# MULTIVARIATE REGRESSION

**Content:** **(7 Lectures)**

- Assumptions of Multivariate Regression Models
- Parameter Estimation
- Multivariate Analysis of Variance and Covariance (MANOVA & MANCOVA)

***Book for Reference: Applied Multivariate Statistics by Wischern and Johnson***

| Simple Regression Model | Multiple Regression Model | Multivariate Regression Model |
| --- | --- | --- |
| One Dependent & one Independent Variable | One Dependent & multiple Independent Variables | Multiple Dependent & Multiple Independent Variables |
| Sales of Ice cream and Temperature | Sales of Ice cream and Temperature<br>& hit of tornado | If it is too hot, ice cream sales increase; If a tornado hits, water and canned foods sales increase while ice cream, frozen foods and meat will decrease; If gas prices increase, prices on all goods will increase |
| Y<br>and<br>X | Y<br>and<br>$X_1, X_2, X_3 .......X_n$ | $Y_1, Y_2, Y_3 ......Y_n$<br>and<br>$X_1, X_2, X_3 .......X_n$ |

## Introduction

- As the name implies, multivariate regression is a technique that estimates a single regression model with more than one outcome variable.
- When there is more than one predictor variable in a multivariate regression model, the model is a multivariate multiple regression.

## Example 1:

A researcher has collected data on three psychological variables, four academic variables (standardized test scores), and the type of educational program the student is in for 600 high school students. She is interested in how the set of psychological variables is related to the academic variables and the type of program the student is in.

## Example 2:

A doctor has collected data on cholesterol, blood pressure, and weight. She also collected data on the eating habits of the subjects (e.g., how many ounces of red meat, fish, dairy products, and chocolate consumed per week). She wants to investigate the relationship between the three measures of health and eating habits.

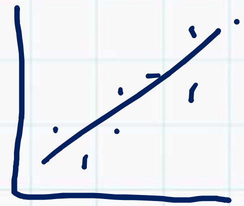
## Example 3:

A researcher is interested in determining what factors influence the health African Violet plants. She collects data on the average leaf diameter, the mass of the root ball, and the average diameter of the blooms, as well as how long the plant has been in its current container. For predictor variables, she measures several elements in the soil, as well as the amount of light and water each plant receives.

## Example 4:

Can a supermarket owner maintain stock of water, ice cream, frozen foods, canned foods and meat as a function of temperature, tornado chance and gas price during tornado season in June? From this question, several obvious assumptions can be drawn: If it is too hot, ice cream sales increase; If a tornado hits, water and canned foods sales increase while ice cream, frozen foods and meat will decrease; If gas prices increase, prices on all goods will increase. A mathematical model, based on multivariate regression analysis will address this and other more complicated questions.

# Simple Linear Regression:



$x \mid y$

$$y = \underbrace{\beta_0 + \beta_1 x}_{\text{Normal eq}^n}$$

$$\sum y = n\beta_0 + \beta_1 \sum x \quad \text{——} \quad (1)$$

$$\sum xy = \beta_0 \sum x + \beta_1 \sum x^2 \quad \text{——} \quad (2)$$

$$(\beta_1, \beta_2 \dots \beta_k)$$

$$y \mid x_1, x_2, x_3 \dots x_k$$

$$\beta_0 \mid \beta_1$$

## Classical Model of Multivariate Regression Model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad i = 1, \cdots, n$$

We can write model in matrix form as,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \qquad \text{or } Y = X\beta + \epsilon,$$

where $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ and $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$

**Where:**

- X is called the design matrix.
- β is the vector of parameters
- ϵ is the error vector
- Y is the response vector

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

**We want to estimate β**

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i.$$

$\rightarrow$ MLR

The subscript i denotes the observational unit from which the observations on Y and the p independent variables were taken. The second subscript designates the independent variable. The sample size is denoted with n, i = 1…….n, and p denotes the number of independent variables.

