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 & 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 and X	Y and $X_1, X_2, X_3 \ldots\ldots X_n$	$Y_1, Y_2, Y_3 \ldots\ldots Y_n$ and $X_1, X_2, X_3 \ldots\ldots X_n$

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:

$x \mid y$

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

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

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

$(\beta_1 \beta_2 \cdots \beta_k)$

$y \mid x_1 \; x_2 \; x_3 \ldots x_k$

$\beta_0 \mid \beta_1$

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

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

We want to estimate β

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

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

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

$\varepsilon = (y - \beta_0 - \beta_1 x_1 - \cdots - \beta_p x_p)_2$

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

The OLS estimate of β is obtained by minimising

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

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

Applied multivariate statistics

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

The normal equations can be written in matrix form as:

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

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 2×2 matrix.

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

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

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

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

$X \quad Y = [\quad]$

X^T

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

$[\quad]$

$X' = [\quad]$

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

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

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

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

Problem No. 1

Let $X = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 6 \end{bmatrix}_{3\times 2}$, $Y = \begin{bmatrix} 3 \\ 6 \\ 7 \end{bmatrix}_{3\times 1}$ 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

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

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

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

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

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

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

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

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

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

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

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

Problem No. 2

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

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. 2

Problem No. 3

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

Student	Test score (y)	IQ (x_1)	Study hours (x_2)
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

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

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

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

Problem No. 3

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

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

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

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

Problem No. 3

Problem No. 3

Parametric Estimation of Multivariate Regression

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

$\sigma^2 \Rightarrow s_\varepsilon^2$

$$MSE = \frac{SSE}{n-k-1}$$

n → no. of obs^n

k → no. of predictors

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

P → Hat Matrix

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

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:

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

$\to (X'X)^{-1}(X'Y)$

a. Find the least squares estimator of $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)'$.

$\to V(\hat{\beta})$

b. Find the variance-covariance matrix of the previous estimator.

c. Compute the estimate s_e^2 of σ^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) $var(\hat{\beta}) = (X'X)^{-1}\sigma^2 = \begin{pmatrix} 0.83529412 & -0.01470588 \\ -0.01470588 & 0.00029412 \end{pmatrix}_{2\times 2}$ $\sigma^2 \rightarrow (7.5)$

var

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

$\rightarrow$ Covariance

Covariance

Variance

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

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

$$= 60$$

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

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 $P_{10\times10} = X_{10\times2} (X'X)^{-1}_{2\times2} X'_{2\times10}$

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

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

Problem No. 4

$$v(\varepsilon_i) = \sigma^2(1 - h_{ii}) = 7.5(1 - 0.218)$$

$$= 7.5 \times 0.782$$

$$= 5.865$$

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

ANOVA

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent groups.

ANOVA Table

Source	Sum of squares	Degree of Freedom	Mean squares	F	F-test
Treatment	SS_T	k-1	$MS_T = \frac{SS_T}{k-1}$	$F = \frac{MS_T}{MS_E}$	$F > F_{\alpha,k-1,N-k}$?
Error	SS_E	N-k	$MS_E = \frac{SS_E}{N-k}$		
Total	TotalSS	N-1			

MANOVA is short

for

Multivariate ANalysis Of VAriance

MANOVA

- Multivariate analysis of variance (MANOVA) is an extension of analysis of variance (ANOVA).
- MANOVA is just an ANOVA with several dependent variables.

Analysis of Variance

$$\underset{\text{(metric)}}{Y_1} = \underset{\text{(nonmetric)}}{X_1 + X_2 + X_3 + \cdots + X_n}$$

Multivariate Analysis of Variance

$$\underset{\text{(metric)}}{Y_1 + Y_2 + Y_3 + \cdots + Y_n} = \underset{\text{(nonmetric)}}{X_1 + X_2 + X_3 + \cdots + X_n}$$

MANOVA

Suppose we are studying three different teaching methods for a course. This variable is our independent variable. We also have student satisfaction scores and test scores. These variables are our dependent variables. We want to determine whether the mean scores for satisfaction and tests differ between the three teaching methods.

