# Foundations of Linear Models, Assignment 1

*Ayder Tanriver Ezgi (1541821), Oana Petrof (1541809), Olusoji Oluwafemi Daniel (1541893), Van Baelen Wessel (1234318) Owokotomo Olajumoke Evangelina (1539654)*

*26 September 2016*

## Problem 1

### (i) Is the matrix of full rank?

Performing elementary row reduction on $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$

$$\begin{array}{c} \xrightarrow{r_2/2} \\ \xrightarrow{r_3 - r_1} \end{array} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \end{bmatrix}$$

$$\xrightarrow{r_3 - 2r_1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

and counting the number of non-zero rows gives a rank of 2. Also, computing the determinant of the matrix $A$,

$$1 \begin{vmatrix} 2 & 2 \\ 2 & 3 \end{vmatrix} - 0 \begin{vmatrix} 0 & 2 \\ 1 & 3 \end{vmatrix} + 1 \begin{vmatrix} 0 & 1 \\ 0 & 2 \end{vmatrix} = 0$$

The two approach showed that the matrix is not of full rank since the matrix has $rank = 2$ which does not span the maximum space in dimension, which is 3. Also, the determinant of the matrix is 0 which implies that one of the rows of the matrix is linearly dependent on another row (row3 = row2 + 1), leading to the same conclusion that the matrix is not of full rank.

### (ii) Determine the eigenvalues and eigenvectors

To obtain the eigen values of the matrix $A$, the equation $|A - \lambda I_3| = 0$ is solved. $\lambda$ is scalar and $I_3$ is a $(3 \times 3)$ identity matrix and 0 is a $(3 \times 1)$ column vector. Solving the system;

$$\begin{vmatrix} 1 & 0 & 1 & & \lambda & 0 & 0 \\ 0 & 2 & 2 & - & 0 & \lambda & 0 \\ 1 & 2 & 3 & & 0 & 0 & \lambda \end{vmatrix} = 0$$

results into the equation $\lambda(6\lambda - \lambda^2 - 6) = 0$ which results in the roots 0, $3 + \sqrt{3}$ and $3 - \sqrt{3}$. Therefore, the eigen values of the matrix are; 4.732, 1.268, 0 and solving the system $(A - \lambda I_3)X = 0$ for each lambda gives the eigen vectors which are;

$$(0.5774, 0.5774, -0.5774) \quad \text{for} \quad 0$$

,

$$(0.2114, 0.5774, 0.7887) \quad \text{for} \quad 3 + \sqrt{3}$$

and

$$(0.7887, -0.5774, 0.2114) \quad \text{for} \quad 3-\sqrt{3}$$

.

Since the determinant of the matrix is 0 and it is as well symmetric, the product of the eigen values should be 0, i.e. $\prod_{i=1}^{3} \lambda_i = 0$ and the sum of the eigen values should give 6 (trace of the matrix), i.e. $\sum_{i=1}^{3} \lambda_i = 6$. $0 \times (3+\sqrt{3}) \times (3-\sqrt{3}) = 0$ and $0 + (3+\sqrt{3}) + (3-\sqrt{3}) = 6$ which validates lemma 2(i).

**(iii) Check for the validity of Lemma 2(iii)**

Lemma 2(iii) is valid if $UDU^{-1} = A$ where $U$ is the eigen vector of the matrix A, $U^{-1}$ is the inverse of the eign vectors of the matrix A and D is a diagonal matrix of the eigenvalues of A.

$$U = \begin{pmatrix} 0.2113 & 0.7887 & 0.5774 \\ 0.5774 & -0.5774 & 0.5774 \\ 0.7887 & 0.2113 & -0.5774 \end{pmatrix}$$

$$D = \begin{pmatrix} 4.732 & 0.000 & 0.000 \\ 0.000 & 1.268 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{pmatrix}$$

and

$$U^{-1} = \begin{pmatrix} 0.2113 & 0.5774 & 0.7887 \\ 0.7887 & -0.5774 & 0.2113 \\ 0.5774 & 0.5774 & -0.5774 \end{pmatrix}$$

Multiplying the matrix, gives;

$$\begin{pmatrix} 0.2113 & 0.7886751 & 0.5774 \\ 0.5774 & -0.5773503 & 0.5774 \\ 0.7887 & 0.2113 & -0.5774 \end{pmatrix} \times \begin{pmatrix} 4.732 & 0.000 & 0.000 \\ 0.000 & 1.268 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{pmatrix} \times \begin{pmatrix} 0.2113 & 0.5774 & 0.7887 \\ 0.7887 & -0.5774 & 0.2113 \\ 0.5774 & 0.5774 & -0.5774 \end{pmatrix}$$

The result of the multiplication gives the matrix below which is exactly A, hence the Lemma holds.

