}}) = \boldsymbol{\mu}_y + \boldsymbol{\beta}^T(\bar{\mathbf{x}}_{\boldsymbol{n}} - \boldsymbol{\mu}_x) \qquad (4 - 51)$$

where

$$E(\overline{\boldsymbol{Y}}_{\mathbf{m}}|\mathbf{x}) = E\left(\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{Y}_j\,|\mathbf{x}_j\right) = \frac{1}{m}\sum_{j=1}^{m}E(\boldsymbol{Y}_j|\mathbf{x}_j) = \frac{1}{m}\sum_{j=1}^{m}[\boldsymbol{\mu}_y + \boldsymbol{\beta}^T(\mathbf{x}_{\boldsymbol{j}} - \boldsymbol{\mu}_x)]$$

$$= \frac{1}{m}\left[m\boldsymbol{\mu}_y + \boldsymbol{\beta}^T\left(\sum_{j=1}^{m}\mathbf{x}_{\mathbf{j}} - m\boldsymbol{\mu}_x\right)\right] = \boldsymbol{\mu}_y + \boldsymbol{\beta}^T(\bar{\mathbf{x}}_{\boldsymbol{m}} - \boldsymbol{\mu}_x) \qquad (4 - 52)$$

So the expectation of $\hat{\boldsymbol{\mu}}_y$ is

$$E(\hat{\boldsymbol{\mu}}_y) = E[E(\hat{\boldsymbol{\mu}}_y|\mathbf{X})] = E[\boldsymbol{\mu}_y + \boldsymbol{\beta}^T(\overline{\mathbf{X}}_{\boldsymbol{n}} - \boldsymbol{\mu}_x)] = \boldsymbol{\mu}_y + \boldsymbol{\beta}^T(\boldsymbol{\mu}_x - \boldsymbol{\mu}_x) = \boldsymbol{\mu}_y \qquad (4 - 53)$$

$\hat{\boldsymbol{\mu}}_y$ is an unbiased estimator.

The conditional covariance matrix of $\hat{\boldsymbol{\mu}}_y$ is

$$\mathrm{Cov}(\hat{\boldsymbol{\mu}}_y|\mathbf{x}) = \mathrm{Cov}\{[\overline{\boldsymbol{Y}}_{\mathbf{m}} - \hat{\boldsymbol{\beta}}^T(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}})]|\mathbf{x}\}$$

$$= \mathrm{Cov}(\overline{\boldsymbol{Y}}_{\mathbf{m}}|\mathbf{x}) + \mathrm{Cov}[\hat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_{\mathbf{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})|\mathbf{x}] - 2\mathrm{Cov}[\overline{\boldsymbol{Y}}_{\mathbf{m}}, \hat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_{\mathbf{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})|\mathbf{x}]$$

$$= \frac{1}{m}\boldsymbol{\Sigma}_e + \boldsymbol{\Sigma}_e\,[(\bar{\mathbf{x}}_{\mathbf{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})^T\mathbf{S}_{\mathrm{xxm}}^{-\mathbf{1}}(\bar{\mathbf{x}}_{\mathbf{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})] \qquad (4 - 54)$$

where

$$\mathrm{Cov}(\overline{\boldsymbol{Y}}_{\mathbf{m}}|\mathbf{x}) = \frac{1}{m}\boldsymbol{\Sigma}_e \qquad (4 - 55)$$

$$\mathrm{Cov}[\hat{\boldsymbol{\beta}}^T(\bar{\mathbf{x}}_{\mathbf{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})|\mathbf{x}] = \mathrm{Cov}\left[(\mathbf{S}_{\mathrm{xxm}}^{-\mathbf{1}}\mathbf{S}_{\mathrm{xym}})^T(\bar{\mathbf{x}}_{\mathbf{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})\right]|\mathbf{x} = \mathrm{Cov}[(\mathbf{S}_{\mathrm{yxm}}\mathbf{S}_{\mathrm{xxm}}^{-\mathbf{1}})(\bar{\mathbf{x}}_{\mathbf{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})]|\mathbf{x}$$

$$= \mathrm{Cov}\left[\sum_{j=1}^{m}(\mathbf{Y}_{\mathrm{j}} - \overline{\mathbf{Y}}_{\mathbf{m}})(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_{\mathbf{m}})^T\right]\mathbf{S}_{\mathrm{xxm}}^{-\mathbf{1}}(\bar{\mathbf{x}}_{\mathbf{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})|\mathbf{x}$$

$$= \mathrm{Cov}\left\{\sum_{j=1}^{m}(\mathbf{Y}_{\mathrm{j}}|\mathbf{x})(\mathbf{x}_{\mathbf{j}} - \bar{\mathbf{x}}_{\mathbf{m}})^T\mathbf{S}_{\mathrm{xxm}}^{-\mathbf{1}}(\bar{\mathbf{x}}_{\mathbf{m}} - \bar{\mathbf{x}}_{\boldsymbol{n}})\right\}$$

$$= \sum_{j=1}^{m} \text{Cov}\left[(\mathbf{Y}_\text{j}|\mathbf{x})(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T \mathbf{S}_\text{xxm}^{-1} (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})\right] \qquad \textit{independent of } \mathbf{Y}_\text{j}$$

$(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T \mathbf{S}_\text{xxm}^{-1} (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})$ is a scalar. Let

$$a_j = (\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T \mathbf{S}_\text{xxm}^{-1} (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})$$

Then

$$a_j^2 = \left[(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T \mathbf{S}_\text{xxm}^{-1} (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})\right]^T \left[(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T \mathbf{S}_\text{xxm}^{-1} (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})\right]$$

$$= \left[(\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})^T \mathbf{S}_\text{xxm}^{-1} (\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})\right] \left[(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T \mathbf{S}_\text{xxm}^{-1} (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})\right]$$

Hence we have

$$\text{Cov}\left[\widehat{\boldsymbol{\beta}}^T (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})|\mathbf{x}\right] = \sum_{j=1}^{m} \text{Cov}\left[(\mathbf{Y}_\text{j}|\mathbf{x}) a_j\right] = \sum_{j=1}^{m} a_j^2 Var\left[(\mathbf{Y}_\text{j}|\mathbf{x})\right] = \sum_{j=1}^{m} a_j^2 \boldsymbol{\Sigma}_e = \boldsymbol{\Sigma}_e \sum_{j=1}^{m} a_j^2$$

$$= \boldsymbol{\Sigma}_e \sum_{j=1}^{m} \left[(\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})^T \mathbf{S}_\text{xxm}^{-1} (\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})\right] \left[(\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T \mathbf{S}_\text{xxm}^{-1} (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})\right]$$

$$= \boldsymbol{\Sigma}_e (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})^T \mathbf{S}_\text{xxm}^{-1} \left[\sum_{\boldsymbol{j=1}}^{\boldsymbol{m}} (\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m}) (\mathbf{x_j} - \bar{\mathbf{x}}_\mathbf{m})^T\right] \mathbf{S}_\text{xxm}^{-1} (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})$$

$$= \boldsymbol{\Sigma}_e (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})^T \mathbf{S}_\text{xxm}^{-1} (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}}) \qquad (4 - 56)$$

$$\text{Cov}\left[\bar{\boldsymbol{Y}}_\mathbf{m}, \widehat{\boldsymbol{\beta}}^T (\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})\right]|\mathbf{x} = \text{Cov}\left[(\bar{\boldsymbol{Y}}_\mathbf{m}, \widehat{\boldsymbol{\beta}}^T)|\mathbf{x}\right](\bar{\mathbf{x}}_\mathbf{m} - \bar{\mathbf{x}}_{\boldsymbol{n}})^T = \mathbf{0} \qquad (4 - 57)$$

To obtain the unconditional covariance matrix of $\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}}$, we use the Law of Total Covariance,

$$\text{Cov}(\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}}) = \text{Cov}\left[E(\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}}|\mathbf{x})\right] + E\left[\text{Cov}(\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}}|\mathbf{x})\right]$$

now

$$\text{Cov}\left[E(\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}}|\mathbf{x})\right] = \text{Cov}\left[\boldsymbol{\mu}_{\boldsymbol{y}} + \boldsymbol{\beta}^T (\bar{\mathbf{X}}_{\boldsymbol{n}} - \boldsymbol{\mu}_{\boldsymbol{x}})\right] = \boldsymbol{\beta}^T \text{Cov}(\bar{\mathbf{X}}_{\boldsymbol{n}}) \boldsymbol{\beta} = \frac{1}{n} \boldsymbol{\beta}^T \boldsymbol{\Sigma}_\text{xx} \boldsymbol{\beta} \qquad (4 - 58)$$

To obtain $E\ [(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}})^T \mathbf{S}_{\mathrm{xxm}}^{-1} (\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}})]$, we need to find the distribution of $\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}}$. Since

$$\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}} = \frac{n-m}{n}(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n-m}}) \tag{4 - 59}$$

$\overline{\mathbf{X}}_m$ and $\overline{\mathbf{X}}_{n-m}$ are independent and normally distributed, and

$$E(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n-m}}) = \boldsymbol{\mu}_{\boldsymbol{x}} - \boldsymbol{\mu}_{\boldsymbol{x}} = \mathbf{0} \tag{4 - 60}$$

$$\mathrm{Cov}(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n-m}}) = \mathrm{Cov}(\overline{\mathbf{X}}_{\mathbf{m}}) + \mathrm{Cov}(\overline{\mathbf{X}}_{\boldsymbol{n-m}})$$

$$= \frac{1}{m}\mathbf{\Sigma}_{\mathrm{xx}} + \frac{1}{n-m}\mathbf{\Sigma}_{\mathrm{xx}} = \frac{n}{m(n-m)}\mathbf{\Sigma}_{\mathrm{xx}} \tag{4 - 61}$$

So we have

$$\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n-m}} \sim N_p\left(\mathbf{0}, \frac{n}{m(n-m)}\mathbf{\Sigma}_{\mathbf{xx}}\right)$$

