F71RA: Machine Learning Lab 1

Table of contents

[1: Matrix Algebra Exercises 1](#_Toc208951304)

[Exercise 1.1: Matrix Creation 1](#_Toc208951305)

[Exercise 1.2: Matrix Multiplication 1](#_Toc208951306)

[Exercise 1.3: Matrix Inversion 2](#_Toc208951307)

[Exercise 1.4: Matrix Transposition 2](#_Toc208951308)

[2: Descriptive and Summary Statistics Exercises 2](#_Toc208951309)

[Exercise 2.1: Data Creation 2](#_Toc208951310)

[Exercise 2.2: Variance and Standard Deviation 2](#_Toc208951311)

[Exercise 2.3: Histogram 2](#_Toc208951312)

[Exercise 2.4: Correlation Matrix 3](#_Toc208951313)

[Exercise 2.5: Skewness and Kurtosis 3](#_Toc208951314)

[3: Linear Regression Exercises 4](#_Toc208951315)

[Exercise 3.1: Data Generation 4](#_Toc208951316)

[Exercise 3.2: Multiple Linear Regression 4](#_Toc208951317)

## 1: Matrix Algebra Exercises

### Exercise 1.1: Matrix Creation

Creating a matrix involves arranging numbers in rows and columns. Create a matrix with random integer values ranging from 1 to 10.

M <- matrix ( sample (1:10 , 9, replace = TRUE ), nrow = 3)

### Exercise 1.2: Matrix Multiplication

Matrix multiplication is a fundamental operation in linear algebra. Generate two random matrices and and compute their product .

A <- matrix ( sample (1:10 , 12, replace = TRUE ), nrow = 4, ncol = 3)  
B <- matrix ( sample (1:10 , 6, replace = TRUE ), nrow = 3, ncol = 2)  
C <- A %\*% B

### Exercise 1.3: Matrix Inversion

Matrix inversion is the process of finding the ”inverse” of a matrix, denoted as . Create a matrix X and calculate its inverse X\_inv if it exists.

X <- matrix (c(2, 1, 1, 2) , nrow = 2)  
X\_inv <- solve (X)

### Exercise 1.4: Matrix Transposition

Matrix transposition involves switching the rows and columns of a matrix. Generate a random matrix and calculate its transpose using the t() function.

D <- matrix(sample(1:10, 8, replace = TRUE), nrow = 4, ncol = 2)  
D\_trans <- t(D)

## 2: Descriptive and Summary Statistics Exercises

### Exercise 2.1: Data Creation

Generate a dataset with 100 random values for .

set.seed(123)  
X <- rnorm(100)

### Exercise 2.2: Variance and Standard Deviation

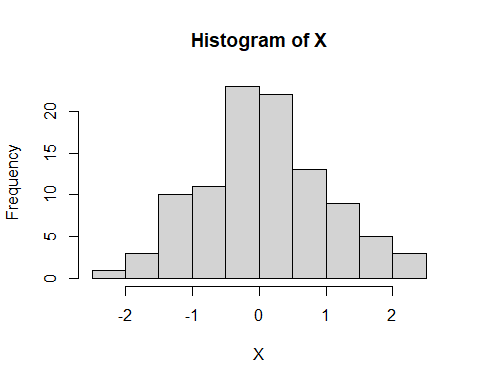
Variance and standard deviation are measures of the spread or dispersion of a dataset. Calculate the variance and standard deviation of the variable.

variance\_X <- var(X)  
sd\_X <- sd(X)

### Exercise 2.3: Histogram

A histogram is a graphical representation of the distribution of data. Create a histogram for the variable with 10 bins using the hist() function.

hist(X, breaks = 10, main = "Histogram of X", xlab = "X")



### Exercise 2.4: Correlation Matrix

Generate a dataset with two variables and and calculate their correlation matrix.

Y <- X + rnorm(100 , mean = 0, sd = 0.5)  
data <- data.frame(X, Y)  
cor\_matrix <- cor(data)

### Exercise 2.5: Skewness and Kurtosis

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. Compute the skewness and kurtosis of the variable.

#install.packages("moments")  
library(moments) # Load the moments package  
skewness\_X <- skewness(X)  
kurtosis\_X <- kurtosis(X)

## 3: Linear Regression Exercises

### Exercise 3.1: Data Generation

Generating a synthetic dataset involves creating a dataset with known properties for testing purposes. Create a synthetic dataset with two variables and for linear regression analysis.

X <- rnorm(100)  
Y <- 2\*X + rnorm(100)  
data <- data.frame(X,Y)

### Exercise 3.2: Multiple Linear Regression

Multiple linear regression involves using multiple predictors to predict a response variable. Extend the dataset to include another predictor and fit a multiple linear regression model using the lm() function. Now try and interpret this model using the summary() function.

Z <- rnorm(100)  
data$Z <- Z  
multi\_model <- lm(Y~X+Z, data = data)  
summary(multi\_model)

Call:  
lm(formula = Y ~ X + Z, data = data)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-2.43862 -0.65637 -0.00031 0.75288 2.52897   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -0.02798 0.10630 -0.263 0.793   
X 1.95038 0.11094 17.581 <2e-16 \*\*\*  
Z -0.02141 0.10651 -0.201 0.841   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 1.048 on 97 degrees of freedom  
Multiple R-squared: 0.7614, Adjusted R-squared: 0.7565   
F-statistic: 154.7 on 2 and 97 DF, p-value: < 2.2e-16