**Four matrices are needed to express the linear *Matrix* model in matrix notation:**

- Y : the n×1 column vector of observations on the dependent variable Yi;
- X: the n × p_ matrix consisting of a column of ones, which is labeled 1, followed by the p column vectors of the observations on the independent variables;
- β: the p_ × 1 vector of parameters to be estimated; and
- ϵ: the n × 1 vector of random errors

With these definitions, the linear model can be written as

$$Y = X\beta + \epsilon$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

Given

← Unknown

Unknown

$$\underset{(n \times 1)}{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}} = \underset{(n \times p')}{\begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \cdots & X_{np} \end{bmatrix}} \underset{(p' \times 1)}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}} + \underset{(n \times 1)}{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}}.$$

# Ordinary Least Squares Regression

Ordinary least squares (OLS) regression is a statistical method of analysis that estimates the relationship between one or more independent variables and a dependent variable; the method estimates the relationship by minimizing the sum of the squares in the difference between the observed and predicted values of the dependent variable configured as a straight line

$y = (\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon$

$\varepsilon = (y - \beta_0 - \beta_1 x_1 - \dots \beta_p x_p)^2$

$\Sigma \varepsilon_i^2 = \Sigma(y - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})$

The OLS estimate of $\beta$ is obtained by minimising

$$\Sigma \hat{\epsilon}_i^2 = \Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Applied multivariate statistics

to get the normal equations (don't do this)

$$\cancel{\begin{aligned} n\hat{\beta}_0 + \Sigma x_i \hat{\beta}_1 &= \Sigma y_i \\ \Sigma x_i \hat{\beta}_0 + \Sigma x_i^2 \hat{\beta}_1 &= \Sigma x_i y_i. \end{aligned}}$$

The normal equations can be written in matrix form as:

$$\underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}}_{X'} \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}}_{\beta} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}}_{X'} \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{y}.$$

Check by multiplying the matrices out.

This is written in matrix notation as

$$X'X\hat{\beta} = X'Y.$$

As the matrix $X'$ is $2 \times n$ and $X$ is $n \times 2$, $X'X$ is a $2 \times 2$ matrix.

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}_{2\times 1}$$

If $(X'X)^{-1}$ exists, we can solve the matrix equation as follows:

$(X'X)\hat{\beta} = X'Y$

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

$$\begin{aligned} X'X\hat{\beta} &= X'Y \\ (X'X)^{-1}(X'X)\hat{\beta} &= (X'X)^{-1}X'Y \\ I\hat{\beta} &= (X'X)^{-1}X'Y \\ \boxed{\hat{\beta} = (X'X)^{-1}X'Y} \end{aligned}$$

$X$ $X^T$ $Y =$ [ ]

$X' =$ [ ]

This is a fundamental result of the OLS theory using matrix notation. The result holds for a multiple linear regression model with $k-1$ explanatory variables in which case $X'X$ is a $k \times k$ matrix.

## Parameter Estimation of Multivariate Regression

Various methods of estimation can be used to determine the estimates of the parameters. Among them, the methods of least squares and maximum likelihood are the popular methods of estimation.

| | |
|---|---|
| Model: | $Y = X\beta + \epsilon$ |
| Normal equations: | $(X'X)\beta = X'Y$ |
| Parameter estimates: | $\widehat{\beta} = (X'X)^{-1}X'Y$ |
| Fitted values: | $\widehat{Y} = X\widehat{\beta}$ |
| | $= PY$, where $P = X(X'X)^{-1}X'$ |
| Residuals: | $e = Y - \widehat{Y}$ |
| | $= (I - P)Y$ |
| Variance of $\widehat{\beta}$ : | $\mathrm{Var}(\widehat{\beta}) = (X'X)^{-1}\sigma^2$ |
| Variance of $\widehat{Y}$ : | $\mathrm{Var}(\widehat{Y}) = P\sigma^2$ |
| Variance of $e$ : | $\mathrm{Var}(e) = (I - P)\sigma^2$ |

## Parameter Estimation of Multivariate Regression

$$\boxed{\hat{\beta} = (X'X)^{-1}.X'Y}$$

## Problem

Determine the linear regression model for fitting a straight line:

| X | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Y | 1 | 4 | 3 | 8 | 9 |

$$y = X\beta + \varepsilon$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} ; \quad X = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{bmatrix} ; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} ; \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$y = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix}_{5\times 1} \qquad X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}_{5\times 2} \qquad X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}$$

**Problem No. 1**

Let $X = \underset{3\times 2}{\begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 6 \end{bmatrix}}$, $Y = \underset{3\times 1}{\begin{bmatrix} 3 \\ 6 \\ 7 \end{bmatrix}}$ with $\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$.