Hence

$$(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}})^T \mathbf{S}_{\mathrm{xxm}}^{-1} (\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}}) = \left(\frac{n-m}{n}\right)^2 (\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n-m}})^T \mathbf{S}_{\mathrm{xxm}}^{-1} (\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n-m}})$$

$$= \left(\frac{n-m}{n}\right)^2 \cdot \frac{n}{m(n-m)} \cdot \frac{1}{m-1} \frac{(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n-m}})^T}{\sqrt{\frac{n}{m(n-m)}}} \left(\frac{\mathbf{S}_{\mathrm{xxm}}}{m-1}\right)^{-1} \frac{(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n-m}})}{\sqrt{\frac{n}{m(n-m)}}}$$

$$= \left(\frac{n-m}{n}\right)^2 \cdot \frac{n}{m(n-m)} \cdot \frac{1}{m-1} T_{p,m-1}^2$$

where

$$T_{p,m-1}^2 = \frac{(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n-m}})^T}{\sqrt{\frac{n}{m(n-m)}}} \left(\frac{\mathbf{S}_{\mathrm{xxm}}}{m-1}\right)^{-1} \frac{(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n-m}})}{\sqrt{\frac{n}{m(n-m)}}} \tag{4 - 62}$$

and

$$E\left(T_{p,m-1}^2\right) = \frac{(m-1)p}{m-p} E\left(F_{p,m-p}\right) = \frac{(m-1)p}{m-p} \cdot \frac{m-p}{m-p-2} = \frac{(m-1)p}{m-p-2}$$

The expectation of the conditional covariance matrix of $\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}}$ is

$$E\left[\mathrm{Cov}(\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}}|\mathbf{x})\right] = E\left\{\frac{1}{m}\boldsymbol{\Sigma}_e + \boldsymbol{\Sigma}_e\left[(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}})^T\mathbf{S}_{\mathrm{xxm}}^{-1}(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}})\right]\right\}$$

$$= \frac{1}{m}\boldsymbol{\Sigma}_e + \boldsymbol{\Sigma}_e E\left[(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}})^T\mathbf{S}_{\mathrm{xxm}}^{-1}(\overline{\mathbf{X}}_{\mathbf{m}} - \overline{\mathbf{X}}_{\boldsymbol{n}})\right]$$

$$= \frac{1}{m}\boldsymbol{\Sigma}_e + \boldsymbol{\Sigma}_e\left(\frac{n-m}{n}\right)^2 \cdot \frac{n}{m(n-m)} \cdot \frac{1}{m-1} \cdot E\left(T_{p,m-1}^2\right)$$

$$= \frac{1}{m}\boldsymbol{\Sigma}_e + \boldsymbol{\Sigma}_e\left(\frac{n-m}{n}\right)^2 \cdot \frac{n}{m(n-m)} \cdot \frac{1}{m-1} \cdot \frac{(m-1)p}{m-p-2}$$

$$= \frac{1}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right]\boldsymbol{\Sigma}_e \qquad (4 - 63)$$

Using $(4-58)$ and $(4-63)$, we have the unconditional covariance matrix of $\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}}$ as

$$\mathrm{Cov}(\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}}) = \frac{1}{m}\left[1 + \frac{(n-m)p}{n(m-p-2)}\right]\boldsymbol{\Sigma}_e + \frac{1}{n}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\mathrm{xx}}\boldsymbol{\beta} \qquad (4 - 64)$$

When sample is large, $\widehat{\boldsymbol{\mu}}_{\boldsymbol{y}}$ is asymptotically normally distributed.

*4.2.5 Estimator of the Conditional Covariance Matrix of **Y** given **x***

We use similar idea for the multiple regression model in Chapter 3. For given $\boldsymbol{x_j}$,

$$\boldsymbol{Y}_j = \boldsymbol{\mu}_y + \boldsymbol{\beta}^T\left(\boldsymbol{x_j} - \boldsymbol{\mu_x}\right) + \boldsymbol{\varepsilon}_j, j = 1,2,\dots,m \qquad (4 - 65)$$

where

$$\mathbf{Y}_{\mathrm{j}} = \left[Y_{j1}, Y_{j2}, \dots, Y_{jq}\right]^T$$

$$\boldsymbol{\mu}_y = \left[\mu_{y1}, \mu_{y2}, \dots, \mu_{yq}\right]^T$$

$$\boldsymbol{\beta}^T = \begin{bmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{p1} \\ \beta_{12} & \beta_{22} & \cdots & \beta_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{1q} & \beta_{2q} & \cdots & \beta_{pq} \end{bmatrix}$$

$$\boldsymbol{\varepsilon}_j = \left[\varepsilon_{j1}, \varepsilon_{j2}, \dots \varepsilon_{jq}\right]^T$$

$$\mathrm{Cov}(\mathbf{Y}_j|\mathbf{x}) = \mathrm{Cov}(\boldsymbol{\varepsilon}_j) = \boldsymbol{\Sigma}_e \tag{4 - 66}$$

$$E(\boldsymbol{\varepsilon}\quad) = E(\mathbf{Y}_j|\mathbf{x}) - \boldsymbol{\mu}_y - \boldsymbol{\beta}^T(\boldsymbol{x_j} - \boldsymbol{\mu_x}) = \mathbf{0} \tag{4 - 67}$$

We have

$$\begin{aligned}
\hat{\boldsymbol{\varepsilon}}_j &= \boldsymbol{Y}_j - \hat{\boldsymbol{\mu}}_y - \hat{\boldsymbol{\beta}}^T(\boldsymbol{x_j} - \hat{\boldsymbol{\mu}}_{\boldsymbol{x}}) \\
&= \boldsymbol{Y}_j - [\overline{\boldsymbol{Y}}_m - \hat{\boldsymbol{\beta}}^T(\overline{\boldsymbol{x}}_{\boldsymbol{m}} - \overline{\boldsymbol{x}}_{\boldsymbol{n}})] - \hat{\boldsymbol{\beta}}^T(\boldsymbol{x_j} - \overline{\boldsymbol{x}}_{\boldsymbol{n}}) \\
&= \boldsymbol{Y}_j - \overline{\boldsymbol{Y}}_m - \hat{\boldsymbol{\beta}}^T(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{\boldsymbol{m}})
\end{aligned} \tag{4 - 68}$$

Hence

$$\hat{\boldsymbol{\Sigma}}_e|\mathbf{x} = \frac{1}{m}\sum_{j=1}^{m}[\boldsymbol{Y_j} - \overline{\boldsymbol{Y}}_m - \hat{\boldsymbol{\beta}}^T(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{\boldsymbol{m}})][\boldsymbol{Y_j} - \overline{\boldsymbol{Y}}_m - \hat{\boldsymbol{\beta}}^T(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{\boldsymbol{m}})]^T = \frac{1}{m}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} \tag{4 - 69}$$

By Results 7.10 in Johnson and Wichern (1998),

$$m\hat{\boldsymbol{\Sigma}}_e|\mathbf{x} = \hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} \sim W_{q,m-p-1}(\boldsymbol{\Sigma}_e)$$

where $W_{q,m-p-1}(\boldsymbol{\Sigma}_e)$ *is* Wishart distribution with $(m-p-1)$ degree of freedom.

The conditional expectation of $\hat{\boldsymbol{\Sigma}}_e$ *is*

$$E(\hat{\boldsymbol{\Sigma}}_e|\mathbf{x}) = \frac{m-p-1}{m}\boldsymbol{\Sigma}_e \tag{4 - 70}$$

and the conditional variance of $\hat{\Sigma}_{e(ij)}$ is

$$Var(\hat{\Sigma}_{e(ij)}|\mathbf{x}) = \frac{m-p-1}{m^2}[\Sigma^2_{e(ij)} + \Sigma_{e(ii)}\Sigma_{e(jj)}] \tag{4 - 71}$$

Both $E(\hat{\boldsymbol{\Sigma}}_e|\mathbf{x})$ and $Var(\hat{\Sigma}_{e(ij)}|\mathbf{x})$ do not involve **X**, so

$$E(\hat{\boldsymbol{\Sigma}}_e) = E[E(\hat{\boldsymbol{\Sigma}}_e|\mathbf{x})] = \frac{m-p-1}{m}\boldsymbol{\Sigma}_e \tag{4 - 72}$$

$\hat{\boldsymbol{\Sigma}}_e$is a biased estimator for $\boldsymbol{\Sigma}_e$**.** If we define

$$\mathbf{S}_e = \frac{1}{m-p-1}\sum_{j=1}^{m}[\boldsymbol{Y_j} - \overline{\boldsymbol{Y}}_m - \hat{\boldsymbol{\beta}}^T(\boldsymbol{X_j} - \overline{\boldsymbol{X}}_{\boldsymbol{m}})][\boldsymbol{Y_j} - \overline{\boldsymbol{Y}}_m - \hat{\boldsymbol{\beta}}^T(\boldsymbol{X_j} - \overline{\boldsymbol{X}}_{\boldsymbol{m}})]^T \tag{4 - 73}$$

Then

$$E(\mathbf{S}_e) = \boldsymbol{\Sigma}_e$$

$\mathbf{S}_e$ is an unbiased estimator for $\mathbf{\Sigma}_e$.