**(iv) Check for the Validity of Lemma 2(iv)**

rank(A) = number of nonzero eigenvalues. This lemma is true for the matrix in question as the number of non-zero eigen values is 2 which is the same as the rank of the matrix.

## Problem 2

$P = X(X'X)^{-1}X'$ is the orthogonal projection onto the column space of the design matrix which maps the vector of reponse values to the vector of the predicted values. $I_n - P$ is the orthogonal to the information contained in the design matrix. Both $P$ and $I_n - P$ gives the solution to the least square problem when the model matrix has linearly independent columns. Intuitively, $P$ projects the response into the plane, and it acts as an identity matrix to the design matrix $X$, since $PX = X(X'X)^{-1}X'X = X$ and $I_n - P$ on the other hand acts as a 0 matrix to X, since $(I_n - P)X = X - PX = X - X = 0$.

The property $(I_n - P)X = 0$ is also crucial to derivation of other theorem as it is used on **page 13**, **line 22** for **part III** of **theorem 5** to show that $\hat{\beta}$ and $MSE$ are independent and also on **page 14**, **line 8** for **part 4** of **theorem 5** to show that $\frac{(n-p)MSE}{\sigma^2} \sim \chi^2_{n-p}$.

## Problem 7

### Correlation Matrix

The correlation matrix for the model $LOS_1 = \beta_0 + \beta_1 Age_1 + \beta_2 IR_1 + \beta_3 AFAS_1 + \epsilon_1$ is;

| | $LOS$ | $Age$ | $IR$ | $AFAS$ |
|---|---|---|---|---|
| $LOS$ | 1.00000 | 0.18891 (0.0451) | 0.53344 ($<.0001$) | 0.35554 (0.0001) |
| $Age$ | 0.18891 (0.0451) | 1.00000 | 0.00109 (0.9908) | $-0.04045$ (0.6705) |
| $IR$ | 0.53344 ($<.0001$) | 0.00109 (0.9908) | 1.00000 | 0.41260 ($<.0001$) |
| $AFAS$ | 0.35554 (0.0001) | $-0.04045$ (0.6705) | 0.41260 ($<.0001$) | 1.00000 |

while that of the model $LOS_1 = \beta_0 + \beta_1 NumofBeds_1 + \beta_2 IR_1 + \beta_3 AFAS_1 + \epsilon_1$ is;

| | $LOS$ | $NumofBeds$ | $IR$ | $AFAS$ |
|---|---|---|---|---|
| $LOS$ | 1.00000 | 0.40927 (0.0451) | 0.53344 ($<.0001$) | 0.35554 (0.0001) |
| $NumofBeds$ | 0.40927 ($<.0001$) | 1.00000 | 0.35977 ($<.0001$) | 0.79452 ($<.0001$) |
| $IR$ | 0.53344 ($<.0001$) | 0.35977 ($<.0001$) | 1.00000 | 0.41260 ($<.0001$) |
| $AFAS$ | 0.35554 (0.0001) | 0.79452 (0.79452) | 0.41260 ($<.0001$) | 1.00000 |

## Appendix (Graphs & SAS Code)

### SAS Codes

### Codes for Question 1

PROC IML;

*DEFINE THE MATRIX WHICH IS A IN THIS CASE; A={1 0 1, 0 2 2, 1 2 3};

RANKOFMATRIX =ROUND(TRACE(GINV(A)*A));* GETTING THE RANK OF THE MATRIX;

DETERMI = DET(A); *DETERMINANT OF MATRIX A;

EVECTOR = EIGVEC(A); *GETTING THE EIGEN VECTOR;

EVALUE = EIGVAL(A); *GETTING THE EIGEN VALUE;

INVEVECTOR = INV(EVECTOR); *GETTING THE INVERSE OF THE EIGENVECTOR;