Show that $\hat{\beta} = \begin{bmatrix} 1.3333 \\ 1.0000 \end{bmatrix}$.

$$\hat{\beta} = (X'X)^{-1} \cdot (X'Y)$$

**Problem No. 1**

$$x = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 6 \end{bmatrix} \quad , \quad x' = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 6 \end{bmatrix}$$

$$x'x = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 12 \\ 12 & 56 \end{bmatrix}_{2\times 2}$$

$|x'x| = 3(56) - 12 \times 12 = 168 - 144 = 24$ ; so $(x'x)^{-1}$ exist

$$(x'x)^{-1} = \frac{1}{24} \begin{bmatrix} 56 & -12 \\ -12 & 3 \end{bmatrix}$$

## Problem No. 1

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

$$= \frac{1}{24}\begin{bmatrix} 56 & -12 \\ -12 & 3 \end{bmatrix} \underset{2\times 3}{\begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 6 \end{bmatrix}} \underset{3\times 1}{\begin{bmatrix} 3 \\ 6 \\ 7 \end{bmatrix}}$$

$$= \frac{1}{24}\underset{2\times 2}{\begin{bmatrix} 56 & -12 \\ -12 & 3 \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} 16 \\ 72 \end{bmatrix}}$$

$$= \frac{1}{24}\begin{bmatrix} 32 \\ 24 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} 32/24 \\ 24/24 \end{bmatrix} = \begin{bmatrix} 1.3333 \\ 1.0000 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

**Parameter Estimation of Multivariate Regression**

Predicted Value: $\hat{Y} = X.\hat{\beta} = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} 1.33 \\ 1.00 \end{bmatrix}$

Residual Error: $\varepsilon = Y - \hat{Y}$

Residual sum of squares $= \varepsilon.\varepsilon^{T}$

sum of square of Error

## Problem No. 2

Determine the least square estimate, predicted value, residual and residual sum of squares of the linear regression model for fitting a straight line:

| X | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Y | 1 | 4 | 3 | 8 | 9 |

$$y = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} ; \quad x = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} ; \quad x' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}$$

$$x'x = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix}$$

## Problem No. 2

$|X'X| = 150 - 100 = 50 \quad (X'X)^{-1} \text{ exist}$

$$(X'X)^{-1} = \frac{1}{50}\begin{bmatrix} 30 & -10 \\ -10 & 5 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix}$$

$$(X'Y) = \begin{bmatrix} 25 \\ 70 \end{bmatrix}$$

$$\hat{\beta} = (X'X)^{-1}(X'Y) = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix}\begin{bmatrix} 25 \\ 70 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

$$\boxed{\hat{y} = 1 + 2X}$$

## Problem No. 2

$$\hat{y} = X\hat{\beta} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}_{\underline{5\times2}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{\underline{2\times1}} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix}$$

So, $\varepsilon = y - \hat{y} = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}$

Residuals sum of squares $\Rightarrow \varepsilon'\varepsilon = \begin{bmatrix} 0 & 1 & -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} = 6$

# Problem No. 3

Consider the table below. It shows three performance measures for five students.

| | $y$ | $x_1$ | $x_2$ |
|---|---|---|---|
| **Student** | **Test score** | **IQ** | **Study hours** |
| **1** | 100 | 110 | 40 |
| **2** | 90 | 120 | 30 |
| **3** | 80 | 100 | 20 |
| **4** | 70 | 90 | 0 |
| **5** | 60 | 80 | 10 |

$$X = \begin{bmatrix} 1 & 110 & 40 \\ 1 & 120 & 30 \\ 1 & 100 & 20 \\ 1 & 90 & 0 \\ 1 & 80 & 10 \end{bmatrix}$$

Using least squares regression, develop a regression equation to predict test score, based on (1) IQ and (2) the number of hours that the student studied.