By the Law of the Total Variance, we have the unconditional variance of $\hat{\Sigma}_{e(ij)}$ as follows

$$Var\left(\hat{\Sigma}_{e(ij)}\right) = E\left[Var\left(\hat{\Sigma}_{e(ij)}|\mathbf{x}\right)\right] + Var\left[E\left(\hat{\Sigma}_{e(ij)}|\mathbf{x}\right)\right]$$

$$= \frac{m-p-1}{m^2}\left[\Sigma_{e(ij)}^2 + \Sigma_{e(ii)}\Sigma_{e(jj)}\right] \tag{4 - 74}$$

Since $m\hat{\mathbf{\Sigma}}_e|\mathbf{x} = \hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}^T$ does not depend on $\mathbf{X}$, so

$$m\hat{\mathbf{\Sigma}}_e \sim W_{q,m-p-1}(\mathbf{\Sigma}_e) \tag{4 - 75}$$

4.3 Prediction

Suppose we have a future observation $\boldsymbol{X}_0 = \left[X_{0,1}, X_{0,2}, \dots, X_{0,p}\right]^T, \boldsymbol{Y}_0 = \left[Y_{0,1}, Y_{0,2}, \dots, Y_{0,q}\right]^T$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$

$$\boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix} \quad \mathbf{\Sigma} = \begin{bmatrix}\mathbf{\Sigma}_{\mathbf{xx}} & \mathbf{\Sigma}_{\mathbf{xy}}\\ \mathbf{\Sigma}_{\mathbf{yx}} & \mathbf{\Sigma}_{\mathbf{yy}}\end{bmatrix}$$

As we do in Chapter 3, We have the following three kinds of prediction interval for $\boldsymbol{Y}_0$:

3) Usual prediction interval for $\boldsymbol{Y}_0$– conditioning on $\mathbf{X} = \mathbf{x}\ and\ \mathbf{X}_0 = \mathbf{x}_0$
4) Prediction interval for $\boldsymbol{Y}_0$– unconditional on $\mathbf{X}$, but conditioning on $\mathbf{X}_0 = \mathbf{x}_0$
5) Unconditional prediction interval for $\boldsymbol{Y}_0$

4.3.1 Usual prediction interval

– Conditioning on $\boldsymbol{X} = \boldsymbol{x}\ and\ \boldsymbol{X}_0 = \boldsymbol{x}_0$

The prediction value of $\boldsymbol{Y}_0$ given $\mathbf{X} = \mathbf{x}\ and\ \mathbf{X}_0 = \mathbf{x}_0$ is

$$\widehat{\boldsymbol{Y}}_0|\mathbf{x},\mathbf{x}_0 = \widehat{\boldsymbol{\mu}}_{\boldsymbol{y}} + \widehat{\boldsymbol{\beta}}^T(\mathbf{x}_0 - \bar{\mathbf{x}}_n) \qquad (4 - 76)$$

The ith response follows the multiple regression model in $(3-64)$ in Chapter 3

$$\widehat{\boldsymbol{Y}}_{0(i)}|\mathbf{x},\mathbf{x}_0 = \widehat{\boldsymbol{\mu}}_{\boldsymbol{y}(i)} + \widehat{\boldsymbol{\beta}}_{(i)}^T(\mathbf{x}_0 - \bar{\mathbf{x}}_n), \;\; i = 1,2,\dots,q \qquad (4 - 77)$$

Hence, the 95% prediction interval for $\boldsymbol{Y}_{0(i)}\; given\; \mathbf{X} = \mathbf{x}\; and\;\; \mathbf{X}_0 = \mathbf{x}_0$ follows

$(3-73)$ too

$$\widehat{\boldsymbol{Y}}_{0(i)}|\mathbf{x},\mathbf{x}_0 \pm t_{0.025,m-p-1}\sqrt{\frac{m\widehat{\Sigma}_{e(ii)}}{m-p-1}\left[1+\frac{1}{m}+(\mathbf{x}_0-\bar{\mathbf{x}}_m)^T\mathbf{S}_{\mathrm{xxm}}^{-1}(\mathbf{x}_0-\bar{\mathbf{x}}_m)\right]} \qquad (4 - 78)$$

$$\mathrm{i} = 1,2,\dots,q$$

4.3.2 Prediction interval

– Unconditional on $\boldsymbol{X}$, but conditioning on $\boldsymbol{X}_0 = \boldsymbol{x}_0$

The prediction value of $\boldsymbol{Y}_0$ given $\mathbf{X}_0 = \mathbf{x}_0$ is

$$\widehat{\boldsymbol{Y}}_0|\mathbf{x}_0 = \widehat{\boldsymbol{\mu}}_{\boldsymbol{y}} + \widehat{\boldsymbol{\beta}}^T(\mathbf{x}_0 - \widehat{\boldsymbol{\mu}}_{\boldsymbol{x}}) \qquad (4 - 79)$$

The ith response follows the multiple regression model in $(3-74)$ in Chapter 3

$$\widehat{\boldsymbol{Y}}_{0(i)}|\mathbf{x}_0 = \widehat{\boldsymbol{\mu}}_{\boldsymbol{y}(i)} + \widehat{\boldsymbol{\beta}}_{(i)}^T(\mathbf{x}_0 - \widehat{\boldsymbol{\mu}}_{\boldsymbol{x}}), \;\; i = 1,2,\dots,q \qquad (4 - 80)$$

The 95% prediction interval for $\boldsymbol{Y}_{0(i)}\; given\; \mathbf{X}_0 = \mathbf{x}_0$ follows $(3-87)$ as

$$\widehat{\boldsymbol{Y}}_{0(i)}|\mathbf{x}_0 \pm z_{0.025}\sqrt{\mathrm{S}_{e(i)}\left[1+\frac{1}{m}+\frac{n-m}{mn(m-3)}+\frac{(\mathbf{x}_0-\bar{\mathbf{x}}_n)^T\mathbf{S}_{\mathbf{xn}}^{-1}(\mathbf{x}_0-\bar{\mathbf{x}}_n)}{m-p-2}\right]} \qquad (4 - 81)$$

where

$$\mathrm{S}_{e(i)} = \frac{1}{m-p-1}\sum_{j=1}^{m}\left\{Y_{j(i)} - \bar{Y}_{m(i)} - \widehat{\boldsymbol{\beta}}_{(i)}^T\left(\mathbf{X}_\mathrm{j} - \bar{\mathbf{X}}_{\boldsymbol{m}}\right)\right\}^2 \qquad (4 - 82)$$

and $\mathbf{S}_{\mathrm{xn}}$ is given in $(4-43)$.

4.3.3 Unconditional prediction interval

The prediction value of $\boldsymbol{Y}_0$ is

$$\widehat{\boldsymbol{Y}}_0 = \widehat{\boldsymbol{\mu}}_{\boldsymbol{y}} + \widehat{\boldsymbol{\beta}}^T(\boldsymbol{X}_0 - \widehat{\boldsymbol{\mu}}_{\mathbf{x}}) \quad (4 - 83)$$

The ith response follows the multiple regression model in $(3-91)$ in Chapter 3

$$\widehat{\boldsymbol{Y}}_{0(i)} = \widehat{\boldsymbol{\mu}}_{\boldsymbol{y}(i)} + \widehat{\boldsymbol{\beta}}_{(i)}^T(\boldsymbol{X}_0 - \widehat{\boldsymbol{\mu}}_{\mathbf{x}}),\ \ i = 1,2, \dots, q \quad (4 - 84)$$

Hence, the 95% prediction interval for $\widehat{\boldsymbol{Y}}_{0(i)}$ follows $(3-102)$ too

$$\widehat{\boldsymbol{Y}}_{0(i)} \pm z_{0.025}\sqrt{\mathrm{S}_{e(i)}\left[1 + \frac{1}{m} + \frac{(n-m)p}{n(m-p-2)}\right]} \quad (4 - 85)$$

where $S_{e(i)}$ is given in $(4-82)$.

Appendix A

Statistical Estimation in Bivariate Normal Distribution without Missing Observations

In this appendix, we will derive the MLE estimators and prediction interval for the bivariate normal distribution without observations missing.

A.1 Maximum Likelihood Estimators

If do not consider extra information $X_{m+1}, X_{m+2}, \ldots X_n$ and only use the first m observations in Chapter 2, then the joint likelihood function is

$L(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)$

$$= (2\pi)^{-m}\sigma_e^{-m}\sigma_x^{-m} \cdot exp\left\{-\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left(y_j - \mu_y - \beta(x_j - \mu_x)\right)^2 - \frac{1}{2\sigma_x^2}\sum_{j=1}^{m}\left(x_j - \mu_x\right)^2\right\}$$

The log of the joint likelihood function is

$l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2) = -m\, ln(2\pi) - mln(\sigma_e) - mln(\sigma_x)$

$$\frac{1}{2\sigma_e^2}\sum_{j=1}^{m}\left(y_j - \mu_y - \beta\left(x_j - \mu_x\right)\right)^2 - \frac{1}{2\sigma_x^2}\sum_{j=1}^{m}\left(x_j - \mu_x\right)^2 \qquad \text{(A - 1)}$$

Similarly, by taking the derivatives of the likelihood function $(A-1)$ to each parameter, then setting it to be zero, we have the following estimating equations:

$$\sum_{j=1}^{m}\left[y_j - \mu_y - \beta\left(x_j - \mu_x\right)\right] = 0 \qquad \text{(A - 2)}$$

$$-\frac{\beta}{\sigma_e^2}\sum_{j=1}^{m}\left[y_j - \mu_y - \beta\left(x_j - \mu_x\right)\right] + \frac{1}{\sigma_x^2}\sum_{j=1}^{m}\left(x_j - \mu_x\right) = 0 \qquad \text{(A - 3)}$$

$$\frac{1}{\sigma_e^2}\sum_{j=1}^{m}\left(y_j - \mu_y - \beta\left(x_j - \mu_x\right)\right)\left(x_j - \mu_x\right) = 0 \qquad \text{(A - 4)}$$

$$-\frac{m}{2\sigma_x^2} + \frac{1}{2\sigma_x^4}\sum_{j=1}^{m}\left(x_j - \mu_x\right)^2 = 0 \qquad \text{(A - 5)}$$

$$-\frac{m}{2\sigma_e^2} + \frac{1}{2\sigma_e^4}\sum_{j=1}^{m}\left(y_j - \mu_y - \beta\left(x_j - \mu_x\right)\right)^2 = 0 \qquad \text{(A - 6)}$$

Simultaneously solve estimating equations $(A-2)$ to $(A-6)$, we obtain the following maximum likelihood estimators:

$$\hat{\mu}_{x\_no} = \frac{1}{m}\sum_{j=1}^{m} X_j = \bar{X}_m \tag{A - 7}$$

$$\hat{\mu}_{y\_no} = \bar{Y}_m \tag{A - 8}$$

$$\hat{\beta}_{\_no} = \frac{\sum_{j=1}^{m}(Y_j - \bar{Y}_m)(X_j - \bar{X}_m)}{\sum_{j=1}^{m}(X_j - \bar{X}_m)^2} \tag{A - 9}$$

$$\hat{\sigma}_{x\_no}^2 = \frac{1}{m}\sum_{j=1}^{m}(X_j - \bar{X}_m)^2 \tag{A - 10}$$

$$\hat{\sigma}_{e\_no}^2 = \frac{1}{m}\sum_{j=1}^{m}[(Y_j - \bar{Y}_m) - \hat{\beta}(X_j - \bar{X}_m)]^2 \tag{A - 11}$$

Since only $(A-7), (A-8)$ and $(A-10)$ are different from corresponding estimators with extra information, and $(A-7)$ and $(A-10)$ are straightforward to derive, so here we give the derivation for $(A-8)$.