You might wish to test the hypothesis that gender and ethnicity interact to influence a set of job-related outcomes including attitudes toward co-workers, attitudes toward supervisors, feelings of belonging in the work environment, and identification with the corporate culture

MANOVA

ONE-WAY MANOVA EXAMPLE
Dependent Variables
(Response)
Independent Variable
(Factor)
Test Score
Level of Education
(High School, College Degree,
or Graduate Degree)
(Response)
Annual Income

MANOVA

TWO-WAY MANOVA EXAMPLE
Independent Variables
(Factor)
Level of Education (High School, College Degree, or Graduate Degree)
(Factor)
Zodiac Sign
Dependent Variables
(Response)
Test Score
(Response)
Annual Income

MANOVA

ANOVA

$H_0 : \mu_1 = \mu_2 = \ldots \mu_k$ vs $H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \cdots \neq \mu_k$

Null hypothesis (H_0) = all the group means are equal, that is, they come from the same population.

MANOVA

$$H_0 : \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \\ \mu_{p1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \\ \mu_{p2} \end{bmatrix} = \ldots\ldots\ldots = \begin{bmatrix} \mu_{1k} \\ \mu_{2k} \\ \\ \mu_{pk} \end{bmatrix}$$

Null hypothesis (H_0) = all the group mean vectors are equal, that is, they come from the same population.

μ_{pk} = means of variable p, group k

1. **Wilks Lambda**

$$\Lambda^* = \frac{|\mathbf{E}|}{|\mathbf{H}+\mathbf{E}|}$$

Here, the determinant of the error sums of squares and cross products matrix **E** is divided by the determinant of the total sum of squares and cross products matrix **T** = **H** + **E**. If **H** is large relative to **E**, then |**H** + **E**| will be large relative to |**E**|. Thus, we will reject the null hypothesis if Wilks lambda is small (close to zero).

2. **Hotelling-Lawley Trace**

$$T_0^2 = trace(\mathbf{HE}^{-1})$$

Here, we are multiplying **H** by the inverse of **E**; then we take the trace of the resulting matrix. If **H** is large relative to **E**, then the Hotelling-Lawley trace will take a large value. Thus, we will reject the null hypothesis if this test statistic is large.

3. **Pillai Trace**

$$V = trace(\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1})$$

Here, we are multiplying **H** by the inverse of the total sum of squares and cross products matrix **T** = **H** + **E**. If **H** is large relative to **E**, then the Pillai trace will take a large value. Thus, we will reject the null hypothesis if this test statistic is large.

4. **Roy's Maximum Root: Largest eigenvalue of HE^{-1}**

Here, we multiply **H** by the inverse of **E**, and then compute the largest eigenvalue of the resulting matrix. If **H** is large relative to **E**, then the Roy's root will take a large value. Thus, we will reject the null hypothesis if this test statistic is large.

Working Procedure:

Step I: Arrange the observations of all groups in matrix form.

Step II: Set the Hypothesis.

Step III: Find the row wise average and grand average.

Step IV: For each observation of all groups, calculate:

$$x_{ni} = \bar{x} + [\bar{x}_n - \bar{x}] + [x_{ni} - \bar{x}_n]$$

Observation Mean Treatment effect Residual

Step V: Write all observations of all groups in matrix form:

Observation=Mean + Treatment effect+ Residual

Working Procedure:

Step VI: Calculate sum of squares for observation, mean, treatment effect and residuals for all groups i.e. SS_{obs}, SS_{mean}, SS_{te} and SS_{res}

[Note: $SS_{obs} = SS_{mean} + SS_{te} + SS_{res}$]

Step VII: Calculate Corrected sum of squares for all groups:

$$SS_t = SS_{obs} - SS_{mean}$$

Step VIII: Obtain cross product contributions, we proceed row by row in the arrays for the two variables for observation, mean, treatment and residuals.

Step IX: Calculate total corrected cross product=Observed cross product- Mean Cross product.