# Problem No. 3

$$y = \begin{bmatrix} 100 \\ 90 \\ 80 \\ 70 \\ 60 \end{bmatrix} \qquad x = \begin{bmatrix} 1 & 110 & 40 \\ 1 & 120 & 30 \\ 1 & 100 & 20 \\ 1 & 90 & 0 \\ 1 & 80 & 10 \end{bmatrix} \qquad x' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 110 & 120 & 100 & 90 & 80 \\ 40 & 30 & 20 & 0 & 10 \end{bmatrix}$$

$$x'x = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 110 & 120 & 100 & 90 & 80 \\ 40 & 30 & 20 & 0 & 10 \end{bmatrix} \begin{bmatrix} 1 & 110 & 40 \\ 1 & 120 & 30 \\ 1 & 100 & 20 \\ 1 & 90 & 0 \\ 1 & 80 & 10 \end{bmatrix} = \begin{bmatrix} 5 & 500 & 100 \\ 500 & 51000 & 10,800 \\ 100 & 10,800 & 3000 \end{bmatrix}$$

$$(x'x)^{-1} = \begin{bmatrix} 101/5 & -7/30 & 1/6 \\ -7/30 & 1/360 & -1/450 \\ 1/6 & -1/450 & 1/310 \end{bmatrix}$$

## Problem No. 3

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

$$= \begin{bmatrix} 101/5 & -7/30 & 1/6 \\ -7/30 & 1/360 & -1/450 \\ 1/6 & -1/450 & 1/360 \end{bmatrix}_{3\times3} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 110 & 120 & 100 & 90 & 80 \\ 40 & 30 & 20 & 0 & 10 \end{bmatrix}_{3\times5} \begin{bmatrix} 100 \\ 90 \\ 80 \\ 70 \\ 60 \end{bmatrix}_{5\times1}$$

$$\hat{\beta} = \begin{bmatrix} 20 \\ 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

$$\hat{y} = 20 + 0.5x_1 + 0.5x_2$$

## Parametric Estimation of Multivariate Regression

$$Var(\hat{\beta}) = (X'X)^{-1}.\sigma^2$$

$\sigma^2 \Rightarrow S^2_{\varepsilon}$

$\sigma^2 \rightarrow MSE = \dfrac{SSE}{n-k-1}$

$n$ ↓ no. of obs$^n$

$k$ ↳ no. of predictors

Hat Matrix

$$Var(\hat{Y}) = P.\sigma^2, where \quad P = X(X'X)^{-1}.X'$$

$$Var(\varepsilon) = (I - P).\sigma^2$$

## Problem No. 4

An auto part is manufactured by a company once a month in lots that vary in size as demand fluctuates. The data below represent observations on lot size (y), and number of man-hours of labour (x) for 10 recent production runs. Suppose that you need to fit the simple regression model. In vector form the data are:

$$\mathbf{Y} = \begin{pmatrix} 73 \\ 50 \\ 128 \\ 170 \\ 87 \\ 108 \\ 135 \\ 69 \\ 148 \\ 132 \end{pmatrix} \begin{array}{l} \to y_1 \\ \to y_2 \\ y_3 \\ \vdots \\ \\ \\ \\ \\ \\ y_{10} \end{array} \text{ and } \mathbf{X} = \begin{pmatrix} 1 & 30 \\ 1 & 20 \\ 1 & 60 \\ 1 & 80 \\ 1 & 40 \\ 1 & 50 \\ 1 & 60 \\ 1 & 30 \\ 1 & 70 \\ 1 & 60 \end{pmatrix}.$$

a. Find the least squares estimator of $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)'$. $\to (X'X)^{-1}(X'Y)$

b. Find the variance-covariance matrix of the previous estimator. $\to V(\hat{\beta})$

c. Compute the estimate $s_e^2$ of $\sigma^2$.

d. Using your answers to parts (b) and (c) find the variances of $\hat{\beta}_0, \hat{\beta}_1$.

e. Find the fitted value $\hat{y}_1$ and its variance.

f. What is the variance of the first residual ($\text{var}(e_1)$)?