The conditional expectation of $\hat{\mu}_y$ given x is

$$E(\hat{\mu}_y|x) = E\{[\bar{Y}_m]|x\} = E\left(\frac{1}{m}\sum_{j=1}^{m} Y_j\,|x_j\right) = \frac{1}{m}\sum_{j=1}^{m}[\mu_y + \beta(x_j - \mu_x)]$$

$$= \frac{1}{m}\left[m\mu_y + \beta\sum_{j=1}^{m} x_j - m\beta\mu_x\right] = \mu_y + \beta(\bar{x}_m - \mu_x) \tag{A - 12}$$

Then we have

$$E(\hat{\mu}_y) = E(E(\hat{\mu}_y|X)) = E[\mu_y + \beta(\bar{X}_m - \mu_x)] = \mu_y + \beta(\mu_x - \mu_x) = \mu_y \tag{A - 13}$$

So, $\hat{\mu}_y$ is an unbiased estimator for μ_y.

Similarly, the conditional variance of $\hat{\mu}_y$ given x is

$$Var(\hat{\mu}_y|x) = Var(\bar{Y}_m|x) = Var\left(\frac{1}{m}\sum_{j=1}^{m} Y_j|x_j\right) = \frac{1}{m^2}\sum_{j=1}^{m} Var\left(Y_j\middle|x_j\right) = \frac{1}{m^2}\cdot m\sigma_e^2 = \frac{\sigma_e^2}{m} \qquad \text{(A - 14)}$$

By the Law of Total Variance,

$$Var(\hat{\mu}_y) = E\left(Var(\hat{\mu}_y|X)\right) + Var(E(\hat{\mu}_y|X))$$

where

$$Var(E(\hat{\mu}_y|X)) = Var\left(\mu_y + \beta(\bar{X}_m - \mu_x)\right) = \beta^2 Var(\bar{X}_m) = \frac{\beta^2\sigma_{xx}}{m} \qquad \text{(A - 15)}$$

$$E\left(Var(\hat{\mu}_y|Y)\right) = E\left(\frac{\sigma_e^2}{m}\right) = \frac{\sigma_e^2}{m}$$

Hence

$$Var(\hat{\mu}_y) = \frac{\sigma_e^2}{m} + \frac{\beta^2\sigma_{xx}}{m} \qquad \text{(A - 16)}$$

A.2 The prediction Interval

A.2.1 Usual prediction interval

The prediction value of Y_0 is

$$\hat{Y}_0|x, x_0 = \hat{\mu}_y + \hat{\beta}(x_0 - \bar{x}_m) \qquad \text{(A - 17)}$$

By $(A - 12)$ and $(A - 16)$,

$$\left(\hat{\mu}_y\middle|x\right) \sim N\left(\mu_y + \beta(\bar{x}_m - \mu_x), \frac{\sigma_e^2}{m}\right) \qquad \text{(A - 18)}$$

So, the expectation of $\hat{Y}_0|x, x_0$ is

$$E\left(\hat{Y}_0\middle|x, x_0\right) = E\left(\hat{\mu}_y\middle|x, x_0\right) + E\left(\hat{\beta}\middle|x, x_0\right)(x_0 - \bar{x}_m)$$

$$= E\left(\hat{\mu}_y\middle|x\right) + E\left(\hat{\beta}\middle|x\right)(x_0 - \bar{x}_m)$$

$$= \mu_y + \beta(\bar{x}_m - \mu_x) + \beta(x_0 - \bar{x}_m)$$

$$= \mu_y + \beta(x_0 - \mu_x) \qquad \text{(A - 19)}$$

The variance of $\hat{Y}_0|x, x_0$ is

$$Var\left(\hat{Y}_0|x, x_0\right) = Var\left(\hat{\mu}_y\middle|x\right) + (x_0 - \bar{x}_m)^2 Var\left(\hat{\beta}\middle|x\right) + 2(x_0 - \bar{x}_m)Cov\left[\left(\hat{\mu}_y, \hat{\beta}\right)|x\right]$$

$$= \frac{\sigma_e^2}{m} + \frac{\sigma_e^2(x_0 - \bar{x}_m)^2}{\sum_{j=1}^{m}\left(x_j - \bar{x}_m\right)^2} - 0$$

$$= \frac{\sigma_e^2}{m} + \frac{\sigma_e^2(x_0 - \bar{x}_m)^2}{\sum_{j=1}^{m}\left(\mathrm{x_j} - \bar{x}_m\right)^2} \qquad \text{(A - 20)}$$

Hence, the expectation of $\left(Y_0 - \hat{Y}_0\right)|x, x_0$ is

$$E\left(Y_0 - \hat{Y}_0\right)|x, x_0 = E(Y_0|x, x_0) - E\left(\hat{Y}_0\middle|x, x_0\right)$$

$$= \mu_y + \beta(x_0 - \mu_x) - \left[\mu_y + \beta(x_0 - \mu_x)\right] = 0 \qquad \text{(A - 21)}$$

And the variance of $\left(Y_0 - \hat{Y}_0\right)|x, x_0$ is

$$Var\left(Y_0 - \hat{Y}_0\right)|x, x_0 = Var(Y_0)|x, x_0 + Var\left(\hat{Y}_0\right)|x, x_0 - 2Cov(\hat{Y}_0, Y_0)|x, x_0$$

$$= \sigma_e^2 + \frac{\sigma_e^2}{m} + \frac{\sigma_e^2(x_0 - \bar{x}_m)^2}{\sum_{j=1}^{m}\left(\mathrm{x_j} - \bar{x}_m\right)^2} - 2 \cdot 0$$

$$= \sigma_e^2\left[1 + \frac{1}{m} + \frac{(x_0 - \bar{x}_m)^2}{\sum_{j=1}^{m}\left(\mathrm{x_j} - \bar{x}_m\right)^2}\right] \qquad \text{(A - 22)}$$

Hence, the 95% prediction interval for $Y_0\ given\ X = x\ and\ \ X_0 = x_0$ is

$$\hat{Y}_0|x, x_0 \pm t_{0.025, m-2}\sqrt{\frac{m\hat{\sigma}_e^2}{m-2}\left[1 + \frac{1}{m} + \frac{(x_0 - \bar{x}_m)^2}{\sum_{j=1}^{m}\left(\mathrm{x_j} - \bar{x}_m\right)^2}\right]} \qquad \text{(A - 23)}$$

A.2.2 Prediction interval

Let

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = \begin{bmatrix} \hat{\mu}_y \\ \hat{\mu}_x \\ \hat{\beta} \end{bmatrix}$$

Then

$$\mathrm{E}(\boldsymbol{Z}|x_0) = \begin{bmatrix} \mu_y \\ \mu_x \\ \beta \end{bmatrix}$$

$$\mathrm{Cov}(\boldsymbol{Z}|x_0) == \begin{bmatrix} \frac{\sigma_e^2}{m} + \frac{\beta^2\sigma_x^2}{m} & \frac{\beta\sigma_x^2}{m} & 0 \\ \frac{\beta\sigma_x^2}{m} & \frac{\sigma_x^2}{m} & 0 \\ 0 & 0 & \frac{\sigma_e^2}{(m-3)\sigma_x^2} \end{bmatrix} \quad \text{(A - 24)}$$

The variance of $\hat{Y}_0|x_0$ is

$$\mathrm{Var}(\hat{Y}_0|x_0) = E\left[\hat{Y}_0|x_0 - E(\hat{Y}_0| \quad _0)\right]^2 \approx E\left[\sum_{j=1}^{3} (\hat{y}'_{0j}|x_0)(Z_j - \mu_{Zj})\right]^2$$

$$= \sum_{j=1}^{3} \left[(\hat{y}'_{0j}|x_0)\right]^2 Var(Z_j) + 2\sum_{i=1}^{3} \sum_{j=1, i\neq j}^{3} (\hat{y}'_{0i}|x_0)(\hat{y}'_{0j}|x_0) Cov(Z_i, Z_j)$$

$$= 1^2 Var(Z_1) + (-\beta)^2 Var(Z_2) + (x_0 - \mu_x)^2 Var(Z_3) + 2\cdot 1 \cdot (-\beta) Cov(Z_1, Z_2)$$

$$= 1^2 Var(\hat{\mu}_y) + (-\beta)^2 Var(\hat{\mu}_x) + (x_0 - \mu_x)^2 Var(\hat{\beta}) + 2\cdot 1 \cdot (-\beta) Cov(\hat{\mu}_y, \hat{\mu}_x)$$

$$= \frac{\sigma_e^2}{m} + \frac{\beta^2\sigma_x^2}{m} + \frac{\beta^2\sigma_x^2}{m} + (x_0 - \mu_x)^2 \frac{\sigma_e^2}{(m-3)\sigma_{xx}} - 2\frac{\beta^2\sigma_x^2}{m}$$

$$= \sigma_e^2 \left[\frac{1}{m} + \frac{(x_0 - \mu_x)^2}{(m-3)\sigma_x^2}\right] \quad \text{(A - 25)}$$