Working Procedure:

Step X: Find B stands for variances of observations between samples

$$B = \begin{bmatrix} SS_{tr}group1 & \text{Total Cross Product} \\ \text{Total Cross Product} & SS_{tr}group2 \end{bmatrix}$$

Step XI: Find W - stands for variances of observations within sample.

$$W = \begin{bmatrix} SS_{res}group1 & \text{Residual Cross Product} \\ \text{Residual Cross Product} & SS_{res}group2 \end{bmatrix}$$

Working Procedure:

Step XII: Manova Table:

Source of Variation	Matrix of SS and cross product	Degree of freedom	Wilk's lambda [equivalent form of the F-test]	Test Statistics
Treatment	B	g-1	$\wedge^* = \frac{\lvert W \rvert}{\lvert W+B \rvert}$	$\left[\frac{1-\sqrt{\wedge^*}}{\sqrt{\wedge^*}}\right]\left[\frac{N-g-1}{g-1}\right]$
Residual	W	N-g		
Total(Corrected)	B+W	N-1		

Working Procedure:

ν_1 ν_2

Step XIII: Find the tabulated/Critical value for F test with degree of freedom 2(g-1) and 2(N-g-1) at LOS 5 %.

Step XIV: Comparison and Conclusion:

$F_{cal} > F_{tab}$

If Test Statistic > Critical Value then reject Null Hypothesis

If Test Statistic < Critical Value then do not reject Null Hypothesis

$F_{cal} < F_{tab}$

Problem 1:

Consider the following dataset where there are 2 dependent variables and 3 groups.

	Group	For DV 1 (y_1)	For DV 2 (y_2)	
x_1	Group 1	9, 6, 9	3,2,7	$n_1 = 3$
x_2	Group 2	0,2	4,0	$n_2 = 2$
x_3	Group 3	3,1,2	8,9,7	$n_3 = 3$

Whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

Step I: Arrange the observations of all groups in matrix form.

$$\begin{pmatrix} \begin{bmatrix} 9 \\ 3 \end{bmatrix} & \begin{bmatrix} 6 \\ 2 \end{bmatrix} & \begin{bmatrix} 9 \\ 7 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 4 \end{bmatrix} & \begin{bmatrix} 2 \\ 0 \end{bmatrix} & \\ \begin{bmatrix} 3 \\ 8 \end{bmatrix} & \begin{bmatrix} 1 \\ 9 \end{bmatrix} & \begin{bmatrix} 2 \\ 7 \end{bmatrix} \end{pmatrix} \begin{array}{l} \rightarrow \text{Group } 1 \\ \rightarrow \text{Group } 2 \\ \rightarrow \text{Group } 3 \end{array}$$

$$\bar{x}_1 = \begin{bmatrix} 8 \\ 4 \end{bmatrix} \quad \bar{x}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\bar{x}_3 = \begin{bmatrix} 2 \\ 8 \end{bmatrix} \quad \bar{x} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

Step III: Find the row wise average and grand average.

$$\bar{x}_1 = \begin{bmatrix} 9+6+9/3 \\ 3+2+7/3 \end{bmatrix} ; \quad \bar{x}_2 = \begin{bmatrix} 0+2/2 \\ 4+0/2 \end{bmatrix} ; \quad \bar{x}_3 = \begin{bmatrix} 3+1+2/3 \\ 8+9+7/3 \end{bmatrix}$$

$$\bar{x} = \begin{bmatrix} \dfrac{9+6+9+0+2+3+1+2}{8} \\ 3+2+7+4+0+8+9+7/8 \end{bmatrix}$$

Step IV: For each observation of all groups, calculate:

$$x_{ni} = \bar{x} + [\bar{x}_n - \bar{x}] + [x_{ni} - \bar{x}_n]$$

Observation | Mean | Treatment effect | Residual

for DV 1

GI → 9 6 9 ⇒ x_{11} x_{12} x_{13}

GII → 0 2 ⇒ x_{21} x_{22}

GIII → 3 1 2 → x_{31} x_{32} x_{33}

$3 = x_{31}$, $\bar{x} = 4$, $\bar{x}_3 = 2$

$3 = x_{31} = 4 + [2 - 4] + [3 - 2]$

$= 4 + [-2] + [1] = 3$

$\bar{x}_1 = 8$, $\bar{x}_2 = 1$, $\bar{x}_3 = 2$

Step V: Write all observations of all groups in matrix form:

$$\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & \\ 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 & \\ 4 & 4 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 & \\ -2 & -2 & -2 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & \\ 1 & -1 & 0 \end{pmatrix}$$

