Practicals 3 (Solution)

Mean, Covariance and Corelation of Bivariate data

Your Name (Reg.No)

Consider the data on the height (in inches) and weight (in lbs.) of 20 college-age males from the file `HeightWeight.txt`.

Questions:

- Plot the scatter plot Height, Weight and Height Vs Weight
- Find the mean vector, covariance matrix and correlation matrix.
- Using matrix operation, create a new dataset where the hei ght and weight information from the observations are measured in metres and kilogr ams, respectively, in the height and weight data of 20 college-age males.
- Calculate the covariance between height and weight in the new dataset. Compare it to the covariance in the original dataset. Do you see a difference?
  - Calculate as well the correlation between height and weight in the new dataset using the matrix operation on the covariance matrix. Compare it to the correlation in the original dataset. Do you now see a difference?

Solutions:

```
# A setup chunk is added to enhance the table printing options for the pdf output
# which uses a library called `printr`.
# Check the markdown file for more details.
```

Loading the Data set:

```
hwdata = read.table('HeightWeight.txt', header=T)
```

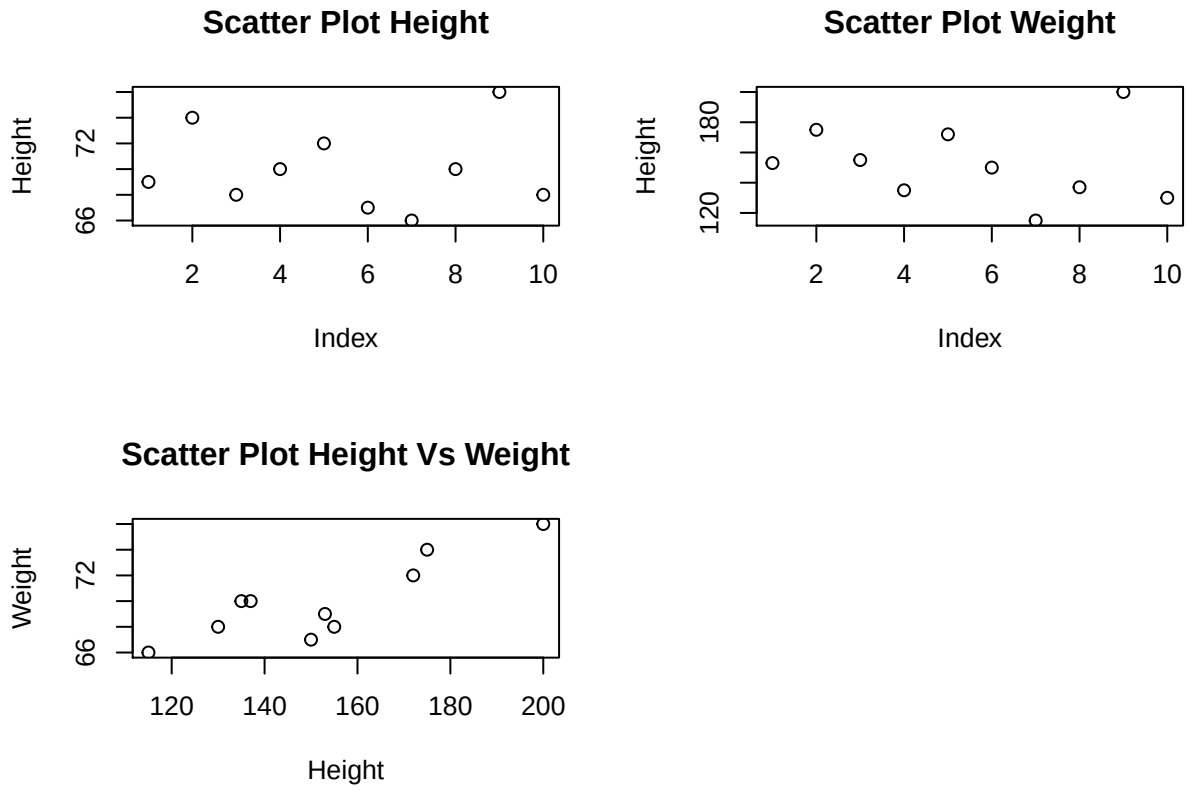
The few rows of the data set is given below:

```
head(hwdata, n = 5)
```