Hence, the variance of $(Y_0 - \hat{Y}_0)|x_0$ is

$$Var(Y_0 - \hat{Y}_0)|x_0 = Var(Y_0)|x_0 + Var(\hat{Y}_0)|x_0 - 2Cov(\hat{Y}_0, Y_0)|x_0$$

$$= \sigma_e^2 + \sigma_e^2 \left[\frac{1}{m} + \frac{(x_0 - \mu_x)^2}{(m-3)\sigma_x^2}\right] - 2 \cdot 0$$

$$= \sigma_e^2 \left[1 + \frac{1}{m} + \frac{(x_0 - \mu_x)^2}{(m-3)\sigma_x^2}\right] \qquad \text{(A - 26)}$$

A.2.3 Unconditional prediction interval

Let

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix} = \begin{bmatrix} \hat{\mu}_y \\ \hat{\mu}_x \\ \hat{\beta} \\ X_0 \end{bmatrix}$$

Then

$$\mathrm{E}(\boldsymbol{Z}) = \boldsymbol{\mu}_Z = \begin{bmatrix} \mu_y \\ \mu_x \\ \beta \\ \mu_x \end{bmatrix}$$

$$\mathrm{Cov}(\boldsymbol{Z}) = \begin{bmatrix} \frac{\sigma_e^2}{m} + \frac{\beta^2\sigma_x^2}{m} & \frac{\beta\sigma_x^2}{m} & 0 & 0 \\ \frac{\beta\sigma_x^2}{m} & \frac{\sigma_x^2}{m} & 0 & 0 \\ 0 & 0 & \frac{\sigma_e^2}{(m-3)\sigma_x^2} & 0 \\ 0 & 0 & 0 & \sigma_x^2 \end{bmatrix} \qquad \text{(A - 27)}$$

The variance of $\hat{Y}_0$ is

$$\mathrm{Var}(\hat{Y}_0) = \mathrm{E}\left[\hat{Y}_0 - E(\hat{Y}_0)\right]^2 \approx \mathrm{E}\left[\sum_{j=1}^{4} \hat{y}'_{0j}(\boldsymbol{\mu}_Z) E(\, Z_j - \mu_{Zj})\right]^2$$

$$= \sum_{j=1}^{4} \left[\hat{y}'_{0j}(\boldsymbol{\mu}_Z)\right]^2 Var(Z_j) + 2 \sum_{i=1}^{4} \sum_{j=1, i \neq j}^{4} \hat{y}'_{0i}(\boldsymbol{\mu}_Z)\hat{y}'_{0j}(\boldsymbol{\mu}_Z) Cov(Z_i, Z_j)$$

$$= 1^2 Var(Z_1) + (-\beta)^2 Var(Z_2) + 0^2 \cdot Var(Z_3) + \beta^2 Var(Z_4) + 2 \cdot 1 \cdot (-\beta) Cov(Z_1, Z_2)$$

$$= 1^2 Var\left(\hat{\mu}_y\right) + (-\beta)^2 Var(\hat{\mu}_x) + 0^2 \cdot Var\left(\hat{\beta}\right) + \beta^2 Var(X_0) + 2 \cdot 1 \cdot (-\beta) Cov\left(\hat{\mu}_y, \hat{\mu}_x\right)$$

$$= \frac{\sigma_e^2}{m} + \frac{\beta^2 \sigma_x^2}{m} + \frac{\beta^2 \sigma_x^2}{m} + 0 + \beta^2 \sigma_x^2 - 2\frac{\beta^2 \sigma_x^2}{m}$$

$$= \frac{\sigma_e^2}{m} + \beta^2 \sigma_x^2 \qquad \text{(A - 28)}$$

Hence, the variance of $\left(Y_0 - \hat{Y}_0\right)$ are

$$Var\left(Y_0 - \hat{Y}_0\right) = Var(Y_0) + Var\left(\hat{Y}_0\right) - 2Cov(\hat{Y}_0, Y_0)$$

$$= \sigma_y^2 + \frac{\sigma_e^2}{m} + \beta^2 \sigma_x^2 - 2 \cdot \beta^2 \sigma_x^2 = \sigma_y^2 + \frac{\sigma_e^2}{m} - \beta^2 \sigma_x^2 = \sigma_e^2 \left[1 + \frac{1}{m}\right] \qquad \text{(A - 29)}$$

The 95% prediction interval for Y_0 is

$$\hat{Y}_0 \pm z_{0.025} \sqrt{S_e^2 \left(1 + \frac{1}{m}\right)} \qquad \text{(A - 30)}$$

Appendix B

Fisher Information Matrix for Bivariate Normal Distribution

Since we have five parameters, so the Fisher Information Matrix $\boldsymbol{I}(\theta)$ is defined by a 5×5 matrix, and the (j,k) entry of $\boldsymbol{I}(\theta)$ is given by

$$I_{\theta_j\theta_k} = -E\left[\frac{\partial^2 l(X,Y,\theta)}{\partial\theta_j\partial\theta_k}\right]$$

Where

$$l(X,Y,\theta) = l\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right)$$

is the log of the joint likelihood function $(2-8)$.

$$I_{\mu \quad \mu_x} = -E\left[\frac{\partial^2 l\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right)}{\partial\mu_x^2}\right]$$

$$= -E\left\{\frac{\partial}{\partial\mu_x}\left[-\frac{\beta}{\sigma_e^2}\sum_{j=1}^{m}\left[Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right] + \frac{1}{\sigma_x^2}\sum_{i=1}^{n}(X_i - \mu_x)\right]\right\}$$

$$= -E\left\{-\frac{\beta}{\sigma_e^2}\sum_{j=1}^{m}\beta + \frac{1}{\sigma_x^2}\sum_{i=1}^{n}(-1)\right\}$$

$$= \frac{m\beta^2}{\sigma_e^2} + \frac{n}{\sigma_x^2}$$

$$I_{\mu_x\mu_y} = -\mathrm{E}\left[\frac{\partial^2 l\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right)}{\partial\mu_x\partial\mu_y}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial\mu_y}\left[-\frac{\beta}{\sigma_e^2}\sum_{j=1}^{m}\left[Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right] + \frac{1}{\sigma_x^2}\sum_{i=1}^{n}(X_i - \mu_x)\right]\right\}$$

$$= \frac{\beta}{\sigma_e^2}\sum_{j=1}^{m}(-1)$$

$$= -\frac{m\beta}{\sigma_e^2}$$

$$I_{\mu_x\beta} = -\mathrm{E}\left[\frac{\partial^2 l\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right)}{\partial\mu_x\partial\beta}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial\beta}\left[-\frac{\beta}{\sigma_e^2}\sum_{j=1}^{m}\left[Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right] + \frac{1}{\sigma_x^2}\sum_{i=1}^{n}(X_i - \mu_x)\right]\right\}$$

$$= -\mathrm{E}\left\{-\frac{1}{\sigma_e^2}\sum_{j=1}^{m}\left[Y_j - \mu_y - 2\beta\left(X_j - \mu_x\right)\right]\right\}$$

$$= \frac{1}{\sigma_e^2} E\left\{\sum_{j=1}^{m}\left[Y_j - \mu_y - 2\beta\left(X_j - \mu_x\right)\right]\right\}$$

$$= 0$$

$$I_{\mu_x\sigma_x^2} = -\mathrm{E}\left[\frac{\partial^2 l\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right)}{\partial\mu_x\partial\sigma_x^2}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial\sigma_x^2}\left[-\frac{\beta}{\sigma_e^2}\sum_{j=1}^{m}\left[Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right] + \frac{1}{\sigma_x^2}\sum_{i=1}^{n}(X_i - \mu_x)\right]\right\}$$

$$= -\mathrm{E}\left\{-\frac{1}{\sigma_x^4}\sum_{i=1}^{n}(X_i - \quad {}_x)\right\}$$

$$= \frac{1}{\sigma_x^4}\sum_{i=1}^{n}[\mathrm{E}(X_i) - \mu_x]$$

$$= 0$$

$$I_{\mu_x\sigma_e^2} = -\mathrm{E}\left[\frac{\partial^2 l\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right)}{\partial\mu_x\partial\sigma_e^2}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial\sigma_e^2}\left[-\frac{\beta}{\sigma_e^2}\sum_{j=1}^{m}\left[Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right] + \frac{1}{\sigma_x^2}\sum_{i=1}^{n}(X_i - \mu_x)\right]\right\}$$

$$= -\mathrm{E}\left\{\frac{\beta}{\sigma_e^4}\sum_{j=1}^{m}\left[Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right]\right\}$$

$$= -\frac{\beta}{\sigma_e^4}\sum_{j=1}^{m}\left[\mathrm{E}(Y_j) - \mu_y - \beta\left(\mathrm{E}(X_j) - \mu_x\right)\right]$$

$$= 0$$

$$I_{\mu_y\mu_y} = -\mathrm{E}\left[\frac{\partial^2 l\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right)}{\partial\mu_y^2}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial \mu_y}\left[\frac{1}{\sigma_e^2}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)\right]\right\}$$

$$= -\mathrm{E}\left\{-\frac{m}{\sigma_e^2}\right\}$$

$$= \frac{m}{\sigma_e^2}$$

$$I_{\mu_y\beta} = -\mathrm{E}\left[\frac{\partial^2 l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \mu_y \partial \beta}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial \beta}\left[\frac{1}{\sigma_e^2}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)\right]\right\}$$

$$= -\mathrm{E}\left[-\tfrac{1}{\sigma_e^2}\textstyle\sum_{j=1}^{m}\left(X_j - \mu_x\right)\right]$$