Obs | mean | treatment effect | residual

Step VI: Calculate sum of squares for observation, mean, treatment effect and residuals for all groups i.e. SS_{obs}, SS_{mean}, SS_{te} and SS_{res}

$$SS_{obs} = 9^2 + 6^2 + 9^2 + 0^2 + 2^2 + 3^2 + 1^2 + 2^2 = 216$$

$$SS_{mean} = 8(4^2) = 128$$

$$SS_{te} = 3(4^2) + 2(-3)^2 + 3(-2)^2 = 78$$

$$SS_{res} = 1^2 + (-2)^2 + (1)^2 + (1)^2 + (-1)^2 = 10$$

$$SS_{obs} = SS_{mean} + SS_{te} + SS_{res} = 128 + 78 + 10 = 216$$

Step VII: Calculate Corrected sum of squares for all groups:

$SS_t = SS_{obs} - SS_{mean}$ $= 216 - 128$

$SS_t = 88$

for group 2,

$$\begin{pmatrix} 3 & 2 & 7 \\ 4 & 0 & \\ 8 & 9 & 7 \end{pmatrix} = \begin{pmatrix} 5 & 5 & 5 \\ 5 & 5 & \\ 5 & 5 & 5 \end{pmatrix} + \begin{pmatrix} -1 & -1 & -1 \\ -3 & -3 & \\ 3 & 3 & 3 \end{pmatrix} + \begin{pmatrix} -1 & -2 & 3 \\ 2 & -2 & \\ 0 & 1 & -1 \end{pmatrix}$$

$SS_{obs} = SS_{mean} + SS_{te} + SS_{residuals}$

$272 = 200 + 48 + 24$

Step VIII: Obtain cross product contributions, we proceed row by row in the arrays for the two variables for observation, mean, treatment and residuals.

$$\text{Mean} = 4(5) + 4(5) + \cdots + 4(5) = 8 \times 4 \times 5 = 160$$

$$\text{Treatment} = 3(4)(-1) + 2(-3)(-3) + 3(-2)(3) = -12 + 18 - 18 = -12$$

$$\text{residuals} = 1(-1) + (-2)(-2) + 1(3) \cdots 0(-1) = 1$$

$$\text{Total} = 9(3) + 6(2) + 9(7) + \cdots + 2(7) = 149$$

Step IX: Calculate total corrected cross product=Observed cross product - Mean Cross product.

$$= 149 - 160$$

$$= -11$$

Step X: Find B stands for variances of observations between samples

$$B = \begin{bmatrix} SS_{tr}group1 & \text{Total Cross Product} \\ \text{Total Cross Product} & SS_{tr}group2 \end{bmatrix}$$

$$B = \begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix}$$

Step XI: Find W- stands for variances of observations within sample.

$$W = \begin{bmatrix} SS_{res}group1 & \text{Residual Cross Product} \\ \text{Residual Cross Product} & SS_{res}group2 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$$

$$Btw = \begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix}$$

Step XII: MANOVA Table:

Source of Variation	Matrix of SS and cross product	Degree of freedom	Wilk's lambda [equivalent form of the F-test]	Test Statistics
Treatment	B	g-1	$\wedge^* = \frac{\lvert W \rvert}{\lvert W+B \rvert}$	$\left[\frac{1-\sqrt{\wedge^*}}{\sqrt{\wedge^*}}\right]\left[\frac{N-g-1}{g-1}\right]$
Residual	W	N-g		
Total(Corrected)	B+W	N-1		