$\hat{y} = 10 + 2x \Rightarrow \boxed{\hat{y}_1 = 70}$

# Problem No. 4

a) $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (X'X)^{-1}(X'Y) = \begin{pmatrix} 0.83529412 & -0.01470588 \\ -0.01470588 & 0.00029412 \end{pmatrix} \begin{pmatrix} 1100 \\ 61800 \end{pmatrix} = \begin{pmatrix} 10 \\ 2 \end{pmatrix}$

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 30 & 20 & 60 & 80 & 40 & 50 & 60 & 30 & 70 & 60 \end{bmatrix} \begin{bmatrix} 1 & 30 \\ 1 & 20 \\ 1 & 60 \\ 1 & 80 \\ 1 & 40 \\ 1 & 50 \\ 1 & 60 \\ 1 & 30 \\ 1 & 70 \\ 1 & 60 \end{bmatrix} = \begin{bmatrix} 10 & 500 \\ 500 & 28400 \end{bmatrix}$$

## Problem No. 4

$var(\hat{\beta}_0) =$ & $v(\hat{\beta}_1) =$

b) $var(\hat{\beta}) = (X'X)^{-1}\sigma^2 = \begin{pmatrix} 0.83529412 & -0.01470588 \\ -0.01470588 & 0.00029412 \end{pmatrix}_{2\times 2} \sigma^2 \rightarrow (7.5)$

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}\sigma_{21} \\ \sigma_{21}\sigma_{12} & \sigma_{22}^2 \end{bmatrix} \rightarrow \text{variance - covariance Matrix}$$

var

covariance

$\rightarrow$ Covariance

variance

c) $\sigma^2 = MSE \qquad S_E^2 = \dfrac{SSE}{n-k-1}$

$$SSE = Y'Y - \hat{\beta}X'Y = 134660 - (10 \quad 2)\begin{pmatrix} 1100 \\ 61800 \end{pmatrix} = 134660 - 134600 = 60$$

$$S_E^2 = \frac{SSE}{n-k-1} = \frac{60}{8} = 7.5$$

# Problem No. 4

$v(\hat{\beta}_0) = 7.5(0.8352) = 6.2647$ ; $v(\hat{\beta}_1) = 7.5(0.00029412) = 0.0022$.

$\hat{y}_1 = 10 + (30 \times 2) = 70$

$v(\hat{y}) = \sigma^2 P$ where $\underset{10\times 10}{P} = \underset{10\times 2}{X}\ \underset{2\times 2}{(X'X)^{-1}}\ \underset{2\times 10}{X'}$

$$\underset{10\times 10}{\begin{bmatrix} \boxed{h_{11}} & h_{12} & h_{13} & \cdots & h_{1,10} \\ \vdots & h_{22} & & & \\ & & \ddots & & \\ & & & \ddots & \\ h_{10\times 1} & & & & h_{10,10} \end{bmatrix}}$$

$h_{11} = (1 \quad 30)\left((X'X)^{-1}\right)\begin{pmatrix}1\\30\end{pmatrix} = 0.218$

$v(\hat{y}_1) = h_{11}\sigma^2 = \sigma^2(0.218) = 7.5(0.218) = 1.635$

## Problem No. 4

$$V(\varepsilon_1) = \sigma^2(1 - h_{11}) = 7.5(1 - 0.218)$$

$$= 7.5 \times 0.782$$

$$= 5.865$$

## Problem No. 5

The data in the accompanying table relate grams plant dry weight Y to percent soil organic matter X1 and Kg of supplemental soil nitrogen added per 1000 square meters X2.

| Y | X1 | X2 |
|---|---|---|
| 78.5 | 7 | 2.6 |
| 74.3 | 1 | 2.9 |
| 104.3 | 11 | 5.6 |
| 111.1 | 13 | 6.3 |
| 115 | 8.1 | 3.1 |

a. Define $X, \hat{Y}, \hat{\beta} \,\&\, \varepsilon$ for the Model.

b. Write the regression equation.

c. Compute $Var(\hat{Y}), Var(\hat{\beta}) \,\&\, Var(\varepsilon)$

# Problem No. 5