$$= \frac{1}{\sigma_e^2}\sum_{j=1}^{m}\left[\mathrm{E}\left(X_j - \mu_x\right)\right]$$

$$= 0$$

$$I_{\mu_y\sigma_x^2} = -\mathrm{E}\left[\frac{\partial^2 l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \mu_y \partial \sigma_{xx}}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial \sigma_x^2}\left[\frac{1}{\sigma_e^2}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)\right]\right\}$$

$$= 0$$

$$I_{\mu_y\sigma_e^2} = -\mathrm{E}\left[\frac{\partial^2 l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \mu_y \partial \sigma_e^2}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial \sigma_e^2}\left[\frac{1}{\sigma_e^2}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)\right]\right\}$$

$$= -\mathrm{E}\left\{-\frac{1}{\sigma_e^4}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)\right\}$$

$$= \frac{1}{\sigma_e^4}\sum_{j=1}^{m}\left\{\mathrm{E}\left[Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right]\right\}$$

$$= 0$$

$$I_{\beta\beta} = -\mathrm{E}\left[\frac{\partial^2 l\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2\right)}{\partial \beta^2}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial \beta}\left[\frac{1}{\sigma_e^2}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)\left(X_j - \mu_x\right)\right]\right\}$$

$$= \frac{1}{\sigma_e^2}\mathrm{E}\left\{\sum_{j=1}^{m}\left(X_j - \mu_x\right)^2\right\}$$

$$= \frac{\sigma_x^2}{\sigma_e^2}\mathrm{E}\left\{\sum_{j=1}^{m}\frac{\left(X_j - \mu_x\right)^2}{\sigma_x^2}\right\} \qquad sinc\ \left\{\sum_{j=1}^{m}\frac{\left(X_j - \mu_x\right)^2}{\sigma_x^2}\right\}\sim\chi^2(m)$$

$$= \frac{m\sigma_x^2}{\sigma_e^2}$$

$$I_{\beta\sigma_x^2} = -\mathrm{E}\left[\frac{\partial^2 l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \beta \partial \sigma_x^2}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial \sigma_x^2}\left[\frac{1}{\sigma_e^2}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)\left(X_j - \mu_x\right)\right]\right\}$$

$$= 0$$

$$I_{\beta\sigma_e^2} = -\mathrm{E}\left[\frac{\partial^2 l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \beta \partial \sigma_e^2}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial \sigma_e^2}\left[\frac{1}{\sigma_e^2}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)\left(X_j - \mu_x\right)\right]\right\}$$

$$= -\mathrm{E}\left\{-\frac{1}{\sigma_e^4}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)\left(X_j - \mu_x\right)\right\}$$

$$= \frac{1}{\sigma_e^4} \mathrm{E}\left\{\sum_{j=1}^{m} \left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)\left(X_j - \mu_x\right)\right\}$$

$$= 0$$

$$I_{\sigma_x^2\sigma_x^2} = -\mathrm{E}\left[\frac{\partial^2 l(\mu \quad, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \sigma_x^4}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial \sigma_x^2}\left[-\frac{n}{2\sigma_x^2} + \frac{1}{2\sigma_x^4}\sum_{i=1}^{n}(X_i - \mu_x)^2\right]\right\}$$

$$= -\mathrm{E}\left\{\frac{n}{2\sigma_x^4} - \frac{1}{\sigma_x^6}\sum_{i=1}^{n}(X_i - \mu_x)^2\right\}$$

$$= -\frac{n}{2\sigma_x^4} + \frac{1}{\sigma_x^4}\mathrm{E}\sum_{i=1}^{n}\frac{(X_i - \mu_x)^2}{{}_x^2} \qquad since \ \sum_{i=1}^{n}\frac{(X_i - \mu_x)^2}{\sigma_x^2} \sim \chi^2(n)$$

$$= -\frac{n}{2\sigma_x^4} + \frac{n}{\sigma_x^4}$$

$$= \frac{n}{2\sigma_x^4}$$

$$I_{\sigma_x^2\sigma_e^2} = -\mathrm{E}\left[\frac{\partial^2 l(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)}{\partial \sigma_x^2 \partial \sigma_e^2}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial \sigma_e^2}\left[-\frac{n}{2\sigma_x^2} + \frac{1}{2\sigma_x^4}\sum_{i=1}^{n}\left(X_i - \mu_y\right)^2\right]\right\}$$

$$= 0$$

$$I_{\sigma_e^2\sigma_e^2} = -\mathrm{E}\left[\frac{\partial^2 l\left(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma\right)}{\partial (\sigma_e^2)^2}\right]$$

$$= -\mathrm{E}\left\{\frac{\partial}{\partial \sigma_e^2}\left[-\frac{m}{2\sigma_e^2} + \frac{1}{2\sigma_e^4}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)^2\right]\right\}$$

$$= -\mathrm{E}\left\{\frac{m}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)^2\right\}$$

$$= -\frac{m}{2\sigma_e^4} + \frac{1}{\sigma_e^6}\mathrm{E}\left\{\sum_{j=1}^{m}\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)^2\right\}$$

$$= -\frac{m}{2\sigma_e^4} + \frac{1}{\sigma_e^6}\sum_{j=1}^{m}\sigma_e^2$$

$$= -\frac{m}{2\sigma_e^4} + \frac{m\sigma_e^2}{\sigma_e^6}$$

$$= \frac{m}{2\sigma_e^4}$$

Where

$$\mathrm{E}(\left(Y_j - \mu_y - \beta\left(x_j - \mu_x\right)\right)^2 |x)$$

$$= \mathrm{E}\left\{\left(Y_j - \mu_y\right)^2 - 2\left(Y_j - \mu_y\right)\beta\left(x_j - \mu_x\right) + \quad {}^2\left(x_j - \mu_x\right)^2\right\}|x$$

$$= \mathrm{E}\left(Y_j - \mu_y\right)^2\middle|x - 2\beta\left(x_j - \mu_x\right)\mathrm{E}\left(Y_j - \mu_y\right)\middle|x + \beta^2\left(x_j - \mu_x\right)^2$$

$$= \sigma_e^2 + \beta^2\left(x_j - \mu_x\right)^2 - 2\beta^2\left(x_j - \mu_x\right)^2 + \beta^2\left(x_j - \mu_x\right)^2$$

$$= \sigma_e^2$$

Hence

$$\mathrm{E}\left(\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)^2\right) = \mathrm{E}\left[\mathrm{E}\left(\left(Y_j - \mu_y - \beta\left(X_j - \mu_x\right)\right)^2 \middle| X\right)\right] = E(\sigma_e^2) = \sigma_e^2$$

The Fisher Information Matrix is

$$\boldsymbol{I}(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2) = \begin{Bmatrix} I_{\mu_y\mu_y} & I_{\mu_y\mu_x} & I_{\mu_y\beta} & I_{\mu_y\sigma_x^2} & I_{\mu_y\sigma_e^2} \\ I_{\mu_x\mu_y} & I_{\mu_x\mu_x} & I_{\mu_x\beta} & I_{\mu_x\sigma_x^2} & I_{\mu_x\sigma_e^2} \\ I_{\beta\mu_y} & I_{\beta\mu_x} & I_{\beta\beta} & I_{\beta\sigma_x^2} & I_{\beta\sigma_e^2} \\ I_{\sigma_x^2\mu_y} & I_{\sigma_x^2\mu_x} & I_{\sigma_x^2\beta} & I_{\sigma_x^2\sigma_x^2} & I_{\sigma_x^2\sigma_e^2} \\ I_{\sigma_e^2\mu_y} & I_{\sigma_e^2\mu_x} & I_{\sigma_e^2\beta} & I_{\sigma_e^2\sigma_x^2} & I_{\sigma_e^2\sigma_e^2} \end{Bmatrix}$$

$$
= \begin{Bmatrix}
\frac{m}{\sigma_e^2} & -\frac{m\beta}{\sigma_e^2} & 0 & 0 & 0 \\
-\frac{m\beta}{\sigma^2} & \frac{m\beta^2}{\sigma_e^2} + \frac{n}{\sigma_x^2} & 0 & 0 & 0 \\
0 & 0 & \frac{m\sigma_x^2}{\sigma_e^2} & 0 & 0 \\
0 & 0 & 0 & \frac{n}{2\sigma_x^4} & 0 \\
0 & 0 & 0 & 0 & \frac{m}{2\sigma_e^4}
\end{Bmatrix}
$$

The inverse of the $\boldsymbol{I}(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2)$ is

$$
\boldsymbol{I}^{-1}(\mu_y, \mu_x, \beta, \sigma_x^2, \sigma_e^2) = \begin{Bmatrix}
\frac{\sigma_e^2}{m} + \frac{\beta^2\sigma_x^2}{n} & \frac{\beta\sigma_x^2}{n} & 0 & 0 & 0 \\
\frac{\beta\sigma_x^2}{n} & \frac{\sigma_x^2}{n} & 0 & 0 & 0 \\
0 & 0 & \frac{\sigma_e^2}{m\sigma_x^2} & 0 & 0 \\
0 & 0 & 0 & \frac{2\sigma_x^4}{n} & 0 \\
0 & 0 & 0 & 0 & \frac{2\sigma_e^4}{m}
\end{Bmatrix}
$$

Appendix C

Some derivation used in the dissertation

In this appendix, we derived some formula used in the dissertation.