Step XII: MANOVA Table:

$$\boxed{\wedge^* = \frac{|W|}{|W+B|}} = \frac{\begin{vmatrix} 10 & 1 \\ 1 & 24 \end{vmatrix}}{\begin{vmatrix} 88 & -11 \\ -11 & 72 \end{vmatrix}} = \frac{10(24) - 1^2}{88(72) - (-11)^2} = 0.0385$$

$$\left[\frac{1-\sqrt{\wedge^*}}{\sqrt{\wedge^*}}\right]\left[\frac{N-g-1}{g-1}\right] = \left[\frac{1-\sqrt{0.0385}}{\sqrt{0.0385}}\right]\left[\frac{8-3-1}{3-1}\right] = 8.19$$

Step XII: MANOVA Table:

Source of Variation	Matrix of SS and cross product	Degree of freedom	Wilk's lambda [equivalent form of the F-test]	Test Statistics
Treatment	$\begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix}$	$g-1$ $=3-1=2$	$\lambda^* = 0.0385$	8.19
Residual	$\begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$	$N-g$ $=8-3=5$		
Total(Corrected)	$\begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix}$	$N-1$ $=8-1=7$		

Step XIII: Find the tabulated/Critical value for F test with degree of freedom 2(g-1) and 2(N-g-1) at LOS 5 %.

$H_0 : T_i = 0 \qquad i = 1, 2 \ldots K$

vs

H_1 : Not H_0 or Not all $T_i = 0$

Test statistics $= 8.19$

$F_{v_1, v_2, \alpha} \Rightarrow v_1 = 2(g-1) = 4$

$v_2 = 2(N-g-1) = 8$

$F_{4,8,5\%} = 7.01$

$TS > F_{tab}$. we reject H_0

If Covariance matrix and mean vectors are given

Grand Mean

$$\boxed{\bar{x} = \frac{n_1 \overline{x_1} + n_2 \overline{x_2} + n_3 \overline{x_3}}{n_1 + n_2 + n_3}}$$

Within samples

$$\boxed{W = (n_1 - 1)S_1 + (n_2 - 1)S_2 + (n_3 - 1)S_3}$$

$$\boxed{B = n_1(\overline{x_1} - \bar{x})(\overline{x_1} - \bar{x})^T + n_2(\overline{x_2} - \bar{x})(\overline{x_2} - \bar{x})^T + n_3(\overline{x_3} - \bar{x})(\overline{x_3} - \bar{x})^T}$$

Between samples

Problem No. 2

Group	Number of observations	Sample mean vectors
$\ell = 1$ (private)	$n_1 = 271$	
$\ell = 2$ (nonprofit)	$n_2 = 138$	$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2.066 \\ .480 \\ .082 \\ .360 \end{bmatrix}; \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 2.167 \\ .596 \\ .124 \\ .418 \end{bmatrix}; \quad \bar{\mathbf{x}}_3 = \begin{bmatrix} 2.273 \\ .521 \\ .125 \\ .383 \end{bmatrix}$
$\ell = 3$ (government)	$n_3 = 107$	
	$\sum_{\ell=1}^{3} n_\ell = 516$	

$$\bar{x} = 271\begin{bmatrix} 2.06 \\ 0.48 \\ 0.82 \\ 0.360 \end{bmatrix} + 138\begin{bmatrix} 2.167 \\ 0.596 \\ 0.124 \\ 0.418 \end{bmatrix} + 107\begin{bmatrix} 2.273 \\ 0.521 \\ 0.125 \\ 0.383 \end{bmatrix}$$

Sample covariance matrices

$$\mathbf{S}_1 = \begin{bmatrix} .291 & -0.01 & 0.002 & 0.018 \\ -.001 & .011 & 0 & 0.003 \\ .002 & .000 & .001 & 0 \\ .010 & .003 & .000 & .010 \end{bmatrix}; \qquad \mathbf{S}_2 = \begin{bmatrix} .561 & 0.011 & 0.001 & 0.37 \\ .011 & .025 & & \\ .001 & .004 & .005 & \\ .037 & .007 & .002 & .019 \end{bmatrix};$$

$$\mathbf{S}_3 = \begin{bmatrix} .261 & & & \\ .030 & .017 & & \\ .003 & -.000 & .004 & \\ .018 & .006 & .001 & .013 \end{bmatrix}$$