DIAGONALEVALUE = DIAG(EVALUE); *CONVERTING THE EIGEN VALUE TO DIAGONAL MATRIX;

ORIGINALMATRIX = EVECTOR * DIAGONALEVALUE * INVEVECTOR; *TESTING THE LEMMA 2(IV) IN THE TEXTBOOK;

PRINT RANKOFMATRIX DETERMI EVALUE EVECTOR INVEVECTOR DIAGONALEVALUE

ORIGINALMATRIX;

QUIT;

**Codes for Question 7**

```
DATA SENIC;
infile "C:\Users\OODOOE\Downloads\Video\Second Year\
First Semester\Foundation\foundation_assignment1\Data\APPENC01.txt";
input IDNUM LOS Age IR RCR RCXR NumOfBeds MSA Region ADC NumberOfNurses AFAS;
label LOS=‘Length of Patient Stay’
  Age='Age of patients'

  IR='Infection Risk'

  AFAS='Available Facilities and Services'

  NumOfBeds='Number of Beds';

run;
data hw1;
set SENIC;
keep IDNUM LOS Age NumOfBeds IR AFAS;
run;
*Scatter plot for each of the predictor variables;
title ‘Stem and Leaf Plot’;
ods listing;
ods graphics off;
ods select Plots SSPlots;
proc univariate data=Senic plot;
var Age IR AFAS NumofBeds;
run;
title;
*Scatter Plot and Correlation Matrix for Model 1;
title ‘Correlation Matrix and Scatter Plots for Model 1’;
ods graphics on;
proc corr data=Senic plots=matrix(histogram);
var LOS Age IR AFAS;
run;
ods graphics off;
title;
*Scatter Plot and Correlation Matrix for Model 2;
```

title ‘Correlation Matrix and Scatter Plots for Model 2’;

ods graphics on;

proc corr data=Senic plots=matrix(histogram);

var LOS IR AFAS NumofBeds;

run;

ods graphics off;

title;

**Stem and Leaf Plot For Predictor Variables**

**Stem and Leaf Plot**

**The UNIVARIATE Procedure**
**Variable: Age (Age of patients)**

```
Stem Leaf                          #              Boxplot
  64 19                            2                 0
  62 29                            2                 |
  60 91                            2                 |
  58 0220569                       7                 |
  56 012335578999922668           18              +-----+
  54 0011224556900788             16              |     |
  52 001112345888802222788899     24              *--+--*
  50 224566790012355677789        21              +-----+
  48 12601355579                  11                 |
  46 12                            2                 |
  44 20257                         5                 |
  42 07                            2                 0
  40
  38 8                             1                 0
     ----+----+----+----+----
```

Figure 1: Age Stem and Leaf Plot

**Stem and Leaf Plot**

The UNIVARIATE Procedure
Variable: IR (Infection Risk)

```
Stem Leaf                        #           Boxplot
   7 678                         3              0
   7
   6 56                          2              |
   6 12334                       5              |
   5 5555666777889              13              |
   5 0000112233344              13           +-----+
   4 5555555666778888999        19           |     |
   4 0111122222333333344444     22           *--+--*
   3 5577778999                 10           +-----+
   3 011244                      6              |
   2 5677899999                 10              |
   2 0013                        4              |
   1 678                         3              |
   1 334                         3              0
     ----+----+----+----+--
```

Figure 2: Infectious Risk Stem and Leaf Plot

**Stem and Leaf Plot**

The UNIVARIATE Procedure
Variable: AFAS (Available Facilities and Services)

```
Stem Leaf                        #           Boxplot
   8 0                           1              |
   7 7                           1              |
   7 14                          2              |
   6 666999                      6              |
   6 0033333                     7              |
   5 7777777                     7              |
   5 11111111444444             14           +-----+
   4 666666666699999999         18           |     |
   4 00000000333                11           *--+--*
   3 777777777                   9           |     |
   3 111114444444               12           +-----+
   2 666669999999               12              |
   2 03333333                    8              |
   1 77                          2              |
   1 14                          2              |
   0 6                           1              |
     ----+----+----+----+
  Multiply Stem.Leaf by 10**+1
```

