DA 2

Abhishek Murthy

21BDS0064

2024-09-12

```
library(nycflights13)
```

```
## Warning: package 'nycflights13' was built under R version 4.3.3
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(mice)
```

```
## Warning: package 'mice' was built under R version 4.3.3
```

```
##
## Attaching package: 'mice'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```
library(VIM)
```

```
## Warning: package 'VIM' was built under R version 4.3.3
```

```
## Loading required package: colorspace
```

```
## Loading required package: grid
```

```
## VIM is ready to use.

## Suggestions and bug-reports can be submitted at:
https://github.com/statistikat/VIM/issues

##
## Attaching package: 'VIM'

## The following object is masked from 'package:datasets':
##
##     sleep

# Load the flights dataset
data("flights")

## COVARIANCE AND CORRELATION
calculate_covariance <- function(x, y) {

  if(length(x) != length(y)) {
    stop("X and Y must have the same length.")
  }


  x_mean <- mean(x)
  y_mean <- mean(y)


  covariance <- sum((x - x_mean) * (y - y_mean)) / (length(x) - 1)

  return(covariance)
}


calculate_correlation <- function(x, y) {

  if(length(x) != length(y)) {
    stop("X and Y must have the same length.")
  }


  covariance <- calculate_covariance(x, y)


  x_sd <- sd(x)
  y_sd <- sd(y)


  correlation <- covariance / (x_sd * y_sd)
```

```r
  return(correlation)
}



x <- flights$dep_delay
y <- flights$arr_delay


valid_indices <- !is.na(x) & !is.na(y)
x_clean <- x[valid_indices]
y_clean <- y[valid_indices]


covariance_value <- calculate_covariance(x_clean, y_clean)
correlation_value <- calculate_correlation(x_clean, y_clean)

cat("Covariance between dep_delay and arr_delay:", covariance_value, "\n")

## Covariance between dep_delay and arr_delay: 1635.908

cat("Correlation between dep_delay and arr_delay:", correlation_value, "\n")

## Correlation between dep_delay and arr_delay: 0.9148028
```