$$(n_2 - 1) S_2 = 137 \begin{bmatrix} 0.561 & \\ 0.011 & \end{bmatrix}$$

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$$

$$= \begin{bmatrix} 2.1359 \\ 0.5195 \\ 0.102 \\ 0.3803 \end{bmatrix}$$

$$W = (n_1-1)S_1 + (n_2-1)S_2 + (n_3-1)S_3$$

$$(n_1-1)S_1 = 270\begin{bmatrix} 0.291 & & & \\ -0.001 & 0.011 & & \\ 0.002 & 0 & 0.001 & \\ 0.010 & 0.003 & 0 & 0.010 \end{bmatrix}$$

$$= \begin{bmatrix} 78.57 & & & \\ -0.27 & 2.97 & & \\ 0.054 & 0 & 0.27 & \\ 2.7 & 0.81 & 0 & 2.7 \end{bmatrix}$$

$$(n_2-1)S_2 = 137\begin{bmatrix} 0.561 & & & \\ 0.011 & 0.025 & & \\ 0.001 & 0.004 & 0.005 & \\ 0.037 & 0.007 & 0.002 & 0.019 \end{bmatrix} = \begin{bmatrix} 76.857 & & & \\ 1.507 & 3.425 & & \\ 0.137 & 0.548 & 0.685 & \\ 5.069 & 0.959 & 0.274 & 2.603 \end{bmatrix}$$

$$(n_3-1)S_3 = 106\begin{bmatrix} 0.261 & & & \\ 0.030 & 0.017 & & \\ 0.003 & 0 & 0.04 & \\ 0.018 & 0.006 & 0.01 & 0.013 \end{bmatrix}$$

$$=$$

$$W = \begin{bmatrix} 182.962 & & & \\ 4.408 & 8.2 & & \\ 1.695 & 0.633 & 1.494 & \\ 9.581 & 2.428 & 0.394 & 6.538 \end{bmatrix}$$

$$B = n_1(\bar{x}_1 - \bar{x})(\bar{x}_1 - \bar{x})' + n_2(\bar{x}_2 - \bar{x})(\bar{x}_2 - \bar{x})' + n_3(\bar{x}_3 - \bar{x})(\bar{x}_3 - \bar{x})'$$

$$n_1(\bar{x}_1 - \bar{x})(\bar{x}_1 - \bar{x})' = 271 \begin{bmatrix} 2.066 - 2.136 \\ 0.48 - 0.579 \\ 0.082 - 0.102 \\ 0.360 - 0.380 \end{bmatrix} \begin{bmatrix} 2.066 - 2.136 & -\|- & 0.360 - 0.380 \end{bmatrix}$$

$$= 271 \begin{bmatrix} -0.07 \\ -0.039 \\ -0.02 \\ -0.02 \end{bmatrix}_{4\times 1} \begin{bmatrix} -0.07 & -0.039 & -0.02 & -0.02 \end{bmatrix}_{1\times 4}$$

$=$

$$B = \begin{bmatrix} 3.475 & & & \\ 1.11 & 1.225 & & \\ 0.821 & 0.453 & 0.253 & \\ 0.584 & 0.610 & 0.23 & 0.304 \end{bmatrix} \qquad W = \begin{bmatrix} 182.962 & & & \\ 4.408 & 8.2 & & \\ 1.695 & 0.63 & 1.484 & \\ 9.581 & 2.428 & 0.354 & 6.538 \end{bmatrix}$$

$H_0: \tau_1 = \tau_2 = \tau_3 = 0$

vs

H_1: Not all $\tau_i = 0$

$$\Lambda^* = \frac{|W|}{|B+W|} = 0.697 / 0.7714$$

$p = 4$, $g = 3$

Test statistics is given by

$$\left(\frac{N-p-2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) = \left(\frac{516-4-2}{4}\right)\left(\frac{1-\sqrt{0.697}}{\sqrt{0.697}}\right)$$

$$= 25.22$$

F_{tab} , $v_1 = 2(g-1) = 4$, $v_2^* = 2(N-g-1) = 2(512) = 1024$

$F_{tab,4,1024,} = 2.51$

$TS > F_{tab}$ $\therefore$ we reject H_0.