Figure 3: Available Facilities and Services and Leaf Plot

**Stem and Leaf Plot**

**The UNIVARIATE Procedure**
**Variable: NumOfBeds (Number of Beds)**

```
Stem Leaf                      #             Boxplot
   8 334                       3                0
   7 57                        2                0
   7
   6
   6 0024                      4                0
   5 5779                      4                |
   5 14                        2                |
   4 6899                      4                |
   4 24                        2                |
   3 5669                      4                |
   3 00000011222              11             +-----+
   2 55567788                  8             |  +  |
   2 0000001244               10             |     |
   1 555666777788888999       18             *-----*
   1 000000001111222333334    21             +-----+
   0 5666777778888899999      19                |
   0 3                         1                |
     ----+----+----+----+-
 Multiply Stem.Leaf by 10**+2
```

Figure 4: Available Facilities and Services and Leaf Plot

## Scatter Plots

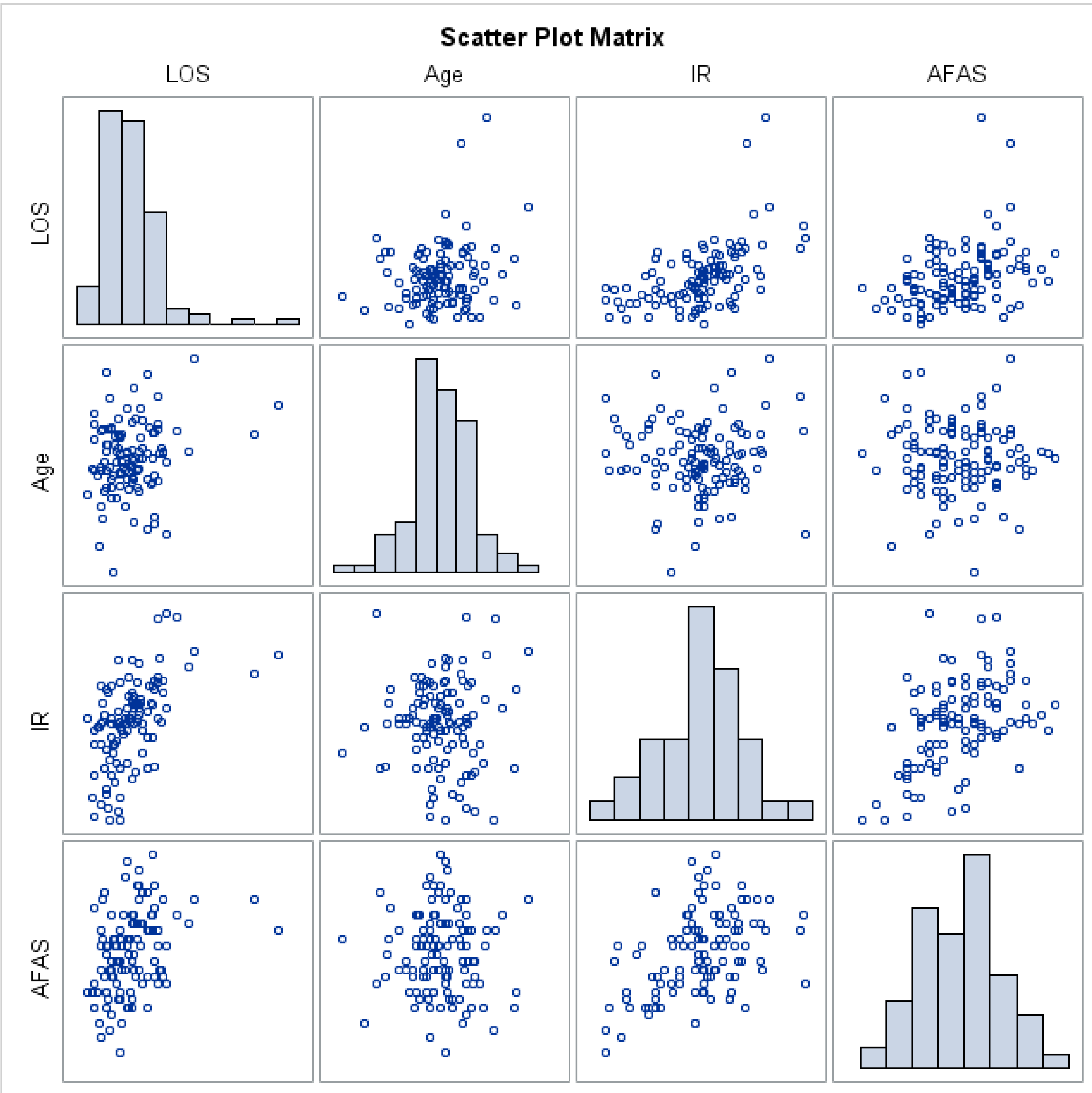


Figure 5: ScatterPlot for Model1

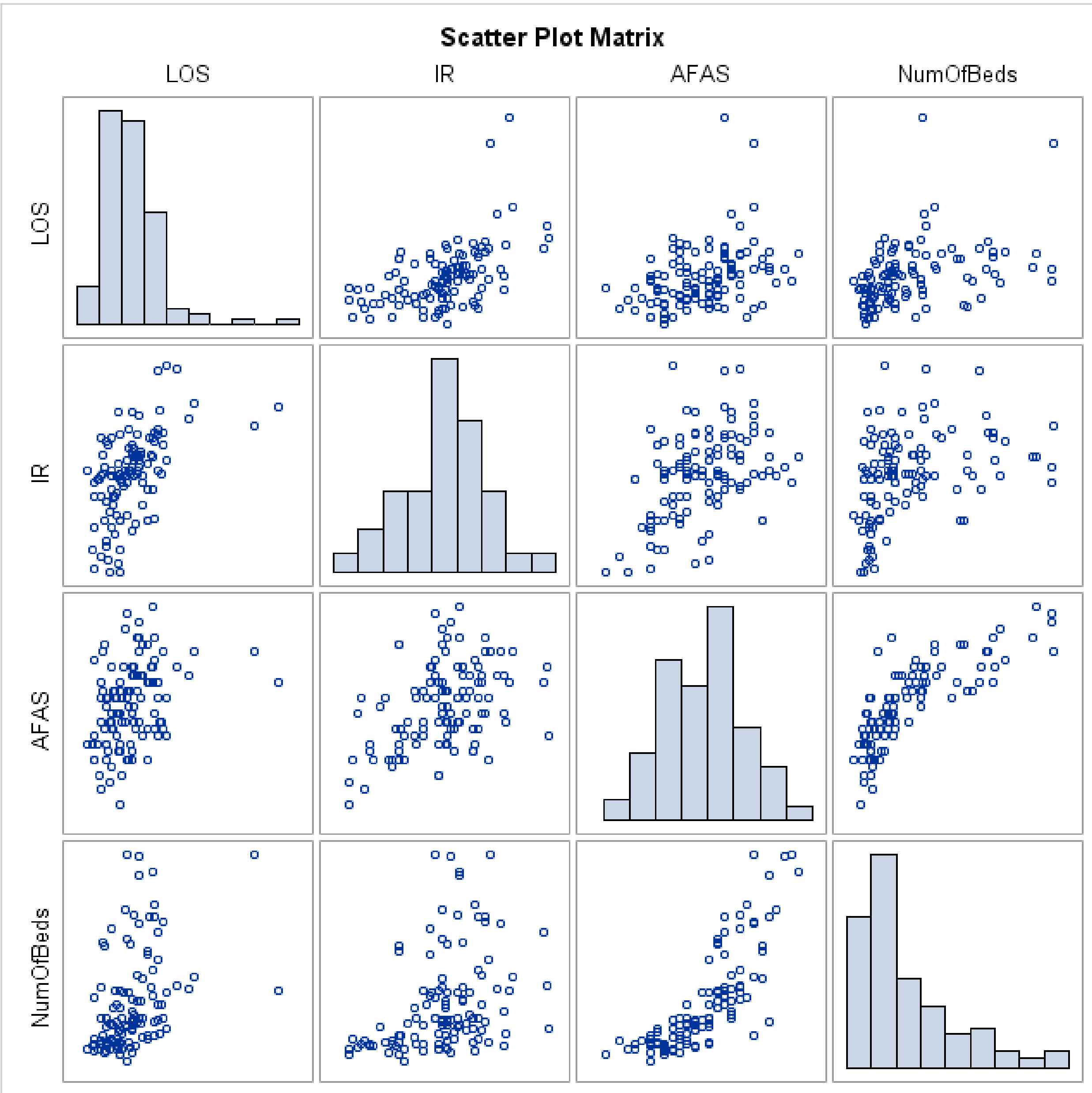


Figure 6: ScattePlot for Model2