$$\sum_{j=1}^{m}(y_j - \bar{y}_m)(x_j - \bar{x}_n) = \sum_{j=1}^{m}(y_j x_j - \bar{y}_m x_j - y_j \bar{x}_n + \bar{y}_m \bar{x}_n)$$

$$= \sum_{j=1}^{m} y_j x_j - \bar{y}_m \sum_{j=1}^{m} x_j - \bar{x}_n \sum_{j=1}^{m} y_j + m\bar{y}_m \bar{x}_n$$

$$= \sum_{j=1}^{m} y_j x_j - \bar{y}_m \sum_{j=1}^{m} x_j - m\bar{x}_n \bar{y}_m + m\bar{y}_m \bar{x}_n$$

$$= \sum_{j=1}^{m} y_j x_j - \bar{y}_m \sum_{j=1}^{m} x_j = \sum_{j=1}^{m} x_j (y_j - \bar{y}_m) = \sum_{j=1}^{m} y_j (x_j - \bar{x}_m) \quad \text{(C - 1)}$$

$$\sum\nolimits_{j=1}^{m} (x_j - \bar{x}_m)(x_j - \bar{x}_n) = \sum_{j=1}^{m}(x_j^2 - \bar{x}_m x_j - x_j \bar{x}_n + \bar{x}_m \bar{x}_n)$$

$$= \sum_{j=1}^{m} x_j^2 - \bar{x}_m \sum_{j=1}^{m} x_j - \bar{x}_n \sum_{j=1}^{m} x_j + m\bar{x}_m \bar{x}_n$$

$$= \sum_{j=1}^{m} x_j^2 - \bar{x}_m \sum_{j=1}^{m} x_j - m\bar{x}_n \bar{x}_m + m\bar{x}_m \bar{x}_n$$

$$= \sum_{j=1}^{m} x_j^2 - \bar{x}_m \sum_{j=1}^{m} x_j = \sum_{j=1}^{m} x_j^2 - \bar{x}_m \sum_{j=1}^{m} x_j - m\bar{x}_m \bar{x}_m + m\bar{x}_m \bar{x}_m$$

$$= \sum_{j=1}^{m} x_j^2 - \bar{x}_m \sum_{j=1}^{m} x_j - \bar{x}_m \sum_{j=1}^{m} x_j + \sum_{j=1}^{m} \bar{x}_m \bar{x}_m = \sum_{j=1}^{m} (x_j - \bar{x}_m)^2 \quad \text{(C - 2)}$$

$$\sum_{j=1}^{m} x_j^2 - m\bar{x}_m \bar{x}_m = \sum_{j=1}^{m} x_j^2 - \sum_{j=1}^{m} x_j \bar{x}_m = \sum_{j=1}^{m} x_j (x_j - \bar{x}_m) \quad \text{(C - 3)}$$

$$\mathrm{Cov}\left[\left(Y_j,\hat{\beta}\right)\middle|x\right] = \mathrm{Cov}(Y_j|x,\frac{\sum_{j=1}^{m} Y_j\left(\mathrm{x_j}-\bar{x}_m\right)}{\sum_{j=1}^{m}\left(\mathrm{x_j}-\bar{x}_m\right)^2}|x)$$

$$= \frac{1}{\sum_{j=1}^{m}\left(x_j-\bar{x}_m\right)^2}\mathrm{Cov}[Y_j|x, Y_1(x_1-\bar{x}_m)|x + Y_2(x_2-\bar{x}_m)|x + \cdots Y_j\left(x_j-\bar{x}_m\right)|x + \cdots Y_m(x_m-\bar{x}_m)|x]$$

$$= \frac{1}{\sum_{j=1}^{m}\left(x_j-\bar{x}_m\right)^2}\mathrm{Cov}\left[Y_j\middle|x, Y_j\left(x_j-\bar{x}_m\right)\middle|x\right] \qquad Y_j|x, Y_k|x\ ae\ independent\ for\ j \neq k$$

$$= \frac{\left(x_j-\bar{x}_m\right)}{\sum_{j=1}^{m}\left(x_j-\bar{x}_m\right)^2}\mathrm{Var}\left(Y_j\middle|x\right) = \frac{\left(x_j-\bar{x}_m\right)\sigma_e^2}{\sum_{j=1}^{m}\left(x_j-\bar{x}_m\right)^2} \qquad \text{(C - 4)}$$

$$\mathrm{Cov}\left[\left(\bar{Y},\hat{\beta}\right)\middle|x\right] = \frac{1}{m}\mathrm{Cov}\left[Y_1|x + Y_2|x \ldots Y_m\middle|x,\hat{\beta}\middle|x\right]$$

$$= \frac{1}{m}\sum_{j=1}^{m}\mathrm{Cov}\left[\left(Y_j,\hat{\beta}\right)\middle|x\right] = \frac{1}{m}\sum_{j=1}^{m}\frac{\left(x_j-\bar{x}_m\right)\sigma_e^2}{\sum_{j=1}^{m}\left(x_j-\bar{x}_m\right)^2} = \frac{\sigma_e^2}{m}\frac{\sum_{j=1}^{m}\left(x_j-\bar{x}_m\right)}{\sum_{j=1}^{m}\left(x_j-\bar{x}_m\right)^2} = 0 \qquad \text{(C - 5)}$$

Hence, $\bar{Y}\middle|x\ and\ \hat{\beta}\middle|x\ are\ NOT\ correlated.$

Appendix D

R codes

In this appendix, we list R codes for the example estimators and variance comparison.

D.1 R Code for the Estimators

```
install.packages("bayesSurv")      #install this package to calculate sample covariance
install.packages("MVN")            #Multivariate normality test
install.packages("usdm")           #Multicollinearity test
library(MVN)

varNames <- c ("Y",
        "GPA",
        "GVerbal",
        "GQuantitative",
        "GAnalytic",
        "TOEFL")

varFormats <- c ("numeric",  # Master GPA
         "numeric",  # Undergraduate GPA
         "numeric",  # GRE Verbal
         "numeric",  # GRE Quantitative
         "numeric",  # GRE Analytic
         "numeric")  # TOEFL

directory <- "D:/PhD Research/Book data"
filename  <- "GPA with Block Missing realdata.txt"
fullFile  <- paste (directory, filename, sep = "/")

mydata <- read.csv (fullFile,
          stringsAsFactors = FALSE,
          nrow = -1,
          col.names = varNames,
          colClasses = varFormats,
          sep = "\t")

mydatas <- head(subset.matrix(mydata,select=GVerbal:TOEFL,(!is.na(mydata[,4]))),40)     #Obtain the sample data

n <- 40
m <- 20

mydatas_m<-subset(mydatas, (!is.na(mydatas[,4])))                        #Ignore TOEFL missing data

# Calculate 5 estimates - Multivariate case
one_n <- as.matrix(rep(1,40))              #vector one 40 by 1
one_m <- as.matrix(rep(1,20))

multi_xndata <- as.matrix(subset(mydatas,select=GVerbal:GAnalytic))

m_xn <- as.matrix(colMeans(multi_xndata))

diff_xn <- multi_xndata - one_n %*% t(m_xn)

ss0 <- diff_xn[1,] %*% t(diff_xn[1,])
ssn <- 0*ss0

for (i in 1:40) {
  temp <- diff_xn[i,] %*% t(diff_xn[i,])
  ssn <- temp + ssn
}

sigma_xxn <- ssn/n             #sigma_xx_n
```

```
library(bayesSurv)
s_xx_n0 <- (n-1)*sampleCovMat(multi_xndata)/n       #for check purpose

ym <- mydatas_m[,4]
y_bar_m <- sum(ym)/m

multi_xmdata <- as.matrix(subset(mydatas_m,select=GVerbal:GAnalytic))
m_xm <- as.matrix(colMeans(multi_xmdata))

diff_xm <- multi_xmdata - one_m %*% t(m_xm)

ssm <- 0*ss0

for (i in 1:20) {
  temp <- diff_xm[i,] %*% t(diff_xm[i,])
  ssm <- temp + ssm
}
sigma_xxm <- ssm/m                    #Sigma_xx_m

s_xx_m0 <- (m-1)*sampleCovMat(multi_xmdata)/m   #Check

d <- m_xm - m_xn             #difference between mean xm and mean xn

diff_ym <- ym - y_bar_m

sxy0 <- diff_ym[1]*diff_xm[1,]
sxy <- 0*sxy0
for (i in 1:20) {
  temp <- diff_ym[i] *diff_xm[i,]
  sxy <- temp + sxy
}

sxy_m <- as.matrix(sxy)
beta_hat <- solve(ssm) %*% sxy_m

mu_yhat <- y_bar_m - t(beta_hat) %*% d          #mu_y_hat

bm <- 0
for (i in 1:20) {
  temp <- (diff_ym[i] - t(beta_hat) %*% diff_xm[i,])^2
  bm <- temp + bm
}

sigma_e_hat <- bm/m                     #Sigma_e^2

#check beta
fit <- lm(TOEFL ~ GVerbal + GQuantitative + GAnalytic, data=mydatas_m)
fit
beta <- coefficients(fit)
beta

##################################################################
cov(mydatas_m[,1],mydatas_m[,4])
cor(mydatas_m[,1],mydatas_m[,4])          #rou=0.3252701 --selected since rou_square>1/18
cor(mydatas_m[,2],mydatas_m[,4])          #rou=-0.07477673
cor(mydatas_m[,3],mydatas_m[,4])          #rou=0.166066

#Calculate 5 estimates for bivariate cases
#1-Verbal score as x
bdata <- subset(mydatas,select=c(GVerbal,TOEFL))
n <- 40
m <- 20
xn <- bdata[,1]
x_bar_n <- sum(xn)/n
```

```
bdata_m <- subset(bdata,(!is.na(bdata[,2])))
xm <- bdata_m[,1]
ym <- bdata_m[,2]
x_bar_m <- sum(xm)/m
y_bar_m <- sum(ym)/m

beta_num <- sum((ym-mean(ym))*(xm-mean(xm)))
beta_den <- sum((xm-mean(xm))^2)
beta <- beta_num/beta_den

muy_hat <- y_bar_m - beta*(x_bar_m - x_bar_n)
sigma_xx_hat_n <- sum((xn-mean(xn))^2)/n
sigma_xx_hat_m <- sum((xm-mean(xm))^2)/m
sigma_e_hat <- sum((ym - y_bar_m - beta*(xm - x_bar_m))^2)/m

biout <- data.frame(x_bar_n,muy_hat,beta,sigma_xx_hat_n,sigma_e_hat)

biout_no <- data.frame(x_bar_m, y_bar_m,beta,sigma_xx_hat_m,sigma_e_hat)
```