Height	Weight
69	153
74	175
68	155
70	135
72	172

Plotting the scatter plot Height, Weight and Height Vs Weight:

```
par(mfrow = c(2,2))
plot(hwdata$Height,
     main = "Scatter Plot Height",
     xlab ="Index",
     ylab = "Height")
plot(hwdata$Weight,
     main = "Scatter Plot Weight",
     xlab ="Index",
     ylab = "Height")
plot(hwdata$Height ~ hwdata$Weight,
     main = "Scatter Plot Height Vs Weight",
     xlab ="Height", ylab = "Weight")
```



Finding the mean vector, covariance matrix and correlation matrix:

Finding the mean vector:

```
one = rep(1, nrow(hwdata))
MeanVec = 1/nrow(hwdata) * t(hwdata) %*% one
```

The mean vector is given by,

```
MeanVec
```

Height	70.0
Weight	152.2

Finding the covariance matrix:

```
Y = hwdata - one %*% t(MeanVec)
Y = as.matrix(Y)
CovMat = 1/(nrow(hwdata) - 1 ) * t(Y) %*% Y
```

The variance-covariance matrix is given by,

```
CovMat
```

	Height	Weight
Height	10.00000	67.88889
Weight	67.88889	623.73333

Finding correlation matrix:

```
var = diag(CovMat)
sd = diag(1/sqrt(var), ncol = 2)
CorMat = sd %*% CovMat %*% sd
```

Therefore the correlation matrix is given by

```
CorMat = data.frame(CorMat)
colnames(CorMat) = colnames(CovMat)
rownames(CorMat) = rownames(CovMat)
CorMat
```

	Height	Weight
Height	1.0000000	0.8596056
Weight	0.8596056	1.0000000

Creating a new data set by converting the units:

```
newHWdata = data.frame(NewHt = 0.0254*hwdata$Height,
                       NewWt = 0.4535*hwdata$Weight)
```

The first few rows of the new data set are,

```
head(newHWdata)
```

NewHt	NewWt
1.7526	69.3855
1.8796	79.3625
1.7272	70.2925
1.7780	61.2225
1.8288	78.0020
1.7018	68.0250

Finding the mean vector, covariance matrix and correlation matrix for the new data set:

Finding the mean vector:

```
one = rep(1, nrow(newHWdata))
NewMeanVec = 1/nrow(newHWdata) * t(newHWdata) %*% one
```

The mean vector is given by,

```
NewMeanVec
```

NewHt	1.7780
NewWt	69.0227

Finding the covariance matrix:

```
Y = newHWdata - one %*% t(NewMeanVec)
Y = as.matrix(Y)
NewCovMat = 1/(nrow(newHWdata) - 1 ) * t(Y) %*% Y
```

The variance-covariance matrix is given by,

```
NewCovMat
```

	NewHt	NewWt
NewHt	0.0064516	0.7820053
NewWt	0.7820053	128.2784007

Comparison of covariance matrices between Original data set and new data set:

```
CovMat - NewCovMat
```

	Height	Weight
Height	9.993548	67.10688
Weight	67.106884	495.45493

Clearly the covariance matrices are **different**.

Finding correlation matrix:

```
newvar = diag(NewCovMat)
newsd = diag(1/sqrt(newvar), ncol = 2)
NewCorMat = newsd %*% NewCovMat %*% newsd
```

Therefore the correlation matrix is given by

```
NewCorMat = data.frame(NewCorMat)
colnames(NewCorMat) = colnames(NewCovMat)
rownames(NewCorMat) = rownames(NewCovMat)
NewCorMat
```

	NewHt	NewWt
NewHt	1.0000000	0.8596056
NewWt	0.8596056	1.0000000

Comparison of covariance matrices between Original data set and new data set:

```
options(digits = 17) # To maintain the digits when knitting to pdf.
CorMat - NewCorMat
```

	Height	Weight
Height	-2.2e-16	2.2e-16
Weight	1.1e-16	0.0e+00

All the entries in the above difference matrix is zero (or approximately zero) and hence it is concluded that there is no difference between correlation matrices from the original data set and new data set.

```
lm(hwdata$Height~hwdata$Weight)
```

```
##
## Call:
## lm(formula = hwdata$Height ~ hwdata$Weight)
##
## Coefficients:
##      (Intercept)     hwdata$Weight
## 53.4341242696309   0.1088428103178
```