Varun Sudhir 21BDS0040

Exploratory Data Anlaysis Lab Exp 6

Write R programs to read a set of X,Y values and find covariance and correlation using functions.

For this experiment,we will utilize the NHANES dataset. The NHANES (National Health and Nutrition Examination Survey) dataset is a large, real-world dataset that comes from a program of studies conducted by the National Center for Health Statistics (NCHS). The NHANES dataset is often used in health and epidemiological studies and contains various missing values across its columns.

We will find the correlation and covariance using 2 of the numeric columns of the dataset

	BPSysAve	BMI
1	113	32.22
2	113	32.22
3	113	32.22
4	NA	15.30
5	112	30.57
6	86	16.82
7	107	20.64
8	118	27.24
9	118	27.24
10	118	27.24
11	111	23.67
12	104	23.69
13	134	26.03
14	104	19.20
15	127	26.22
16	142	26.60
17	93	27.40
18	128	28.54
19	152	25.84

Code:

```r
# Varun Sudhir 21BDS0040
library(NHANES)


# For this example, we will choose 'BPSysAve' (average systolic blood
pressure) and 'BMI' (Body Mass Index)
X <- NHANES$BPSysAve
Y <- NHANES$BMI

# Function to impute missing values with the mean of the column
impute_missing_values <- function(data) {
  data[is.na(data)] <- mean(data, na.rm = TRUE)
  return(data)
}



# Impute missing values in X and Y
X_imputed <- impute_missing_values(X)
Y_imputed <- impute_missing_values(Y)

# Function to calculate covariance
calculate_covariance <- function(X, Y) {
  return(cov(X, Y))
}

# Function to calculate correlation
calculate_correlation <- function(X, Y) {
  return(cor(X, Y))
}

# Calculate and print the covariance
covariance_result <- calculate_covariance(X_imputed, Y_imputed)
cat("Covariance between BPSysAve and BMI (with imputation): ",
covariance_result, "\n")

# Calculate and print the correlation
correlation_result <- calculate_correlation(X_imputed, Y_imputed)
cat("Correlation between BPSysAve and BMI (with imputation): ",
correlation_result, "\n")
cat("Varun Sudhir 21BDS0040")
```

Output:

```
> # Calculate and print the covariance
> covariance_result <- calculate_covariance(X_imputed, Y_imputed)
> cat("Covariance between BPSysAve and BMI (with imputation): ", covariance_result, "\n")
Covariance between BPSysAve and BMI (with imputation):  26.72422
> # Calculate and print the correlation
> correlation_result <- calculate_correlation(X_imputed, Y_imputed)
> cat("Correlation between BPSysAve and BMI (with imputation): ", correlation_result, "\n")
Correlation between BPSysAve and BMI (with imputation):  0.2314194
> cat("Varun Sudhir 21BDS0040")
Varun Sudhir 21BDS0040
> |
```