Similar codes for Quantitative score, Analytic score as x

```
#Bivariate normal test
mydatas1 <- subset.matrix(mydatas_m,select=c(GVerbal,TOEFL))
res1 <- mardiaTest(mydatas1)              #Henze-Zirkler's Multivariate Normality Test
mvnPlot(res1, type = "persp", default = TRUE)          # Perspective Plot
mvnPlot(res1, type = "contour", default = TRUE)        # Contour Plot

mydatas2 <- subset.matrix(mydatas_m,select=c(GQuantitative,TOEFL))
res2 <- hzTest(mydatas2)              #Henze-Zirkler's Multivariate Normality Test
mvnPlot(res2, type = "persp", default = TRUE)

mydatas3 <- subset.matrix(mydatas_m,select=c(GAnalytic,TOEFL))
res3 <- hzTest(mydatas3)              #Henze-Zirkler's Multivariate Normality Test
mvnPlot(res3, type = "persp", default = TRUE)

##Multivariate normality test
hzTest(mydatas,cov = TRUE, qqplot = TRUE)                #Henze-Zirkler's Multivariate Normality Test
mardiaTest(mydatas, cov = TRUE, qqplot = TRUE)           #Mardia's Multivariate Normality Test

hzTest(multi_xndata,cov = TRUE, qqplot = FALSE)              #Henze-Zirkler's Multivariate Normality Test
mardiaTest(multi_xndata, cov = TRUE, qqplot = TRUE)          #Mardia's Multivariate Normality Test

#Collinearity test -- if VIF>4 then assume multicollinearity then remove
library(usdm)
xn_data <- data.frame(multi_xndata)                #Have to use data frame to use VIF
vif(xn_data)
```

D.2 R Code of Simulation for Variance Comparison

```
library(MASS)

#1-Bivariate
mu <- c(420,540)                #use GRE Verbal and TOEFL means
Sigma <- matrix(c(3450,600,600,970),2,2)      #Verbal and TOEFL covariance
n0 <- 2000

#Sigma <- matrix(c(3450,1200,1200,970),2,2)      #high correlation example

set.seed(2017612)
fobs <- mvrnorm(n0,mu=mu,Sigma=Sigma)        #x0 and y0
cov(fobs)
x0 <- fobs[,1]
y0 <- fobs[,2]
```

```
fob_head <- head(fobs,1)
x00 <- fob_head[,1]
y00 <- fob_head[,2]

#Obtain estimates
n <- 40
m <- 20      #m=20 - miss 50%; m=28 miss 30%; m=36 miss 10%;

sig_xx_n <- 0
sig_xx_m <- 0
xbar_m <- 0
xbar_n <- 0
ybar_m <- 0
mu_y <- 0
beta <- 0
sig_e <- 0

set.seed(20176132)
for (i in 1:10000)
{
  simdata<- mvrnorm(n,mu=mu,Sigma=Sigma)
  x <- simdata[,1]
  y <- simdata[,2]

  xbar_n[i] <- mean(x)
  sig_xx_n[i] <- sum((x-mean(x))^2)/n

  sub_data <- simdata[1:m,1:2]
  xm <- sub_data[,1]
  ym <- sub_data[,2]
  sig_xx_m[i] <- sum((xm-mean(xm))^2)/m

  xbar_m[i] <- mean(xm)
  ybar_m[i] <- mean(ym)

  beta_num <- sum((ym-mean(ym))*(xm-mean(xm)))
  beta_den <- sum((xm-mean(xm))^2)
  beta[i] <- beta_num/beta_den
  mu_y[i] <- ybar_m[i] - beta[i]*(xbar_m[i]-xbar_n[i])

  sig_e[i] <- sum((ym - ybar_m[i] - beta[i]*(xm - xbar_m[i]))^2)/m
}

beta_xbar_nCov <- cov(xbar_n,beta)                    #Check covariance
beta_xbar_mCov <- cov(xbar_m,beta)
mu_y_mu_x_cov <- cov(mu_y,xbar_n)
beta_mu_yCov <- cov(beta,mu_y)


#Comparison of variance mu_y_hat
c3 <- 1 + ((n-m)/(n*(m-3)))
var_muy_hat <- (c3*sig_e)/(m-2) + ((beta^2)*sig_xx_n)/(n-1)
var_muy_hat_m <- sig_e/(m-2) + ((beta^2)*sig_xx_m)/(m-1)

m_var_muy_hat <- mean(var_muy_hat)
var_muy_hat_sd <- (var(var_muy_hat))^0.5

m_var_muy_hat_m <- mean(var_muy_hat_m)
var_muy_hat_m_sd <- (var(var_muy_hat_m))^0.5

len_out_var <- data.frame(n,m,m_var_muy_hat,var_muy_hat_sd,m_var_muy_hat_m,var_muy_hat_m_sd)
```

References

1. Allison, P. D. (2002), *Missing Data*, SAGE University Papers
2. Anderson, T.W. (1957) *Maximum Likelihood Estimates for A Multivariate Normal Distribution when Some Observations Are Missing*, Journal of the American Statistical Association, Vol. 52, No. 278 (Jun., 1957), pp. 200-203
3. Anderson, T.W. (2015) *An Introduction to Multivariate Statistical Analysis*, Third Edition. Wiley. Reprint 2015.
4. Chung, Hie-Choon and Han,Chien-Pai.(2000) *Discriminant Analysis When A Block of Observations Is Missing,* Annals of the Institute of Statistical Mathematics, Vol. 52, No. 3, 544-556.
5. Edgett, G. L. (1956) *Multiple Regression with Missing Observations Among the Independent Variables,* Journal of the American Statistical Association, Vol. 51, No. 273 (Mar., 1956), pp. 122-131
6. Han, Chien-Pai and Li, Yan. (2011) *Regression Analysis with Block Missing Values and Variables Selection,* Parkistan Journal of Statistics and Operation Research, 7, 391-400.
7. Hogg, R.V., McKean, J.W. and Craig, A.T. (2013) *Introduction to Mathematical Statistics*, 7th Edition, Pearson Education, Inc.
8. Howell, D.C. (2008) *The analysis of missing data*. In Outhwaite, W. & Turner, S. Handbook of Social Science Methodology. London: Sage
9. Johnson, Richard A. and Wichern, Dean W. (1998) *Applied Multivariate Statistical Analysis*, Fourth Edition, Prentice-Hall, Inc. NJ.
10. Kendall, M.G. and Stuart, A. (1945) *The Advanced Theory of Statistics*, Vol.1, 382-393, London.

11. Korkmaz, S., Goksuluk, D. and Zararsiz, G. (2016) *Package 'MVN'.* URL http://www.biosoft.hacettepe.edu.tr/MVN/

12. Kutner, Michael H., Nachtsheim, Christopher J., Neter, John and Li, William. (2005) *Applied Linear Statistical Models.* Fifth Edition. McGraw-Hill Irwin.

13. Little, R. J. A., and Rubin, D. B. (2002), *Statistical Analysis With Missing Data* ,Second Edition,New York: Wiley

14. Little, R.J. and Zhang, Nanhua, (2011) *Subsample Ignorable Likelihood for Regression Analysis with Missing Data*, Journal of Royal Statistical Society, Appl. Statist. (2011) 60, Part 4, 591-605.

15. Loan, Charles Van. (2009) *The Kronecker Product - A Product of the Times.* URL: https://www.cs.cornell.edu/cv/ResearchPDF/KPHist.pdf, Presented at the SIAM Conference on Applied Linear Algebra, Monterey, California, October 26, 2009

16. Nadarajah, S. and Gupta, A. K. (2005) *A Skewed Truncated Pearson Type VII Distribution,* J. Japan Statist. Soc. Vol. 35 No. 1 2005 61–71

17. Nydick, Steven W. (2012) *The Wishart and Inverse Wishart Distributions*, URL: https://pdfs.semanticscholar.org/ac51/ee74af59c432d493da98bd950cc6f856a0ca.pdf, May 25, 2012

18. Papanicolaou, A. (2009) *Taylor Approximation and the Delta Method*. April 28, 2009.

19. Petersen, K. B. and Pedersen, M.S. (2012) *The Matrix Cookbook*, http://matrixcookbook.com/

20. Sinsomboonthong, J. (2011) *Jackknife Maximum Likelihood Estimates for a Bivariate Normal Distribution with Missing Data*, Thailand Statistician, July 2011; 9(2): 151 -169

21. Rawlings, John O., Pantula, Ssatry G. and Dickey, David A. (1998) *Applied Regression Analysis: A Research Tool.* Second Edition. Springer, New York, 1998.

22. Rubin, D.B. (1976) *Inference and Missing Data*, Biometrika, Vol 63, Issue 3 (Dec., 1976), 581 - 592

23. Sun, J., et al.(2010) *Robust mixture clustering using Pearson type VII distribution*. Pattern Recognition Lett. (2010), doi:10.1016/ j.patrec.2010.07.015

Biographical Information

Yi Liu was born in Sichuan, China in 1963. She obtained her B.S. degree in Physics and M.S. degree in Theoretical Physics from Beijing Normal University in 1984 and 1987, respectively. She worked for China Aerospace Engineering Consultation Center from 1987 to 2008.

She enrolled in Department of Mathematics in University of Texas at Arlington in 2012, and obtained her M.S. and PhD degrees in Statistics from University of Texas at Arlington in 2014 and 2017, respectively. She worked for Thomas J. Stephens & Associates as a Biostatistician for clinical research from 2014 to 2015. She began to work for Sabre as data analytics since September 2015. Her current interests are big data and machine learning.